

Measurement of inclusive $B \rightarrow X_u \ell \nu$ partial branching fractions and $|V_{ub}|$ at Belle II

M. Abumusab , I. Adachi , K. Adamczyk , L. Aggarwal , H. Ahmed , Y. Ahn , H. Aihara , N. Akopov , S. Alghamdi , M. Alhakami , A. Aloisio , N. Althubiti , K. Amos , N. Anh Ky , C. Antonioli , D. M. Asner , H. Atmacan , T. Aushev , R. Ayad , V. Babu , H. Bae , N. K. Baghel , S. Bahinipati , P. Bambade , Sw. Banerjee , M. Barrett , M. Bartl , J. Baudot , A. Beaubien , F. Becherer , J. Becker , J. V. Bennett , F. U. Bernlochner , V. Bertacchi , M. Bertemes , E. Bertholet , M. Bessner , S. Bettarini , V. Bhardwaj , B. Bhuyan , F. Bianchi , T. Bilka , D. Biswas , A. Bobrov , D. Bodrov , G. Bonvicini , J. Borah , A. Boschetti , A. Bozek , M. Bračko , P. Branchini , R. A. Briere , T. E. Browder , A. Budano , S. Bussino , Q. Campagna , M. Campajola , L. Cao , G. Casarosa , C. Cecchi , P. Chang , P. Cheema , L. Chen , B. G. Cheon , C. Cheshta , H. Chetri , K. Chilikin , J. Chin , K. Chirapattpimol , H.-E. Cho , K. Cho , S.-J. Cho , S.-K. Choi , S. Choudhury , S. Chutia , J. Cochran , J. A. Colorado-Caicedo , I. Consigny , L. Corona , J. X. Cui , E. De La Cruz-Burelo , S. A. De La Motte , G. De Nardo , G. De Pietro , R. de Sangro , M. Destefanis , S. Dey , A. Di Canto , J. Dingfelder , Z. Doležal , I. Domínguez Jiménez , T. V. Dong , X. Dong , M. Dorigo , G. Dujany , P. Ecker , J. Eppelt , R. Farkas , P. Feichtinger , T. Ferber , T. Fillinger , C. Finck , G. Finocchiaro , F. Forti , B. G. Fulsom , A. Gabrielli , A. Gale , E. Ganiev , M. Garcia-Hernandez , R. Garg , G. Gaudino , V. Gaur , V. Gautam , A. Gaz , A. Gellrich , G. Ghevondyan , D. Ghosh , H. Ghumaryan , G. Giakoustidis , R. Giordano , A. Giri , P. Gironella Gironell , A. Glazov , B. Gobbo , R. Godang , O. Gogota , P. Goldenzweig , W. Gradl , M. Graf-Schreiber , E. Graziani , D. Greenwald , Y. Guan , K. Gudkova , I. Haide , Y. Han , H. Hayashii , S. Hazra , C. Hearty , M. T. Hedges , A. Heidelbach , G. Heine , I. Heredia de la Cruz , M. Hernández Villanueva , T. Higuchi , M. Hoek , M. Hohmann , R. Hoppe , P. Horak , X. T. Hou , C.-L. Hsu , A. Huang , T. Humair , T. Iijima , K. Inami , N. Ipsita , A. Ishikawa , R. Itoh , M. Iwasaki , P. Jackson , D. Jacobi , W. W. Jacobs , E.-J. Jang , S. Jia , Y. Jin , A. Johnson , M. Kaleta , A. B. Kaliyar , J. Kandra , K. H. Kang , S. Kang , G. Karyan , F. Keil , C. Ketter , M. Khan , C. Kiesling , D. Y. Kim , J.-Y. Kim , K.-H. Kim , H. Kindo , K. Kinoshita , P. Kodyš , T. Koga , S. Kohani , K. Kojima , A. Korobov , S. Korpar , E. Kovalenko , R. Kowalewski , P. Križan , P. Krokovny , T. Kuhr , Y. Kulii , D. Kumar , K. Kumara , T. Kunigo , Y.-J. Kwon , S. Lacaprara , T. Lam , L. Lanceri , J. S. Lange , T. S. Lau , M. Laurens , R. Leboucher , F. R. Le Diberder , H. Lee , M. J. Lee , C. Lemettais , P. Leo , P. M. Lewis , C. Li , H.-J. Li , L. K. Li , Q. M. Li , W. Z. Li , Y. Li , Y. B. Li , Y. P. Liao , J. Libby , J. Lin , S. Lin , Z. Liptak , M. H. Liu , Q. Y. Liu , Z. Liu , D. Liventsev , S. Longo , A. Lozar , T. Lueck , C. Lyu , J. L. Ma , Y. Ma , M. Maggiora , S. P. Maharana , R. Maiti , G. Mancinelli , R. Manfredi , E. Manoni , M. Mantovano , D. Marcantonio , M. Marfoli , C. Marinas , C. Martellini , A. Martens , T. Martinov , L. Massaccesi , M. Masuda , D. Matvienko , S. K. Maurya , M. Maushart , J. A. McKenna , Z. Mediankin Gruberová , R. Mehta , F. Meier , D. Meleshko , M. Merola , C. Miller , M. Mirra , K. Miyabayashi , H. Miyake , R. Mizuk , G. B. Mohanty , S. Moneta , A. L. Moreira de Carvalho , H.-G. Moser , M. Mrvar , H. Murakami , R. Mussa , I. Nakamura , M. Nakao , Y. Nakazawa , M. Naruki , Z. Natkaniec , A. Natochii , M. Nayak , M. Neu , S. Nishida , R. Nomaru , A. Novosel , S. Ogawa , R. Okubo , H. Ono , F. Otani , G. Pakhlova , A. Panta , S. Pardi , K. Parham , J. Park , K. Park , S.-H. Park , A. Passeri , S. Patra , S. Paul , T. K. Pedlar , R. Pestotnik , M. Piccolo , L. E. Piilonen , P. L. M. Podesta-Lerma , T. Podobnik , C. Praz , S. Prell , E. Prencipe , M. T. Prim , H. Purwar , P. Rados , G. Raeuber , S. Raiz , K. Ravindran , J. U. Rehman , M. Reif , S. Reiter , L. Reuter , D. Ricalde Herrmann , I. Ripp-Baudot , G. Rizzo , S. H. Robertson , J. M. Roney

