



*symmetry*

IMPACT  
FACTOR  
2.7

CITESCORE  
4.9

Review

---

# Analysis of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ Decays at the Large Hadron Collider

---

Kai-Feng Chen, Titus Mombächer and Umberto De Sanctis

Special Issue

Symmetries and Anomalies in Flavour Physics

Edited by

Dr. Stefania Ricciardi, Dr. Thomas Blake and Dr. Farvah Nazila Mahmoudi



<https://doi.org/10.3390/sym16020251>

Review

# Analysis of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ Decays at the Large Hadron Collider

Kai-Feng Chen <sup>1,\*</sup>, Titus Mombächer <sup>2,\*</sup> and Umberto De Sanctis <sup>3,4,\*</sup>

<sup>1</sup> Department of Physics, National Taiwan University, Taipei City 106319, Taiwan

<sup>2</sup> CERN, 1211 Geneva, Switzerland

<sup>3</sup> Department of Physics, University of Roma Tor Vergata, 00133 Rome, Italy

<sup>4</sup> INFN Sezione di Roma Tor Vergata, 00133 Rome, Italy

\* Correspondence: kai-feng.chen@cern.ch (K.-F.C.); titus.mombacher@cern.ch (T.M.); umberto.de.sanctis@uniroma2.it (U.D.S.)

† The authors contributed equally to this work.

**Abstract:** This article reviews the most recent measurements of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decay properties at the Large Hadron Collider (LHC) which are the most precise to date. The measurements of the branching fraction and effective lifetime of the  $B_s^0 \rightarrow \mu^+ \mu^-$  decay by the ATLAS, CMS, and LHCb collaborations, as well as the search for  $B^0 \rightarrow \mu^+ \mu^-$  decays, are summarized with a focus on the experimental challenges. Furthermore, prospects are given for these measurements and new observables that become accessible with the foreseen amounts of data by the end of the LHC.

**Keywords:**  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ; FCNC; rare decay; flavor physics; ATLAS; CMS; LHCb; LHC

## 1. Introduction

This review summarizes the most recent measurements related to the  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^-$  decays performed with the ATLAS, CMS, and LHCb experiments.

The  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^-$  decays belong to the category of Flavor Changing Neutral Current (FCNC) processes and are therefore highly suppressed in the Standard Model (SM). This makes them important tools in the search for New Physics (NP) since they can provide indirect constraints on NP processes that interfere with the SM processes and alter the rates and decay properties sizeably. They are even sensitive to particles that are out of the kinematic range accessible by particle colliders, including the Large Hadron Collider (LHC). The  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays are among the most sensitive FCNC processes due to their small theoretical uncertainty and clean experimental signature [1–5]. In the SM, the decays  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  are forbidden at leading-order and can proceed only via loop diagrams. In addition, they are also suppressed by the helicity conservation and the presence of off-diagonal CKM matrix elements, leading to a very small expected decay time of integrated branching fractions. Additional interest in  $B_s^0 \rightarrow \mu^+ \mu^-$  decays comes from the simple description in effective field theory [6,7]. The decays can only proceed via axial-vector (Wilson coefficient  $C_{10}$ ), scalar ( $C_S$ ), or pseudo-scalar ( $C_P$ )  $b \rightarrow s \ell \bar{\ell}$  currents, where the scalar and pseudo-scalar currents are forbidden in the SM. Thus, measurements of  $B_s^0 \rightarrow \mu^+ \mu^-$  properties are crucial inputs for global fits of the parameters that govern  $b \rightarrow s \ell \bar{\ell}$  transitions.

The most up-to-date SM predictions for the  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B^0 \rightarrow \mu^+ \mu^-$  branching fractions are calculated in Ref. [5] and yield

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (3.66 \pm 0.14) \times 10^{-9} \quad \text{and} \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &= (1.03 \pm 0.05) \times 10^{-10}. \end{aligned} \quad (1)$$

The values of these branching fractions are  $CP$  averaged and time-integrated and include final state radiation effects, so that they can be readily compared with the experi-



**Citation:** Chen, K.-F.; Mombächer, T.; De Sanctis, U. Analysis of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  Decays at the Large Hadron Collider. *Symmetry* **2024**, *16*, 251. <https://doi.org/10.3390/sym16020251>

Academic Editor: Tianjun Li

Received: 18 January 2024

Revised: 7 February 2024

Accepted: 15 February 2024

Published: 18 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

mental measurements, which do not distinguish between  $B_{(s)}^0$   $CP$  eigenstates and take final state radiation effects into account. Next-to-leading order electroweak corrections and next-to-next-to-leading order QCD corrections are also included in the calculations. Recently, several progresses in lattice quantum chromodynamics (QCD) [8–12], in the calculation of electroweak effects at next-to-leading order [2], and QCD effects at next-to-next-to-leading order [3] helped significantly in reducing the theoretical uncertainties on both branching fractions. Enhanced electromagnetic contributions from virtual photon exchange were also proven to produce larger corrections to theoretical uncertainties than previously assumed [4,5]. These predictions also take into account the finite width differences measured in the  $B_s^0$  system that apply in the experimental measurements where the data samples are untagged (see Refs. [13,14]). Alternative predictions are also available. They are obtained using the relation between  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays and  $\Delta m_{d(s)}$ , plus the mass difference of the  $B_{(s)}^0$  mass eigenstates [15,16]. In addition, it has been recently pointed out [17] that the current way to calculate  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$  could be affected by the presence of NP effects. Therefore, a calculation based on  $\Delta m_{d(s)}$  and  $|\epsilon_K|$  considering only the SM contribution has been proposed. In both cases, the resulting values for  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$  are slightly different than the ones shown in Equation (1), but still compatible within the theoretical uncertainties of the calculation, that are, on their own, still smaller than the experimental precision. Since all collaborations used the values reported in Equation (1) to assess the level of compatibility of the measurements with the SM predictions, in the remainder of this article, the values quoted in Equation (1) will be used as reference values for the  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$  SM predictions.

While the mentioned reference does not quote a value for the ratio of the two branching fractions, this can be easily calculated as follows:

$$\mathcal{R} = \frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)} = \frac{\tau_{B^0}}{1/\Gamma_H^s} \left( \frac{f_{B^0}}{f_{B_s^0}} \right)^2 \left| \frac{V_{td}}{V_{ts}} \right|^2 \frac{M_{B^0} \sqrt{1 - \frac{4m_\mu^2}{M_{B^0}^2}}}{M_{B_s^0} \sqrt{1 - \frac{4m_\mu^2}{M_{B_s^0}^2}}} = 0.0281 \pm 0.0016 \quad (2)$$

where  $\tau_{B^0}$  and  $1/\Gamma_H^s$  are the lifetimes of the  $B^0$  and of the heavy mass eigenstate of the  $B_s^0$ ;  $M_{B_s^0}$  and  $M_{B^0}$  are the masses and  $f_{B_s^0}$  and  $f_{B^0}$  the meson decay constants of the  $B_s^0$  and  $B^0$  mesons respectively;  $V_{td}$  and  $V_{ts}$  are the elements of the CKM matrix and  $m_\mu$  is the mass of the muon. Using the same input values as Ref. [5], the numerical value in Equation (2) is obtained. It is worth noting that the ratio has a theoretical uncertainty which is smaller than the single branching fractions due to the cancellation of most of the factors. In particular, this ratio has the same value in all theories obeying the Minimal Flavor Violation (MFV) paradigm (including the SM) and as such it is a test of the latter. It is therefore of additional interest to evaluate  $\mathcal{R}$  also in upcoming measurements.

A second observable of the  $B_s^0 \rightarrow \mu^+ \mu^-$  decay considered in the latest experimental results is its effective lifetime  $\tau_{\mu\mu}$ . This observable is complementary to the branching fraction because it is sensitive to potential New Physics (NP) effects which are flavor-dependent [13]. In fact, in the SM, only the heavy  $CP$  odd heavy-mass eigenstate component of the  $B_s^0$ - $\bar{B}_s^0$  system contributes to the  $B_s^0 \rightarrow \mu^+ \mu^-$  decay amplitude: an assumption which does not hold in every NP scenario. Therefore, the measurement of this quantity could reveal the presence of NP effects which do not affect the branching fraction measurement.  $\tau_{\mu\mu}$  is simply defined as the mean lifetime of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays

$$\begin{aligned} \tau_{\mu\mu} &\equiv \frac{\int_0^\infty t \langle \Gamma(B_s^0 \rightarrow \mu^+ \mu^-) \rangle dt}{\int_0^\infty \langle \Gamma(B_s^0 \rightarrow \mu^+ \mu^-) \rangle dt} \\ &= \frac{\tau_{B_s^0}}{1 - y_s^2} \left[ \frac{1 + 2\mathcal{A}_{\Delta\Gamma} y_s + y_s^2}{1 + \mathcal{A}_{\Delta\Gamma} y_s} \right], \end{aligned} \quad (3)$$

where  $t$  is the proper decay time of the  $B_s^0$  meson and  $y_s$  and the  $CP$  parameter  $\mathcal{A}_{\Delta\Gamma}$  are defined as

$$y_s \equiv \frac{\Delta\Gamma_s}{2\Gamma_s} \quad \mathcal{A}_{\Delta\Gamma} \equiv \frac{R_H^{\mu^+\mu^-} - R_L^{\mu^+\mu^-}}{R_H^{\mu^+\mu^-} + R_L^{\mu^+\mu^-}} \quad (4)$$

and  $R_H^{\mu^+\mu^-}$  and  $R_L^{\mu^+\mu^-}$  denote the contributions of the heavy and light mass eigenstates of the  $B_s^0$  system to the untagged  $B_s^0 \rightarrow \mu^+\mu^-$  decay rate. Since the  $\mu^+\mu^-$  final state is  $CP$ -odd, the SM  $\mathcal{A}_{\Delta\Gamma} = +1$  and the effective lifetime is equal to the lifetime of the heavy-mass  $B_s^0$  eigenstate  $\tau_{\mu\mu}^{SM}$ . The  $CP$  asymmetry  $\mathcal{A}_{\Delta\Gamma}$  can receive contributions from NP effects, particularly from scalar and pseudoscalar operators, even in cases where the branching fraction  $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$  is not modified. The most recent  $\tau_{\mu\mu}^{SM}$  value is  $1.624 \pm 0.009$  ps [18], which can be slightly different from that used by the various collaborations depending on the publication time of their most recent measurement.

All experimental results described in this review assume the SM hypothesis  $\mathcal{A}_{\Delta\Gamma} = 1$  in the calculation of efficiencies and acceptance for the  $B_s^0 \rightarrow \mu^+\mu^-$  decay and thus for its branching fractions. However, the ATLAS, CMS, and LHCb collaborations estimated the impact of such assumptions on the  $B_s^0 \rightarrow \mu^+\mu^-$  branching fraction, which spans from 4% to 10% depending on the  $\mathcal{A}_{\Delta\Gamma}$  varying in the interval  $[-1, 1]$ .

This article is structured as follows: in Sections 2–4, the measurements by the ATLAS, CMS, and LHCb collaborations are reported respectively, while in Section 5, the results obtained by the latest official LHC combination are presented. Section 6 provides a summary of the status of the measurements and prospects of the three collaborations for the HL-LHC phase.

## 2. The ATLAS $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$ and Effective Lifetime Measurements

### 2.1. The $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$ Measurement

The measurement of the branching fractions of the  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B^0 \rightarrow \mu^+\mu^-$  decays performed by the ATLAS collaboration is described and documented in Ref. [19]. The analysis uses  $26.1 \text{ fb}^{-1}$  of Run-2 data collected at  $\sqrt{s} = 13 \text{ TeV}$ , and combines the result with the previously published Run-1 analysis [20] on  $4.7 \text{ fb}^{-1}$  of data at  $\sqrt{s} = 7 \text{ TeV}$  and  $20.3 \text{ fb}^{-1}$  at  $\sqrt{s} = 8 \text{ TeV}$ .

