Observation of the semileptonic decay $B^+ \to p\bar{p}\mu^+\nu_\mu$

The LHCb collaboration

E-mail: mark.smith@cern.ch

Abstract: The Cabibbo-suppressed semileptonic decay $B^+ \to p\bar{p}\mu^+\nu_\mu$ is observed for the first time using a sample of $pp$ collisions corresponding to an integrated luminosity of 1.0, 2.0 and 1.7 fb$^{-1}$ at centre-of-mass energies of 7, 8 and 13 TeV, respectively. The differential branching fraction is measured as a function of the $p\bar{p}$ invariant mass using the decay mode $B^+ \to J/\psi K^+$ for normalisation. The total branching fraction is measured to be

$$B(B^+ \to p\bar{p}\mu^+\nu_\mu) = (5.27^{+0.23}_{-0.24} \pm 0.21 \pm 0.15) \times 10^{-6},$$

where the first uncertainty is statistical, the second systematic and the third is from the uncertainty on the branching fraction of the normalisation channel.

Keywords: B physics, Branching fraction, Flavor physics, Hadron-Hadron scattering (experiments)

ArXiv ePrint: 1911.08187
1 Introduction

Studies of semileptonic $B$ meson decays have recently generated interest due to a number of anomalies in experimental results. Measurements of the observables $R(D)$ and $R(D^*)$ [1–6] have shown hints of lepton non-universality with a combined significance of over $3\sigma$ [7]. To probe the flavour structure of possible new physics contributions to these decay modes, it is desirable to make analogous measurements for decays involving different quark-level processes, such as $b \to u$ transitions. To that end, the decay mode $B^+ \to p\bar{p}\ell^+\nu_\ell$ is promising experimentally, particularly when performing the measurement at a hadron collider. The requirement of a proton anti-proton pair in the final state should significantly reduce combinatorial background, which would otherwise be significant for final states with pions.

Semileptonic decays of $B$ mesons to a final state containing multiple baryons are as yet unobserved. A theoretical model of $B^+ \to p\bar{p}\ell^+\nu_\ell$ decays has been constructed with perturbative QCD (pQCD) [8]. This model is based on studies of several fully hadronic $B \to Y\bar{Y}'X$ decays where $Y$ and $Y'$ represent baryons and $X$ one or more mesons. By fitting the angular distributions and decay rates of the hadronic modes the authors of refs. [8–10] estimate the differential rate of $B^+ \to p\bar{p}\ell^+\nu_\ell$ decays. They also predict the total branching fraction of the $B^+ \to p\bar{p}\ell^+\nu_\ell$ decay to be $(1.04\pm 0.38) \times 10^{-4}$ for $l = \mu, e$ leptons. This prediction motivated a search by the Belle collaboration for this channel that lead to evidence for the $B^+ \to p\bar{p}e^+\nu_e$ decay mode with $3.0\sigma$ significance [11]. The branching fraction was measured to be $(8.2^{+3.7}_{-3.2} \pm 0.6) \times 10^{-6}$, one order of magnitude smaller than the prediction.
The measurements of the fully hadronic modes show features that merit further investigation. It is surprising that the branching fractions of decays of $B$ mesons to final states comprising only two baryons are suppressed compared to those of two baryons and one or more extra final state particles [12]. For example, the branching fraction of $B^0 \rightarrow p\bar{p}$ is two orders of magnitude smaller than that of the similar $B^0 \rightarrow p\bar{p}\pi^+\pi^-$ decay [12, 13]. Furthermore, the invariant-mass distributions of the baryon pair in $B \rightarrow Y\bar{Y}X$ decays show a characteristic accumulation at low values, called the threshold enhancement effect [14–17]. Measurements of $B^+ \rightarrow p\bar{p}\ell^+\nu_\ell$ semileptonic decays provide the ideal environment for understanding the $\langle Y\bar{Y} | (\bar{q}b)_{V-A} | B \rangle$ matrix element that contributes to hadronic decay modes.

In this paper, the first observation of the decay $B^+ \rightarrow p\bar{p}\mu^+\nu_\mu$ is presented. As the dynamics of the transition are not known, the branching fraction is measured in bins of $p\bar{p}$ invariant mass. These bins are then summed to obtain a measurement of the total branching fraction. The decay $B^+ \rightarrow J/\psi K^+$, with $J/\psi \rightarrow \mu^+\mu^-$, is chosen as the normalisation mode as it is fully reconstructed and can pass similar selection requirements to the signal. The branching fraction within a bin $i$ is

$$B_i(B^+ \rightarrow p\bar{p}\mu^+\nu_\mu) = \frac{N_i(B^+ \rightarrow p\bar{p}\mu^+\nu_\mu)}{N(B^+ \rightarrow J/\psi K^+)} \times \frac{\epsilon(B^+ \rightarrow J/\psi K^+)}{\epsilon_i(B^+ \rightarrow p\bar{p}\mu^+\nu_\mu)} \times \mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-),$$

where $N_i(B^+ \rightarrow p\bar{p}\mu^+\nu_\mu)$ is the yield of $B^+ \rightarrow p\bar{p}\mu^+\nu_\mu$ candidates in bin $i$, $N(B^+ \rightarrow J/\psi K^+)$ is the total yield of $B^+ \rightarrow J/\psi K^+$ candidates and $\epsilon$ represents the product of the detector acceptance and the reconstruction and selection efficiencies of the two modes. The branching fractions of $B^+ \rightarrow J/\psi K^+$ and $J/\psi \rightarrow \mu^+\mu^-$ decays are taken from ref. [12].