M.-Z. Wang , A. Warburton , S. Watanuki , C. Wessel , E. Won , X. P. Xu , B. D. Yabsley , W. Yan , W. Yan , J. Yelton , K. Yi , J. H. Yin , K. Yoshihara , C. Z. Yuan , J. Yuan , Y. Yusa , L. Zani , F. Zeng , M. Zeyrek , B. Zhang , V. Zhilich , J. S. Zhou , Q. D. Zhou , L. Zhu , and R. Žlebčík

(The Belle II Collaboration)

A sample of 365 fb^{-1} of $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ data collected by the Belle II experiment is used to measure the partial branching fractions of charmless semileptonic B meson decays and determine the magnitude of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{ub}|$. Events containing a signal electron or muon ℓ and a fully reconstructed hadronic B decay that constrains the signal kinematics are selected, while the rest of the event defines the hadronic system X_u associated with the signal. To discriminate the signal from the 50-times larger background originating from CKM-favored semileptonic B decays, a template fit is performed in both signal and control regions after applying an optimized selection. The partial branching fraction measured for lepton energies greater than 1 GeV in the signal B meson rest frame is $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu) = (1.54 \pm 0.08 \text{ (stat.)} \pm 0.12 \text{ (syst.)}) \times 10^{-3}$. From this measurement, using the Gambino, Giordano, Ossola, Uraltsev theoretical framework, $|V_{ub}| = (4.01 \pm 0.19^{+0.07}_{-0.08}) \times 10^{-3}$ is determined, where the uncertainties are experimental and theoretical, respectively. This value is consistent with the world average obtained from previous inclusive measurements. Different theoretical predictions and partial branching fractions measured in other phase-space regions, defined by additional selections on the X_u and leptonic system masses, are also used to determine $|V_{ub}|$. This allows for a comparison of the resulting values across theoretical frameworks and phase-space regions.

