Search for $B^+ \to \mu^+ \nu_\mu$ and $B^+ \to \mu^+ N$ with inclusive tagging

M. T. Prim,[32] F. U. Bernlochner,[3] P. Goldenzweig,[32] M. Heck,[32] I. Adachi,[18,15] K. Adamczyk,[62] H. Aihara,[85] S. Al Said,[79,34] D. M. Asner,[4] H. Atmacan,[76] V. Aulchenko,[5,65] T. Aushev,[54] R. Ayad,[79] V. Babu,[9] A. M. Bakich,[78] V. Bansal,[67] P. Behera,[25] C. Beleño,[14] V. Bhardwaj,[22] B. Bhuyan,[23] T. Bilka,[6] J. Biswal,[31] A. Bobrov,[5,65] A. Bozek,[62] M. Bračko,[48,31] N. Braun,[32] T. E. Browder,[17] M. Campajola,[29,57] L. Cao,[32] D. Červenkov,[6] P. Chang,[61] V. Chekelian,[49] A. Chen,[59] B. G. Cheon,[16] K. Chilikin,[42] H. E. Cho,[16] K. Cho,[36] Y. Choi,[77] S. Choudhury,[24] D. Cinabro,[89] S. Cunliffe,[9] Z. Doležal,[6] S. Eidelman,[5,65,42] D. Epifanov,[5,65] J. E. Fast,[67] T. Ferber,[9] B. G. Fulsom,[67] R. Garg,[68] V. Gaur,[88] A. Garmash,[5,65] A. Giri,[24] O. Grzymkowska,[62] Y. Guan,[8] J. Haba,[18,15] T. Hara,[18,15] K. Hayasaka,[64] H. Hayashii,[58] W.-S. Hou,[61] T. Iijima,[56,55] K. Inami,[55] G. Inguglia,[9] A. Ishikawa,[18] M. Iwasaki,[66] Y. Iwasaki,[18] S. Jia,[2] Y. Jin,[85] D. Joffe,[33] K. K. Joo,[7] A. B. Kaliyar,[25] G. Karyan,[9] T. Kawasaki,[35] H. Kichimi,[18] C. Kiesling,[49] C. H. Kim,[16] D. Y. Kim,[75] K. T. Kim,[37] S. H. Kim,[16] K. Kinoshita,[8] P. Kodyš,[6] S. Korpar,[48,31] D. Kotchetkov,[17] P. Križan,[44,31] R. Kroeger,[51] P. Krokovny,[5,65] T. Kuhr,[45] R. Kulasiri,[33] R. Kumar,[70] T. Kumita,[87] A. Kuzmin,[5,65] Y.-J. Kwon,[91] K. Lalwani,[47] J. S. Lange,[12] I. S. Lee,[16] J. K. Lee,[73] J. Y. Lee,[73] S. C. Lee,[39] P. Lewis,[17] C. H. Li,[43] L. Li Gioi,[49] J. Libby,[25] K. Lieret,[45] D. Liventsev,[88,18] P.-C. Lu,[61] T. Luo,[11] J. MacNaughton,[52] M. Masuda,[84] T. Matsuda,[52] D. Matvienko,[5,65,42] M. Merola,[29,57] F. Metzner,[32] K. Miyabayashi,[58] R. Mizuk,[42,54] G. B. Mohanty,[80] R. Mussa,[30] M. Nakao,[18,15] G. De Nardo,[29,57] K. J. Nath,[23] Z. Natkaniec,[62] M. Nayak,[89,18] M. Niiyama,[38] N. K. Nisar,[69] S. Nishida,[18,15] K. Nishimura,[17] S. Ogawa,[82] H. Ono,[63,64] Y. Onuki,[85] P. Pakhlov,[42,53] G. Pakhlova,[42,54] B. Pal,[4] S. Pardi,[29] H. Park,[39] S.-H. Park,[91] S. Patra,[22] S. Paul,[81] T. K. Pedlar,[46] R. Pestotnik,[31] L. E. Piilonen,[88] V. Popov,[42,54] E. Prencipe,[20] M. Ritter,[45] A. Rostomyan,[9] M. Rozanska,[62] G. Russo,[57] D. Sahoo,[80] Y. Sakai,[18,15] L. Santelj,[18] V. Savinov,[69] O. Schneider,[41] G. Schnell,[1,21] J. Schueler,[17] C. Schwanda,[27] Y. Seino,[64] K. Senyo,[90] M. E. Sevior,[50] V. Shebalin,[17] J.-G. Shiu,[61] B. Shwartz,[5,65] F. Simon,[49] A. Sokolov,[28] E. Solovieva,[42] M. Starič,[31] J. F. Strube,[67] M. Sumihama,[13] T. Sumiyoshi,[87] W. Sutcliffe,[32] M. Takizawa,[74,19,71] U. Tamponi,[30] Y. Tao,[10] F. Tenchini,[9] K. Trabelsi,[40] M. Uchida,[86] T. Uglov,[42,54] Y. Unno,[16] S. Uno,[18,15] Y. Ushiroda,[18,15] Y. Usov,[5,65] S. E. Vahsen,[17] R. Van Tonder,[32] G. Varner,[17] K. E. Varvell,[78] A. Vinokurova,[5,65] B. Wang,[49] C. H. Wang,[60] M.-Z. Wang,[61] P. Wang,[26] S. Watanuki,[83] E. Won,[37] S. B. Yang,[37] H. Ye,[9] J. H. Yin,[26] Y. Yusa,[64] Z. P. Zhang,[72] V. Zhilich,[5,65] V. Zhukova,[42] and V. Zhulanov[5,65]

(Belle Collaboration)

[1]*University of the Basque Country UPV/EHU, 48080 Bilbao*
[2]*Beihang University, Beijing 100191*
[3]*University of Bonn, 53115 Bonn*
[4]*Brookhaven National Laboratory, Upton, New York 11973*
[5]*Budker Institute of Nuclear Physics SB RAS, Novosibirsk 630090*
[6]*Faculty of Mathematics and Physics, Charles University, 121 16 Prague*
[7]*Chonnam National University, Kwangju 660-701*
[8]*University of Cincinnati, Cincinnati, Ohio 45221*
[9]*Deutsches Elektronen–Synchrotron, 22607 Hamburg*
[10]*University of Florida, Gainesville, Florida 32611*
[11]*Key Laboratory of Nuclear Physics and Ion-beam Application (MOE) and Institute of Modern Physics, Fudan University, Shanghai 200443*
[12]*Justus-Liebig-Universität Gießen, 35392 Gießen*
[13]*Gifu University, Gifu 501-1193*
[14]*II. Physikalisches Institut, Georg-August-Universität Göttingen, 37073 Göttingen*
[15]*SOKENDAI (The Graduate University for Advanced Studies), Hayama 240-0193*
[16]*Hanyang University, Seoul 133-791*
[17]*University of Hawaii, Honolulu, Hawaii 96822*
[18]*High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801*
[19]*J-PARC Branch, KEK Theory Center, High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801*
[20]*Forschungszentrum Jülich, 52425 Jülich*
[21]*IKERBASQUE, Basque Foundation for Science, 48013 Bilbao*
[22]*Indian Institute of Science Education and Research Mohali, SAS Nagar, 140306*
[23]*Indian Institute of Technology Guwahati, Assam 781039*
[24]*Indian Institute of Technology Hyderabad, Telangana 502285*
[25]*Indian Institute of Technology Madras, Chennai 600036*

[26]*Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049*
[27]*Institute of High Energy Physics, Vienna 1050*
[28]*Institute for High Energy Physics, Protvino 142281*
[29]*INFN—Sezione di Napoli, 80126 Napoli*
[30]*INFN—Sezione di Torino, 10125 Torino*
[31]*J. Stefan Institute, 1000 Ljubljana*
[32]*Institut für Experimentelle Teilchenphysik, Karlsruher Institut für Technologie, 76131 Karlsruhe*
[33]*Kennesaw State University, Kennesaw, Georgia 30144*
[34]*Department of Physics, Faculty of Science, King Abdulaziz University, Jeddah 21589*
[35]*Kitasato University, Sagamihara 252-0373*
[36]*Korea Institute of Science and Technology Information, Daejeon 305-806*
[37]*Korea University, Seoul 136-713*
[38]*Kyoto University, Kyoto 606-8502*
[39]*Kyungpook National University, Daegu 702-701*
[40]*LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay*
[41]*École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015*
[42]*P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow 119991*
[43]*Liaoning Normal University, Dalian 116029*
[44]*Faculty of Mathematics and Physics, University of Ljubljana, 1000 Ljubljana*
[45]*Ludwig Maximilians University, 80539 Munich*
[46]*Luther College, Decorah, Iowa 52101*
[47]*Malaviya National Institute of Technology Jaipur, Jaipur 302017*
[48]*University of Maribor, 2000 Maribor*
[49]*Max-Planck-Institut für Physik, 80805 München*
[50]*School of Physics, University of Melbourne, Victoria 3010*
[51]*University of Mississippi, University, Mississippi 38677*
[52]*University of Miyazaki, Miyazaki 889-2192*
[53]*Moscow Physical Engineering Institute, Moscow 115409*
[54]*Moscow Institute of Physics and Technology, Moscow Region 141700*
[55]*Graduate School of Science, Nagoya University, Nagoya 464-8602*
[56]*Kobayashi-Maskawa Institute, Nagoya University, Nagoya 464-8602*
[57]*Università di Napoli Federico II, 80055 Napoli*
[58]*Nara Women's University, Nara 630-8506*
[59]*National Central University, Chung-li 32054*
[60]*National United University, Miao Li 36003*
[61]*Department of Physics, National Taiwan University, Taipei 10617*
[62]*H. Niewodniczanski Institute of Nuclear Physics, Krakow 31-342*
[63]*Nippon Dental University, Niigata 951-8580*
[64]*Niigata University, Niigata 950-2181*
[65]*Novosibirsk State University, Novosibirsk 630090*
[66]*Osaka City University, Osaka 558-8585*
[67]*Pacific Northwest National Laboratory, Richland, Washington 99352*
[68]*Panjab University, Chandigarh 160014*
[69]*University of Pittsburgh, Pittsburgh, Pennsylvania 15260*
[70]*Punjab Agricultural University, Ludhiana 141004*
[71]*Theoretical Research Division, Nishina Center, RIKEN, Saitama 351-0198*
[72]*University of Science and Technology of China, Hefei 230026*
[73]*Seoul National University, Seoul 151-742*
[74]*Showa Pharmaceutical University, Tokyo 194-8543*
[75]*Soongsil University, Seoul 156-743*
[76]*University of South Carolina, Columbia, South Carolina 29208*
[77]*Sungkyunkwan University, Suwon 440-746*
[78]*School of Physics, University of Sydney, New South Wales 2006*
[79]*Department of Physics, Faculty of Science, University of Tabuk, Tabuk 71451*
[80]*Tata Institute of Fundamental Research, Mumbai 400005*
[81]*Department of Physics, Technische Universität München, 85748 Garching*
[82]*Toho University, Funabashi 274-8510*
[83]*Department of Physics, Tohoku University, Sendai 980-8578*
[84]*Earthquake Research Institute, University of Tokyo, Tokyo 113-0032*
[85]*Department of Physics, University of Tokyo, Tokyo 113-0033*

[86]Tokyo Institute of Technology, Tokyo 152-8550
[87]Tokyo Metropolitan University, Tokyo 192-0397
[88]Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061
[89]Wayne State University, Detroit, Michigan 48202
[90]Yamagata University, Yamagata 990-8560
[91]Yonsei University, Seoul 120-749

(Received 10 November 2019; accepted 2 January 2020; published 21 February 2020)

We report the result for a search for the leptonic decay of $B^+ \to \mu^+\nu_\mu$ using the full Belle dataset of 711 fb^{-1} of integrated luminosity at the $\Upsilon(4S)$ resonance. In the Standard Model leptonic B-meson decays are helicity and Cabibbo-Kobayashi-Maskawa suppressed. To maximize sensitivity an inclusive tagging approach is used to reconstruct the second B meson produced in the collision. The directional information from this second B meson is used to boost the observed μ into the signal B-meson rest frame, in which the μ has a monochromatic momentum spectrum. Though its momentum is smeared by the experimental resolution, this technique improves the analysis sensitivity considerably. Analyzing the μ momentum spectrum in this frame we find $\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (5.3 \pm 2.0 \pm 0.9) \times 10^{-7}$ with a one-sided significance of 2.8 standard deviations over the background-only hypothesis. This translates to a frequentist upper limit of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 8.6 \times 10^{-7}$ at 90% confidence level. The experimental spectrum is then used to search for a massive sterile neutrino, $B^+ \to \mu^+ N$, but no evidence is observed for a sterile neutrino with a mass in a range of 0–1.5 GeV. The determined $B^+ \to \mu^+\nu_\mu$ branching fraction limit is further used to constrain the mass and coupling space of the type II and type III two-Higgs-doublet models.

DOI: 10.1103/PhysRevD.101.032007

I. INTRODUCTION

Precision measurements of leptonic decays of B mesons offer a unique tool to test the validity of the Standard Model (SM) of particle physics. Produced by the annihilation of the $\bar{b} - u$ quark pair and the subsequent emission of a virtual W^+ boson decaying into a antilepton and neutrino, this process is both Cabibbo-Kobayashi-Maskawa (CKM) and helicity suppressed in the SM. The branching fraction of the $B^+ \to \ell^+\nu_\ell$ [1] process is given by

$$\mathcal{B}(B^+ \to \ell^+\nu_\ell) = \frac{G_F^2 m_B m_\ell^2}{8\pi}\left(1 - \frac{m_\ell^2}{m_B^2}\right)^2 f_B^2 |V_{ub}|^2 \tau_B, \quad (1)$$

with G_F denoting Fermi's constant, m_B and m_ℓ the B meson and lepton masses, respectively, and $|V_{ub}|$ the relevant CKM matrix element of the process. Further, τ_B denotes the B-meson lifetime and the decay constant f_B parametrizes the $b - u$ annihilation process,

$$\langle 0|A^\mu|B(p)\rangle = ip^\mu f_B, \quad (2)$$

with $A^\mu = \bar{b}\gamma^\mu\gamma^5 u$ the corresponding axial-vector current and p^μ the B-meson four-momentum. The value of f_B has to be determined using nonperturbative methods, such as lattice QCD [2] or QCD sum-rule calculations [3,4].