In order to remove the dependence from the knowledge of the  $b$ -quark production cross section and minimize the systematic uncertainties, the branching fractions are measured relative to a reference channel. For its abundance and well-measured branching ratio, the  $B^+ \rightarrow J/\psi(\mu^+\mu^-)K^+$  decay channel has been chosen for this purpose. As a consequence, the procedure to extract the  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$  takes into account the difference in the fragmentation fractions  $f_{u,d,s}$  of  $b$ -quarks to form, respectively, a  $B^+$ ,  $B_s^0$ , or  $B^0$  meson. Also, the different acceptances and efficiencies between the signal and the reference channels are taken into account. Hence, the branching fractions  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$  are expressed as follows:

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-) = \frac{N_{B_{(s)}^0}}{N_{B^+}} \frac{f_u}{f_{d(s)}} \frac{\epsilon_{\text{tot}}^{B^+}}{\epsilon_{\text{tot}}} \mathcal{B}(B^+ \rightarrow J/\psi(\mu^+\mu^-)K^+), \quad (5)$$

where  $N_{B_{(s)}^0}$  ( $N_{B^+}$ ) is the measured yield of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  ( $B^+ \rightarrow J/\psi K^+$ ) events, and  $\epsilon_{\text{tot}}^{B^+}$  ( $\epsilon_{\text{tot}}^{B_{(s)}^0}$ ) is the total signal ( $B^+ \rightarrow J/\psi K^+$ ) efficiency. Events from  $B_s^0 \rightarrow J/\psi\phi$  decay, with  $J/\psi \rightarrow \mu\mu$  and  $\phi \rightarrow KK$ , are also used as a control sample for the signal kinematic variables exploited in the analysis.

The signal selection starts with a hardware dimuon trigger requiring one muon with transverse momentum  $p_T > 4 \text{ GeV}$  and the other with  $p_T > 6 \text{ GeV}$ . In the offline analysis, both muons are required to have the same  $p_T$  thresholds as in the trigger selection, to have pseudo-rapidity  $|\eta| < 2.5$ , and to pass stringent track-quality requirements (Tight

muons). Signal candidates are formed with two muons with opposite electric charges. Kaon candidates for the reference channel are reconstructed in the tracking system and are required to have  $p_T > 1$  GeV and  $|\eta| < 2.5$ .

$B$ -meson kinematic observables are reconstructed imposing quality requirements on the dimuon vertex for the signal, or on the vertex formed by the dimuon system and one track for the reference channel. The reconstructed  $B$  candidates are also required to fall within a fiducial volume defined as  $p_T(B) > 8$  GeV and  $|\eta(B)| < 2.5$ .

The analysis uses mainly the  $B$ -candidate invariant mass to characterize the selected events.  $B$ -candidates with a mass in the 4766–5966 MeV interval are considered. A blind analysis is performed where the dimuon invariant mass signal region between 5166 and 5526 MeV is not used until the analysis criteria and strategies are finalized.

The main backgrounds for this analysis can be split into three categories: continuum background, partially reconstructed  $B$  decays (PRD), and peaking background. The continuum background consists mainly of muons produced in uncorrelated hadron decays. It is the dominant background for the analysis and it is several orders of magnitude larger than the signal. Therefore, a Boosted Decision Tree [21] (c-BDT) is employed to efficiently separate the signal from this background type. The c-BDT is based on 15 kinematic variables with high discriminating power which describe the kinematics of the  $B$ -meson candidate, the secondary vertex displacement, the kinematic properties of the muons, and the rest of the event (such as the isolation of the  $B$  candidate and that of the two muon tracks with respect to the rest of the event). The c-BDT is trained and validated on the data mass sidebands.

The PRD background is made of decays where the two muons in the final state come from one of the following topologies: (a) ‘cascade’ transitions with the muons coming from the same ancestor (e.g.,  $b \rightarrow c\mu\nu \rightarrow s\mu\mu\nu\nu$ ), and labeled same-side muons (SS); (b) from the same decay chain (e.g.,  $B \rightarrow J/\psi X$  or  $B^0 \rightarrow \mu\mu K^*$ ) and labeled same vertex muons (SV); (c) from  $B_c \rightarrow J/\psi\mu\nu$  decays; (d) from semileptonic  $B$  decays where a hadron  $h$  ( $\pi$ ,  $K$  or proton) is misidentified as a muon (e.g.,  $B \rightarrow \mu h\nu$ ). All these types of backgrounds populate the low-mass sideband with contributions also into the dimuon mass signal region.

The peaking background consists of charmless two-body decays  $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$  ( $h^{(\prime)}$  being a pion or a kaon) that are reconstructed as signal events due to the hadrons being misidentified as muons. This background component falls in the signal region and presents the same features of the  $B^0 \rightarrow \mu^+\mu^-$  signal. Its contribution has been studied with the help of a dedicated MC sample and validated in data in a region enriched by hadrons misidentified as muons. The resulting peaking background contribution is estimated to be  $2.9 \pm 2.0$  events in the signal region.

To extract the  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$  using Equation (5), the yield of the reference channel and the efficiency ratio between the two channels needs to be computed. The  $B^+ \rightarrow J/\psi K^+$  yield  $N_{B^+}$  is obtained by an unbinned extended maximum-likelihood fit to the  $\mu\mu K^+$  invariant mass distribution, where the shape parameters are fitted simultaneously in data and simulation.

The efficiency ratio between signal and reference channels is computed from appropriate simulation samples within the fiducial volume of the analysis. These samples are reweighted in such a way that they reproduce the distributions of the number of primary vertices (and therefore pile-up),  $p_T(B)$ ,  $|\eta(B)|$  and trigger efficiencies (as a function of  $p_T(\mu)$  and  $|\eta(\mu)|$ ) as measured in data. Furthermore, a correction to the  $B_s^0$  lifetime in the simulated signal sample is applied to match the distribution of the heavy  $B_s^0$  mass eigenstate, because the  $B_s^0 \rightarrow \mu^+\mu^-$  decay proceeds in the SM exclusively through the heavy  $B_s^0$  mass eigenstate, as described in Section 1.

The yields of signal events  $N_{B_{(s)}^0}$  are extracted simultaneously from an unbinned extended maximum-likelihood fit to the dimuon invariant mass distribution  $m_{\mu\mu}$ . In order to enhance the sensitivity of the analysis, four bins in the c-BDT output (in increasing order of signal-over-background ratio) are defined in order to have constant signal efficiency equal to 18% in each bin. The fit is performed simultaneously in the four c-BDT bins.

The first c-BDT bin, which has the lowest signal-over-background ratio, is dominated by the main backgrounds. It is introduced in the fit to improve the background modeling and reduce the systematic uncertainties related to them. The  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  signals are parameterized by a double Gaussian function to take into account different resolutions in the dimuon invariant mass depending on the different regions of the ATLAS detector. The shape and the relative signal efficiencies are assumed to be the same in all c-BDT bins. The continuum background is described by a first order polynomial, while the background coming from the  $SS$  and  $SV$  events is parameterized with an exponential function. These backgrounds are fluctuated independently in each c-BDT bin. Finally, the description of the peaking background is based on the same model used to describe the signal, with a constraint on the total yield of  $2.9 \pm 2.0$  equally distributed in the c-BDT bins.

The  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$  values are extracted through a simultaneous unbinned extended maximum-likelihood fit using the components written in Equation (5) and the  $N_{B_{(s)}^0}$  event yields extracted from the invariant mass fits just described. The  $\mathcal{B}(B^+ \rightarrow J/\psi(\mu^+ \mu^-)K^+)$  value is taken as the world average from the PDG [22], while the hadronisation probability ratio  $\frac{f_u}{f_{d(s)}} = 0.256 \pm 0.013$  is taken from the HFLAV average [23].

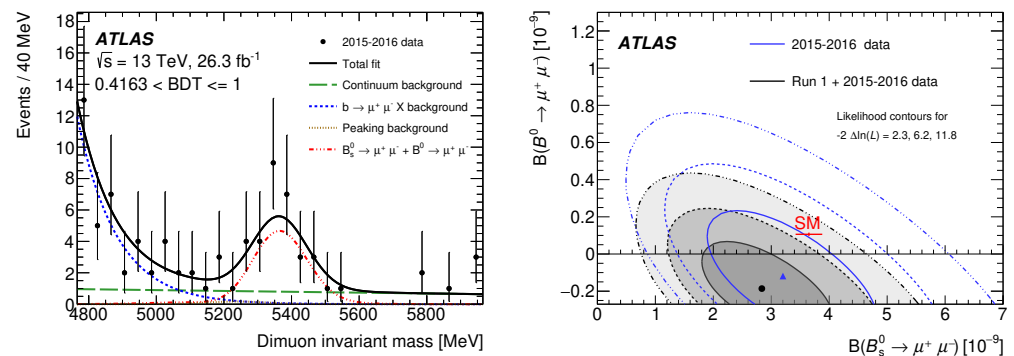
The measurements are dominated by statistical uncertainties, with the most prominent sources of systematic uncertainty coming from the fit uncertainties (where the largest contributors are the mass scale and the  $b \rightarrow \mu^+ \mu^- X$  background parameterisation), the  $\frac{f_u}{f_{d(s)}}$  ratio (only for the  $\text{BR}(B_s^0)$  measurement), and the reference channel yield. All systematic uncertainties are described in the likelihood as Gaussian constraints.

A Neyman construction [24] is employed to extract the 68.3%, 95.5%, and 99.7% confidence intervals in the  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  plane. The likelihood function from the described Run-2 result is then combined with the likelihood function from the Run-1 result [20]. The only common parameters in the combination are the fitted  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$  and the external inputs ( $\mathcal{B}(B^+ \rightarrow J/\psi(\mu^+ \mu^-)K^+)$  and  $\frac{f_u}{f_{d(s)}}$ ). All remaining nuisance parameters are treated as uncorrelated between the two results.

The ATLAS results, obtained by combining  $25 \text{ fb}^{-1}$  from Run1 and  $26.1 \text{ fb}^{-1}$  from Run2 LHC campaigns, are [19]:

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (2.8^{+0.8}_{-0.7}) \times 10^{-9} \quad , \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &= (-1.9 \pm 1.6) \times 10^{-10} \quad , \end{aligned} \quad (6)$$

with a significance for the  $B_s^0 \rightarrow \mu^+ \mu^-$  signal of 4.6 standard deviations ( $\sigma$ ). The 95% confidence level (CL) upper limit for the  $B^0 \rightarrow \mu^+ \mu^-$  signal is  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 2.1 \times 10^{-10}$ , as obtained with the Neyman procedure described in Ref. [24]. Figure 1 shows the dimuon invariant mass distribution in the highest-score BDT bin (left) and the likelihood contours in the  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  plane (right).



**Figure 1.** (Left): Dimuon invariant mass distributions in data for the bin with the highest scores of BDT output. Superimposed is the result of the maximum-likelihood fit. The total fit is shown as a



continuous line, with the dashed lines corresponding to the observed signal component, the  $b \rightarrow \mu^+ \mu^- X$  background, and the continuum background. The signal components are grouped in one single curve, including both the  $B_s^0 \rightarrow \mu^+ \mu^-$  and the (negative)  $B^0 \rightarrow \mu^+ \mu^-$  component. The curve representing the peaking  $B_{(s)}^0 \rightarrow hh'$  background lies very close to the horizontal axis. **(Right)**: Likelihood contours for the combination of the Run 1 and 2015 to 2016 Run 2 results (shaded areas). The contours are obtained from the combined likelihoods of the two analyses, for values of  $-2\Delta\ln(L)$  equal to 2.3, 6.2, and 11.8. The empty contours represent the result from 2015 to 2016 Run 2 data alone. The SM prediction with uncertainties is also indicated. Figures from Ref. [19].