I. INTRODUCTION

Precise measurements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix elements [1, 2] are crucial to test the quark weak interactions described by the standard model of particle physics. For three fermion generations the CKM matrix is unitary, which imposes strict relations between its elements. The precision of experimental tests of several of these relations is limited by the current uncertainty in the determination of the strength of the b -to- u quark coupling, $|V_{ub}|$. B -meson decays to either an electron or a muon, ℓ , a neutrino, ν , and a charmless hadronic system, X_u , depend on $|V_{ub}|$ as illustrated in Figure 1. Such decays are denoted as $B \rightarrow X_u \ell \nu$. Despite the experimental challenges due to the neutrino in the final state,¹ charmless semileptonic decays offer the best sensitivity to $|V_{ub}|$. Determinations from fully hadronic B meson decays suffer from significant QCD uncertainties [3], while fully leptonic decays of charged B mesons are extremely challenging to study with the available data sets as the light-lepton channels are helicity suppressed and the tau channel suffers from large backgrounds [4].

There are two complementary strategies to measure $|V_{ub}|$ via semileptonic decays: exclusive and inclusive measurements. Exclusive measurements select a specific decay mode, such as $B \rightarrow \pi \ell \nu$, while inclusive measurements place little or no restriction on the content of the X_u system. Both strategies have been applied extensively at B factories over the past two decades (see Ref. [5] for a summary of all measurements).

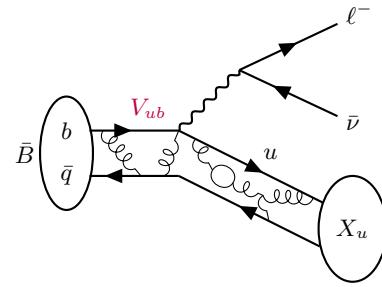


FIG. 1. Feynman diagram of the leading $B \rightarrow X_u \ell \nu$ transition. The X_u system can either represent a specific meson or a non-resonant system comprising several hadrons produced by the fragmentation of the $u\bar{q}$ system.

The two most recent averages of $|V_{ub}|$ extracted from exclusive and inclusive decays read [5]

$$|V_{ub}|_{\text{excl}} = (3.43 \pm 0.12) \times 10^{-3},$$

$$|V_{ub}|_{\text{incl}} = (4.06 \pm 0.16) \times 10^{-3}, \quad (1)$$

where the total (experimental and theoretical) uncertainty is given. The two values disagree by about three standard deviations, which limits the precision of the CKM unitarity test. This discrepancy calls for a deeper investigation of the theoretical and experimental factors that may contribute to it. The measurement presented here uses the inclusive method.

The decay width of inclusive charmless semileptonic decays can be described by the *heavy quark expansion* (HQE) [6, 7], an expansion in powers of the strong coupling constant α_s and inverse powers of the b quark mass m_b . Decay rate predictions in the HQE framework

¹ Semitauonic decays are not considered, being even more challenging due to multiple neutrinos in the final state.

rely on quark-hadron duality which states that predictions made at the parton level, when integrated over sufficiently large portions of the phase space, should yield reliable predictions of hadronic observables [8]. However, selections on kinematic variables such as the lepton energy or the X_u mass are necessary in experimental analyses to suppress backgrounds. The most important source of background emerges from CKM-favored semileptonic decays to a hadronic system containing a charm quark. Such decays, denoted as $B \rightarrow X_c \ell \nu$, are enhanced by about a factor 50 compared to $B \rightarrow X_u \ell \nu$ decays. These kinematic selections break the inclusivity requirement of the HQE [9] and, instead, predictions of differential decay rates are necessary. These predictions in restricted kinematical regions rely on an effective description of the b -quark motion inside the meson through a non-perturbative *shape function*, whose development has been a major theoretical challenge.² Treatment of such non-perturbative effects has led to the development of several theoretical frameworks that yield different predictions of the partial decay rate $\Delta\tilde{\Gamma}(B \rightarrow X_u \ell \nu)$ ($\Delta\tilde{\Gamma}(B \rightarrow X_u \ell \nu) = \Delta\Gamma(B \rightarrow X_u \ell \nu)/|V_{ub}|^2$) over a given region of phase space [9–11]. A value of $|V_{ub}|$ can be extracted from the measurement of the inclusive partial branching fraction $\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)$ via

$$|V_{ub}| = \sqrt{\frac{\Delta\mathcal{B}(B \rightarrow X_u \ell \nu)}{\tau_B \Delta\tilde{\Gamma}(B \rightarrow X_u \ell \nu)}}, \quad (2)$$

where τ_B is the B meson lifetime.