The signal yields are extracted from fits to a variable called the corrected mass, which accounts for the unreconstructed neutrino in the signal decay. It is defined as [18]

$$m_{\text{corr}} = |p_\perp| + \sqrt{|p_\perp|^2 + m_{p\bar{p}\mu}^2},$$

(1.1)

where $|p_\perp|$ is defined as the magnitude of the reconstructed $p\bar{p}\mu^+$ momentum transverse to the $B$ flight direction and $m_{p\bar{p}\mu}^2$ is the square of the $p\bar{p}\mu^+$ invariant mass.

This study uses the data collected with the LHCb detector in proton-proton collisions in 2011, 2012 and 2016. This corresponds to integrated luminosities of 1.0, 2.0 and 1.7 fb$^{-1}$ at centre-of-mass energies of 7, 8 and 13 TeV, respectively. The 2011 and 2012 data sets are treated together and collectively referred as the Run 1 data set. Charge conjugate processes are implied throughout this paper.

2 Detector and simulation

The LHCb detector [19, 20] is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$, designed for the study of particles containing $b$ or $c$ quarks. The detector includes a high-precision tracking system consisting of a silicon-strip vertex detector surrounding the $pp$ interaction region [21], a large-area silicon-strip
detector located upstream of a dipole magnet with a bending power of about 4 Tm, and three stations of silicon-strip detectors and straw drift tubes [22, 23] placed downstream of the magnet. The tracking system provides a measurement of the momentum, \( p \), of charged particles with a relative uncertainty that varies from 0.5% at low momentum to 1.0% at 200 GeV/c. The minimum distance of a track to a primary vertex (PV), the impact parameter (IP), is measured with a resolution of \( (15 + 29/p_T) \mu \text{m} \), where \( p_T \) is the component of the momentum transverse to the beam in GeV/c. Different types of charged hadrons are distinguished using information from two ring-imaging Cherenkov (RICH) detectors [24]. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers [25].

The online event selection is performed by a trigger [26], which consists of a hardware stage that performs some basic selection, followed by a software stage, which applies a full event reconstruction. At the first level, a track consistent with being a muon with significant \( p_T \) is required to be present in the event. Subsequently in the software stage, two tracks are required to form a secondary vertex with significant displacement from a \( pp \) interaction vertex. A multivariate algorithm [27] is used to identify vertices that are consistent with the decay of a \( b \) hadron.

Simulation is used to determine the efficiency of the signal mode and estimate the shapes of the signal and several backgrounds modes in the fits to the \( m_{\text{corr}} \) distribution. In the simulation, \( pp \) collisions are generated using PYTHIA [28, 29] with a specific LHCb configuration [30]. Decays of unstable particles are described by EvtGen [31], in which final-state radiation is generated using Photos [32]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit [33, 34], as described in ref. [35]. The generated \( B \) meson \( p \) and \( p_T \) spectra are corrected to match the data distributions. A boosted decision tree (BDT) weighter [36] is trained on samples of \( B^+ \rightarrow J/\psi K^+ \) data and simulation, independent of those used for the normalisation of the branching fraction. This is then used to correct all the simulation samples used in the analysis.

### 3 Selection

Signal candidates are constructed from three charged tracks which are required to be of good quality and well separated from any PV. The tracks must also have particle identification consistent with their particle hypothesis. The requirement for positive proton identification enforces a minimum value of \( p \) of 18 GeV/c such that the protons are above the threshold for radiating in the RICH. Similarly, the muons must have \( p \) above 3 GeV/c to propagate through the muon stations. All the tracks must have \( p_T \) larger than 1.5 GeV/c and form a good-quality vertex significantly displaced from the PV with which the candidate is associated. The signal muon must have fired the hardware trigger and the reconstructed \( B^+ \) candidate formed by the three tracks must be consistent with the object that fired the software trigger. Potential decays of \( \eta_c \), \( J/\psi \) and \( \psi(2S) \) mesons to \( p\bar{p} \) are
removed with vetoes in the $p\bar{p}$ invariant mass of $\pm 50 \text{MeV}/c^2$ around their respective known masses \cite{12}.

The selection of the $B^+ \rightarrow J/\psi K^+$ normalisation mode is aligned with that of the signal to reduce systematic uncertainties. The selection criteria for the signal protons are applied to the kaon and the muon of opposite sign ($K^+\mu^-$), with the exception of the particle identification criteria. The selection of the other muon is the same as that of the muon in the signal decay.

Further selection is used to reduce several sources of backgrounds relative to the signal. In total there are five variables to which selection is applied, with the chosen criteria on each optimised together. These variables, and the backgrounds targeted by them are described in the following paragraphs.

The largest background contribution comes from a mixture of partially reconstructed decays producing two protons and a muon in the final state. It is expected that the largest among these originates in $b \rightarrow c$ quark transitions. The most pernicious is $B \rightarrow \Lambda_c^- p \mu^+ \nu_\mu X$ decays, where $X$ represents any number of charged or neutral pions (including none) and the $\Lambda_c^-$ baryon decays to a final state including one proton. The other major background arises from $B \rightarrow p\bar{p}D X$ decays, where the $D$ meson may be of any variety ($D^0$, $D^-$, $D^{*-}$, etc.) and ultimately decays to a final state with a muon. The contribution of $B \rightarrow p\bar{p}\Lambda_c^- X$ decays with the $\Lambda_c^-$ baryon decaying semileptonically is comparatively small, as the semileptonic branching fraction is dominated by $\Lambda_c^- \rightarrow \Lambda^- \nu_\ell$ decays. The $\Lambda$ baryon flies a sufficient distance within the detector before decaying such that the resulting proton is not associated with the $B$ decay vertex. Another source of partially reconstructed background is formed of $B \rightarrow p\bar{p}\mu^+ \nu_\mu X$ decays, where $X$ denotes one or more charged or neutral pions. These decays may proceed with intermediate $N^*$ or $\Delta$ resonances and could naively be expected to have similar branching fractions to the signal.