## 2.2. The $B_s^0 \rightarrow \mu^+ \mu^-$ Effective Lifetime Measurement

Using the same dataset and the same configurations for the event selection and the simulated samples, ATLAS has subsequently performed a measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime  $\tau_{\mu\mu}$  [25]. As explained in Section 1, the measurement of this quantity is complementary to the branching fraction measurement in the searches for NP phenomena. The only difference between the two analyses lies in the different selection applied to the c-BDT output. A requirement for the c-BDT output to be larger than 0.365 is applied to the dataset, while all other requirements are the same as the BR analysis. The value of this requirement was selected after an optimization procedure based on the maximization of the  $S/\sqrt{S+B}$  figure-of-merit.

The  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime is measured using a binned  $\chi^2$  fit to the proper decay time distribution of the  $B_s^0 \rightarrow \mu^+ \mu^-$  signal component after the subtraction of the background. The proper decay time  $\tilde{t}_{\mu^+ \mu^-}$  is defined as  $\tilde{t}_{\mu^+ \mu^-} = \frac{L_{xy} m_{B_s^0}^{\text{PDG}}}{p_T^{B_s^0}}$ , where  $L_{xy}$  is the decay length projected along the reconstructed  $B_s^0$  momentum in the transverse plane,  $m_{B_s^0}^{\text{PDG}}$  is the world averaged mass of  $B_s^0$  mesons from Ref. [26], and  $p_T^{B_s^0}$  is the magnitude of the candidate's reconstructed transverse momentum. To extract  $\tau_{\mu\mu}$ , three main steps have been completed:

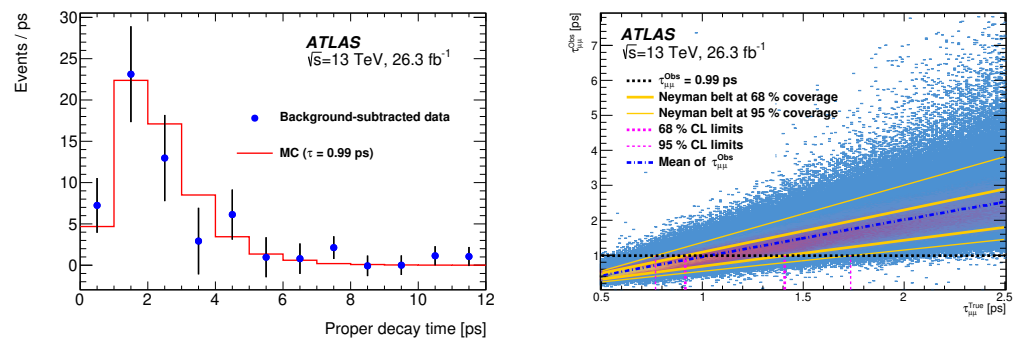
- A fit to the dimuon invariant mass, in the same range as for the BR analysis
- The extraction of the  $\tilde{t}_{\mu^+ \mu^-}$  distribution of the  $B_s^0 \rightarrow \mu^+ \mu^-$  component using the *sPlot* technique [27]
- A binned  $\chi^2$  fit to  $\tilde{t}_{\mu^+ \mu^-}$  distribution comparing Monte-Carlo simulated effective lifetime templates corresponding to different values of  $\tau_{\mu\mu}$ .

In the first step, the dimuon invariant mass distribution, after all selection cuts described above, is fit using a five parameters model made of three Probability Density Functions (PDF): a double Gaussian to describe the  $B_s^0 \rightarrow \mu^+ \mu^-$  component, a linear function to describe the combinatorial (or continuum) background component, and an exponential function to describe the PRD component. Additional resonant and non-resonant backgrounds (such as  $B \rightarrow hh'$ ,  $B_c^\pm$  and semileptonic  $B$  decays), as well as the  $B^0 \rightarrow \mu^+ \mu^-$  component, are neglected in this fit and considered as sources of systematic uncertainties whose impact on  $\tau_{\mu\mu}$  is evaluated through MC pseudo-experiments (as described later in the text). The fit yields  $58 \pm 13$  events in the  $B_s^0 \rightarrow \mu^+ \mu^-$  mass window.

The second step exploits the *sPlot* statistical technique to extract the  $B_s^0 \rightarrow \mu^+ \mu^-$  signal proper-decay time component from the invariant mass fit. The signal proper-decay time distribution is background-subtracted by means of per-event weights computed using the result of the invariant mass fit described above.

The third and final step consists of a binned- $\chi^2$  fit to the proper-decay time distribution extracted in the previous step. This distribution is considered in the interval of 0 to 12 ps and divided in twelve equal width bins. Pure signal proper-decay time simulated templates in the same interval and binning scheme corresponding to different values of  $\tau_{\mu\mu}$  are generated, and a  $\chi^2$ -binned fit is performed with respect to background-subtracted data. The  $\chi^2$  calculation takes both the statistical uncertainty on the weight-corrected MC and

the Poissonian uncertainty in each data bin as expected from the predicted MC content for that bin into account. The template minimizing the  $\chi^2$  corresponds to an observed lifetime  $\tau_{\mu\mu}^{\text{Obs}}$  of 1.07 ps. MC pseudo-experiments studies, generated for a lifetime of 1.624 ps (i.e., the SM predicted value), showed that the lifetime extraction procedure had a bias of 82 fs due to the low-statistics regime of the fit. This bias is found to be constant in the  $B_s^0$  lifetime range considered in the analysis. Therefore, the quoted value for  $\tau_{\mu\mu}^{\text{Obs}}$  has been corrected for this effect. The statistical uncertainty on  $\tau_{\mu\mu}^{\text{Obs}}$  is instead extracted using an MC pseudo-experiments-based Neyman construction, yielding to a value of  $\tau_{\mu\mu}^{\text{Obs}} = 0.99^{+0.42}_{-0.07}(\text{stat.})$  ps. Figure 2 shows the signal proper decay time distribution extracted from data superimposed with the MC template minimizing the  $\chi^2$  distribution (left) and the MC pseudo-experiments-based Neyman construction used to estimate the statistical uncertainty of the measurement (right).



**Figure 2.** (Left): Signal proper decay time distribution extracted with the *sPlot* background subtraction procedure applied to the  $B_s^0 \rightarrow \mu^+ \mu^-$  invariant mass fit. The superimposed signal MC template is the result of the lifetime fit procedure discussed in the text. The uncertainties on the data points are calculated as Poisson fluctuations on the MC yield prediction (continuous red histogram) in the corresponding bin. (Right): 68% and 95% CL bands obtained with a Neyman construction based on MC pseudo-experiments for the signal and background components. The yellow lines interpolate the band boundaries in order to smooth the effects of the limited number of MC pseudo-experiments used. The dashed-dotted blue line corresponds to the average expected  $\tau_{\mu\mu}^{\text{Obs}}$  value at a given  $\tau_{\mu\mu}$  value used to generate MC pseudo-experiments. The horizontal dashed black line corresponds to the experimentally observed value of  $\tau_{\mu\mu}^{\text{Obs}} = 0.99$  ps, yielding a 68% CL band of [0.92, 1.41] ps (thick vertical dashed purple lines) and a 95% CL band of [0.77, 1.73] ps (thin vertical dashed purple lines). The same construction at the  $\tau_{\mu\mu}^{\text{Obs}}$  corresponding to  $\tau_{\mu\mu} = 1.624$  ps (the SM prediction) yields [1.44, 2.26] ps as 68% CL band. Figures from Ref. [25].

The dominant systematic uncertainties for this measurement are related to the data-MC discrepancies (134 fs evaluated in data by repeating, under the same statistical regime as the  $B_s^0 \rightarrow \mu^+ \mu^-$  signal case, the same fit procedure in the  $B^\pm \rightarrow J/\psi K^\pm$  channel), to the background mass and lifetime models (86 fs), to the fit dependence from the lifetime used in MC pseudo-experiments generation and the  $B_s^0$  eigenstates admixture (15 fs evaluated by generating MC pseudo-experiments in the  $\tau_L^{\text{SM}} - \tau_H^{\text{SM}}$  lifetime interval) and the neglected resonant and non-resonant backgrounds (12 fs). The total systematic uncertainty is then obtained by summing in quadrature and symmetrising the impact on  $\tau_{\mu\mu}$  of all single sources. This yields to an observed value of  $\tau_{\mu\mu}^{\text{Obs}}$  of  $0.99^{+0.42}_{-0.07}(\text{stat.}) \pm 0.17(\text{syst.})$ . The value is compatible with the SM prediction of 1.624 ps ( $\mathcal{A}_{\Delta\Gamma} = 1$ ) as well as with the other experimental results described in this article.

### 3. Measurement of $B_s^0 \rightarrow \mu^+ \mu^-$ Decay Properties and Search for $B^0 \rightarrow \mu^+ \mu^-$ Decay at CMS

The latest analysis by the CMS collaboration is based on the LHC Run-2 data collected from 2016 to 2018 at a center-of-mass energy of 13 TeV, corresponding to an integrated



luminosity of  $140 \text{ fb}^{-1}$  [28]. The studies based on LHC Run-1 samples collected from 2011 to 2012 can be found in the earlier publications [29]. There is no attempt to combine the latest publication with the results from the 2011 to 2012 data as the expected gain in sensitivity is modest. In this section, the latest CMS measurement of  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ , the search for  $B^0 \rightarrow \mu^+ \mu^-$  decay, and the effective lifetime measurement using  $B_s^0 \rightarrow \mu^+ \mu^-$  events are discussed.

### 3.1. Measurement of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and Search for $B^0 \rightarrow \mu^+ \mu^-$ Decay

The characteristic signal  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  comprises two muons originating from a single displaced vertex, isolated from other activities, with momentum aligned with the flight direction, and an invariant mass peaking at  $M(B_s^0)$  or  $M(B^0)$ . The primary contributors to the background comprise combinatorial events, involving instances where the two muons originate from different heavy quarks, partially reconstructed semileptonic decays wherein both muons emanate from the same  $B$  meson (with one of the muons from a misidentified charged hadron), and the background arising from peaking charmless two-body hadronic  $B$  meson decays.

The data events were collected with a set of dimuon triggers for this study: the L1 trigger required two oppositely charged muons within the range of  $|\eta| < 1.5$ , while at HLT the dimuons should form a secondary vertex and are required to be within specific mass ranges. The dimuon candidates are used to reconstruct  $B$  mesons for the signal and normalization  $B^+ \rightarrow J/\psi K^+$  and  $B_s^0 \rightarrow J/\psi \phi$  channels. The selections are reserved to be as similar as possible for partial cancellation of systematic effects. Muons at offline analysis are required to have a high-quality track fit at tracker, a transverse momentum at least 4 GeV and  $|\eta| < 1.4$ . To suppress misidentified muons from charged pion and kaon decays, a multivariate-analysis(MVA)-based algorithm has been introduced. Extra kaons are required in the reconstruction for the normalization channels. A trajectory representing the  $B$  candidate is built from the decay vertex and  $B$  candidate's momentum, and is extrapolated to the closest point for each reconstructed primary vertex; the primary vertex with the smallest distance to the extrapolated point is selected for the analysis.

How to reduce the combinatorial and partially reconstructed backgrounds are the main challenges for this study. To enhance the analysis sensitivity, a dedicated MVA discriminator, combining various discriminating observables into a single score distribution ( $d_{\text{MVA}}$ ) using a boosted decision tree algorithm, is introduced. The inputs for  $d_{\text{MVA}}$  include pointing angles, defined as the angles between the  $B$  momentum and the direction connecting the primary and secondary vertices, observables related to the secondary vertex such as quality of the vertex finding, and observables that are designed to identify nearby decay products in semi-leptonic decays of  $b$  and  $c$  hadrons. The  $d_{\text{MVA}}$  training is employed by an advanced gradient boosting algorithm, supported by the XGBoost library [30]. The training utilizes a mix of  $B_s^0 \rightarrow \mu^+ \mu^-$  signal events and background events selected from the data sidebands. Following a fine-tuning of input observables to align the kinematics of  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B^+ \rightarrow J/\psi K^+$  decays (considering variations in the uncertainties of the dimuon vertex position), the control decay  $B^+ \rightarrow J/\psi K^+$  channel is employed to evaluate the performance of the  $d_{\text{MVA}}$  in data.