In this paper an improved search for $B^+ \to \mu^+\nu_\mu$ using the full Belle dataset is presented. Using the results of $f_B = 184 \pm 4$ MeV [2] and either inclusive or exclusive world averages for $|V_{ub}|$ [5] one finds an expected SM branching fraction of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (4.3 \pm 0.8) \times 10^{-7}$ or $\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (3.8 \pm 0.4) \times 10^{-7}$, respectively. This implies an expected total of approximately 300 signal events in the entirety of the Belle dataset of 711 fb^{-1} of integrated luminosity recorded at the $\Upsilon(4S)$ resonance. Thus it is imperative to maximize the overall selection efficiency, which rules out the use of exclusive tagging algorithms,[1] as even advanced machine learning based implementations such as Ref. [6] only achieve efficiencies of a few percent. Events containing a high-momentum muon candidate are identified as potential signal events, and the additional charged particles and neutral energy depositions in the rest of the event (ROE) are used to reconstruct the second B meson produced in the collision process. With such an inclusive reconstruction one reduces the background due to nonresonant $e^+e^- \to q\bar{q}$ $(q = u, d, s, c)$ continuum processes, and, after a dedicated calibration, it is possible to deduce the direction of the signal B meson. This is used to carry out the search in the signal B rest frame, in which the $B^+ \to \mu^+\nu_\mu$ decay

[1]Exclusive here is referring to the explicit reconstruction of the second B meson in the $\Upsilon(4S) \to B^+[\to \mu^+\nu_\mu]B^-$ decay in a known hadronic or semileptonic decay chain.

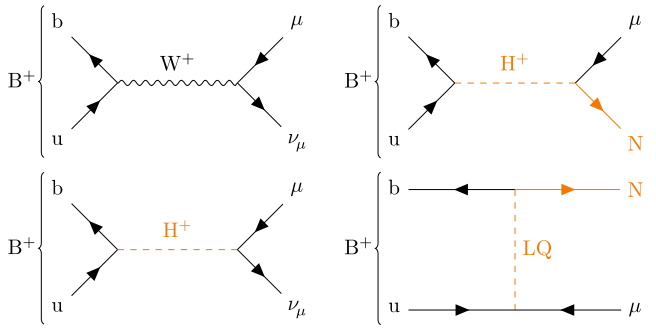

FIG. 1. The SM leptonic $B^+ \to \mu^+\nu_\mu$ decay process and possible BSM processes with and without a sterile neutrino N in the final state are shown.

produces a muon with a monochromatic momentum of $p_\mu^B = 2.64$ GeV. The experimental resolution on the boost vector reconstructed from ROE information broadens this signal signature. The use of this frame, which enhances the expected sensitivity of the search, is the main improvement over the preceding analysis, published in Ref. [7]. Further, the modeling of the crucial $b \to u\ell\nu_\ell$ semileptonic and continuum backgrounds has been improved with respect to the preceding analysis. In Ref. [7] a 90% confidence interval of $[2.9, 10.7] \times 10^{-7}$ for the $B^+ \to \mu^+\nu_\mu$ branching fraction was determined, while the most stringent 90% upper limit for this quantity that has been determined is 1×10^{-6} [8].

In the presence of new physics interactions or particles, the CKM and helicity suppression of the $B^+ \to \mu^+\nu_\mu$ decay can be lifted: the presence of, for instance, a charged Higgs boson, favored in many supersymmetric extensions of the SM, could strongly enhance the observed $B^+ \to \ell^+\nu_\ell$ branching fractions. Leptoquarks could have a similar effect. Another interesting exotic particle whose existence can be investigated with this decay are sterile neutrinos. This hypothetical particle acts as a singlet under the fundamental symmetry group of the SM, i.e., they carry no color charge, no weak isospin, nor weak hypercharge quantum numbers. Further, sterile neutrinos do not couple to the gauge bosons of the SM, but their existence could explain, for instance, the dark matter content of the Universe [9] or the smallness of the neutrino mass terms [10]. The only possibility for a sterile neutrino N to occur in a $B^+ \to \mu^+ N$ final state is due to the existence of a non-SM mediator. Further, the mass of the sterile neutrino has to be $m_N < 5.17$ GeV $= m_B - m_\mu$ and in the present analysis we are able to probe a mass range of $m_N \in [0, 1.5)$ GeV. In Fig. 1 the SM and a selection of beyond the SM (BSM) processes are shown.

The rest of this paper is organized as follows: Sec. II summarizes the used dataset, simulated samples and reconstruction steps. Section III outlines the inclusive tag reconstruction and calibration of its direction. In addition, the employed background suppression strategies and the used categorization are summarized. In Sec. IV the validation of the inclusive tag reconstruction and calibration using $B^+ \to \bar{D}^0\pi^+$ decays is described. Section V introduces the statistical methods used to determine the $B^+ \to \mu^+\nu_\mu$ signal yield. In Sec. VI systematic uncertainties of the measurement are discussed and Sec. VII documents sideband studies to validate the modeling of the crucial $b \to u\ell\nu_\ell$ semileptonic and continuum backgrounds. Section VIII presents the main findings of the paper. Finally, Sec. IX contains a summary and our conclusions.

II. DATA SET AND SIMULATED SAMPLES

We analyze the full Belle dataset of $(772 \pm 10) \times 10^6$ B-meson pairs, produced at the KEKB accelerator complex [11] with a center-of-mass energy (c.m.) of $\sqrt{s} = 10.58$ GeV at the $\Upsilon(4S)$ resonance. In addition, we use 79 fb^{-1} of collisions recorded 60 MeV below the $\Upsilon(4S)$ resonance peak to derive corrections and carry out cross-checks.

The Belle detector is a large-solid-angle magnetic spectrometer that consists of a silicon vertex detector, a 50-layer central drift chamber (CDC), an array of aerogel threshold Čerenkov counters (ACC), a barrel-like arrangement of time-of-flight (TOF) scintillation counters, and an electromagnetic calorimeter comprised of CsI(Tl) crystals (ECL) located inside a superconducting solenoid coil that provides a 1.5 T magnetic field. An iron flux return located outside of the coil is instrumented to detect K_L^0 mesons and to identify muons (KLM). A more detailed description of the detector, its layout and performance can be found in Ref. [12] and in references therein.

Charged tracks are identified as electron or muon candidates by combining the information of multiple subdetectors into a lepton identification likelihood ratio, $\mathcal{L}_{\text{LID}}$. For electrons the identifying features are the ratio of the energy deposition in the ECL with respect to the reconstructed track momentum, the energy loss in the CDC, the shower shape in the ECL, the quality of the geometrical matching of the track to the shower position in the ECL, and the photon yield in the ACC [13]. Muon candidates are identified from charged track trajectories extrapolated to the outer detector. The identifying features are the difference between expected and measured penetration depth as well as the transverse deviation of KLM hits from the extrapolated trajectory [14]. Charged tracks are identified as pions or kaons using a likelihood classifier which combines information from the CDC, ACC, and TOF subdetectors. In order to avoid the difficulties understanding the efficiencies of reconstructing K_L^0 mesons, they are not explicitly reconstructed in what follows.

Photons are identified as energy depositions in the ECL without an associated track. Only photons with an energy deposition of $E_\gamma > 100$, 150, and 50 MeV in the forward

endcap, backward endcap and barrel part of the calorimeter, respectively, are considered.

We carry out the entire analysis in the Belle II analysis software framework [15]: to this end the recorded Belle collision data and simulated Monte Carlo (MC) samples were converted using the software described in Ref. [16]. MC samples of B-meson decays and continuum processes are simulated using the EvtGen generator [17]. The used sample sizes correspond to approximately ten and six times the Belle collision data for B-meson and continuum decays, respectively. The interactions of particles traversing the detector are simulated using GEANT3 [18]. Electromagnetic final-state radiation (FSR) is simulated using the PHOTOS [19] package. The efficiencies in the MC are corrected using data-driven methods.

Signal $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \mu^+ N$ decays are simulated as two-body decays of a scalar initial-state meson to a lepton and a massless antineutrino. The effect of the nonzero sterile neutrino mass is incorporated by adjusting the kinematics of the simulated events.

The most important background processes are semileptonic $b \to u\ell\nu_\ell$ decays and continuum processes, which both produce high-momentum muons in a momentum range similar to the $B^+ \to \mu^+\nu_\mu$ process. Charmless semileptonic decays are produced as a mixture of specific exclusive modes and nonresonant contributions and are normalized to their corresponding world average branching fractions from Ref. [5]: Semileptonic $B \to \pi\ell^+\nu_\ell$ decays are simulated using the Bourrely-Caprini-Lellouch (BCL) parametrization [20] with form factor central values and uncertainties from the global fit carried out by Ref. [21]. The processes of $B \to \rho\ell^+\nu_\ell$ and $B \to \omega\ell^+\nu_\ell$ are modeled using the BCL form factor parametrization. We fit the measurements of Refs. [22–24] in combination with the light-cone sum-rule predictions of Ref. [25] to determine a set of form factor central values and uncertainties. The subdominant processes of $B \to \eta\ell^+\nu_\ell$ and $B \to \eta'\ell^+\nu_\ell$ are modeled using the ISGW2 model [26]. In addition to these narrow resonances, we produce nonresonant $b \to u\ell\nu_\ell$ decays with at least two pions in the final state using the De Fazio-Neubert (DFN) model [27]. In this model, the triple differential rate is regarded as a function of the four-momentum transfer squared (q^2), the lepton energy (E_ℓ^B), and the hadronic invariant mass squared (m_X^2) at next-to-leading order precision in the strong coupling constant α_s. The triple differential rate is convolved with a nonperturbative shape function using an *ad-hoc* exponential model. The free parameters in this model are the b quark mass in the $1S$ scheme, $m_b^{1S} = (4.69 \pm 0.04)$ GeV and a nonperturbative parameter $a = 1.9 \pm 0.5$. The values of these parameters were determined in Ref. [21] from a fit to $b \to c\ell\nu_\ell$ information. The nonperturbative parameter a is related to the average momentum squared of the b quark inside the B meson and controls the second moment of the shape function. It is defined as $a = \frac{3\bar{\Lambda}^2}{-\lambda_1} - 1$ with the binding energy $\bar{\Lambda} = m_B - m_b^{1S}$ and the hadronic matrix element expectation value λ_1. Hadronization of parton-level DFN predictions for the $b \to u\ell\nu_\ell$ process is accomplished using the JETSET algorithm [28] to produce two or more final-state mesons. The inclusive and exclusive $b \to u\ell\nu_\ell$ predictions are combined using a so-called "hybrid" approach, which is a method originally suggested by Ref. [29]: to this end we combine both predictions such that the partial branching fractions in the triple differential rate of the inclusive ($\Delta\mathcal{B}_{ijk}^{\text{incl}}$) and combined exclusive ($\Delta\mathcal{B}_{ijk}^{\text{excl}}$) predictions reproduce the inclusive values. This is achieved by assigning weights to the inclusive contributions w_{ijk} such that

$$\Delta\mathcal{B}_{ijk}^{\text{incl}} = \Delta\mathcal{B}_{ijk}^{\text{excl}} + w_{ijk} \times \Delta\mathcal{B}_{ijk}^{\text{incl}}, \tag{3}$$

with i, j, k denoting the corresponding bin in the three dimensions of q^2, E_ℓ^B, and m_X:

$$\begin{aligned} q^2 &= [0, 2.5, 5, 7.5, 10, 12.5, 15, 20, 25]\ \text{GeV}^2, \\ E_\ell^B &= [0, 0.5, 1, 1.25, 1.5, 1.75, 2, 2.25, 3]\ \text{GeV}, \\ m_X &= [0, 1.4, 1.6, 1.8, 2, 2.5, 3, 3.5]\ \text{GeV}. \end{aligned}$$

To study the model dependence of the DFN shape function and possible effects of next-to-next-to-leading order corrections in α_s, we also determine weights using the Bosch-Lange-Neubert-Paz (BLNP) model of Ref. [30].

The modeling of simulated continuum background processes is corrected using a data-driven method, which was first proposed in Ref. [31]: a boosted decision tree (BDT) is trained to distinguish between simulated continuum events and the recorded off-resonance data sample. This allows the BDT to learn differences between both samples, and a correction weight, $w = p/(1-p)$, accounting for differences in both samples can be derived directly from the classifier output p. As input for the BDT we use the same variables used in the continuum suppression approach (which is further detailed in Sec. III) and, additionally, the signal-side muon momentum in the signal B-meson frame.

The semileptonic background from $b \to c\ell\nu_\ell$ decays is dominated by $B \to D\ell^+\nu_\ell$ and $B \to D^*\ell^+\nu_\ell$ decays. $B \to D\ell^+\nu_\ell$ decays are modeled using the Boyd-Grinstein-Lebed (BGL) parametrization [32] with form factor central values and uncertainties taken from the fit in Ref. [33]. For $B \to D^*\ell^+\nu_\ell$ we use the BGL implementation proposed by Refs. [34,35] with form factor central values and uncertainties from the fit of the preliminary measurement of Ref. [36]. Both backgrounds are normalized to the isospin-combined world average branching fractions of Ref. [21]. The measurement is insensitive to the precise details of the modeling of $b \to c\ell\nu_\ell$ involving higher charm resonances.

TABLE I. Used branching fractions for the main background processes are listed. All values are from Ref. [5], with the exception of the $b \to c\ell\nu_\ell$ branching fractions, which are taken from the isospin-combined world average from Ref. [21].