If any of these partially reconstructed decay modes produces charged tracks in addition to the $p\bar{p}\mu^+$ candidate, it can be efficiently suppressed with an isolation technique. Once a signal candidate has been constructed, the other tracks in the event close to the $B$ decay vertex are examined. A BDT is used to identify those nearby tracks that can be associated with the signal candidate decay vertex. If the candidate is truly signal, there should be few other tracks that can be associated with it and the BDT should classify them with a low score. On the other hand, the extra track(s) from a partially reconstructed decay returns a high score if such tracks are found. The isolation algorithm returns the BDT output for the four tracks most likely to have come from the $B$ vertex. These four numbers are themselves combined into a single BDT classifier, known as the charged-isolation BDT. This BDT is trained on simulation to discriminate signal from $B^+ \rightarrow \Lambda_c^- p \mu^+ \nu_\mu$ decays, which is expected to be the largest mode with extra charged tracks. The efficacy of this BDT in reducing such background is shown in figure 1(a). The indicated requirement on the charged BDT score rejects 80% of the major background decay $B \rightarrow \Lambda_c^- p \mu^+ \nu_\mu X$ (with all possible decay modes of the $\Lambda_c^-$ considered), whilst retaining 93% of the signal.

For those partially reconstructed final states with only additional neutral particles, further suppression is achieved by considering the kinematics of the decays. An additional BDT, the so called part-reco BDT, considers 11 variables: the impact parameter significance of the three final-state tracks, the $p\bar{p}$ pair and the $B^+$ candidate with respect to the
Figure 1. Result of training (a) the charged-isolation BDT and (b) the part-reco BDT. The chosen selection on the classifier outputs are indicated by the dashed red line. For some candidates there are no additional tracks near the $B$-decay vertex; these candidates are accepted and do not appear in the charged-isolation BDT output. The background samples shown here have the $\Xi_c^-$ and $D^0$ hadrons decaying via $\Xi_c^- \rightarrow pK^+\pi^-$ and $D^0 \rightarrow \mu^+X$. The part-reco BDT is trained on a mixture of background modes with only one shown here for illustration.

PV; the impact parameters of the tracks with respect to the fitted $B^+$ decay vertex; the $\chi^2$ of the $B^+$ vertex fit; the angle between the $B^+$ candidate momentum and displacement vectors; and the difference between the $p$ and $\bar{p}$ momenta. The part-reco BDT is trained on simulation in order to discriminate signal from a mixture of all the considered background modes. The result of this training is shown in figure 1(b). The selection on the part-reco BDT output removes 18% of the decays $B \rightarrow p\bar{p}D$ and keeps 98% of the signal.

An additional background arises from particles that are misidentified as protons (misID). The particle identification requirements on the proton tracks are therefore further tightened. Background due to hadrons being misidentified as muons is considered and reduced to a negligible amount with a loose particle identification requirement. A background occurs due to the combination of two tracks from the decay of a heavy hadron with a track from elsewhere in the event. This is referred to as combinatorial background. This component is expected to have a small contribution due to the tight vertex requirements on the $p\bar{p}\mu^+$ candidate and the requirement for positively identifying two protons. Therefore no additional selection is employed specifically to reduce it.

In addition to the two BDTs and proton identification criteria, one further quantity is considered. The uncertainty on the corrected mass of the candidate may be used to improve the separation between signal and background [37]. It is calculated from the estimated uncertainties on the positions of the $B^+$ primary and secondary vertices, and the momenta of the tracks. Selecting lower values of the corrected-mass uncertainty produces a sharper peak for the signal mode in the corrected mass distribution, which will aid the discrimination of the signal from background in the fit to determine the yield.

In total the selection uses five quantities (two BDTs, the proton PID, the muon PID and the corrected-mass uncertainty). In order to ascertain the optimum selection, a five dimensional grid search is performed using pseudoexperiments. Data sets are generated from the simulation samples with the expected proportions of each background. The
expected signal amount is taken from the central value of the $B^+ \to \rho \pi^+ \nu_\tau$ branching fraction reported by the Belle collaboration [11]. For the backgrounds, the current averages for the branching fractions are used if they have been measured. For those backgrounds that have not been measured, their branching fractions are estimated relative to that expected for the signal, accounting for different CKM matrix elements and the available phase space. For each point in the grid, the selection is applied to the simulation to estimate the efficiency for each component. The efficiency of the PID requirements on the simulation is estimated with a method based on data control samples [38]. For each data set the $m_{\text{corr}}$ variable is simulated and the expected relative uncertainty on the signal yield is found by a fit to the simulated pseudodata. These fits are not binned in $m(J/\psi K)$ but consider the entire sample. The selection that produces the smallest relative uncertainty on the signal yield is chosen.

4 Signal and normalisation yields

The yields of the signal and normalisation modes are ascertained with unbinned extended maximum-likelihood fits. In the case of the normalisation mode, the invariant mass distribution of the $J/\psi K^+$ candidates is fitted. The 2011, 2012 and 2016 data sets are fitted separately and then the yields combined. The fit to the 2016 data set is shown in figure 2.

For the signal mode, the corrected mass is fitted. The distribution of this variable peaks at the $B^+$ mass for candidates where one massless particle is not reconstructed. On the other hand, candidates from partially reconstructed decays that are missing one or more massive particles in addition to the neutrino have wide distributions concentrated at lower corrected mass values. The Run 1 and 2016 data are combined and fitted together to improve the fit stability.