Charmless two-body decays  $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ , like  $B^0 \rightarrow K^+ \pi^-$  and  $B_s^0 \rightarrow K^+ K^-$ , can mimic the signal when both charged hadrons are misidentified as muons. The misidentification probabilities in data are calculated by utilizing  $K_S^0 \rightarrow \pi^+ \pi^-$ ,  $\phi(1020) \rightarrow K^+ K^-$ , and  $\Lambda \rightarrow p \pi^-$  decays, restricting the decay distance of  $K_S^0$  and  $\Lambda$  candidates to align with the lifetime of the  $B$  meson. Misidentification of pions and kaons primarily originates from their decays into muons. An agreement between the observed data and simulations is observed for both pions and kaons. The proton misidentifying rate is much smaller and hence the contributions from the associated processes are totally negligible. After a stringent multivariate-based muon identification requirement, the charmless two-body backgrounds reduce to a negligible level.

Because of the limited precision in measuring the b-quark production cross section at the LHC, directly determining the branching fraction ( $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ ) could introduce significant uncertainty. As a common practice, the signal branching fraction is assessed by normalizing it to the  $B^+ \rightarrow J/\psi K^+$  decay channel. In addition to the  $B_s^0 \rightarrow J/\psi \phi$  decays,  $J/\psi \rightarrow \mu\mu$  and  $\phi \rightarrow KK$  are considered as a cross-check, and might become more precise if the  $\mathcal{B}(B_s^0 \rightarrow J/\psi \phi(1020))$  is further improved by future B-factory studies. Another advantage of measuring branching fractions in a relative manner is the potential cancellation of systematic uncertainties common in the selection of the signal and normalization channels. The exact formulae for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  branching fractions are similar to those used in Equation (5):

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \mathcal{B}(B^+ \rightarrow J/\psi K^+) \frac{N_{B_s^0 \rightarrow \mu^+ \mu^-} \epsilon_{B^+ \rightarrow J/\psi K^+} f_u}{N_{B^+ \rightarrow J/\psi K^+} \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-} f_s}, \quad (7)$$

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \mathcal{B}(B_s^0 \rightarrow J/\psi \phi) \frac{N_{B_s^0 \rightarrow \mu^+ \mu^-} \epsilon_{B^+ \rightarrow J/\psi \phi}}{N_{B_s^0 \rightarrow J/\psi \phi} \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}}, \quad (8)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = \mathcal{B}(B^+ \rightarrow J/\psi K^+) \frac{N_{B^0 \rightarrow \mu^+ \mu^-} \epsilon_{B^+ \rightarrow J/\psi K^+} f_u}{N_{B^+ \rightarrow J/\psi K^+} \epsilon_{B^0 \rightarrow \mu^+ \mu^-} f_d}, \quad (9)$$

where the yields and the selection efficiencies for each process are denoted by  $N_X$  and  $\epsilon_X$  ( $X = B_s^0 \rightarrow \mu^+ \mu^-$ ,  $B^0 \rightarrow \mu^+ \mu^-$ ,  $B^+ \rightarrow J/\psi K^+$ , or  $B_s^0 \rightarrow J/\psi \phi$ ). The production fractions for the  $B^+$ ,  $B^0$ , and  $B_s^0$  mesons are represented by  $f_u$ ,  $f_d$ , and  $f_s$ . The ratio  $\frac{f_u}{f_d}$  is set to unity due to isospin symmetry, while the ratio  $\frac{f_u}{f_s}$ , together with  $\mathcal{B}(B^+ \rightarrow J/\psi K^+)$  and  $\mathcal{B}(B_s^0 \rightarrow J/\psi \phi)$ , are external inputs.

The results are obtained with simultaneous unbinned maximum likelihood fits in multiple categories. For the measurement of branching fractions, a two-dimensional fit using the dimuon invariant mass and its uncertainty as observables is introduced. The events are categorized according to a data-taking period, signal purity based on  $d_{MVA}$ , and  $|\eta|$  of the most-forward muon. The likelihood function includes five components:  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B^0 \rightarrow \mu^+ \mu^-$  signals, semileptonic background, peaking two-body decays, and the combinatorial events. The signal components are represented using Crystal Ball functions for the dimuon mass. The width of these Crystal Ball functions is parameterized based on the per-event mass resolution. To model the mass resolution, a kernel estimation approach is employed, utilizing Gaussian kernels. The semileptonic background is modeled by a Gaussian with free parameters in the fit to the data, while the peaking background is modeled by a sum of Gaussian and Crystal Ball functions with the shape parameters determined from simulated events. The yields of the semileptonic and peaking background components are first derived and then included in the fit with uncertainties from the hadron to muon misidentifying rate as constrained nuisance parameters. The combinatorial background is modeled by a linear function with yields and slope free to vary in the fits.

For the branching fraction measurements, the experimental uncertainties include signal efficiency corrections due to mismodeling of  $d_{MVA}$ , the charged kaon efficiency in the normalization channels, trigger efficiencies, and fitting bias, while the rest of the uncertainties are smaller than 1%. The mismodeling of the  $d_{MVA}$  distribution has been investigated through two distinct studies with  $B^+ \rightarrow J/\psi K^+$  events. In the first study, a direct comparison is conducted between background-subtracted data, with the *sPlot* technique [27] on the  $B^+ \rightarrow J/\psi K^+$  invariant mass distribution, and the simulated distributions. The second study involves reweighting of the simulated samples to align with the data distributions, employing the XGBoost tool. The disparity between the two studies is quantified as a systematic uncertainty. The systematic uncertainty arising from the selection of background models is derived through pseudo-experiments, incorporating variations in the fit. The uncertainties in the input branching fractions of the normalization channels and the  $\frac{f_u}{f_{d(s)}}$  ratio are implemented as external uncertainties.

The resulting branching fractions for  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B^0 \rightarrow \mu^+\mu^-$  are as follows:

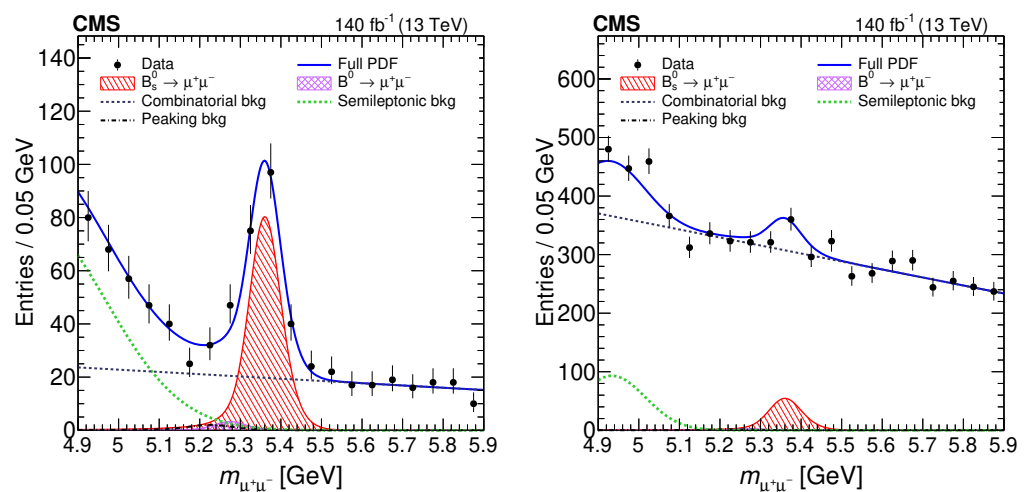
$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = \left[ 3.83_{-0.36}^{+0.38} (\text{stat})_{-0.16}^{+0.19} (\text{syst})_{-0.13}^{+0.14} (f_s/f_u) \right] \times 10^{-9}, \quad (10)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = \left[ 0.37_{-0.67}^{+0.75} (\text{stat})_{-0.09}^{+0.08} (\text{syst}) \right] \times 10^{-10}. \quad (11)$$

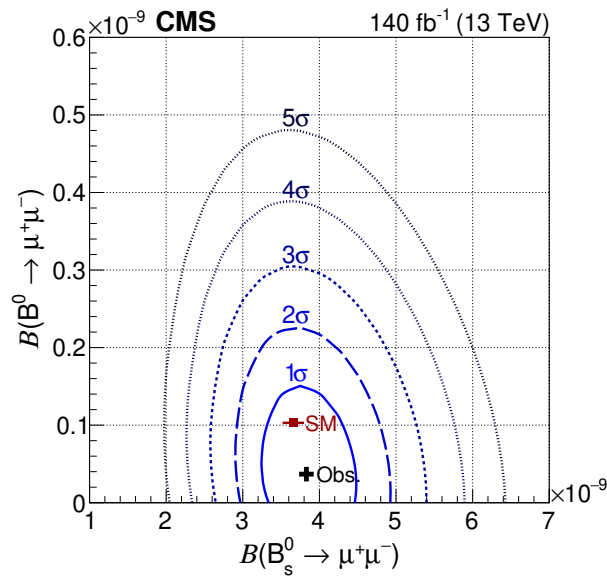
The results incorporate external inputs, specifically  $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.020 \pm 0.019) \times 10^{-3}$ ,  $\mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) = (5.961 \pm 0.033) \times 10^{-2}$ , and  $f_s/f_u = 0.231 \pm 0.008$ . The input  $f_s/f_u$  value is derived from the  $p_T$ -dependent measurement by LHCb [31] and the  $p_T$  distribution observed in this measurement. Figure 3 shows the dimuon invariant mass distributions from the categories with different signal purity whereby the results of the fit are superimposed. The profile likelihood contours enclosing the regions with different coverage are shown in Figure 4. Alternatively, the  $B_s^0 \rightarrow \mu^+\mu^-$  branching fraction is measured using the  $B_s^0 \rightarrow J/\psi\phi$  decays as the normalization, which leads to

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = \left[ 4.02_{-0.38}^{+0.40} (\text{stat})_{-0.23}^{+0.28} (\text{syst})_{-0.15}^{+0.18} (\mathcal{B}) \right] \times 10^{-9}, \quad (12)$$

where the last uncertainty arises from the uncertainty in the  $B_s^0 \rightarrow J/\psi\phi$  branching fraction ( $\mathcal{B}(B_s^0 \rightarrow J/\psi\phi) = (1.04 \pm 0.040) \times 10^{-3}$ ). The lifetime of the  $B_s^0$  meson has a significant impact on the  $B_s^0 \rightarrow \mu^+\mu^-$  branching fraction too; a scaling factor is provided on the resulting branching fraction ( $1.577 - 0.358 \cdot \tau_{B_s^0}$ , where  $\tau_{B_s^0}$  is  $B_s^0$  lifetime in ps) for alternative lifetime hypotheses other than the SM value (1.61 ps). The upper limit on the  $B^0 \rightarrow \mu^+\mu^-$  decay is calculated to be  $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 1.9 \times 10^{-10}$  at 95% confidence level, using the  $\text{CL}_s$  method [32].



**Figure 3.** The dimuon invariant mass distributions for the candidates with  $d_{\text{MVA}} > 0.99$  (left) and  $0.99 > d_{\text{MVA}} > 0.90$  (right) categories. The solid blue curves are the projections of fit model, while the individual components of the fit are also presented. Figures from Ref. [28].



**Figure 4.** The profile likelihood in  $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$  versus  $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$  2D plane, while the contours enclose the regions with  $1\sigma$ – $5\sigma$  coverage. Figure from Ref. [28].