$\mathcal{B}$	Value B^+	Value B^0
$b \to u\ell\nu_\ell$		
$B \to \pi\ell^+\nu_\ell$	$(7.8 \pm 0.3) \times 10^{-5}$	$(1.5 \pm 0.06) \times 10^{-4}$
$B \to \eta\ell^+\nu_\ell$	$(3.9 \pm 0.5) \times 10^{-5}$	⋯
$B \to \eta'\ell^+\nu_\ell$	$(2.3 \pm 0.8) \times 10^{-5}$	⋯
$B \to \omega\ell^+\nu_\ell$	$(1.2 \pm 0.1) \times 10^{-4}$	⋯
$B \to \rho\ell^+\nu_\ell$	$(1.6 \pm 0.1) \times 10^{-4}$	$(2.9 \pm 0.2) \times 10^{-4}$
$B \to X_u\ell^+\nu_\ell$	$(2.2 \pm 0.3) \times 10^{-3}$	$(2.0 \pm 0.3) \times 10^{-3}$
$b \to c\ell\nu_\ell$		
$B \to D\ell^+\nu_\ell$	$(2.3 \pm 0.1) \times 10^{-2}$	$(2.1 \pm 0.1) \times 10^{-2}$
$B \to D^*\ell^+\nu_\ell$	$(5.3 \pm 0.1) \times 10^{-2}$	$(4.9 \pm 0.1) \times 10^{-2}$
$B^+ \to \mu^+\nu_\mu\gamma$	$(1.0 \pm 1.3) \times 10^{-6}$	⋯
$b \to s/d$		
$B \to K_L^0\pi^+$	$(2.4 \pm 0.1) \times 10^{-5}$	⋯
$B \to K^+\pi^0$	$(1.3 \pm 0.1) \times 10^{-5}$	⋯
$B \to \rho^+\pi^-$	⋯	$(2.3 \pm 0.2) \times 10^{-5}$

For the contributions of $B^+ \to \mu^+\nu_\mu\gamma$ we use the recent experimental bounds of Ref. [37]. In this process, structure-dependent corrections, which are suppressed by the electromagnetic coupling constant $\alpha_{\rm em}$, lift the helicity suppression of the $B^+ \to \mu^+\nu_\mu$ decay. We simulate this process using the calculation of Ref. [38] and only allow daughter photons with $E_\gamma > 300$ MeV, to avoid overlap with the FSR corrections simulated by PHOTOS as corrections to the $B^+ \to \mu^+\nu_\mu$ final state. In the following, we treat these two processes separately.

The small amount of background from rare $b \to s/d$ processes is dominated by $B^+ \to K_L^0\pi^+$ decays. Subdominant contributions are given by the decays $B^+ \to K^+\pi^0$ and $B^0 \to \rho^+\pi^-$. We adjust those branching fractions to the latest averages of Ref. [5].

Table I summarizes the branching fractions used for all important background processes.

III. ANALYSIS STRATEGY, INCLUSIVE TAG RECONSTRUCTION AND CALIBRATION

We select $B\bar{B}$ candidate events by requiring at least three charged particles to be reconstructed and a significant fraction of the c.m. energy to be deposited in the ECL. We first reconstruct the signal side: a muon candidate with a momentum of $p_\mu^* > 2.2$ GeV in the c.m. frame of the colliding e^+e^- pair. For $B^+ \to \mu^+\nu_\mu$ signal decays we expect p_μ^* to range from approximately 2.4–2.8 GeV, cf. Fig. 2, due to the nonzero B-meson momentum. The candidate is required to have a distance of closest approach to the nominal interaction point transverse to and along the beam axis of $\mathrm{d}r < 0.5$ cm and $|\mathrm{d}z| < 2$ cm, respectively. This initial selection results in a signal-side efficiency of ≈82.2%. After this the remaining charged tracks and neutral depositions are used to reconstruct the ROE to allow us to boost this signal muon candidate into the rest frame of the signal-side B meson. A looser selection on the ROE tracks is imposed, $\mathrm{d}r < 10$ cm and $|\mathrm{d}z| < 20$ cm, to also include charged particle candidates which are displaced from the interaction region. All ROE charged particles are treated as pions and no further particle identification is performed. Track candidates with a transverse momentum of $p_T < 275$ MeV do not leave the CDC, but curl back into the detector. To avoid double counting of those tracks, we check if such are compatible with another track. If the track parameters indicate that this is the case, we veto the lower momentum track. When we combine the momentum information with ROE photon candidates (reconstructed as described in Sec. II) we determine the three-momentum ($\mathbf{p}_{\rm tag}^{\rm lab}$) and energy ($E_{\rm tag}^{\rm lab}$) of the tag-side B meson in the laboratory frame as

$$\mathbf{p}_{\rm tag}^{\rm lab} = \sum_i^{\rm tracks} \mathbf{p}_i^{\rm lab} + \sum_j^{\rm photons} \mathbf{E}_j^{\rm lab},$$
$$E_{\rm tag}^{\rm lab} = \sqrt{(\mathbf{p}_{\rm tag}^{\rm lab})^2 + m_B^2}. \quad (4)$$

Here $\mathbf{p}_i^{\rm lab}$ and $\mathbf{E}_j^{\rm lab}$ denote the three-momentum of tracks and photons in the ROE. We proceed by boosting the tag-side four-vector into the c.m. frame of the e^+e^- collision. Due to the two-body nature of the $\Upsilon(4S) \to B\bar{B}$ decay, we have precise knowledge of the magnitude of tag- and signal-side B meson in this frame: $|\mathbf{p}_B^*| = 330$ MeV. We thus correct after the boost the energy component of the tag-side four-vector to be exactly

$$E_{\rm tag}^* = \sqrt{(\mathbf{p}_B^*)^2 + m_B^2} = \sqrt{(330\ \mathrm{MeV})^2 + m_B^2}, \quad (5)$$

keeping the direction of the three-momentum unchanged. This improves the resolution with respect to using the boosted absolute three-momentum $\mathbf{p}_{\rm tag}^*$. Due to the asymmetric beam energies of the colliding e^+e^- pair, all produced B-meson decay products are boosted in the positive z direction in the laboratory frame. Thus it is more likely that charged and neutral particles escape the Belle detector acceptance in the forward region and bias the inclusive tag reconstruction. This bias degrades the resolution of the reconstructed z component of the $\mathbf{p}_{\rm tag}^*$ momentum vector. The resolution is significantly improved by applying a calibration function derived from simulated $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ decays, where one B decays into a $\mu\nu_\mu$ pair. The goal of this function is to map the reconstructed mean momentum z component, $(\mathbf{p}_{\rm tag}^*)_z$, to the mean of the simulated true distribution. The functional dependence between the reconstructed and true momentum z component is shown in Fig. 3. In addition, an overall

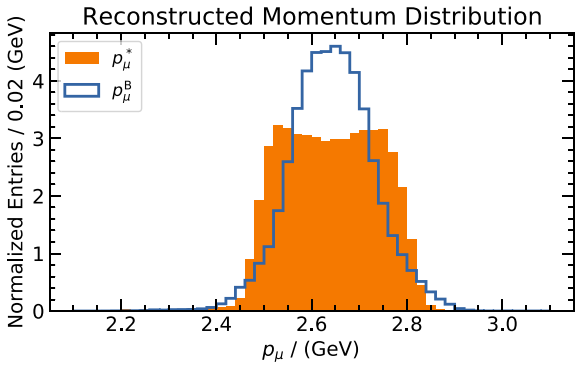

FIG. 2. The signal resolution of $B^+ \to \mu^+\nu_\mu$ is compared for signal events reconstructed in the e^+e^- c.m. (p^*_μ) and the signal B rest frame (p^B_μ).

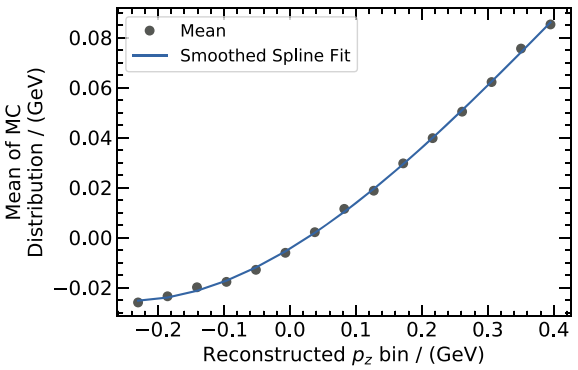

FIG. 3. The functional dependence between the reconstructed and true momentum z component of the inclusively reconstructed tag-side B meson is shown.

correction factor ζ is applied to the calibrated three-momentum, chosen such that the difference between the corrected and the simulated three-momentum becomes minimal. The corrected tag-side z and transverse momentum components are then

$$(\mathbf{p}^*_{\mathrm{tag,cal}})_z = \zeta f[(\mathbf{p}^*_{\mathrm{tag}})_z],$$
$$(\mathbf{p}^*_{\mathrm{tag,cal}})_T = \zeta\sqrt{(\mathbf{p}^*_{\mathrm{tag}})^2 - (\mathbf{p}^*_{\mathrm{tag,corr}})^2_z}, \quad (6)$$

with f the calibration function. The absolute difference between corrected and simulated three-momentum is found to be minimal for $\zeta = 0.58$. Using the calibrated tag-side B-meson three-momentum $\mathbf{p}^*_{\mathrm{tag,cal}}$, we boost the signal-side muon candidate into the signal-side B-meson rest frame using

$$\mathbf{p}_{\mathrm{sig}} = -\mathbf{p}^*_{\mathrm{tag,cal}}. \quad (7)$$

Figure 2 compares the muon momentum spectrum for signal $B^+ \to \mu^+\nu_\mu$ decays in the e^+e^- c.m. frame with the obtained resolution in the B rest frame (further denoted as p^B_μ) using the calibrated momentum vector. Carrying out the boost into the approximated B-meson rest frame improves the resolution of the reconstructed muon momentum by 7% with respect to the resolution in the c.m. frame.

To reduce the sizable background from continuum processes, a multivariate classifier using an optimized implementation of gradient BDTs [39] is used and trained to distinguish $B^+ \to \mu^+\nu_\mu$ signal decays from continuum processes. The BDT exploits the fact that the event topology for nonresonant e^+e^--collision processes differ significantly from the resonant $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ process. Event shape variables, such as the magnitude of the thrust of final-state particles from both B mesons, the reduced Fox-Wolfram moment R_2, the modified Fox-Wolfram moments [40] and CLEO cones [41], are highly discriminating. To these variables we add as additional inputs to the BDT the number of tracks in the ROE, the number of leptons (electrons or muons) in the ROE, the normalized beam-constrained mass of the tag-side B meson defined as

$$\hat{m}^{\mathrm{tag}}_{bc} = \sqrt{s/4 - (\mathbf{p}^*_{\mathrm{tag,cal}})^2}/(\sqrt{s}/2), \quad (8)$$

and the normalized missing energy defined as

$$\Delta\hat{E} = (E^*_{\mathrm{tag,reco}} - \sqrt{s}/2)/(\sqrt{s}/2), \quad (9)$$

with $E^*_{\mathrm{tag,reco}}$ denoting the energy from boosting the ROE four-vector from the laboratory into the c.m. frame. This list of variables and p^B_μ are used in the data-driven correction described in Sec. II to correct the simulated continuum events. We apply a loose set of ROE preselection cuts: only events with at least two tracks, fewer than three leptons, $\hat{m}^{\mathrm{tag}}_{bc} > 0.96$, $\Delta\hat{E} \in [-0.5, 0.1)$, and $R_2 < 0.5$ are further considered. Figure 4 compares the classifier output C_{out} and p^B_μ distributions of the simulated and corrected continuum contribution with recorded off-resonance collision events. Both variables show good agreement.

Using this classifier and the cosine of the angle between the calibrated signal B meson in the c.m. system and the muon in the B rest frame ($\cos\Theta_{B\mu}$) we define four mutually exclusive categories. The first two of these are signal enriched categories with $C_{\mathrm{out}} \in [0.98, 1)$ and split with respect to their $\cos\Theta_{B\mu}$ values. For $B^+ \to \mu^+\nu_\mu$ signal decays no preferred direction in $\cos\Theta_{B\mu}$ is expected. For the semileptonic and continuum background events, which pass the selection, the muons are emitted more frequently in the direction of the reconstructed B-meson candidate. The second two categories have $C_{\mathrm{out}} \in [0.93, 0.98)$, and they help separate $b \to u\ell\nu_\ell$ and continuum processes from $B^+ \to \mu^+\nu_\mu$ signal decays. Table II summarizes the four categories. The chosen cut values were determined using a grid search and by fits to Asimov datasets (using the fit procedure further described in Sec. V).

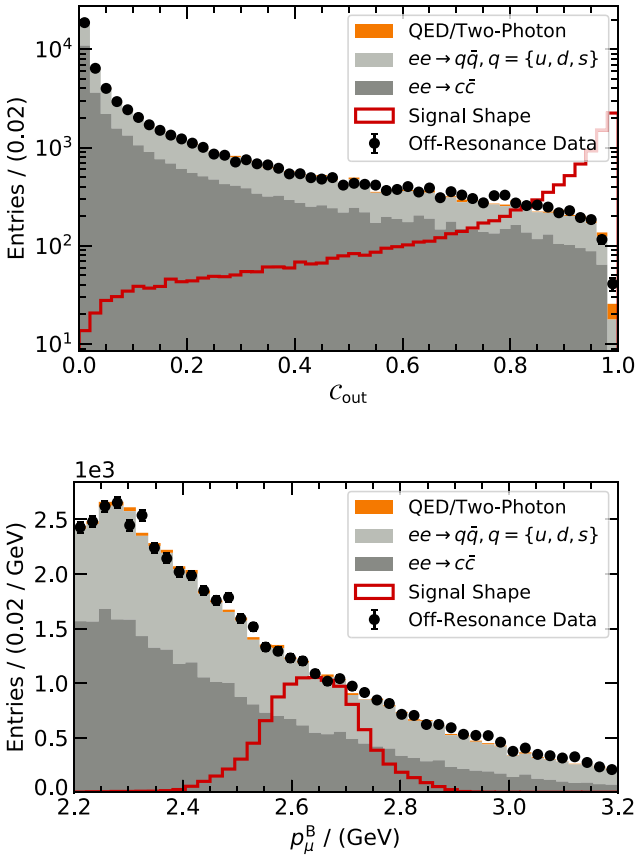

FIG. 4. The classifier output C_{out} and the p_μ^B distribution of off-resonance data are compared to the simulated continuum background after applying the correction described in Sec. II.