The shapes for the signal component and contributions from partially reconstructed decays are determined using simulation. The shape of the signal probability density function (PDF) is parametrised by the sum of four bifurcated Gaussian functions with a common mean. The parameters of the Gaussian functions as well as their relative fractions are
all fixed in the fit. All of the background PDFs are accounted for with kernel density estimation [39].

The shape of the proton misID background comes from a separate independent data sample in which the particle identification requirements on one of the protons have been removed. In this sample the true number of each hadron species can be unfolded and so the probability of a hadron being misidentified as a proton can be estimated. These probabilities are used to weight this sample to estimate both the template shape for the fit component and the yield of misID events.

A background component due to random combinations of protons and muons, referred to as the combinatorial background, is included in the fit. A sample of data for which the $B^+$ decay vertex quality selection has been reversed is used to estimate the shape of this background.

The yields of the signal, proton misID, combinatorial and partially reconstructed decays are determined by the fit, as are the relative fractions of each partially reconstructed mode. All of the fit parameters are free to vary with the exception of the misID yield which is constrained.

The fit in each $m(p\bar{p})$ bin is performed independently. The $m_{\text{corr}}$ distributions in each bin, and the resulting fits are shown in figure 3. In each bin the fits are validated using pseudoexperiments. An ensemble of $10^5$ data sets is generated and fitted with the component yields taken from the fits to data. Some small biases on the signal yield are found and these are considered as a source of systematic uncertainty.

5 Efficiency

The efficiencies for the signal and normalisation modes to be reconstructed and selected are both assessed with simulation. Corrections are applied to account for known differences between data and simulation in the track-reconstruction efficiency [40] and the efficiency of the hardware trigger [41]. The efficiency of the particle identification requirements on each track is evaluated with data [38] and applied to the simulation.

The binning in $m(p\bar{p})$ reduces the dependence on the model of the $B^+$ decay when calculating the efficiency of the signal mode. However, as there are selection requirements on kinematic quantities of the candidates, there is still some residual dependence on the dynamics of the decay. The simulation is therefore weighted to represent the pQCD model of ref. [8] as the current best estimate of how the decay proceeds. This weighting corrects the distribution of the invariant mass of the $\mu^+\nu_\mu$ system. The variation of the parameters of this model is considered as a source of systematic uncertainty.

The ratio of selection efficiencies between the signal and normalisation modes in each bin of $m(p\bar{p})$ is shown in table 1. These efficiencies are presented separately for the Run 1 and 2016 samples. They are combined to form an overall efficiency ratio, accounting for the difference in sample sizes between Run 1 and 2016. This combination is calculated using the measured $B^+$ production cross-sections [42] and integrated luminosities of each data set.
Figure 3. Distributions of $m_{\text{corr}}$ in each $m(p\bar{p})$ bin with the fit results shown.
6 Systematic uncertainties

The systematic uncertainties can be split into two categories: those that affect the calculation of the ratio of efficiencies of the signal and normalisation modes and those that may change the determination of the signal yield in the fit. For the former, each of the corrections to the simulation contributes a source of uncertainty both from the limited sizes of the samples used to derive the corrections and from the method of deriving them. The method of correcting the $p$ and $p_T$ distributions of the $B^+$ mesons in the simulation may give rise to a systematic uncertainty. The parameters of the BDT weighter used to correct these distributions are altered and the relative efficiencies recalculated, with the difference to the nominal relative efficiency being the assigned uncertainty. An additional uncertainty due to any residual differences between data and simulation is determined using the $B^+ \to J/\psi K^+$ decay mode. The difference in efficiency due to the selection on the two BDTs and corrected-mass uncertainty is compared between data and simulation.

To account for the uncertainty in the correction of the simulation for the reconstruction efficiency of each track, the applied weights are varied within their uncertainties and the relative efficiencies recalculated. Similarly, an uncertainty is assessed for the particle-identification weights applied to each track. The uncertainty due to the limited simulation sample sizes used to calculate the efficiencies is also included.

A further uncertainty is due to the physics model that the simulation is weighted to represent. The model affects the kinematic distributions of the final state tracks which feeds into the efficiency calculation as these distributions are biased by the selection requirements. Since the model is unproven a conservative uncertainty is taken. New sets of weights for the simulation are created that sample extreme variations of the model parameters ($\pm 5\sigma$), and for each variation the efficiency is recalculated. Despite this extreme test, the systematic uncertainty due to the physics model is not dominant, which reflects the flat selection efficiency over the kinematic ranges in which the final-state particles lie within each bin of $m(p\bar{p})$. Finally, the uncertainties on the $B^+$ production cross-section [42] and integrated luminosities of the data samples are combined to give the systematic uncertainty on the averaging of the efficiencies when combining Run 1 and 2016.