### 3.2. Measurement of $B_s^0 \rightarrow \mu^+\mu^-$ Effective Lifetime

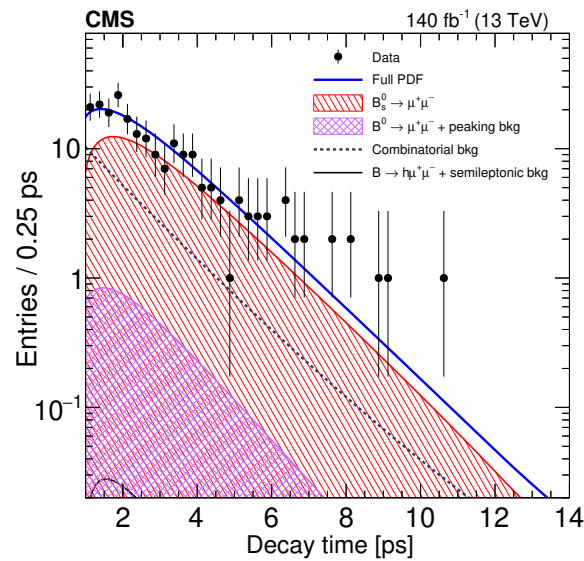
The  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime  $\tau_{\mu\mu}$  is extracted with an unbinned maximum likelihood in three-dimensions including dimuon invariant mass, decay time, and decay time uncertainty. The decay time  $t_{\mu^+\mu^-}$ , which is calculated for each event, is defined by the product of the flight length and the invariant mass of the  $B$  candidate, divided by the magnitude of the  $B$  candidate momentum. The events are also categorized in the data-taking period, purity based on  $d_{MVA}$ , and the pseudorapidity of the most forward muon. The dimuon invariant mass distribution is modeled with the same functions introduced for the branching fraction measurements, while the decay time distribution for signal events is modeled by an exponential function convoluted with the decay time resolution function. The decay time resolution function is parameterized with the measured decay time uncertainty. The acceptance as a function of the decay time is obtained from simulated events and corrected with the  $B^+ \rightarrow J/\psi K^+$  events from data. The decay time distribution for combinatorial background decays is obtained from high-mass sideband events. The decay time uncertainty models used in the fit are obtained from simulation samples and mass sideband data as well.

The systematic uncertainties in the lifetime measurement are mostly driven by the correlations between the  $d_{MVA}$  and the decay time, as the key input variables for the  $d_{MVA}$  classifier: the pointing angle of the  $B$  candidate and its associated uncertainty are strongly correlated with the decay time observable. Any mismodeling in the simulation results in significant impacts on the decay time distribution. A correction as a ratio of the decay time distributions for different  $d_{MVA}$  requirements is derived from  $B^+ \rightarrow J/\psi K^+$  events. This method introduced a bias up to 0.1 ps for the data recorded in 2016, and reduced in the later data sets. The possible bias arising in fitting and modeling is also tested with  $B^+ \rightarrow J/\psi K^+$  events, but with a relaxed selection criterion. Other systematic uncertainties are minor, estimated to be smaller than 0.01 ps.

The resulting effective lifetime for  $B_s^0 \rightarrow \mu^+\mu^-$  events is as follows:

$$\tau_{\mu\mu} = 1.83_{-0.20}^{+0.23} (\text{stat})_{-0.04}^{+0.04} (\text{syst}) \text{ ps}, \quad (13)$$

which is consistent with the SM prediction and the other experimental results described in this article. The decay time distribution for the candidates in the region of  $5.28 < m_{\mu^+\mu^-} < 5.48$  GeV is shown in Figure 5.



**Figure 5.** The proper decay time distribution for the candidates in the region  $5.28 < m_{\mu^+\mu^-} < 5.48$  GeV, with the result of the fit superimposed. The solid blue curve is the sum of all fit component, while the shaded areas are the background components. Figure from Ref. [28].

#### 4. Analysis of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ Decays with LHCb

The most recent analysis of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  with the LHCb experiment [33,34] was performed with the full  $pp$ -collision data collected in the LHC Run 1 and Run 2 campaigns. The total integrated luminosity corresponds to  $1 \text{ fb}^{-1}$  at  $\sqrt{s} = 7$  TeV,  $2 \text{ fb}^{-1}$  at  $\sqrt{s} = 8$  TeV, and  $6 \text{ fb}^{-1}$  at  $\sqrt{s} = 13$  TeV.

In total, the analysis comprises the search and branching fraction measurements of the decays  $B_s^0 \rightarrow \mu^+\mu^-$ ,  $B^0 \rightarrow \mu^+\mu^-$ , and  $B_s^0 \rightarrow \mu^+\mu^-\gamma$  with initial state radiation ( $B_s^0 \rightarrow \mu^+\mu^-\gamma$  was only investigated in the region  $m(\mu^+\mu^-) > 4.9 \text{ GeV}/c^2$ ), as well as the measurement of the effective lifetime of the  $B_s^0 \rightarrow \mu^+\mu^-$  decay. A precise branching fraction measurement is achieved by normalizing the signal decay with two high-statistics decay modes,  $B^0 \rightarrow K^+\pi^-$  and  $B^+ \rightarrow J/\psi K^+$  with  $J/\psi \rightarrow \mu^+\mu^-$ , similarly to what is performed by the ATLAS and CMS collaborations and shown in Equation (5) for the  $B^+ \rightarrow J/\psi K^+$  channel. The decay modes  $B^0 \rightarrow K^+\pi^-$  and  $B_s^0 \rightarrow K^+K^-$  are used as control modes for the effective lifetime measurement as well.

Dominant background processes mimicking the signal, on the one hand, arise from random combinations of two muons from two different  $b$ -hadron decays in the same event. On the other hand, they can come from  $b$ -hadron decays where one or more final state particles have been wrongly identified as a muon. Furthermore,  $b$ -hadron decays where part of the decay products have not been reconstructed can constitute a background. The selection of the signal decays largely inherits from previous analyses of a subset of the data [35] and targets particularly the selecting of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays over aforementioned backgrounds, whereas the measurement of  $B_s^0 \rightarrow \mu^+\mu^-\gamma$  is a byproduct of the analysis. The LHCb detector, as used to collect the above mentioned data, employed a two-staged online selection. Firstly, events are selected by a hardware trigger that requires at least one muon with a high transverse momentum. Secondly, a two-staged software trigger is applied, which performs a full event reconstruction. In the software trigger, events fulfilling minimum requirements on the muon momentum and its impact parameter, are kept. Also, events are kept where these requirements are met by non-signal candidates to maximize signal efficiency.

In the offline selection, candidate  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays are selected by combining two well-reconstructed oppositely charged particles identified as muons [36] with a transverse momentum in the range of  $0.25 < p_T < 40 \text{ GeV}/c$ . The common vertex is required to



have a good vertex fit quality and be clearly separated from the associated  $pp$ -collision vertex. The resulting  $B_{(s)}^0$  candidate is required to have a transverse momentum greater than  $0.5 \text{ GeV}/c$ . Candidates in the full instrumented pseudorapidity region  $2 < \eta < 5$  are retained for analysis. A preliminary selection based on a Boosted Decision Tree (BDT) is applied to remove a large fraction of combinatorial background while maintaining a high signal efficiency. The BDT is trained with variables related to the decay topology of two particles originating from a vertex displaced with respect to the primary vertex. A highly efficient veto on the combination of a signal muon with another particle in the event identified as a muon that results in a dimuon mass close to the  $J/\psi$  mass allows for effectively removing  $B_c^+ \rightarrow J/\psi \mu^+ \nu$  decays. A selection on a combination of particle identification algorithms is performed and tuned to maximize the  $B^0 \rightarrow \mu^+ \mu^-$  significance ( $h = K, \pi$ ), suppressing  $B_{(s)}^0 \rightarrow h^+ h^-$  and  $\Lambda_b^0 \rightarrow p \mu^- \nu$  decays. The final selection is performed on a second BDT, called in the following s-BDT. This s-BDT includes, apart from variables related to the decay topology, notably isolation classifiers—specifically developed for this analysis—that inspect the closeness of the signal muon tracks to other tracks in the event that are either reconstructed in all tracking detector stations or only in the detector closest to the collision region. The  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  yields are measured by fitting the dimuon invariant mass distribution in bins of this final selection s-BDT, discarding only the lowest bin (that corresponds to about 25% of the signal) in order to maximize the signal sensitivity. The samples of  $B^0 \rightarrow K^+ \pi^-$  and  $B^+ \rightarrow J/\psi K^+$  are selected in a similar way except for trigger and particle identification criteria for the  $B^0 \rightarrow K^+ \pi^-$  mode and removing the  $J/\psi$  veto. For  $B^0 \rightarrow K^+ \pi^-$ , the muon identification criteria are replaced by hadron identification and a trigger selection independent of the candidate is required to achieve an unbiased selection.

#### 4.1. Measurement of the Branching Fractions of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^- \gamma$

In order to achieve unbiased branching fraction estimates, efficiencies are calculated either on corrected simulation or directly on data. Importantly, the fractions of the s-BDT bins are determined from  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  simulation, where the  $B_{(s)}^0$  quantities and the number of tracks in the event are reweighted from data-simulation comparisons in high-statistics  $B^+ \rightarrow J/\psi K^+$  and  $B_s^0 \rightarrow J/\psi \phi$  samples. The resulting corrected s-BDT fractions are then independently cross-checked with  $B^0 \rightarrow K^+ \pi^-$  data samples, corrected by the different trigger and particle identification response. Measuring the branching fractions relative to two modes,  $B^0 \rightarrow K^+ \pi^-$  and  $B^+ \rightarrow J/\psi K^+$ , allows for a stringent cross check of the efficiencies by calculating the ratio between the estimated branching fractions of the two and comparing it to the ratio of the published branching fractions [37]. An excellent agreement is found.

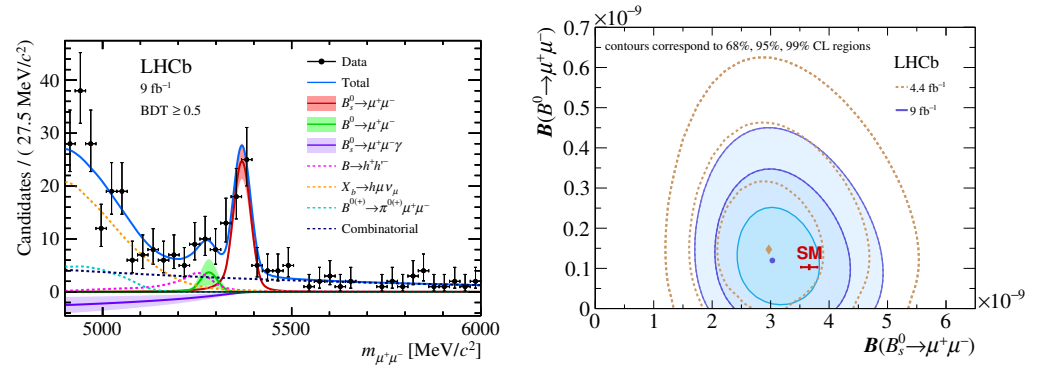
The invariant mass shape of signal  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays is described with two-sided Crystal Ball functions [38], where the mean of the Gaussian core is calibrated from  $B_s^0 \rightarrow K^+ K^-$  and  $B^0 \rightarrow K^+ \pi^-$  data samples. The mass resolution of about  $22 \text{ MeV}/c^2$  is determined from the interpolation of the measured resolutions of charmonium and bottomonium resonances. The tail parameters are estimated from simulation. Small differences in the resolution and the tail parameters are found to appear across the s-BDT bins and are accounted for in the final fit.