In Sec. VII the signal-depleted region of $C_{\text{out}} \in [0.9, 0.93)$ is analyzed and simultaneous fits in two categories, $\cos\Theta_{B\mu} < 0$ and $\cos\Theta_{B\mu} > 0$ are carried out

TABLE II. The definition of the four signal categories is shown.

Category	C_{out}	$\cos\Theta_{B\mu}$	Signal efficiency
I	[0.98, 1.00)	[−0.13, 1.00)	6.5%
II	[0.98, 1.00)	[−1.00, −0.13)	5.9%
III	[0.93, 0.98)	[0.04, 1.00)	7.1%
IV	[0.93, 0.98)	[−1.00, 0.04)	8.3%

TABLE III. The cumulative selection efficiencies of $B^+ \to \mu^+\nu_\mu$ signal decays and dominant background processes throughout the selection is listed. See text for details about the various selection steps.

Efficiency	$B^+ \to \mu^+\nu_\mu$	$b \to u\ell\nu_\ell$	Continuum
$B\bar{B}$ & Muon reco.	82%	10%	0.9%
ROE Presel.	55%	1.4%	0.03%
C_{out} cut	28%	0.2%	0.001%

to validate the modeling of the important $b \to u\ell\nu_\ell$ background and to extract a value of the inclusive $\mathcal{B}(B \to X_u\ell^+\nu)$ branching fraction. The selection efficiencies of $B^+ \to \mu^+\nu_\mu$ signal and the background processes are summarized in Table III.

IV. INCLUSIVE TAG VALIDATION USING $B^+ \to \bar{D}^0\pi^+$ DECAYS

In order to validate the quality of the inclusive tag reconstruction and rule out possible biases introduced by the calibration method, we study the hadronic two-body decay of $B^+ \to \bar{D}^0\pi^+$ with $\bar{D}^0 \to K^+\pi^-$. Due to the absence of any neutrino in this decay, we are able to fully reconstruct the B^+ four-vector and boost the prompt π^+ into the B^+ rest frame. Alternatively, we use the ROE, as outlined in the previous section, to reconstruct the very same information. Comparing the results from both allows us to determine if the calibration introduces potential biases and also to validate the signal resolution predicted in the simulation. In addition, we use this dataset to test the validity of the continuum suppression and the data-driven continuum corrections outlined in Sec. II.

We reconstruct the $B^+ \to \bar{D}^0\pi^+$ with $\bar{D}^0 \to K^+\pi^-$ using the same impact parameter requirements used in the $B^+ \to \mu^+\nu_\mu$ analysis. For the prompt π^+ candidate we require a momentum of more than 2.1 GeV in the c.m. frame. For the $\bar{D}^0$ decay product candidates a looser requirement is imposed, selecting charged tracks with a three-momentum of at least 0.3 GeV in the laboratory frame. To identify the kaon and pion candidates, we use the particle identification methods described in Sec. II. To further suppress contributions from background processes we require that the reconstructed $\bar{D}^0$ mass is to be within 50 MeV of its expected value. Using the reconstructed four-vector of the $B^+ \to \bar{D}^0\pi^+$ candidate we impose additional cuts to enhance the purity of the selected sample by using the beam-constrained mass and energy difference:

$$m_{bc} = \sqrt{s/4 - (\mathbf{p}^*_{B^+})^2} > 5.2 \text{ GeV},$$
$$|\Delta E| = |E^*_{B^+} - \sqrt{s}/2| < 0.2 \text{ GeV}. \quad (10)$$

Here $\mathbf{p}^*_{B^+}$ and $E^*_{B^+}$ denote the reconstructed B^+ three-momentum and energy in the c.m. frame of the colliding e^+e^- pair, respectively. The inclusive tag is reconstructed in the same way as outlined in the previous section and Fig. 5 shows the reconstructed prompt π^+ absolute three-momentum p_π^B after using the inclusive tag information to boost into the B^+-meson frame of rest. The simulated and reconstructed $B^+ \to \bar{D}^0\pi^+$ decays show good agreement. Using the signal-side information, we also reconstruct the residual $\Delta p_\pi^B = p_\pi^B - p_\pi^{B_{\text{sig}}}$, with $p_\pi^{B_{\text{sig}}}$ denoting the absolute three-momentum in the B^+ rest frame when reconstructed using the signal-side B^+ decay chain. The mean

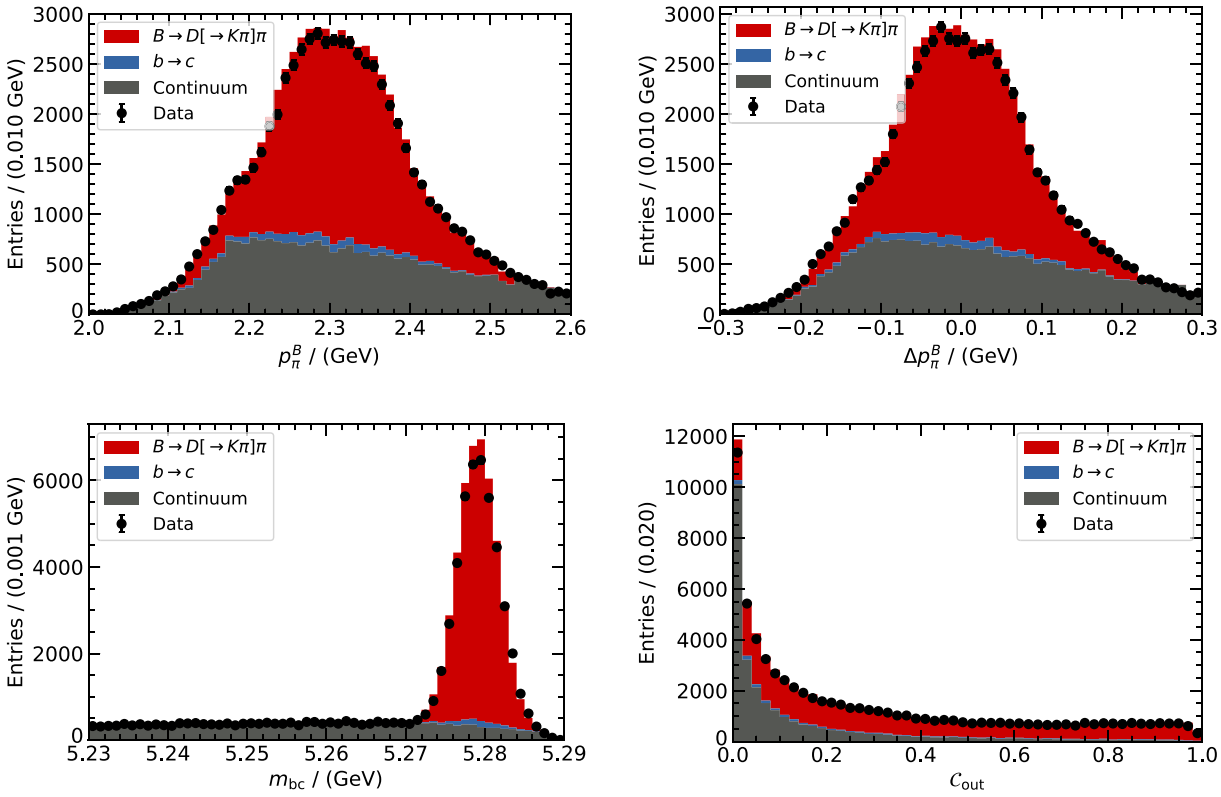

FIG. 5. The p_π^B distribution and the residual Δp_π^B for $B^+ \to \bar{D}^0\pi^+$ decays with $\bar{D}^0 \to K^+\pi^-$ are shown in the reconstructed rest frame of the B^+ meson. The p_π^B distribution is derived from the inclusive tag reconstruction method described in the text and the residual shows the difference with respect to using the full B^+ decay chain to determine the same information. In addition, m_{bc} and the continuum classifier of simulated and reconstructed collision events are compared.

and variance of this distribution between simulated and reconstructed samples show good agreement and are compatible within their statistical uncertainties. We obtain a data-driven estimate for the inclusive tag resolution for p_π^B of 0.11 GeV. In addition, Fig. 5 shows the reconstructed m_{bc} distribution.

To validate the response of the multivariate classifier used to suppress continuum events, we remove the reconstructed $\bar{D}^0$ decay products from the signal side to emulate the $B^+ \to \mu^+\nu_\mu$ decay topology. Using the same BDT weights as for $B^+ \to \mu^+\nu_\mu$ we then recalculate the classifier output $C_{\rm out}$. Its distribution is shown in Fig. 5 and simulated and reconstructed events are in good agreement. In Table IV we further compare the selection efficiency denoted as ϵ between simulated and reconstructed events for the four signal selection categories of the $B^+ \to \mu^+\nu_\mu$ analysis. The efficiency is defined as the fraction of reconstructed candidates with $C_{\rm out} > 0.93$ or 0.98, respectively, with respect to the total number of reconstructed candidates. The efficiency from simulated and reconstructed events are in agreement within their statistical uncertainty and we do not assign additional corrections or uncertainties to the $B^+ \to \mu^+\nu_\mu$ analysis in the following.

TABLE IV. Selection efficiencies of the category cuts defined in Table II of simulated and reconstructed data events. The quoted uncertainty is the statistical error.

Categories	I − IV	I + II	III + IV
$\epsilon^{\rm Data}$	0.030 ± 0.001	0.0047 ± 0.0003	0.024 ± 0.001
$\epsilon^{\rm MC}$	0.030 ± 0.001	0.0051 ± 0.0003	0.025 ± 0.001

V. STATISTICAL ANALYSIS AND LIMIT SETTING PROCEDURE

In order to determine the $B^+ \to \mu^+\nu_\mu$ or $B^+ \to \mu^+ N$ signal yield and to constrain all background yields, we perform a simultaneous binned likelihood fit to the p_μ^B spectra using the four event categories defined in Sec. III. The total likelihood function we consider has the form

$$\mathcal{L} = \prod_c \mathcal{L}_c \times \prod_k \mathcal{G}_k, \tag{11}$$

with the individual category likelihoods $\mathcal{L}_c$ and nuisance-parameter (NP) constraints $\mathcal{G}_k$. The product in Eq. (11) runs over all categories c and fit components k, respectively. The role of the NP constraints is detailed in Sec. VI.

Each category likelihood $\mathcal{L}_c$ is defined as the product of individual Poisson distributions $\mathcal{P}$,

$$\mathcal{L}_c = \prod_i^{\text{bins}} \mathcal{P}(n_i; \nu_i), \tag{12}$$

with n_i denoting the number of observed data events and ν_i the total number of expected events in a given bin i. We divide the muon momentum spectrum into 22 equal bins of 50 MeV, ranging over $p_\mu^B \in [2.2, 3.3)$ GeV, and the number of expected events in a given bin, ν_i is estimated using simulated collision events. It is given by

$$\nu_i = \sum_k^{\text{processes}} f_{ik}\eta_k, \tag{13}$$

with η_k the total number of events from a given process k with the fraction f_{ik} of such events being reconstructed in the bin i.

The likelihood Eq. (11) is numerically maximized to fit the value of four different components η_k using the observed events using the sequential least squares programming method implementation of Ref. [42].

The four components we determine are:

(1) Signal $B^+ \to \mu^+\nu_\mu$ events.
(2) Background $b \to u\ell\nu_\ell$ events; simulated as described in Sec. II.
(3) Background $b \to c\ell\nu_\ell$ events, dominated by $B \to D^{(*)}\ell^+\nu_\ell$ decays and simulated as described in Sec. II.
(4) Background continuum events, dominated by $e^+e^- \to q\bar{q}$ and $e^+e^- \to \tau^+\tau^-$ processes.

Two additional background components, $B^+ \to \mu^+\nu_\mu\gamma$ and other rare $b \to s$ processes, are constrained in the fit to the measurement of Ref. [37] and world averages of Ref. [5]. Both mimic the signal shape and are allowed to vary in the fit within their corresponding experimental uncertainties. Further details on how this is implemented are found in Sec. VI.

We construct confidence levels for the components using the profile likelihood ratio method. For a given component η_k the ratio is

$$\Lambda(\eta_k) = -2\ln\frac{\mathcal{L}(\eta_k, \hat{\hat{\boldsymbol{\eta}}}_{\eta_k}, \hat{\hat{\boldsymbol{\theta}}}_{\eta_k})}{\mathcal{L}(\hat{\eta}_k, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}})}, \tag{14}$$

where $\hat{\eta}_k$, $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\theta}}$ are the values of the component of interest, the remaining components, and a vector of nuisance parameters that unconditionally maximize the likelihood function, whereas the remaining components $\hat{\hat{\boldsymbol{\eta}}}_{\eta_k}$ and nuisance parameters $\hat{\hat{\boldsymbol{\theta}}}_{\eta_k}$ maximize the likelihood under the condition that the component of interest is kept fixed at a given value η_k. In the asymptotic limit, the test statistic Eq. (14) can be used to construct approximate confidence intervals (CI) through

$$1 - \text{CI} = \int_{\Lambda(\eta_k)}^{\infty} f_{\chi^2}(x; 1\,\text{d.o.f.})\text{d}x, \tag{15}$$

with $f_{\chi^2}(x; 1\,\text{d.o.f.})$ denoting the χ^2 distribution with a single degree of freedom. In the absence of a significant signal, we determine frequentist and Bayesian limits. For the frequentist one-sided (positive) limit, we modify our test statistic according to Refs. [43,44] to

$$q_0(\eta_k) = \begin{cases} \Lambda(\eta_k) & \eta_k \geq 0 \\ -\Lambda(\eta_k) & \eta_k < 0 \end{cases}, \tag{16}$$

to maximize our sensitivity. This test statistic is asymptotically distributed as

$$f(q_0) = \frac{1}{2}f_{\chi^2}(-q_0; 1\,\text{d.o.f.}) + \frac{1}{2}f_{\chi^2}(q_0; 1\,\text{d.o.f.}) \tag{17}$$

and with an observed value q_0^{obs} we evaluate the (local) probability of an observed signal, p_0, as

$$p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_0)\text{d}q_0. \tag{18}$$

For the Bayesian limit, we convert the likelihood Eq. (11) using a vector of observed event yields in the given bins of all categories $\boldsymbol{n}$ [denoted as $\mathcal{L} = \mathcal{L}(\boldsymbol{n}|\eta_k)$ in the following] into a probability density function $\mathcal{F}$ of the parameter of interest η_k using a flat prior $\pi(\eta_k)$ to exclude unphysical negative branching fractions. This $\mathcal{L}$ is numerically maximized for given values of the parameter of interest η_k, by floating the other components and nuisance parameters. The probability density function $\mathcal{F}$ is then given by

$$\mathcal{F}(\eta_k|\boldsymbol{n}) = \frac{\mathcal{L}(\boldsymbol{n}|\eta_k)\pi(\eta_k)}{\int_0^\infty \mathcal{L}(\boldsymbol{n}|\eta_k)\pi(\eta_k)\text{d}\eta_k}, \tag{19}$$

with the prior $\pi(\eta_k) = \text{constant}$ for $\eta_k \geq 0$ and zero otherwise.