In the corrected-mass fit, uncertainties arise from potential variations in the shapes of the components. This variation is assessed with pseudoexperiments. Data sets are generated with the nominal fit model and then fitted with both the nominal model and an

| $m(p\bar{p})$, [GeV/c$^2$] | $\epsilon(B \to p\bar{p}\mu\nu)_{/}/\epsilon(B \to J/\psi K)$ |
|----------------------------|---------------------------------|
| Bin 1: 1.87 − 2.0          | 0.37 ± 0.02                    |
| Bin 2: 2.0 − 2.2           | 0.37 ± 0.02                    |
| Bin 3: 2.2 − 2.4           | 0.36 ± 0.02                    |
| Bin 4: 2.4 − 2.6           | 0.36 ± 0.02                    |
| Bin 5: 2.6 − 5.0           | 0.35 ± 0.02                    |

Table 1. Relative efficiencies for Run 1 and 2016 and the weighted combination of both.
Table 2. Summary of the systematic uncertainties on the differential branching fractions. The contributions pertaining to the efficiency estimate are first, those for the yield extraction are below. The particle identification and tracking efficiency uncertainties are assumed to be 100% correlated between Run 1 and 2016. The total correlations of the uncertainties between the bins are shown in table 4.

| Source                                | Relative uncertainties on $B$ [%] |
|---------------------------------------|-----------------------------------|
|                                       | Bin 1   | Bin 2   | Bin 3   | Bin 4   | Bin 5   |
| Kinematic weighting                   | 0.7     | 0.6     | 0.4     | 0.5     | 0.4     |
| Data-simulation agreement             | 0.4     | 0.4     | 0.4     | 0.4     | 0.4     |
| Tracking efficiency                   | 2.7     | 2.7     | 2.7     | 2.7     | 2.7     |
| Particle identification               | 1.0     | 0.7     | 1.3     | 1.0     | 1.7     |
| Simulation sample size                | 3.6     | 3.2     | 3.2     | 3.1     | 3.0     |
| Physics model                         | 0.3     | 0.6     | 0.6     | 0.4     | 0.3     |
| Run 1 and 2016 combination            | 2.1     | 1.6     | 1.7     | 1.7     | 1.6     |
| Kernel smoothing                      | 0.0     | 1.1     | 2.7     | 7.9     | 3.5     |
| Signal model                          | 0.6     | 2.0     | 3.0     | 4.8     | 9.9     |
| Simulation sample size                | 0.3     | 0.0     | 0.3     | 2.4     | 5.2     |
| misID model                           | 0.9     | 0.1     | 0.6     | 5.2     | 13.5    |
| Combinatorial model                   | 0.9     | 1.2     | 1.2     | 8.5     | 4.7     |
| Fit bias                              | 0.2     | 0.1     | 0.9     | 2.5     | 7.8     |
| Total systematic uncertainty          | 5.3     | 5.2     | 6.5     | 15.6    | 20.8    |
| Total statistical uncertainty         | 9.1     | 5.5     | 12.5    | 25.3    | 29.8    |

alternative. The width of the distribution of differences between the nominal and alternative fits is taken as the uncertainty. For those components that rely on kernel density estimators, a systematic uncertainty is assessed for the choice of smoothing parameter by varying it. The uncertainty due to the choice of model for the signal shape is found by replacing the nominal PDF with one constructed with kernel density estimators. The uncertainty due to the limited sizes of the simulation samples is determined by generating new simulation from the nominal fit PDFs with the same sample sizes and making alternative PDFs with those samples. Similarly, an estimate of the uncertainty on the shape of the proton misID background component is assessed. For the shape of the combinatorial background component, an alternative data sample is trialled which requires the two protons to be of the same charge. Finally, the small biases in the fit noted in section 4 are included.

A summary of the systematic uncertainties is presented in table 2. They are given as relative uncertainties on the branching fraction with the combination accounting for the correlation of the uncertainties between the two data sets. The correlations of the total uncertainties between the bins are shown in table 4 and the covariance matrix is presented in table 5, in the appendix.
Table 3. Number of observed $B^+ \to p\bar{p}\mu^+\nu_\mu$ candidates and differential branching fraction in each bin of $m(p\bar{p})$. The uncertainties on the signal yields are statistical only. For the differential branching fractions the first uncertainties are statistical, the second systematic and the third from the uncertainties on the branching fractions of the normalisation channel.

| $m(p\bar{p})$ [GeV/c$^2$] | Signal Yield | $d\mathcal{B}(B^+ \to p\bar{p}\mu^+\nu_\mu)/dm(p\bar{p})$ [$\times 10^{-6}$ GeV$^{-1}$c$^2$] |
|-----------------------------|-------------|----------------------------------------------------------------------------------|
| Bin 1: 1.87–2.0 | 1210 ± 110 | 12.9 ± 1.2 ± 0.7 ± 0.4 |
| Bin 2: 2.0–2.2 | 1830 ± 110 | 12.9 ± 0.7 ± 0.7 ± 0.4 |
| Bin 3: 2.2–2.4 | 530 ± 70 | 3.8 ± 0.5 ± 0.2 ± 0.1 |
| Bin 4: 2.4–2.6 | 150 ± 40 | 1.04 ± 0.30 ± 0.16 ± 0.03 |
| Bin 5: 2.6–5.0 | 88 ± 26 | 0.054 ± 0.016 ± 0.011 ± 0.002 |

7 Results

The fitted yields for the signal mode are presented in table 3. The extracted yields of the normalisation channel are $14,930 \pm 260$ for 2011, $31,380 \pm 190$ for 2012 and $49,270 \pm 250$ for 2016. Combining these with the efficiency ratios from section 5, the differential branching fraction in each $m(p\bar{p})$ bin is calculated. The results are presented in table 3. The relative differential branching fractions are summed over the bins, with the correlation of the systematic uncertainties between the bins accounted for, to give the total relative branching fraction of

$$B(B^+ \to p\bar{p}\mu^+\nu_\mu) = (5.27^{+0.23}_{-0.24} \pm 0.21 \pm 0.15) \times 10^{-6},$$

where the first uncertainty is statistical and the second systematic. Multiplying this by the current average of the normalisation branching fraction [12] leads to

$$B(B^+ \to p\bar{p}\mu^+\nu_\mu) = (8.75 \pm 0.39 \pm 0.35) \times 10^{-2},$$

where the third uncertainty is from the normalisation branching fraction. Finally, the absolute differential branching fraction as a function of $m(p\bar{p})$ is shown in figure 4, where the indicated uncertainties include statistical, systematic and normalisation uncertainty contributions. As expected from the theory model and the analogous hadronic decays, the differential distribution peaks at a very low value and falls off sharply. The measured total branching fraction agrees with the previous measurement from the Belle collaboration and represents the first observation of the $B^+ \to p\bar{p}\mu^+\nu_\mu$ decay mode.