Exclusive background decays remaining in the fully selected samples have been carefully studied with a simulation and calibrated in data. A large focus in the most recent analysis is laid on the correct estimation of the misidentification of charged hadrons as muons. Decays of the form  $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$  ( $h = K, \pi$ ) with both charged hadrons misidentified create a peaking structure very close to the  $B^0 \rightarrow \mu^+ \mu^-$  peak and therefore form the most relevant remaining background component. Misidentification occurs in the detector dominantly because the hadrons decay in-flight into muons. The hadron misidentification rate is estimated with a dedicated procedure using  $D^0 \rightarrow K^- \pi^+$  from  $D^{*+} \rightarrow D^0 \pi^+$  decays from simulation and data. This procedure takes explicitly into



account that the  $D^0 \rightarrow K^- \pi^+$  invariant mass shape deforms significantly with hadrons decaying in-flight. As an additional cross check, the misidentification rate is investigated from  $B^0 \rightarrow K^+ \pi^-$  data samples by determining the  $B^0 \rightarrow K^+ \pi^-$  yield in  $\pi \mu$ ,  $K \mu$ , and  $\pi K$  mass distributions.

A summary of the final mass fit to obtain the signal branching fractions is displayed in Figure 6.



**Figure 6.** (Left) mass distribution of the selected  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  candidates (black dots) with s-BDT  $> 0.5$ . The result of the fit is overlaid and the different components are detailed:  $B_s^0 \rightarrow \mu^+ \mu^-$  (red solid line),  $B^0 \rightarrow \mu^+ \mu^-$  (green solid line),  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  (violet solid line), combinatorial background (blue dashed line),  $B_{(s)}^0 \rightarrow h^+ h^-$  (magenta dashed line),  $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ ,  $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ ,  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  and  $\Lambda_b^0 \rightarrow p \mu^- \nu_\mu$  (orange dashed line), and  $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$  (cyan dashed line). The solid bands around the signal shapes represent the variations of the branching fractions by their total uncertainty. (Right) two-dimensional profile likelihood of the branching fractions for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays. The measured central values of the branching fractions are indicated with a blue dot. The profile likelihood contours for 68%, 95%, and 99% CL regions of the result are shown as blue contours, while the brown contours indicate the previous measurement [35] and the red cross shows the SM prediction. Figures from Ref. [34].

A precise measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction is obtained as

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left( 3.09_{-0.43}^{+0.46+0.15} \right) \times 10^{-9}, \quad (14)$$

where the first uncertainties are of statistical nature and the second uncertainties are of systematic nature. The systematic uncertainties are dominated by the knowledge of the ratio of fragmentation fractions  $\frac{f_u}{f_{d(s)}}$  of  $B_s^0$  and  $B^0$  mesons which enters the normalization equation because the decay is measured relative to  $B^0$  and  $B^+$  decays. The  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  decays are not observed and consequently upper limits on their branching fractions are set to

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.2 \times 10^{-10}, \quad \text{and} \quad (15)$$

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^- \gamma) < 2.0 \times 10^{-9} \quad (16)$$

at 95% CL, respectively. Similarly, an upper limit on the branching fraction ratio  $\mathcal{R}$  was determined at 95% CL to

$$\mathcal{R} < 0.095 \quad (17)$$

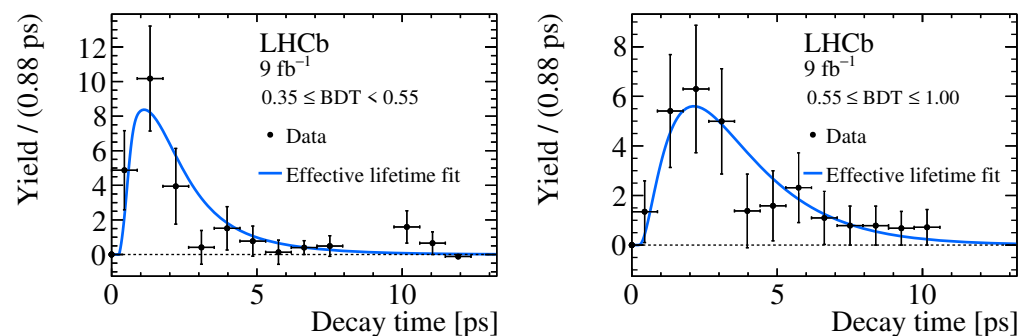
These values include systematic uncertainties, which are dominated by the knowledge of the background components that include misidentified hadrons. A correlation of 11% is observed between the measurement of the  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^-$  components.

#### 4.2. Measurement of the Effective Lifetime of the $B_s^0 \rightarrow \mu^+ \mu^-$ Decay

The effective lifetime of the  $B_s^0 \rightarrow \mu^+ \mu^-$  decay has been measured on the same sample with a slightly different selection. Since there is effectively no background from hadron-muon misidentification in the  $B_s^0 \rightarrow \mu^+ \mu^-$  mass peak region, the dimuon mass window is adapted to exclude these backgrounds and the particle identification requirements are loosened to increase the signal yield. The conditions of triggered events are required to be met either from the signal candidate itself or the remainder of the event, which facilitates the modeling of the acceptance. Furthermore, the data are analyzed in only two bins of the final selection s-BDT, chosen to maximize the sensitivity to the effective lifetime. The mass distributions in each s-BDT region are fitted independently to extract background-subtracted decay time distributions with the *sPlot* technique [27]. A simultaneous fit to the two background-subtracted decay-time distributions as shown in Figure 7 is employed to extract the effective lifetime. In order to extract an unbiased lifetime measurement, the acceptance effects of the reconstruction selection requirements have to be modeled. The decay time acceptance is modeled by fitting parametric functions to the efficiency distribution in simulation, where the simulation has been weighted to improve data simulation differences. The procedure is validated by measuring the lifetimes of  $B_s^0 \rightarrow K^+ K^-$  and  $B^0 \rightarrow K^+ \pi^-$  in the data, finding good agreement with the world average values [37]. The uncertainty of the measurement of the  $B_s^0 \rightarrow K^+ K^-$  lifetime is taken as systematic uncertainty. Further systematic effects like the sample contamination from  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_{(s)}^0 \rightarrow h^+ h^-$  decays, acceptance modeling, uncertainties in the background decay time distributions, and  $B_s^0 \text{-}\bar{B}_s^0$ -production asymmetries are investigated and are found to have only sub-leading to negligible effects. The measured effective lifetime is found to be

$$\tau_{\mu^+ \mu^-} = 2.07 \pm 0.29 \pm 0.03 \text{ ps}, \quad (18)$$

where the first uncertainty is statistical and the second systematic. This value is outside the lifetime interval defined by the  $B_s^0$  light ( $\mathcal{A}_{\Delta\Gamma} = -1$ ) and heavy ( $\mathcal{A}_{\Delta\Gamma} = 1$ ) mass eigenstates, but is consistent with these values at the level of 2.2 and 1.5 standard deviations, respectively.



**Figure 7.** The background-subtracted decay-time distributions with the fit model used to determine the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime superimposed. The distributions in the low and high BDT regions are shown in the left and right plot, respectively. Figures from Ref. [34].

#### 5. Combination of the Measurements by the LHC Experiments

The latest combination of the measurements from LHC experiments is presented in Ref. [39]. It includes the results from ATLAS [19], CMS [29], and LHCb [35]. The combination is performed based on the two-dimensional profile likelihoods obtained by each experiment from the fits to the dimuon events. Such a method allows for properly taking into account the correlations between  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  and  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ . Plus, the upper limit of  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  can be evaluated using the same inputs. Note the previous combination paper [40] based on CMS and LHCb data collected during LHC Run-1 is based on an unbinned maximum likelihood fit simultaneously to the events from both experiments.

ATLAS results, described in Section 2, are extracted from the data samples of  $25 \text{ fb}^{-1}$  collected at the center-of-mass energies  $\sqrt{s} = 7$  and  $8 \text{ TeV}$  and  $26.3 \text{ fb}^{-1}$  at  $\sqrt{s} = 13 \text{ TeV}$ . The resulting branching fractions were obtained:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left[ 2.8_{-0.7}^{+0.8} \right] \times 10^{-9}, \quad (19)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = [-1.9 \pm 1.6] \times 10^{-10}, \quad (20)$$

where the systematic uncertainties are included in the evaluation. The corresponding significance for the  $B_s^0 \rightarrow \mu^+ \mu^-$  signal is  $4.6\sigma$ , while the upper limit for the  $B^0 \rightarrow \mu^+ \mu^-$  branching fraction is  $< 2.1 \times 10^{-10}$  at 95% CL.

CMS analysis is based on the data samples of  $5 \text{ fb}^{-1}$ ,  $20 \text{ fb}^{-1}$ , and  $36 \text{ fb}^{-1}$ , collected at the center-of-mass energies of  $\sqrt{s} = 7, 8,$  and  $13 \text{ TeV}$ , respectively [29]. The resulting branching fractions and effective lifetime for  $B_s^0 \rightarrow \mu^+ \mu^-$  are as follows:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left[ 2.9_{-0.6}^{+0.7} (\text{exp}) \pm 0.2 (\text{frag}) \right] \times 10^{-9}, \quad (21)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = \left[ 0.8_{-1.3}^{+1.4} \right] \times 10^{-10}, \quad (22)$$

$$\tau_{B_s^0 \rightarrow \mu^+ \mu^-} = 1.70_{-0.43}^{+0.60} (\text{stat}) \pm 0.09 (\text{syst}) \text{ ps}. \quad (23)$$

The signals for  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B^0 \rightarrow \mu^+ \mu^-$  yield a significance of  $5.6\sigma$  and  $1.0\sigma$ , respectively. The upper limit for the  $B^0 \rightarrow \mu^+ \mu^-$  branching fraction is evaluated as  $< 3.6 \times 10^{-10}$  at 95% CL. The first uncertainty of the  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  combined statistical and systematic uncertainties is from the analysis, while the second uncertainty is from the uncertainty in the fragmentation ratio  $f_d/f_s$ .

LHCb studies are performed on the data samples of  $1 \text{ fb}^{-1}$ ,  $2 \text{ fb}^{-1}$ , and  $1.4 \text{ fb}^{-1}$ , collected at the center-of-mass energies  $\sqrt{s} = 7, 8,$  and  $13 \text{ TeV}$ , respectively [35]. The analysis yields the following results:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left[ 3.0 \pm 0.6 (\text{stat})_{-0.2}^{+0.3} (\text{syst}) \right] \times 10^{-9}, \quad (24)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = \left[ 1.5_{-1.0}^{+1.2} (\text{stat})_{-0.1}^{+0.2} (\text{syst}) \right] \times 10^{-10}, \quad (25)$$

$$\tau_{B_s^0 \rightarrow \mu^+ \mu^-} = 2.04 \pm 0.44 (\text{stat}) \pm 0.05 (\text{syst}) \text{ ps}, \quad (26)$$

with signal significances of  $7.8\sigma$  and  $1.6\sigma$  for  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B^0 \rightarrow \mu^+ \mu^-$  decays, respectively. An upper limit  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.4 \times 10^{-10}$  at 95% CL is obtained.

For the combination of decay branching fractions, profiled likelihoods are computed in the two-dimensional grid of the  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  and  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  plane from each experiment and the SM  $B_s^0 \rightarrow \mu^+ \mu^-$  lifetime is assumed. As the current measurements are dominated by statistical uncertainties, the systematic uncertainties are treated independently for the three measurements, except for the common nuisance parameter, which is the  $f_d/f_s$  ratio. The  $f_d/f_s$  uncertainty is profiled separately in each likelihood and retained only in the LHCb experiment. To test the impact of this correlation, the  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$  are evaluated with and without the  $f_d/f_s$  uncertainty in ATLAS and CMS likelihoods. The impact is found to be negligible. Additionally, the dependence of  $f_d/f_s$  on the transverse momentum is checked and is found to be consistent within the assigned uncertainties.