To quote the significance over the background-only hypothesis for the search for $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \mu^+N$ we adapt Eq. (16) and set $\eta_k = 0$. For the search for a heavy sterile neutrino we do not account for the look-elsewhere effect.

We validate the fit procedure using ensembles of pseudoexperiments generated for different input branching fractions for $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \mu^+N$ decays and observe no biases, undercoverage or overcoverage of CI.

Using a SM branching fraction of $\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (4.3 \pm 0.8) \times 10^{-7}$, calculated assuming an average value of $|V_{ub}| = (3.94 \pm 0.36) \times 10^{-3}$ [5], we construct Asimov

datasets for all four categories. These are used to determine the median expected significance of our analysis. We find a value of $2.4^{+0.8}_{-0.9}$ standard deviations incorporating all systematic uncertainties and $2.6^{+1.0}_{-0.9}$ standard deviations if we only consider statistical uncertainties. The quoted uncertainties on the median expected significance correspond to the 68% confidence level (CL) intervals.

VI. SYSTEMATIC UNCERTAINTIES

There are several systematic uncertainties that affect the search for $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \mu^+ N$. The most important uncertainty stems from the modeling of the dominant semileptonic $b \to u\ell\nu_\ell$ background decays. As we determine the overall normalization of these decays directly from the measured collision events, we only need to evaluate shape uncertainties. The most important here stem from the modeling of the $B \to \pi\ell^+\nu_\ell$, $B \to \rho\ell^+\nu_\ell$, and $B \to \omega\ell^+\nu_\ell$ form factors, the branching fractions for these processes, $B \to \eta\ell^+\nu_\ell$, $B \to \eta'\ell^+\nu_\ell$ and inclusive $b \to u\ell\nu_\ell$ decays. The uncertainty of the nonresonant $b \to u\ell\nu_\ell$ contributions in the hybrid model approach is estimated by changing the underlying model from DFN to BLNP. In addition, the uncertainty on the DFN parameters m_b^{1S} and a are included in the shape uncertainty (see Sec. II). There is no sizable shape uncertainty contribution owing to either muon identification or track reconstruction. The second most important uncertainty for the reported results is from the shape of the continuum template: the off-resonance data sample, which was used to correct the simulated continuum events, introduces additional statistical uncertainties. We evaluate the size of these using a bootstrapping procedure. The $b \to c\ell\nu_\ell$ background near the kinematic endpoint for such decays is dominated by $B \to D\ell^+\nu_\ell$ and $B \to D^*\ell^+\nu_\ell$ decays. We evaluate the uncertainties in the used BGL form factors and their branching fractions for both channels. For the $B^+ \to \mu^+\nu_\mu$ signal, and the fixed backgrounds from $B^+ \to \mu^+\nu_\mu\gamma$ and rare $b \to s$ processes, we also evaluate the impact on the efficiency of the lepton-identification uncertainties, the number of produced B-meson pairs in the Belle dataset, and the overall tracking efficiency uncertainty. In addition, we propagate the experimental uncertainty on the used $B^+ \to \mu^+\nu_\mu\gamma$ branching fraction. The rare $b \to s/d$ template is dominated by $B^+ \to K_L^0\pi^+$ events (which make up about 32% of all selected events) and we assign an uncertainty on the measured branching fraction and the two next-most occurring decay channels, $B^+ \to K^+\pi^0$ (5%) and $B^0 \to \rho^+\pi^-$ (4%), in the template. The statistical uncertainty on the generated MC samples is also evaluated and taken into account. A full listing of the systematic uncertainties is found in Table V.

The effect of systematic uncertainties is directly incorporated into the likelihood function. For this we introduce a vector of NPs, $\boldsymbol{\theta}_k$, for each fit template k. Each vector element represents one bin of the fitted p_μ^B spectrum in all four categories. The NPs are constrained in the likelihood Eq. (11) using multivariate Gaussian distributions $\mathcal{G}_k = \mathcal{G}_k(\mathbf{0}; \boldsymbol{\theta}_k, \Sigma_k)$, with Σ_k denoting the systematic covariance matrix for a given template k. The systematic covariance is constructed from the sum over all possible uncertainty sources affecting a template k, i.e.,

$$\Sigma_k = \sum_s^{\text{error sources}} \Sigma_{ks}, \tag{20}$$

with Σ_{ks} the covariance matrix of error source s which depends on an uncertainty vector $\boldsymbol{\sigma_{ks}}$. The vector elements of $\boldsymbol{\sigma_{ks}}$ represent the absolute error in bins of p_μ^B of fit template k across the four event categories. We treat uncertainties from the same error source either as fully correlated, or, for MC or other statistical uncertainties as uncorrelated, such that $\Sigma_{ks} = \boldsymbol{\sigma_{ks}} \otimes \boldsymbol{\sigma_{ks}}$ or $\Sigma_{ks} = \text{Diag}(\boldsymbol{\sigma_{ks}}^2)$. The impact of nuisance parameters is included in Eq. (13) as follows. First, the fractions f_{ik} for all templates are rewritten as

$$f_{ik} = \frac{\eta_{ik}^{\text{MC}}}{\sum_j \eta_{jk}^{\text{MC}}} \to \frac{\eta_{ik}^{\text{MC}}(1+\theta_{ik})}{\sum_j \eta_{jk}^{\text{MC}}(1+\theta_{jk})}, \tag{21}$$

TABLE V. The fractional uncertainty on the extract $B^+ \to \mu^+\nu_\mu$ branching fraction is shown. For definitions of additive and multiplicative errors please see text.

Source of uncertainty	Fractional uncertainty $B^+ \to \mu^+\nu_\mu$
Additive uncertainties	
$b \to u\ell\nu_\ell$ modeling	11%
$B \to \pi\ell^+\nu_\ell$ FFs	4.8%
$B \to \rho\ell^+\nu_\ell$ FFs	3.4%
$B \to \omega\ell^+\nu_\ell$ FFs	3.0%
$\mathcal{B}(B \to \pi\ell^+\nu_\ell)$	3.4%
$\mathcal{B}(B \to \rho\ell^+\nu_\ell)$	3.2%
$\mathcal{B}(B \to \omega\ell^+\nu_\ell)$	3.1%
$\mathcal{B}(B \to X_u\ell^+\nu)$	4.0%
DFN parameters	4.0%
Hybrid model	4.2%
MC statistics	2.6%
Continuum modeling	13%
Shape correction	4.1%
MC statistics	12.2%
$b \to c\ell\nu_\ell$ modeling	2.5%
$B^+ \to \mu^+\nu_\mu$ MC statistics	1.0%
$\mathcal{B}(b \to s)$ processes	1.0%
$\mathcal{B}(B^+ \to \mu^+\nu_\mu\gamma)$	0.1%
Multiplicative uncertainties	
$\mathcal{L}_{\text{LID}}$ efficiency	2.0%
$N_{\text{B}\bar{\text{B}}}$	1.4%
Tracking efficiency	0.3%
Total syst. uncertainty	17%

to take into account shape uncertainties. These uncertainties are listed as "Additive uncertainties" in Table V. Here θ_{ik} represents the NP vector element of bin i and η_{ik}^{MC} the expected number of events in the same bin for event type k as estimated from the simulation. Note that this notation absorbs the size of the absolute error into the definition of the NP. Second, we include for the $B^+ \to \mu^+\nu_\mu$ signal template and fixed background templates overall efficiency and luminosity related uncertainties: this is achieved by rewriting the relevant fractions as

$$\eta_k \to \eta_k(1 + \theta_{ks}), \tag{22}$$

with θ_{ks} the NP parametrizing the uncertainty in question. The uncertainty sources treated this way include the overall lepton identification and track reconstruction efficiency uncertainty and the uncertainty on the number of B-meson pairs produced in the full Belle dataset and are labeled as "Multiplicative uncertainties" in Table V. For the fixed background templates the corresponding uncertainties from branching fractions are also included this way.

VII. $b \to u\ell\nu_\ell$ AND OFF-RESONANCE CONTROL REGION

To test the simulation of the crucial semileptonic $b \to u\ell\nu_\ell$ background, we construct a signal-depleted region with moderate continuum contamination. This is achieved by selecting events with continuum suppression classifier values of $C_{\mathrm{out}} \in [0.9, 0.93)$. In this sample, the region of high muon momentum p_μ^B is used to test the validity of the continuum description and the region with a muon momentum between 2.2 and 2.6 GeV is dominated by semileptonic $b \to u\ell\nu_\ell$ and $b \to c\ell\nu_\ell$ decays. To also test the modeling of both backgrounds with respect to the employed signal categorization exploiting the angle between the muon and the signal B meson, we further split the selected events using $\cos\Theta_{B\mu} > 0$ and $\cos\Theta_{B\mu} < 0$. The full likelihood fit procedure including all systematic uncertainties detailed in Secs. V and VI is then carried out. Figure 6 depicts the fit result: the individual contributions are shown as histograms and the recorded collision events are displayed as data points. The size of the systematic uncertainties is shown on the histograms as a hatched band. In the fit the signal $B^+ \to \mu^+\nu_\mu$ yield was fixed to the SM expectation and in both categories we expect about 15 $B^+ \to \mu^+\nu_\mu$ events. Both the $b \to u\ell\nu_\ell$ and $b \to c\ell\nu_\ell$, and continuum dominated regions are described well by the fit templates. Assuming that for most bins the statistical uncertainty is approximately Gaussian, we calculate a χ^2 of 30.4 over 41 degrees of freedom by comparing predicted and observed yields in each bin and by taking into account the full systematic uncertainties. This approximation is justified for most of the p_μ^B region, but breaks down for the high-momentum bins due to low statistics. The value still

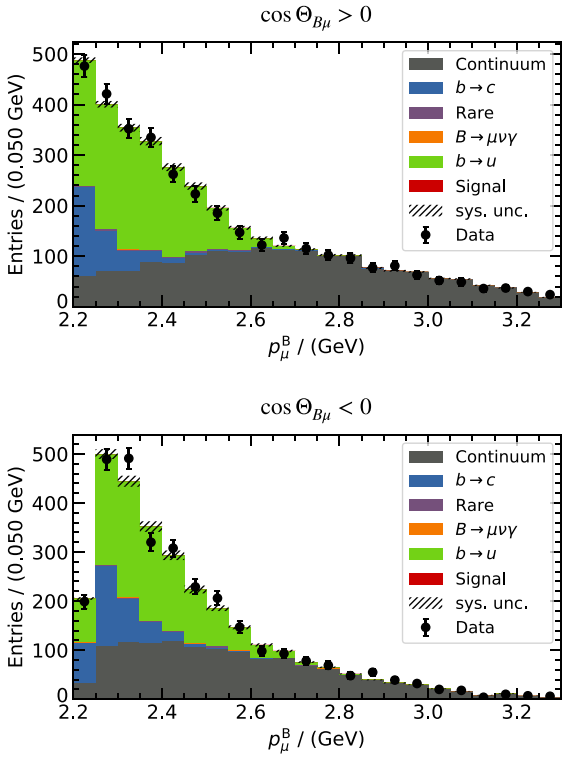

FIG. 6. The $b \to u\ell\nu_\ell$ control region fit is shown.

gives an indication that the fit model is able to describe the observed data well. We also carry out a fit in which the $B^+ \to \mu^+\nu_\mu$ signal template is allowed to float: We determine a value of -37 ± 61 events, which is compatible with the SM expectation.

For the inclusive $b \to u\ell\nu_\ell$ branching fraction in which the signal template is kept fixed at its SM expectation, we find

$$\mathcal{B}(B \to X_u\ell^+\nu) = (2.04 \pm 0.10) \times 10^{-3}, \tag{23}$$

where the uncertainty corresponds to the statistical error. The central value is compatible with the world average of Ref. [5], $\mathcal{B}(B \to X_u\ell^+\nu) = (2.13 \pm 0.31) \times 10^{-3}$. Note that Ref. [5] inflated the quoted uncertainty to account for incompatibilities between the measurements used in the average.

We also apply the signal continuum classifier selection of $C_{\mathrm{out}} \in [0.93, 1)$ on the recorded off-resonance data. With these events we carry out a two-component fit, determining the yields of $B^+ \to \mu^+\nu_\mu$ signal and continuum events. This allows us to determine whether the classifier selection could cause a sculpting of the background shape, which in turn would result in a spurious signal. The low number of events passing the selection does not allow further categorization of the events using angular information as only 39 off-resonance events pass the selection. We fit 37 ± 10 background events and 2 ± 7 signal events.