Acknowledgments

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
Figure 4. Differential branching fraction as a function of the $p\bar{p}$ invariant mass. The $\eta_c \to p\bar{p}$, $J/\psi \to p\bar{p}$ and $\psi(2S) \to p\bar{p}$ vetoes are indicated by the (left) red, (middle) blue and (right) pink bands, respectively. The red dashed line represents the prediction of the pQCD model normalised to the observed branching fraction [8].

(China); RFBR, RSF and Yandex LLC (Russia); GVA, XuntaGal and GENCAT (Spain); the Royal Society and the Leverhulme Trust (United Kingdom).
A Correlation and covariance matrices

| $m(p\bar{p})$ [GeV/c^2] | Bin 1  | Bin 2  | Bin 3  | Bin 4  | Bin 5  |
|--------------------------|--------|--------|--------|--------|--------|
| 1.87-2.0                 | 1.00   | 0.19   | 0.11   | 0.05   | 0.04   |
| 2.0-2.2                  | —      | 1.00   | 0.15   | 0.07   | 0.06   |
| 2.2-2.4                  | —      | —      | 1.00   | 0.04   | 0.04   |
| 2.4-2.6                  | —      | —      | —      | 1.00   | 0.02   |
| 2.6-5.0                  | —      | —      | —      | —      | 1.00   |

Table 4. Correlations in the uncertainties between bins of $m(p\bar{p})$.

| $m(p\bar{p})$ [GeV/c^2] | Bin 1  | Bin 2  | Bin 3  | Bin 4  | Bin 5  |
|--------------------------|--------|--------|--------|--------|--------|
| 1.87-2.0                 | $2.0 \times 10^{-12}$ | $2.8 \times 10^{-13}$ | $8.5 \times 10^{-14}$ | $2.3 \times 10^{-14}$ | $1.2 \times 10^{-15}$ |
| 2.0-2.2                  | —      | $1.1 \times 10^{-12}$ | $8.3 \times 10^{-14}$ | $2.3 \times 10^{-14}$ | $1.2 \times 10^{-15}$ |
| 2.2-2.4                  | —      | —      | $2.9 \times 10^{-13}$ | $6.9 \times 10^{-15}$ | $3.8 \times 10^{-16}$ |
| 2.4-2.6                  | —      | —      | —      | $9.6 \times 10^{-14}$ | $1.0 \times 10^{-16}$ |
| 2.6-5.0                  | —      | —      | —      | —      | $3.9 \times 10^{-16}$ |

Table 5. Covariance matrix for bins of $m(p\bar{p})$.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] LHCb collaboration, Measurement of the ratio of the $B^0 \to D^{*+} \tau^+ \nu_\tau$ and $B^0 \to D^{*+} \mu^+ \nu_\mu$ branching fractions using three-prong $\tau$-lepton decays, Phys. Rev. Lett. 120 (2018) 171802 [arXiv:1708.08856] [inSPIRE].

[2] LHCb collaboration, Measurement of the ratio of branching fractions $B(B^0 \to D^{*+} \tau^- \bar{\nu}_\tau)/B(B^0 \to D^{*+} \mu^- \bar{\nu}_\mu)$, Phys. Rev. Lett. 115 (2015) 111803 [Erratum ibid. 115 (2015) 159901] [arXiv:1506.08614] [inSPIRE].

[3] Belle collaboration, Measurement of $R(D)$ and $R(D^*)$ with a semileptonic tagging method, arXiv:1904.08794 [inSPIRE].

[4] Belle collaboration, Measurement of the $\tau$ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^* \tau^- \bar{\nu}_\tau$, Phys. Rev. Lett. 118 (2017) 211801 [arXiv:1612.00529] [inSPIRE].

[5] Belle collaboration, Measurement of the branching ratio of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)} e^- \bar{\nu}_e$ decays with hadronic tagging at Belle, Phys. Rev. D 92 (2015) 072014 [arXiv:1507.03233] [inSPIRE].

[6] BaBar collaboration, Evidence for an excess of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ decays, Phys. Rev. Lett. 109 (2012) 101802 [arXiv:1205.5442] [inSPIRE].
[7] HFLAV collaboration, Averages of b-hadron, c-hadron and τ-lepton properties as of 2018, arXiv:1909.12524 [insPIRE].

[8] C.Q. Geng and Y.K. Hsiao, Semileptonic $B^{-} \rightarrow p\bar{p}\ell^{-}\bar{\nu}_{\ell}$ decays, Phys. Lett. B 704 (2011) 495 [arXiv:1107.0801] [insPIRE].

[9] C.Q. Geng and Y.K. Hsiao, Angular distributions in three-body baryonic $B$ decays, Phys. Rev. D 74 (2006) 094023 [hep-ph/0606141] [insPIRE].

[10] C.-H. Chen, H.-Y. Cheng, C.Q. Geng and Y.K. Hsiao, Charmful three-body baryonic $B$ decays, Phys. Rev. D 78 (2008) 054016 [arXiv:0806.1108] [insPIRE].

[11] BELLE collaboration, Evidence for semileptonic $B^{-} \rightarrow p\bar{p}\ell^{-}\bar{\nu}_{\ell}$ decays, Phys. Rev. D 89 (2014) 011101 [arXiv:1306.3353] [insPIRE].