The profiled likelihood for each experiment is then modeled with a two-dimensional variable-width Gaussian, which describes asymmetric likelihoods (and asymmetric uncertainties) and also the correlation between the two branching fractions. This analytical function is found to be consistent with the original likelihood for each experiment. The log-likelihoods from the three measurements are summed across the  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ - $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  grid points and then fitted using the variable-width Gaussians. By

maximizing the modeled likelihood function, the combined branching fractions and the associated uncertainties are derived:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = [2.69^{+0.37}_{-0.35}] \times 10^{-9}, \quad (27)$$

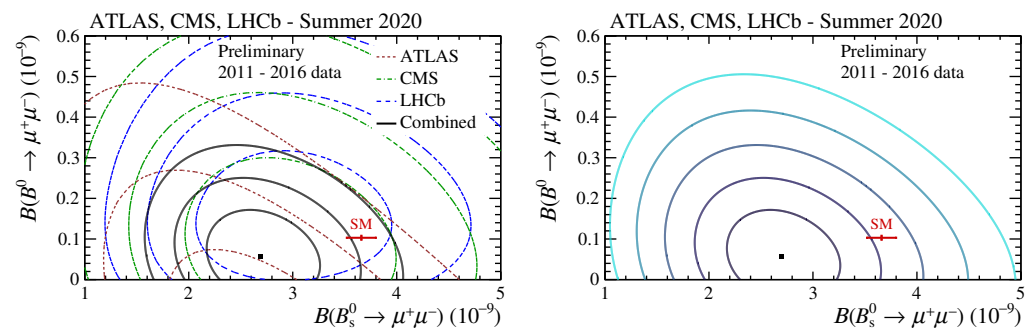
$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = [0.6 \pm 0.7] \times 10^{-10}. \quad (28)$$

The upper limit on  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  is evaluated as  $<1.6$  ( $1.9$ )  $\times 10^{-10}$  at 90% (95%) CL, which is calculated under the positive  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  hypothesis by renormalizing the likelihood in the interested region. The combined  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  branching fraction is found to be lower than any single result, which is due to the strong anti-correlation between two branching fractions. The individual profiled likelihood (left) and the combined likelihood in the  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  -  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  plane (right) are shown in Figure 8.

The compatibility with the SM predictions is estimated to be  $2.4\sigma$  for  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ ,  $0.64\sigma$  for  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ , and  $2.1\sigma$  if computed in the  $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  -  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$  plane. These values are calculated assuming Wilks' theorem and with theoretical uncertainties included. In addition to the individual branching fractions, a combined estimation on the ratio of branching fractions  $\mathcal{R}$  (see Equation (2)) is also derived:

$$\mathcal{R} = \frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)} = 0.021^{+0.030}_{-0.025}, \quad (29)$$

where the corresponding upper limit is evaluated to be  $\mathcal{R} < 0.052$  ( $0.060$ ) at 90% (95%) CL.



**Figure 8.** (Left plot) the two-dimensional likelihood contours for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays from ATLAS (red dashed line), CMS (green dot-dashed line), and LHCb (blue long-dashed line) experiments, together with contours for their combination (continuous line). The likelihood contours are corresponding to the values of  $-2\Delta \ln \mathcal{L} = 2.3, 6.2,$  and  $11.8,$  respectively. (Right plot) the likelihood contours for the combination of the three results, corresponding to the values of  $-2\Delta \ln \mathcal{L} = 2.3, 6.2, 11.8, 19.3,$  and  $30.2,$  or 1 to 5  $\sigma$  levels in a bidimensional Gaussian approximation. Figures from Ref. [39].

The  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime is measured in the last analysis iteration by all three experiments, as reported in Sections 2–4. However, at the time when the combination was performed, only the CMS and LHCb collaborations had a measurement of this quantity. Therefore, a combination has been carried out based only on their results, exploiting a similar method as for the  $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$  combination. The LHCb analysis is carried out with a bin-likelihood fit to the background-subtracted decay time distribution, while the CMS measurement is carried out with a two-dimensional likelihood fit to the decay time and dimuon invariant mass distributions. As the analyses are fully dominated by the statistical uncertainties, the combination is performed by describing CMS and LHCb likelihoods (as a function of effective lifetime  $\tau_{B_s^0 \rightarrow \mu^+ \mu^-}$ ) with variable-width Gaussians, and then, to determine their combined value, the two measurements are assumed to

be uncorrelated. The resulting  $\tau_{B_s^0 \rightarrow \mu^+ \mu^-}$  value and the corresponding uncertainty are as follows:

$$\tau_{B_s^0 \rightarrow \mu^+ \mu^-} = 1.91_{-0.35}^{+0.37} \text{ ps.} \quad (30)$$

Both CMS and LHCb collaborations have recently released updated analyses, as discussed in Sections 3 and 4. Another iteration of the combination is foreseen in the near future, incorporating the results from all three experiments based on the full Run 2 LHC campaign data.

## 6. Conclusions and Prospects

In recent years, the ATLAS, CMS, and LHCb collaborations made a push towards precision measurements of the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction, which resulted in measurements that reach a precision of down to 10% relative uncertainty. These measurements are the most precise to date. At the same time, all three collaborations have begun measuring the effective lifetime of the decay to understand the  $CP$  structure of the decay. Contrary to initial evidence in the first combination of CMS and LHCb measurements [40], the  $B^0 \rightarrow \mu^+ \mu^-$  decay has not been confirmed yet. All results are in good agreement with the SM, strongly constraining potential NP scenarios. To achieve even higher sensitivities, a community effort is ongoing to combine the results of all three experiments. The results of the previous combination have been presented in this review, but have been superseded by the legacy measurements of the CMS and LHCb collaborations. Once the measurement with the full Run 2 data of the ATLAS collaboration is published as well, this combination will be repeated to have the most precise picture possible with the harvest of Run 2 data.

After the LHC Run 2, in 2021 the experiments began to take data again with increased instantaneous luminosity until the end of 2025. After that, the High-Luminosity LHC phase will begin, which will have increased pile-up conditions for all experiments and a massively increased total luminosity. The ATLAS and CMS experiments will strongly upgrade their detectors to cope with the increased pile-up conditions. However, they also target a significant dimuon mass resolution improvement by 20 to 30% (ATLAS) and 40 to 50% (CMS), respectively. The LHCb experiment is planning to follow and go through a major upgrade in 2031 to begin taking data with the LHC Run 5. By the end of the LHC lifetime, ATLAS and CMS aim to have collected  $3000 \text{ fb}^{-1}$ , while LHCb is estimating  $300 \text{ fb}^{-1}$ . Under these conditions and assuming the central values as predicted by the Standard Model, the ATLAS, CMS, and LHCb collaborations made extrapolations to the expected sensitivity of future measurements [41,42]. For the ATLAS experiment, the sensitivity strongly depends on the trigger conditions for dimuon events with the upgraded detector. In the most conservative scenario, the expected statistical-only (statistical and systematic) uncertainties reach 19% (23%) relative to the central value for the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction and 134% (135%) for the  $B^0 \rightarrow \mu^+ \mu^-$  branching fraction, while in the most optimistic scenario they reach 5% (13) for  $B_s^0 \rightarrow \mu^+ \mu^-$  and 25% (26%) for  $B^0 \rightarrow \mu^+ \mu^-$ . The dominant systematic uncertainties in these projections arise from external inputs like the uncertainty of the fragmentation fraction ratio  $f_s/f_d$  and the branching fraction uncertainty of the normalization channel.

The CMS collaboration expects to reach uncertainties of 7% on the branching fraction of  $B_s^0 \rightarrow \mu^+ \mu^-$  and 16% on the branching fraction of  $B^0 \rightarrow \mu^+ \mu^-$ . The expected uncertainty on the effective  $B_s^0 \rightarrow \mu^+ \mu^-$  lifetime is 0.05 ps. This precision will allow for stringent constraints on the parameter  $A_{\Delta\Gamma}^{\mu^+ \mu^-}$  and in particular break the degeneracy between possible scalar and pseudoscalar contributions beyond the SM to this decay.

The LHCb collaboration expects to reach a statistical uncertainty on the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction of 1.8%; however, the analysis will be systematically limited by the external inputs of the fragmentation fraction ratios and the normalization branching fractions, which are estimated to become 4% by then. On the contrary, the ratio  $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)/\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$  is not expected to become systematically limited and is expected



to reach a relative precision of 10%. The measurement of the effective  $B_s^0 \rightarrow \mu^+ \mu^-$  lifetime is expected to reach a precision of 0.033 ps. Both the CMS and LHCb collaborations expect to establish the  $B^0 \rightarrow \mu^+ \mu^-$  decay signal at more than  $5\sigma$  level.

The expected large yield of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays will also allow for accessing the  $CP$  parameter  $S_{\mu^+ \mu^-}$ , which describes the time-dependent  $CP$ -violation in the decay [43]. Adding this parameter will complete the base of  $CP$  observables and provide complementary constraints to physics beyond the SM that are not constrained by the other observables. A nonzero value of this parameter will be an immediate sign for a  $CP$ -violating phase beyond the SM. This parameter can only be determined by measuring the decay-time distribution of  $B_s^0$  and  $\bar{B}_s^0$  decays separately and thus requires the tagging of the  $B_s^0$  flavor. Assuming a similar performance of the flavor tagging as in Run 2, the LHCb collaboration expects to reach a precision of 0.2 for this parameter. Provided a sufficient flavor tagging performance can be achieved, this analysis could potentially be performed by the CMS and ATLAS experiments.

To achieve the projected sensitivities discussed in this section and possibly surpass them, it will be important to maintain the basic assumptions. For the ATLAS and CMS experiments, it will be crucial to design trigger strategies that allow for keeping the muon transverse momentum thresholds as low as possible in the high pile-up environment. Furthermore, the level of backgrounds from random combinations must be maintained or decreased, which might be achieved through the tracking detectors, the fast timing information in the reconstruction, and the improvement of current selection algorithms based on Machine Learning tools. Fast timing information to disentangle  $pp$ -collision points will also facilitate the analysis of LHCb data and enable the flavor tagging of the  $B_s^0$  mesons. Further improvements over the projected sensitivities in this section—especially on  $B^0 \rightarrow \mu^+ \mu^-$  measurements—might be achieved by improvements for the muon identification and the momentum resolution, which will have a significant impact on the dimuon mass resolution.

**Author Contributions:** U.D.S. summarised the ATLAS analysis, K.-F.C. the CMS analysis, T.M. the LHCb analysis, writing—original draft preparation, K.-F.C., T.M. and U.D.S.; writing—review and editing, K.-F.C., T.M. and U.D.S. and the remaining sections have been edited jointly. All authors have read and agreed to the published version of the manuscript.

**Funding:** K.-F. Chen is supported by grant number 112-2112-M-002-026 from the National Science and Technology Council, Taiwan.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** The authors thank the ATLAS, CMS, and LHCb physics working groups for cross reading the draft.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Bobeth, C.; Gorbahn, M.; Hermann, T.; Misiak, M.; Stamou, E.; Steinhauser, M.  $B_{s,d} \rightarrow \ell^+ \ell^-$  in the Standard Model with Reduced Theoretical Uncertainty. *Phys. Rev. Lett.* **2014**, *112*, 101801. [[CrossRef](#)] [[PubMed](#)]
- Bobeth, C.; Gorbahn, M.; Stamou, E. Electroweak Corrections to  $B_{s,d} \rightarrow \ell^+ \ell^-$ . *Phys. Rev. D* **2014**, *89*, 034023. [[CrossRef](#)]
- Hermann, T.; Misiak, M.; Steinhauser, M. Three-loop QCD corrections to  $B_s \rightarrow \mu^+ \mu^-$ . *J. High Energy Phys.* **2013**, *12*, 097. [[CrossRef](#)]
- Beneke, M.; Bobeth, C.; Szafron, R. Enhanced electromagnetic correction to the rare B-meson decay  $B_{s,d} \rightarrow \mu^+ \mu^-$ . *Phys. Rev. Lett.* **2018**, *120*, 011801. [[CrossRef](#)] [[PubMed](#)]
- Beneke, M.; Bobeth, C.; Szafron, R. Power-enhanced leading-logarithmic QED corrections to  $B_q \rightarrow \mu^+ \mu^-$ . *J. High Energy Phys.* **2019**, *10*, 232. [[CrossRef](#)]
- Altmannshofer, W.; Paradisi, P.; Straub, D.M. Model-Independent Constraints on New Physics in  $b \rightarrow s$  Transitions. *J. High Energy Phys.* **2012**, *4*, 8. [[CrossRef](#)]
- Beaujean, F.; Bobeth, C.; van Dyk, D.; Wacker, C. Bayesian Fit of Exclusive  $b \rightarrow s \bar{\ell} \ell$  Decays: The Standard Model Operator Basis. *J. High Energy Phys.* **2012**, *8*, 30. [[CrossRef](#)]
- Aoki, S.; Aoki, Y.; Becirevic, D.; Blum, T.; Colangelo, G.; Collins, S.; Morte, M.D.; Dimopoulos, P.; Dür, S.; Fukaya, H.; et al. FLAG Review 2019. *arXiv* **2019**, arXiv:1902.08191.