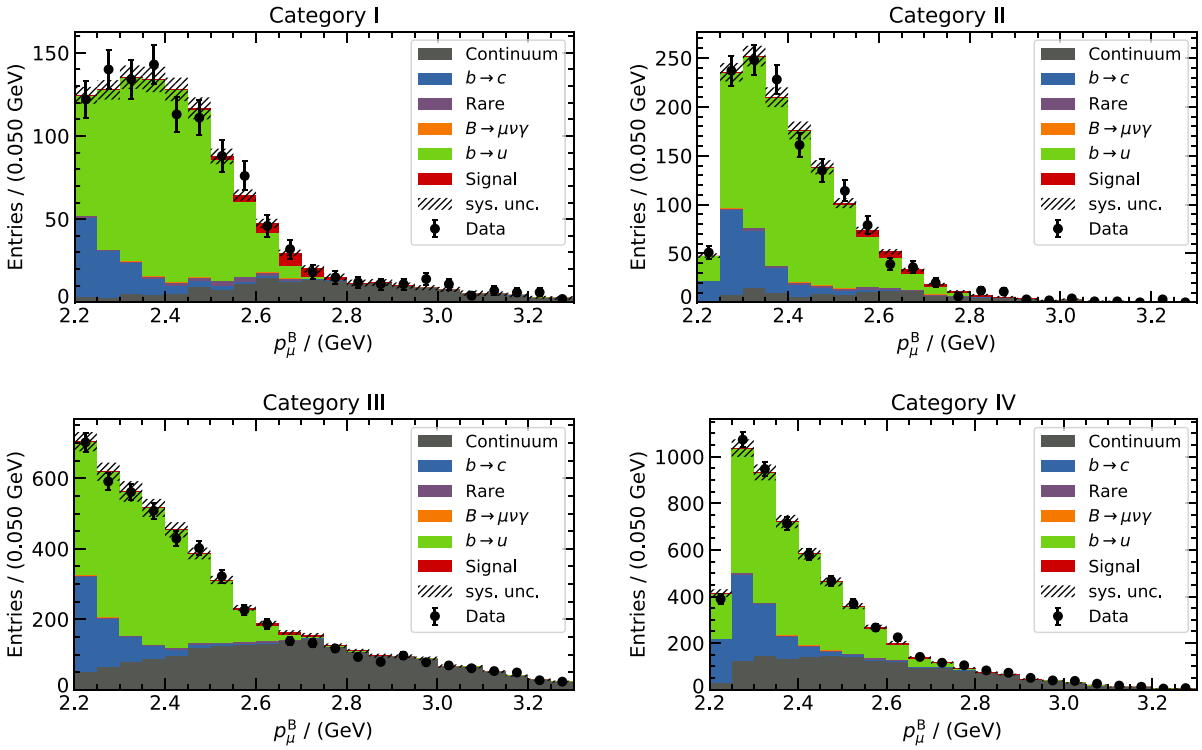

FIG. 7. The fitted distribution of p_μ^B for the four signal categories described in the text. The signal and background templates are shown as histograms and the recorded collision events as data points with uncertainties. The systematic uncertainties on the signal and background templates are shown as a hatched band.

VIII. RESULTS

In Fig. 7 the muon momentum spectrum in the B rest frame p_μ^B for the four signal categories is shown. The selected data events were used to maximize the likelihood Eq. (11): in total 4×22 bins with 4×132 NPs parametrizing systematic uncertainties are determined. In Appendix A a full breakdown of the NP pulls is given. The recorded collision data are shown as data points and the fitted $B^+ \to \mu^+\nu_\mu$ signal and background components are displayed as colored histograms. The size of the systematic uncertainties is shown on the histograms as a hatched band. We observe for the $B^+ \to \mu^+\nu_\mu$ branching fraction a value of

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (5.3 \pm 2.0 \pm 0.9) \times 10^{-7}, \quad (24)$$

with the first uncertainty denoting the statistical error and the second is from systematics. Figure 8 shows the profile likelihood ratio $\Lambda(\nu_{\rm sig})$ [cf. Eq. (14)]. Assuming that all bins are described with approximately Gaussian uncertainty and including systematics with their full covariance, we calculate a χ^2 value of 58.8 with 84 degrees of freedom using the predicted and observed bin values. The observed significance over the background-only hypothesis using the one-sided test statistics Eq. (16) is 2.8 standard deviations. This is in agreement with the median SM expectation of $2.4^{+0.8}_{-0.9}$ standard deviations, cf. Sec. V.

From the observed branching fraction we determine in combination with the B-meson decay constant f_B a value for the CKM matrix element $|V_{ub}|$. Using $f_B = 184 \pm 4$ MeV [2] we find

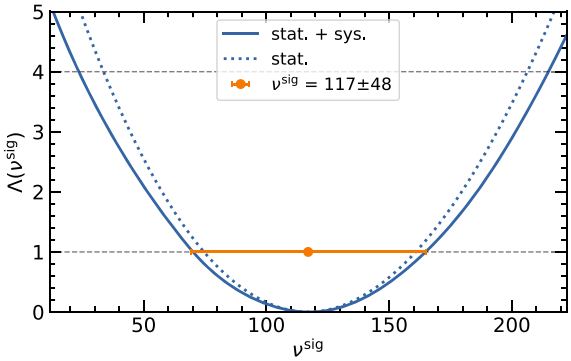

FIG. 8. The likelihood ratio contour $\Lambda(\nu_{\rm sig})$ as a function of the number of $B^+ \to \mu^+\nu_\mu$ signal events is shown: the dotted curve shows the contour incorporating only the statistical uncertainty with all systematic nuisance parameters fixed at their best-fit value. The solid curve shows full likelihood contour including all systematic and statistical uncertainties. The orange data point and errors shows the determined best-fit value and the 1σ (statistical + systematic) uncertainty.

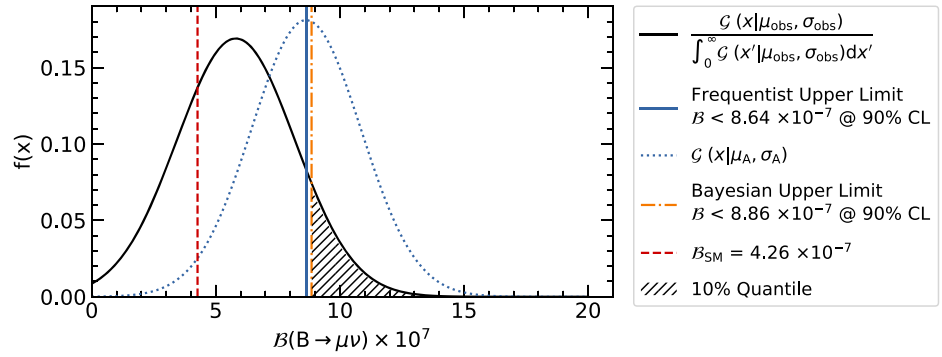

FIG. 9. The observed Bayesian (yellow dash-dotted) and frequentist (blue) upper limits at 90% CL are shown, along with the SM expectation of the $B^+ \to \mu^+\nu_\mu$ branching fraction and the Bayesian and frequentist PDFs.

$$|V_{ub}| = (4.4^{+0.8}_{-0.9} \pm 0.4 \pm 0.1) \times 10^{-3}, \tag{25}$$

where the first uncertainty is the statistical error, the second from systematics and the third from theory. This value is compatible with both exclusive and inclusive measurements of $|V_{ub}|$ [5].

Due to the low significance of the observed $B^+ \to \mu^+\nu_\mu$ signal, we calculate Bayesian and frequentist upper limits of the branching fraction. We convert the likelihood into a Bayesian probability density function (PDF) using the procedure detailed in Sec. V and Eq. (19): Figure 9 shows the one-dimensional PDF, which was obtained using a flat prior in the partial branching fraction. The resulting Bayesian upper limit for $B^+ \to \mu^+\nu_\mu$ at 90% CL is

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 8.9 \times 10^{-7} \text{ at } 90\% \text{ CL}. \tag{26}$$

The frequentist upper limit is determined using fits to ensembles of Asimov datasets with NPs shifted to the observed best-fit values. Figure 9 shows the corresponding frequentist likelihood, for convenience also converted into a PDF (blue dotted line) and the resulting upper limit at 90% CL is

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 8.6 \times 10^{-7} \text{ at } 90\% \text{ CL}. \tag{27}$$

The observed branching fraction is used to constrain the allowed parameter space of the two-Higgs-doublet model (2HDM) of type II and type III. In these models the presence of charged Higgs bosons as a new mediator with specific couplings would modify the observed branching fraction, cf. Fig. 1. The effect of the charged Higgs boson in the type II model is included in the expected $B^+ \to \mu^+\nu_\mu$ branching fraction by modifying Eq. (1) according to Ref. [45] to

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = \mathcal{B}^{\text{SM}} \times \left(1 - \frac{m_B^2 \tan^2\beta}{m_{H^+}^2}\right)^2, \tag{28}$$

with $\mathcal{B}^{\text{SM}}$ denoting the SM branching fraction, $\tan\beta$ being the ratio of the vacuum expectation values of the two Higgs fields in the model, and m_{H^+} the mass of the charged Higgs boson. The type III model further generalizes the couplings to [46,47]

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = \mathcal{B}^{\text{SM}} \times \left|1 + \frac{m_B^2}{m_b m_\mu}\left(\frac{C_{\text{R}}^\mu}{C_{\text{SM}}} - \frac{C_{\text{L}}^\mu}{C_{\text{SM}}}\right)\right|^2, \tag{29}$$

with m_b denoting the b quark mass and the $C_{R/L}^\mu$ are the coefficients encoding the new physics contribution.

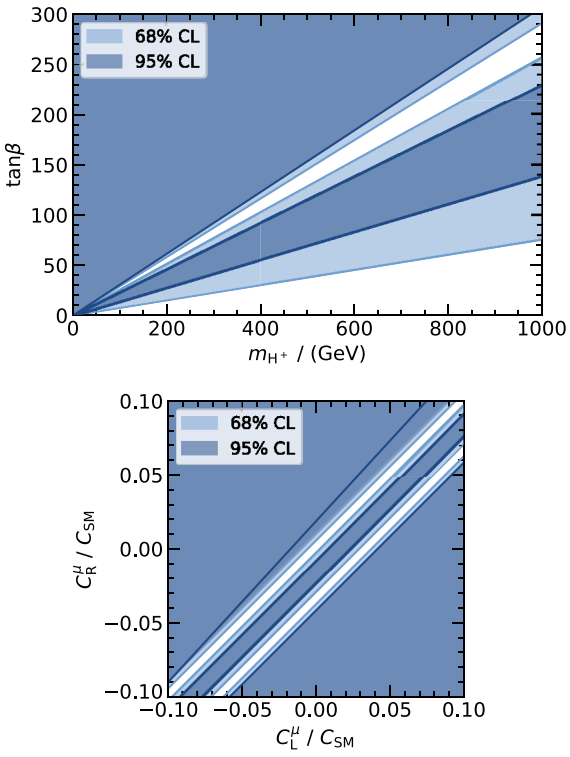

FIG. 10. The 68% and 95% CL excluded model parameter space for the 2HDM type II ($\tan\beta$, m_{H^+}) and type III (C_{L}^μ, C_{R}^μ) is shown. The coefficients C_{L}^μ and C_{R}^μ are assumed to be real.

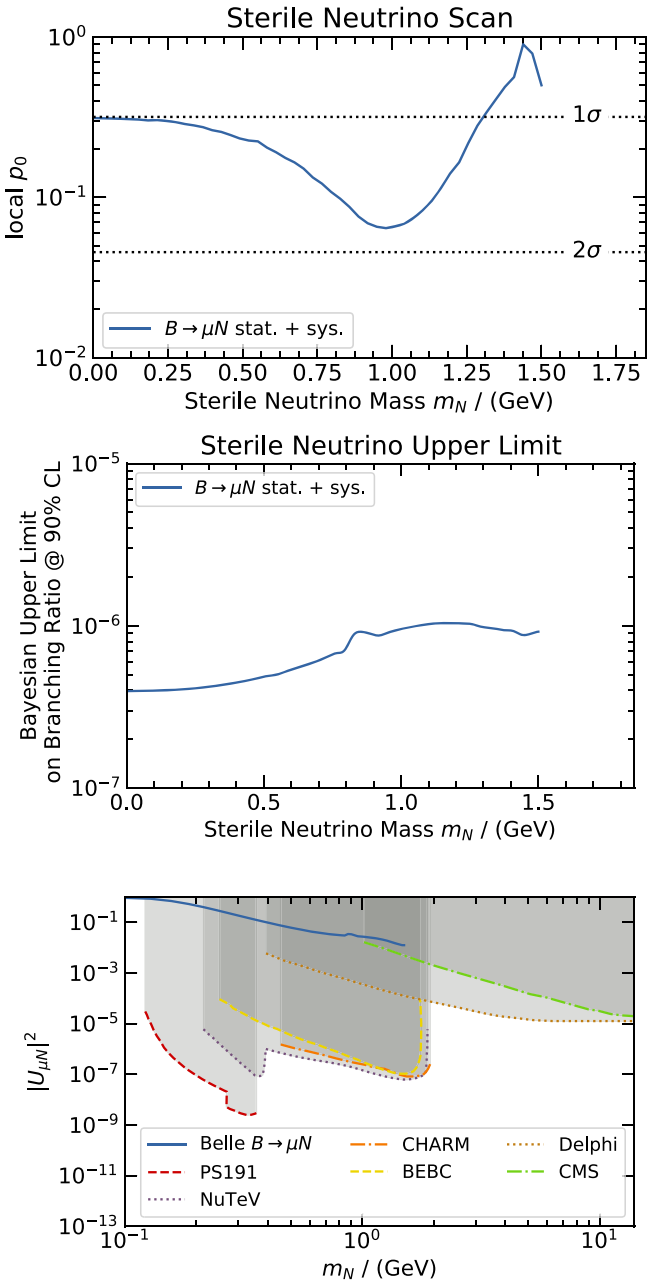

FIG. 11. (Top) The observed local p_0 values for the sterile neutrino search $B^+ \to \mu^+ N$ are shown with the SM process $B^+ \to \mu^+\nu_\mu$ included. If the SM process is accounted for, no significant excess is observed. The largest deviation from the background-only hypothesis is at $m_N = 1$ GeV. No correction for the look-elsewhere effect is included. (Center) The Bayesian upper limit on the branching fraction as calculated from the sterile neutrino signal yield. The $B^+ \to \mu^+\nu_\mu$ process is fixed to its SM expectation. (Bottom) The excluded area in the coupling-mass plane from this search in comparison to previous searches for sterile neutrinos.