[12] PARTICLE DATA GROUP collaboration, Review of particle physics, Phys. Rev. D 98 (2018) 030001 [insPIRE].

[13] LHCb collaboration, First observation of the rare purely baryonic decay $B^{0} \rightarrow p\bar{p}$, Phys. Rev. Lett. 119 (2017) 232001 [arXiv:1709.01156] [insPIRE].

[14] BABAR collaboration, Evidence for the $B^{0} \rightarrow p\bar{p}K^{*0}$ and $B^{+} \rightarrow \eta_{c}K^{*+}$ decays and study of the decay dynamics of $B$ meson decays into $p\bar{p}h$ final states, Phys. Rev. D 76 (2007) 092004 [arXiv:0707.1648] [insPIRE].

[15] BABAR collaboration, Measurements of the Decays $B^{0} \rightarrow D^{0}p\bar{p}$, $B^{0} \rightarrow D^{*0}p\bar{p}$, $B^{0} \rightarrow D^{-}p\pi^{+}$ and $B^{0} \rightarrow D^{*-}p\pi^{+}$, Phys. Rev. D 74 (2006) 051101 [hep-ex/0607039] [insPIRE].

[16] BELLE collaboration, Observation of $B^{+} \rightarrow p\pi^{+}$, $B^{0} \rightarrow pK^{0}$ and $B^{+} \rightarrow pK^{*+}$, Phys. Rev. Lett. 92 (2004) 131801 [hep-ex/0310018] [insPIRE].

[17] LHCb collaboration, Evidence for CP-violation in $B^{+} \rightarrow p\bar{p}K^{+}$ decays, Phys. Rev. Lett. 113 (2014) 141801 [arXiv:1407.5907] [insPIRE].

[18] SLD collaboration, A measurement of $R_{b}$ using a vertex mass tag, Phys. Rev. Lett. 80 (1998) 660 [hep-ex/9708015] [insPIRE].

[19] LHCb collaboration, The LHCb detector at the LHC, 2008 JINST 3 S08005 [insPIRE].

[20] LHCb collaboration, LHCb detector performance, Int. J. Mod. Phys. A 30 (2015) 1530022 [arXiv:1412.6352] [insPIRE].

[21] R. Aaij et al., Performance of the LHCb Vertex Locator, 2014 JINST 9 P09007 [arXiv:1405.7808] [insPIRE].

[22] R. Arink et al., Performance of the LHCb Outer Tracker, 2014 JINST 9 P01002 [arXiv:1311.3893] [insPIRE].

[23] P. d’Argent et al., Improved performance of the LHCb Outer Tracker in LHC Run 2, 2017 JINST 12 P11016 [arXiv:1708.00819] [insPIRE].

[24] M. Adinolfi et al., Performance of the LHCb RICH detector at the LHC, Eur. Phys. J. C 73 (2013) 2431 [arXiv:1211.6759] [insPIRE].

[25] A.A. Alves, Jr. et al., Performance of the LHCb muon system, 2013 JINST 8 P02022 [arXiv:1211.1346] [insPIRE].

[26] R. Aaij et al., The LHCb trigger and its performance in 2011, 2013 JINST 8 P04022 [arXiv:1211.305S] [insPIRE].
[27] V.V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, 2013 *JINST* **8** P02013 [arXiv:1210.6861] [insPIRE].

[28] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [insPIRE].

[29] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [insPIRE].

[30] I. Belyaev et al., *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, *J. Phys. Conf. Ser.* **331** (2011) 032047 [insPIRE].

[31] D.J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A 462** (2001) 152 [insPIRE].

[32] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, *Eur. Phys. J. C* **45** (2006) 97 [hep-ph/0506026] [insPIRE].

[33] GEANT4 collaboration, *GEANT4 developments and applications*, *IEEE Trans. Nucl. Sci.* **53** (2006) 270.

[34] GEANT4 collaboration, *GEANT4: a simulation toolkit*, *Nucl. Instrum. Meth.* **A 506** (2003) 250 [insPIRE].

[35] M. Clemencic et al., *The LHCb simulation application, Gauss: design, evolution and experience*, *J. Phys. Conf. Ser.* **331** (2011) 032023 [insPIRE].

[36] A. Rogozhnikov, *Reweighting with boosted decision trees*, *J. Phys. Conf. Ser.* **762** (2016) 012036 [arXiv:1608.05806] [insPIRE].

[37] LHCb collaboration, *Determination of the quark coupling strength |V_{ub}| using baryonic decays*, *Nature Phys.* **11** (2015) 743 [arXiv:1504.01568] [insPIRE].

[38] L. Anderlini et al., *The PIDCalib package*, LHCb-PUB-2016-021 (2016).

[39] K.S. Cranmer, *Kernel estimation in high-energy physics*, *Comput. Phys. Commun.* **136** (2001) 198 [hep-ex/0011057] [insPIRE].

[40] LHCb collaboration, *Measurement of the track reconstruction efficiency at LHCb*, 2015 *JINST* **10** P02007 [arXiv:1408.1251] [insPIRE].

[41] S. Tolk, J. Albrecht, F. Dettori and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039 (2014).