9. Bazavov, A.; Bernard, C.; Brown, N.; DeTar, C.; El-Khadra, A.X.; Gámiz, E.; Gottlieb, S.; Heller, U.M.; Komijani, J.; Kronfeld, A.S.; et al. B- and D-meson leptonic decay constants from four-flavor lattice QCD. *Phys. Rev. D* **2018**, *98*, 074512. [[CrossRef](#)]
10. Bussone, A.; Carrasco, N.; Dimopoulos, P.; Frezzotti, R.; Lami, P.; Lubicz, V.; Picca, E.; Riggio, L.; Rossi, G.; Simula, S.; et al. Mass of the b quark and B meson decay constants from  $N_f=2+1+1$  twisted-mass lattice QCD. *Phys. Rev. D* **2016**, *93*, 114505. [[CrossRef](#)]
11. Dowdall, R.J.; Davies, C.T.H.; Horgan, R.R.; Monahan, C.J.; Shigemitsu, J. B-meson decay constants from improved lattice nonrelativistic QCD with physical u, d, s, and c quarks. *Phys. Rev. Lett.* **2013**, *110*, 222003. [[CrossRef](#)] [[PubMed](#)]
12. Hughes, C.; Davies, C.T.H.; Monahan, C.J. New methods for B meson decay constants and form factors from lattice NRQCD. *Phys. Rev. D* **2018**, *97*, 054509. [[CrossRef](#)]
13. De Bruyn, K.; Fleischer, R.; Knegjens, R.; Koppenburg, P.; Merk, M.; Pellegrino, A.; Tuning, N. Probing new physics via the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime. *Phys. Rev. Lett.* **2012**, *109*, 041801. [[CrossRef](#)]
14. Bruyn, K.D.; Fleischer, R.; Knegjens, R.; Koppenburg, P.; Merk, M.; Tuning, N. Branching Ratio Measurements of  $B_s^0$  Decays. *Phys. Rev.* **2012**, *D86*, 014027. [[CrossRef](#)]
15. Buras, A.J. Relations between  $\Delta M(s, d)$  and  $B(s, d) \rightarrow \mu \bar{\mu}$  in models with minimal flavor violation. *Phys. Lett. B* **2003**, *566*, 115–119. [[CrossRef](#)]
16. King, D.; Lenz, A.; Rauh, T.  $B_s$  mixing observables and  $|V_{td}/V_{ts}|$  from sum rules. *J. High Energy Phys.* **2019**, *5*, 34. [[CrossRef](#)]
17. Buras, A.J. Standard Model Predictions for Rare K and B Decays without New Physics Infection. *Eur. Phys. J. C* **2023**, *83*, 66. [[CrossRef](#)]
18. HFLAV Group; Amhis, Y.; Banerjee, S.W.; Ben-Haim, E.; Bertholet, E.; Bernlochner, F.U.; Bona, M.; Bozek, A.; Bozzi, C.; Brodzicka, J.; et al. Averages of b-hadron, c-hadron, and  $\tau$ -lepton properties as of 2021. *Phys. Rev. D* **2023**, *107*, 052008. [[CrossRef](#)]
19. ATLAS Collaboration. Study of the rare decays of  $B_s^0$  and  $B^0$  mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector. *J. High Energy Phys.* **2019**, *4*, 098. [[CrossRef](#)]
20. ATLAS Collaboration. Study of the rare decays of  $B_s^0$  and  $B^0$  into muon pairs from data collected during the LHC Run 1 with the ATLAS detector. *Eur. Phys. J. C* **2016**, *76*, 513. [[CrossRef](#)]
21. Hoecker, A.; Speckmayer, P.; Stelzer, J.; Therhaag, J.; von Toerne, E.; Voss, H. TMVA: Toolkit for Multivariate Data Analysis. *arXiv* **2007**, arXiv:physics/0703039.
22. Tanabashi, M.; Hagiwara, K.; Hikasa, K.; Nakamura, K.; Sumino, Y.; Takahashi, F.; Tanaka, J.; Agashe, K.; Aielli, G.; AMSler, C.; et al. Review of Particle Physics. *Phys. Rev. D* **2018**, *98*, 030001. [[CrossRef](#)]
23. HFLAV Group; Amhis, Y.; Banerjee, S.; Ben-Haim, E.; Bernlochner, F.; Bozek, A.; Bozzi, C.; Chrzyszcz, M.; Dingfelder, J.; Duell, S.; et al. Averages of b-hadron, c-hadron, and  $\tau$ -lepton properties as of summer 2016. *Eur. Phys. J. C* **2017**, *77*, 895. [[CrossRef](#)]
24. Neyman, J. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Phil. Trans. R. Soc. Lond. A* **1937**, *236*, 333–380. Available online: <http://rsta.royalsocietypublishing.org/content/236/767/333> (accessed on 1 January 2024).
25. ATLAS Collaboration. Measurement of the  $B_s^0 \rightarrow \mu \mu$  Effective Lifetime with the ATLAS Detector. *J. High Energy Phys.* **2023**, *9*, 199. [[CrossRef](#)]
26. Workman, R.L.; Burkert, V.D.; Crede, V.; Klempt, E.; Thoma, U.; Tiator, L.; Agashe, K.; Aielli, G.; Allanach, B.C.; AMSler, C.; et al. Review of Particle Physics. *Prog. Theor. Exp. Phys.* **2022**, *2022*, 083C01; 2023 update. [[CrossRef](#)]
27. Pivk, M.; Le Diberder, F. sPlot: A statistical tool to unfold data distributions. *Nucl. Instrum. Meth. A* **2005**, *555*, 356–369. [[CrossRef](#)]
28. CMS Collaboration. Measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  decay properties and search for the  $B^0 \rightarrow \mu^+ \mu^-$  decay in proton–proton collisions at  $\sqrt{s} = 13$  TeV. *Phys. Lett. B* **2022**, *842*, 137955. [[CrossRef](#)]
29. CMS Collaboration; Sirunyan, A.M.; Tumasyan, A.; Adam, W.; Ambrogio, F.; Bergauer, T.; Brandstetter, J.; Dragicevic, M.; Erö, J.; Valle, E.D.; et al. Measurement of properties of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays and search for  $B^0 \rightarrow \mu^+ \mu^-$  with the CMS experiment. *J. High Energy Phys.* **2020**, *4*, 188. [[CrossRef](#)]
30. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016. [[CrossRef](#)]
31. LHCb Collaboration; Aaij, R.; Beteta, C.A.; Ackernley, T.; Adeva, B.; Adinolfi, M.; Afsharnia, H.; Aidala, C.A.; Aiola, S.; Ajaltouni, Z.; et al. Precise measurement of the  $f_s/f_d$  ratio of fragmentation fractions and of  $B_s^0$  decay branching fractions. *Phys. Rev. D* **2021**, *104*, 032005. [[CrossRef](#)]
32. Read, A.L. Presentation of search results: The CL(s) technique. *J. Phys.* **2002**, *G28*, 2693–2704. [[CrossRef](#)]
33. LHCb Collaboration; Aaij, R.; Abellán Beteta, C.; Ackernley, T.; Adeva, B.; Adinolfi, M.; Afsharnia, H.; Aidala, C.A.; Aiola, S.; Ajaltouni, Z.; et al. Measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  decay properties and search for the  $B^0 \rightarrow \mu^+ \mu^-$  and  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  decays. *Phys. Rev. D* **2022**, *105*, 012010. [[CrossRef](#)]
34. LHCb Collaboration; Aaij, R.; Abellán Beteta, C.; Ackernley, T.; Adeva, B.; Adinolfi, M.; Afsharnia, H.; Aidala, C.A.; Aiola, S.; Ajaltouni, Z.; et al. Analysis of Neutral B-Meson Decays into Two Muons. *Phys. Rev. Lett.* **2022**, *128*, 041801. [[CrossRef](#)] [[PubMed](#)]
35. LHCb Collaboration; Aaij, R.; Adeva, B.; Adinolfi, M.; Ajaltouni, Z.; Akar, S.; Albrecht, J.; Alessio, F.; Alexander, M.; Ali, S.; et al. Measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction and effective lifetime and search for  $B^0 \rightarrow \mu^+ \mu^-$  decays. *Phys. Rev. Lett.* **2017**, *118*, 191801. [[CrossRef](#)] [[PubMed](#)]
36. Archilli, F.; Baldini, W.; Bencivenni, G.; Bondar, N.; Bonivento, W.; Cadeddu, S.; Campana, P.; Cardini, A.; Ciambone, P.; Vidal, X.C.; et al. Performance of the muon identification at LHCb. *J. Instrum.* **2013**, *8*, P10020. [[CrossRef](#)]
37. Zyla, P.A.; Barnett, R.M.; Beringer, J.; Dahl, O.; Dwyer, D.A.; Groom, D.E.; Lin, C.-J.; Lugovsky, K.S.; Pianori, E.; Robinson, D.J.; et al. Review of Particle Physics. *Prog. Theor. Exp. Phys.* **2020**, *2020*, 083C01. [[CrossRef](#)]

38. Skwarnicki, T. A Study of the Radiative Cascade Transitions between the Upsilon-Prime and Upsilon Resonances. Ph.D. Thesis, Institute of Nuclear Physics, Krakow, Poland, 1986. Available online: <https://inspirehep.net/literature/230779> (accessed on 1 January 2024).
39. ATLAS, CMS and LHCb Collaborations. *Combination of the ATLAS, CMS and LHCb Results on the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  Decays*; Technical Report; ATLAS-CONF-2020-049, CMS-PAS-BPH-20-003, LHCb-CONF-2020-002; CERN: Geneva, Switzerland, 2020.
40. CMS and LHCb Collaborations; Khachatryan, V.; Khachatryan, V.; Sirunyan, A.M.; Tumasyan, A.; Adam, W.; Bergauer, T.; Dragicevic, M.; Erö, J.; Friedl, M.; et al. Observation of the rare  $B_s^0 \rightarrow \mu^+ \mu^-$  decay from the combined analysis of CMS and LHCb data. *Nature* **2015**, *522*, 68–72. [[CrossRef](#)]
41. ATLAS and CMS Collaborations. Snowmass White Paper Contribution: Physics with the Phase-2 ATLAS and CMS Detectors. Technical Report; ATL-PHYS-PUB-2022-018. 2022. Available online: <https://cds.cern.ch/record/2805993> (accessed on 17 January 2024).
42. LHCb Collaboration; Aaij, R.; Adeva, B.; Adinolfi, M.; Aidala, C.; Ajaltouni, Z.; Akar, S.; Albicocco, P.; Albrecht, J.; Alessio, F.; et al. Physics case for an LHCb Upgrade II: Opportunities in flavour physics, and beyond, in the HL-LHC era. *arXiv* **2018**, arXiv:1808.08865. [[CrossRef](#)]
43. Buras, A.J.; Fleischer, R.; Girsbach, J.; Knegjens, R. Probing New Physics with the  $B_s \rightarrow \mu + \mu^-$  Time-Dependent Rate. *J. High Energy Phys.* **2013**, *7*, 77. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.