Figure 10 shows the allowed and excluded parameter regions at 68% (light blue) and 95% (dark blue) CL as calculated using the observed branching fraction Eq. (24) and by constructing a χ^2 test. For the SM branching fraction prediction we use $\mathcal{B}^{\rm SM} = (4.3 \pm 0.8) \times 10^{-7}$ calculated assuming an average value of $|V_{ub}| = (3.94 \pm 0.36)) \times 10^{-3}$ from Ref. [5]. Due to the explicit lepton mass dependence in the type III model, the constructed bounds on $C^{\mu}_{\rm L/R}$ are more precise than any existing limits on $C^{\tau}_{\rm L/R}$ based on results from studying $B^+ \to \tau^+\nu_\tau$ decays.

To search for sterile neutrinos in $B^+ \to \mu^+ N$ we fix the $B^+ \to \mu^+\nu_\mu$ contribution to its SM value ($\mathcal{B}^{\rm SM}$) and search simultaneously in the four categories for an excess in the p^B_μ distributions. From the observed yields and our simulated predictions we calculate local p_0 values using the test statistic Eq. (16). The observed p_0 values are shown in Fig. 11 for sterile neutrino masses ranging from 0–1.5 GeV, and no significant excess over the background-only SM hypothesis is observed. The largest deviation is seen at a mass of $m_N = 1$ GeV with a significance of 1.8σ. The result does not account for any corrections for the look-elsewhere effect. We also calculate the Bayesian upper limit on the branching fraction from the extracted signal yield of the $B^+ \to \mu^+ N$ process with the $B^+ \to \mu^+\nu_\mu$ contribution fixed to its SM value. The upper limit as a function of the sterile neutrino mass is also shown in Fig. 11. To compare the upper limit from the $B^+ \to \mu^+ N$ process to previous searches [48–55] for sterile neutrinos we calculate the excluded values of the coupling $|U_{\mu N}|^2$ and the sterile neutrino mass m_N using [56]

$$\frac{\mathcal{B}(B^+ \to \mu^+ N)}{\mathcal{B}(B^+ \to \mu^+\nu_\mu)} = |U_{\mu N}|^2 \frac{m_N^2 + m_\mu^2}{m_\mu^2} \frac{\sqrt{\lambda(r_{N\rm B}, r_{\mu\rm B})}}{\sqrt{\lambda(0, r_{\mu\rm B})}} \times \frac{1 - (r_{N\rm B}^2 - r_{\mu\rm B}^2)^2/(r_{N\rm B}^2 + r_{\mu\rm B}^2)}{1 - r_{\mu\rm B}^2}, \quad (30)$$

with $r_{XY} = m_X/m_Y$ and the Källén function $\lambda(x, y) = (1 - (x - y)^2)(1 - (x + y)^2)$. The excluded values from this and the previous searches are shown in Fig. 11.

IX. SUMMARY AND CONCLUSIONS

In this paper results for the improved search of the $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \mu^+ N$ processes using the full Belle dataset and an inclusive tag approach are shown. The measurement supersedes the previous result of Ref. [7] as it has a higher sensitivity and a more accurate modeling of the crucial semileptonic $b \to u\ell\nu_\ell$ background. The analysis is carried out in the approximate B rest frame of the signal $B^+ \to \mu^+\nu_\mu$ decay, reconstructed from the remaining charged and neutral particles of the collision event. These are combined and calibrated to reconstruct the second B meson produced in the collision. In combination with the known beam

properties the four-momentum of the signal B meson is then reconstructed and used to boost the reconstructed signal muon in the reference frame, where the signal B meson is at rest. This results in a better signal resolution and improved sensitivity in contrast to carrying out the search in the c.m. frame of the colliding e^+e^- pair. The analysis is carried out in four analysis categories using the continuum suppression classifier and angular information of the B meson and the muon. The branching fraction is determined using a binned maximum likelihood fit of the muon momentum spectrum. Shape and normalization uncertainties from the signal and background templates are directly incorporated into the likelihood. We report an observed branching fraction of

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = (5.3 \pm 2.0 \pm 0.9) \times 10^{-7}, \quad (31)$$

with a significance of 2.8 standard deviations over the background-only hypothesis. We also quote the corresponding 90% upper limit using Bayesian and Frequentist approaches and use the observed branching fraction to set limits on type II and type III two-Higgs-doublet models. We find $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 8.9 \times 10^{-7}$ and $\mathcal{B}(B^+ \to \mu^+\nu_\mu) < 8.6 \times 10^{-7}$ at 90% CL for the Bayesian and Frequentist upper limits, respectively. The type III constraints are the most precise determined to date. In addition, we use the reconstructed muon spectrum to search for the presence of a sterile neutrino created through the process of $B^+ \to \mu^+ N$ and via a new mediator particle. No significant excess is observed for masses in the probed range of $m_N \in [0, 1.5)$ GeV. The largest excess is seen at a sterile neutrino mass of 1 GeV with a local significance of 1.8 standard deviations.

ACKNOWLEDGMENTS

We thank Marumi Kado and Günter Quast for discussions about one-sided test statistics and Ulrich Nierste and Dean Robinson for discussions about the sterile neutrino scenario. We thank the KEKB group for the excellent operation of the accelerator; the KEK cryogenics group for the efficient operation of the solenoid; and the KEK computer group, and the Pacific Northwest National Laboratory (PNNL) Environmental Molecular Sciences Laboratory (EMSL) computing group for strong computing support; and the National Institute of Informatics, and Science Information NETwork 5 (SINET5) for valuable network support. We acknowledge support from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, the Japan Society for the Promotion of Science (JSPS), and the Tau-Lepton Physics Research Center of Nagoya University; the Australian Research Council including Grants No. DP180102629, No. DP170102389, No. DP170102204, No. DP15010 3061, No. FT130100303; Austrian Science Fund (FWF); the National Natural Science Foundation of China under Contracts No. 11435013, No. 11475187, No. 11521505, No. 11575017, No. 11675166, No. 11705209; Key Research Program of Frontier Sciences, Chinese Academy of Sciences (CAS), Grant No. QYZDJ-SSW-SLH011; the CAS Center for Excellence in Particle Physics (CCEPP); the Shanghai Pujiang Program under Grant No. 18PJ1401000; the Ministry of Education, Youth and Sports of the Czech Republic under Contract No. LTT 17020; the Carl Zeiss Foundation, the Deutsche Forschungsgemeinschaft, the Excellence Cluster Universe, and the VolkswagenStiftung; the Department of Science and Technology of India; the Istituto Nazionale di Fisica Nucleare of Italy; National Research Foundation (NRF) of Korea Grants No. 2015H1A2A1033649, No. 2016R1D1 A1B01010135, No. 2016K1A3A7A09005 603, No. 2016 R1D1A1B02012900, No. 2018R1A2B3003 643, No. 2018 R1A6A1A06024970, No. 2018R1D1 A1B07047294; Radiation Science Research Institute, Foreign Large-size Research Facility Application Supporting project, the Global Science Experimental Data Hub Center of the Korea Institute of Science and Technology Information and KREONET/GLORIAD; the Polish Ministry of Science and Higher Education and the National Science Center; the Grant of the Russian Federation Government, Agreement No. 14.W03.31.0026; the Slovenian Research Agency; Ikerbasque, Basque Foundation for Science, Spain; the Swiss National Science Foundation; the Ministry of Education and the Ministry of Science and Technology of Taiwan; and the United States Department of Energy and the National Science Foundation. F. B. was supported by the DFG Emmy-Noether Grant No. BE 6075/1-1.

APPENDIX A: NUISANCE-PARAMETER PULL DISTRIBUTIONS

The summary of the systematic nuisance parameters of the $B^+ \to \mu^+\nu_\mu$ fit is shown in Fig. 12: pull distributions are displayed [defined as $(\theta - 0)/\sigma_\theta$] for each NP θ with uncertainty σ_θ. In total $528 + 3$ NPs were fitted, one for each fit template and bin. The same processes were correlated over the four categories and constraints were incorporated using multidimensional Gaussian PDFs. We observe mild pulls to adjust the $b \to u\ell\nu_\ell$ and continuum background shapes.

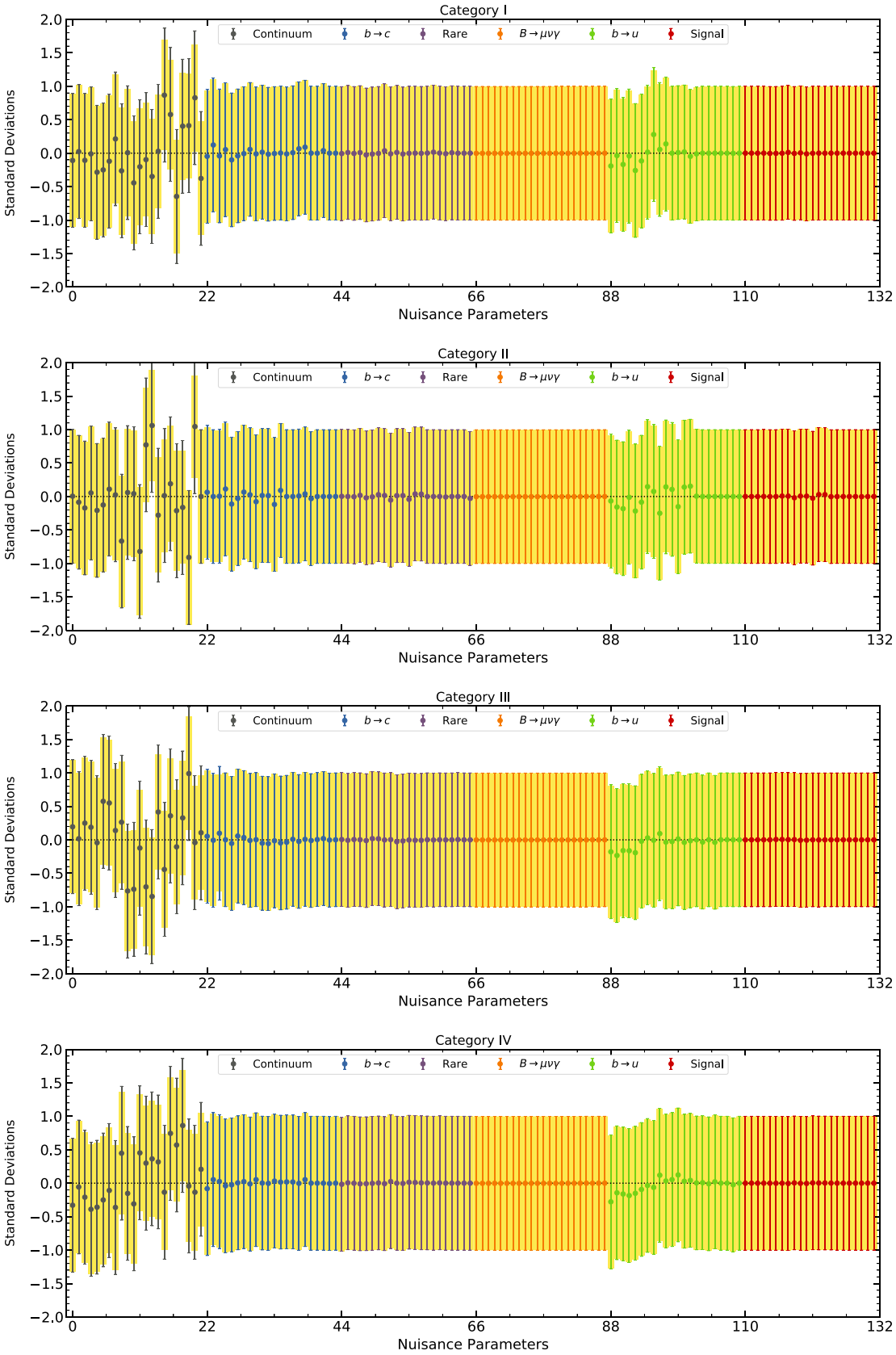

FIG. 12. The post-fit nuisance-parameter distribution for each category is shown.

APPENDIX B: $b \to u\ell\nu_\ell$ HYBRID MODEL DETAILS

Figure 13 shows the $b \to u\ell\nu_\ell$ hybrid model: inclusive and exclusive decays are merged, such that for a given bin in the three-dimensional space the total inclusive branching fraction is recovered. This is done by scaling down the inclusive prediction.