[42] LHCb collaboration, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, *Phys. Rev. Lett.* **118** (2017) 052002 [Erratum *ibid.* **119** (2017) 169901] [arXiv:1612.05140] [insPIRE].
N.S. Nolte, C. Nunez, A. Oblakowska-Mucha, V. Obraztsov, S. Ogilvy, D.P. O’Hanlon, R. Oldeman, C.J.G. Onderwater, J. D. Osborn, A. Ossowska, J.M. Otalora Goicoechea, T. Ovsianikov, P. Owen, A. Oyanguren, P.R. Pais, T. Pajero, A. Palano, M. Palutan, G. Panšin, A. Papaleontis, M. Pappagallo, L.L. Pappalardo, C. Papenheimmer, W. Parker, M. Parkes, G. Passaleva, A. Pastore, M. Patel, M. Patrignani, A. Pearce, A. Pellegrino, M. Pepe Altarelli, S. Perazzini, D. Pereima, P. Perrot, L. Pescatore, K. Petridis, A. Petrolini, L. Petrucci, M. Petruzzi, B. Pietrzyk, G. Pietrzyk, M. Pikies, M. Pill, D. Pinci, J. Pinzino, F. Pisani, A. Pucci, V. Placinta, S. Playfer, J. Plews, M. Plo Casasus, F. Polci, M. Poli Lenzi, M. Poliakova, A. Poluektov, N. Polukhina, I. Polyakov, E. Polycarpou, G.J. Pomery, S. Ponce, A. Popov, D. Popov, S. Poslavskii, K. Prasanth, L. Promberger, C. Prouve, V. Pugatch, A. Puig Navarro, H. Pullen, G. Punzi, W. Qian, J. Qin, R. Quagliani, B. Quintana, N.V. Raab, R.I. Rabadan, T. Ranchi, J.H. Raedemaker, M. Rama, M. Ramos Peras, M.S. Rangeil, F. Ratnikov, G. Raven, M. Rebold, F. Redi, F. Reiss, C. Remon Alepu, Z. Ren, V. Renaudin, S. Ricciardi, S. Richards, K. Rinnert, P. Robbe, A. Robert, A.B. Rodrigues, E. Rodrigues, J.A. Rodriguez Lopez, M. Roehrken, S. Roiser, A. Rollings, V. Romanovskiy, M. Romero Lamas, A. Romero Vidal, J.D. Roth, M. Rotondo, M.S. Rudolph, T. Ruf, J. Ruiz Vidal, J. Ryzak, J.J. Saborido Silva, N. Sagioda, B. Saitta, C. Sanchez Gras, C. Sanchez Mayordomo, B. Sammartin, S. Sammut, R. Santacesaria, C. Santamarina Rios, M. Santamaria, E. Santovetti, G. Sapris, A. Sarti, C. Satriano, A. Satta, M. Saul, D. Savrina, L.G. Scantlebury Smead, S. Schaeb, M. Schellenberg, M. Schiller, H. Schindler, M. Schmoll, T. Schmelzer, B. Schmidt, A. Schopper, H.F. Schreiner, M. Schubiger, S. Schulte, M.H. Schune, R. Schwenner, B. Sciascia, A. Sciubba, S. Sellan, A. Semenikov, A. Sergi, N. Serra, J. Serrano, L. Sestini, A. Seuth, P. Seyfert, D.M. Shangase, M. Shapkin, T. Shears, L. Shekhitman, V. Shevchenko, E. Shmnin, J.D. Shupperd, B.G. Siddi, R. Silva Coutinho, L. Silva de Oliveira, G. Simi, S. Simone, I. Skiba, N. Skidmore, T. Skwarnicki, M.W. Slater, J.G. Smeaton, A. Smektina, E. Smith, I.T. Smith, M. Smith, A. Snoch, M. Soares, L. Soares Lavra, M.D. Sokoloff, F.J.P. Soler, B. Souza De Paula, B. Spaan, E. Spadaro Norella, P. Spradlin, F. Stagni, M. Stahl, S. Stahl, P. Steffke, S. Stefkova, O. Steinkamp, S. Stemmler, O. Stenyakin, M. Stepanova, H. Stevens, S. Stone, S. Stracka, M.E. Stramaglia, M. Straticiuc, S. Strokov, J. Sun, L. Sun, Y. Sun, P. Svishtov, K. Swientek, M. Szalewski, T. Szumlak, M. Szymanski, S. Taneja, T. Tekampe, G. Tellarini, F. Teubert, E. Thomas, K.A. Thomson, M.J. Tilley, V. Tisserand, S. T’Jampens, M. Tobin, S. Tolk, L. Tomassetti, D. Tonelli, D.Y. Tou, E. Tournefier, M. Traill, M.T. Tran, C. Tripple, A. Trisovic, A. Tsaregorodtsev, G. Tuci, A. Tully, N. Tuning, A. Ukleja, A. Usachov, A. Ustyuzhanin, U. Uwer, A. Vagner, V. Vagnoni, A. Valassi, G. Valenti, M. van Beuzekom, H. Van Hecke, E. Van Herwijnen, C.B. Van, M. Van Veghel, R. Vazquez Gomez, P. Vazquez Regueiro, C. Vázquez Sierra, S. Vecchi, J.J. Velthuis, M. Veltri, A. Venkateswaran, M. Vernet, M. Veronesi, M. Vesterinen, J.V. Viana Barbosa, D. Vieira, M. Vieites Diaz, H. Viemann, X. Vilasis-Cardona, A. Vitkovskiy, V. Volkov, A. Vollhardt, D. Vom Bruch, A. Vorobyev, V. Vorobyev, N. Voropaev, R. Wald, J. Walsh, J. Wang, J. Wang, M. Wang, Y. Wang, Z. Wang, D.R. Ward, H.M. Wark, N.K. Watson, D. Websdale, A. Weidt, C. Weissler, B.D.C. Westhenry, D.J. White, M. Whitehead, D. Wiedner, G. Wilkinson, M. Wilkinson, I. Williams, M. Williams.