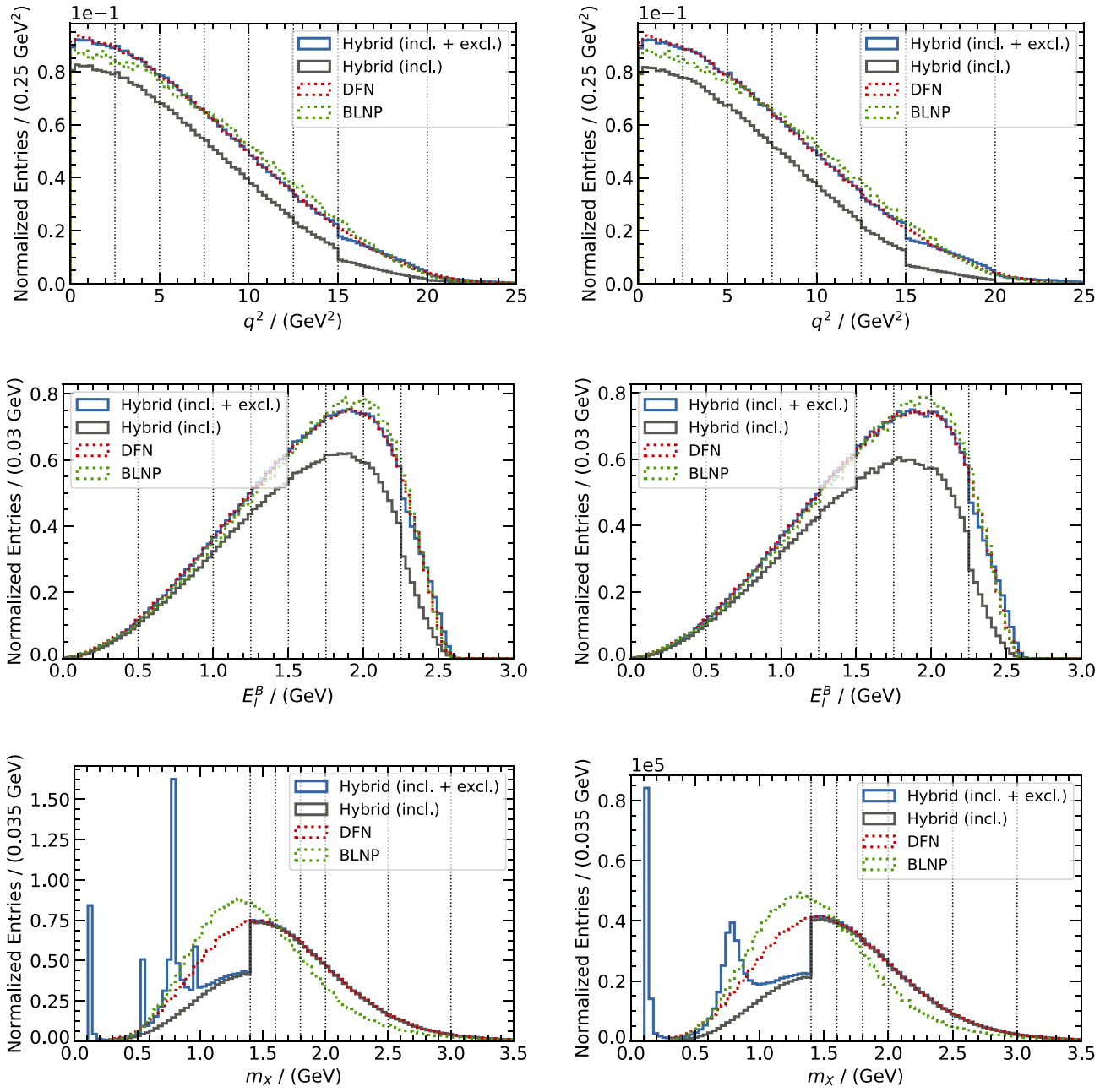

FIG. 13. The hybrid model predictions for semileptonic $b \to u\ell\nu_\ell$ decays for B^+ (left) and B^0 (right) as a function of q^2, E_ℓ^B and m_X are shown. The dashed lines show the chosen hybrid binning.

APPENDIX C: DATA VS MC REWEIGHTING

Figure 14 shows the effect of the data vs MC reweighting using the off-resonance collision events for three selected variables used in the training.

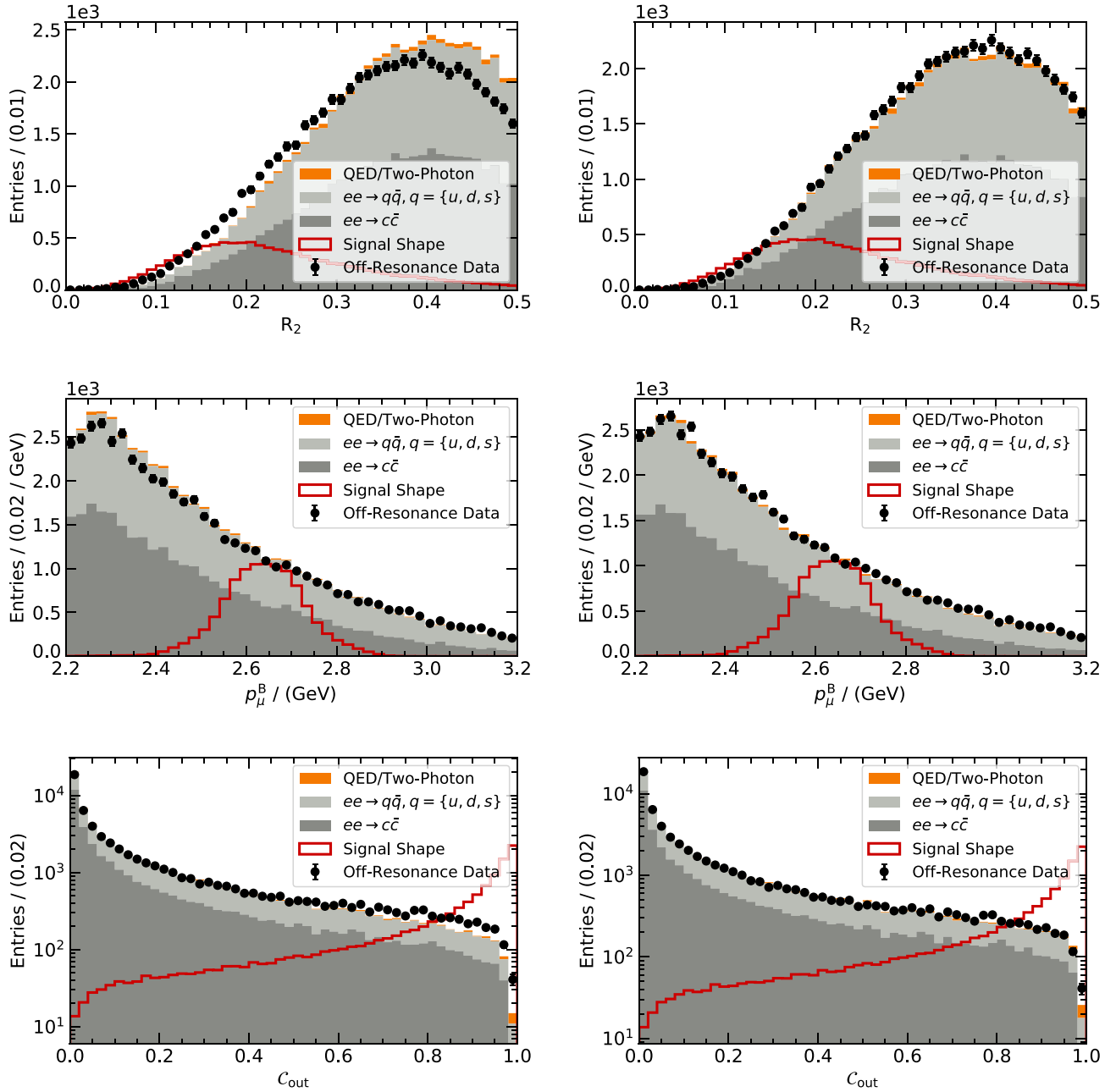

FIG. 14. A selection of variables used as input in the data vs MC reweighting before (left) and after (right) the weights are applied. The simulated data is shown as histogram and the recorded off-resonance collision events as data points with uncertainties.

[1] Charge conjugation is implied throughout this manuscript. Further we use natural units $c = \hbar = 1$ throughout.
[2] S. Aoki *et al.*, Eur. Phys. J. C **77**, 112 (2017).
[3] M. J. Baker, J. Bordes, C. A. Dominguez, J. Penarrocha, and K. Schilcher, J. High Energy Phys. 07 (2014) 032.
[4] P. Gelhausen, A. Khodjamirian, A. A. Pivovarov, and D. Rosenthal, Phys. Rev. D **88**, 014015 (2013); **91**, 099901(E) (2015).
[5] M. Tanabashi *et al.* (Particle Data Group), Phys. Rev. D **98**, 030001 (2018).
[6] T. Keck *et al.*, Comput. Softw. Big Sci. **3**, 6 (2019).
[7] A. Sibidanov *et al.* (Belle Collaboration), Phys. Rev. Lett. **121**, 031801 (2018).
[8] B. Aubert *et al.* (*BABAR* Collaboration), Phys. Rev. D **79**, 091101 (2009).
[9] A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens, and O. Ruchayskiy, Prog. Part. Nucl. Phys. **104**, 1 (2019).
[10] E. K. Akhmedov, arXiv:hep-ph/0001264.
[11] S. Kurokawa and E. Kikutani, Nucl. Instrum. Methods A **499**, 1 (2003), and other papers included in this Volume; T. Abe *et al.*, Prog. Theor. Exp. Phys. (**2013**), 03A001 and references therein.
[12] A. Abashian *et al.*, Nucl. Instrum. Methods A **479**, 117 (2002); also see detector section in J. Brodzicka *et al.*, Prog. Theor. Exp. Phys. (**2012**), 04D001.
[13] K. Hanagaki, H. Kakuno, H. Ikeda, T. Iijima, and T. Tsukamoto, Nucl. Instrum. Methods A **485**, 490 (2002).
[14] A. Abashian *et al.*, Nucl. Instrum. Methods A **491**, 69 (2002).
[15] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun (Belle II Framework Software Group), Comput. Softw. Big Sci. **3**, 1 (2019).
[16] M. Gelb *et al.*, Comput. Softw. Big Sci. **2**, 9 (2018).
[17] D. J. Lange, Nucl. Instrum. Methods Phys. Res., Sect. A **462**, 152 (2001).
[18] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, Report No. CERN-DD-EE-84-1, 1987.
[19] E. Barberio, B. van Eijk, and Z. Was, Comput. Phys. Commun. **66**, 115 (1991).
[20] C. Bourrely, I. Caprini, and L. Lellouch, Phys. Rev. D **79**, 013008 (2009); **82**, 099902(E) (2010).
[21] Y. Amhis *et al.* (Heavy Flavor Averaging Group), Eur. Phys. J. C **77**, 895 (2017).
[22] A. Sibidanov *et al.* (Belle Collaboration), Phys. Rev. D **88**, 032005 (2013).
[23] J. P. Lees *et al.* (*BABAR* Collaboration), Phys. Rev. D **87**, 032004 (2013); **87**, 099904(E) (2013).
[24] P. del Amo Sanchez *et al.* (*BABAR* Collaboration), Phys. Rev. D **83**, 032007 (2011).
[25] A. Bharucha, J. High Energy Phys. 05 (2012) 092.
[26] D. Scora and N. Isgur, Phys. Rev. D **52**, 2783 (1995).
[27] F. De Fazio and M. Neubert, J. High Energy Phys. 06 (1999) 017.
[28] T. Sjöstrand, Comput. Phys. Commun. **82**, 74 (1994).
[29] C. Ramirez, J. F. Donoghue, and G. Burdman, Phys. Rev. D **41**, 1496 (1990).
[30] B. O. Lange, M. Neubert, and G. Paz, Phys. Rev. D **72**, 073006 (2005).
[31] D. Martschei, M. Feindt, S. Honc, and J. Wagner-Kuhr, J. Phys. **368**, 012028 (2012).
[32] C. G. Boyd, B. Grinstein, and R. F. Lebed, Phys. Rev. Lett. **74**, 4603 (1995).
[33] R. Glattauer *et al.* (Belle Collaboration), Phys. Rev. D **93**, 032006 (2016).
[34] B. Grinstein and A. Kobach, Phys. Lett. B **771**, 359 (2017).
[35] D. Bigi, P. Gambino, and S. Schacht, Phys. Lett. B **769**, 441 (2017).
[36] A. Abdesselam *et al.* (Belle Collaboration), arXiv:1702.01521.
[37] M. Gelb *et al.* (Belle Collaboration), Phys. Rev. D **98**, 112016 (2018).
[38] G. P. Korchemsky, D. Pirjol, and T.-M. Yan, Phys. Rev. D **61**, 114510 (2000).
[39] T. Keck, Comput. Softw. Big Sci. **1**, 2 (2017).
[40] G. C. Fox and S. Wolfram, Phys. Rev. Lett. **41**, 1581 (1978); K. Abe *et al.* (Belle Collaboration), Phys. Rev. Lett. **87**, 101801 (2001); Phys. Lett. B **511**, 151 (2001).
[41] D. M. Asner *et al.* (CLEO Collaboration), Phys. Rev. D **53**, 1039 (1996).
[42] E. Jones, T. Oliphant, P. Peterson *et al.*, SciPy: Open source scientific tools for Python, http://www.scipy.org/ (2001).
[43] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. D **86**, 032003 (2012).
[44] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur. Phys. J. C **71**, 1554 (2011).
[45] W.-S. Hou, Phys. Rev. D **48**, 2342 (1993).
[46] C.-H. Chen and T. Nomura, Phys. Rev. D **98**, 095007 (2018).
[47] A. Crivellin, C. Greub, and A. Kokulu, Phys. Rev. D **86**, 054014 (2012).
[48] G. Bernardi, G. Carugno, J. Chauveau, F. Dicarlo, M. Dris, J. Dumarchez, M. Ferro-Luzzi, J.-M. Levy, D. Lukas, J.-M. Perreau, Y. Pons, A.-M. Touchard, and F. Vannucci, Phys. Lett. B **203**, 332 (1988).
[49] A. Vaitaitis *et al.* (NuTeV, E815 Collaborations), Phys. Rev. Lett. **83**, 4943 (1999).
[50] J. Dorenbosch *et al.*, Phys. Lett. B **166**, 473 (1986).
[51] J. Orloff, A. N. Rozanov, and C. Santoni, Phys. Lett. B **550**, 8 (2002).
[52] P. Vilain, G. Wilquet, S. Petrak *et al.*, Phys. Lett. B **343**, 453 (1995).
[53] A. Cooper-Sarkar, S. Haywood, M. Parker, S. Sarkar *et al.*, Phys. Lett. B **160**, 207 (1985).
[54] DELPHI Collaboration, Z. Phys. C **74**, 57 (1997).
[55] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **120**, 221801 (2018).
[56] D. Robinson (private communication).