In this paper, an inclusive measurement of partial branching fractions of $B \rightarrow X_u \ell \nu$ ($\ell = e, \mu$) decays using $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ data collected by the Belle II experiment is presented. Three variables are used to describe the kinematics of inclusive semileptonic decays: the lepton energy E_ℓ – with the superscript B added when the energy is calculated in the signal B meson rest frame; the squared momentum transfer defined as $q^2 = (p_\ell + p_\nu)^2$ where p_ℓ and p_ν are the charged lepton and neutrino four momenta; and the hadronic system mass M_X . The measurement is performed in three phase-space regions covering about 87%, 57% and 32% of the full $B \rightarrow X_u \ell \nu$ phase space and defined by three sets of selections on E_ℓ^B , q^2 and M_X designed to suppress the $B \rightarrow X_c \ell \nu$ background. For each measured partial branching fraction, $|V_{ub}|$ is extracted from Equation 2 using different decay rate predictions.

The partner B meson produced in the $e^+e^- \rightarrow B\bar{B}$ process is reconstructed in specific hadronic decay channels. This method, referred to as *hadronic tagging*, constrains the kinematics of the inclusively reconstructed

hadronic system X_u despite the neutrino being undetected. The partner B is referred to as the tag B or B_{tag} .

The remainder of this manuscript is organized as follows: the collected and simulated data sets used for the measurement are presented in Section II; the event reconstruction and signal selections are outlined in Section III; checks on the modeling of the $B \rightarrow X_c \ell \nu$ background and associated corrections are detailed in Section IV; the signal extraction procedure is discussed in Section V; finally, the results of the measurement are presented in Section VI and discussed in Section VII. Throughout this article, charge conjugation is implied and natural units $\hbar = c = 1$ are used.

II. DATA SET AND SIMULATION

The measurement is performed using a 365 fb^{-1} sample of electron-positron collision data recorded at a center-of-mass energy of $\sqrt{s} = 10.58 \text{ GeV}$ corresponding to the $\Upsilon(4S)$ resonance mass (*on-resonance* data). This sample contains 387×10^6 $\Upsilon(4S)$ events. In addition, 43 fb^{-1} of data collected at a center-of-mass energy 60 MeV below the $\Upsilon(4S)$ resonance (*off-resonance* data) are used to study background from $e^+e^- \rightarrow q\bar{q}$ ($q \in \{u, d, s, c\}$) production (*continuum* background).

The Belle II detector is situated at the SuperKEKB asymmetric-energy electron–positron collider, based at KEK in Tsukuba, Japan [12]. It is a large solid-angle cylindrical detector consisting of several layers of sub-detectors around the e^+e^- interaction point (IP). The z axis is defined as the symmetry axis of the detector in the direction of the electron beam, the y axis corresponds to the vertical axis, pointing upwards and the x axis points towards the outside of the accelerator ring. At the innermost part of Belle II is the pixel vertex detector (PXD), which consists of two layers of silicon pixel sensors located at 1.4 cm and 2.2 cm from the IP, which provides precise vertexing capabilities. In the data set used in this measurement, the second pixel layer is installed in only one-sixth of the solid angle. Surrounding the PXD is the silicon vertex detector (SVD), which comprises four layers of silicon strip sensors. The SVD covers polar angles ranging from 17° to 150° , which further enhances the vertex reconstruction and improves the overall tracking performance. The central drift chamber (CDC) surrounds the SVD and is specifically designed for charged particle tracking. It contains 14,336 sense wires operating in a helium-ethane gas mixture and also covers polar angles ranging from 17° to 150° . The CDC provides sub-percent accuracy in measuring charged particle momenta, essential for precise kinematic reconstruction. A time-of-propagation (TOP) detector positioned outside the CDC in the barrel region uses Cherenkov light to identify charged particles. In the forward end-cap, this role is fulfilled

² Beyond the leading order in the HQE, additional shape functions arise.

by a ring-imaging Cherenkov detector equipped with an aerogel radiator (ARICH). Beyond the TOP and ARICH lies the electromagnetic calorimeter (ECL), composed of 8,736 homogeneous thallium-doped caesium iodide crystals. The ECL is divided into three regions: the barrel and two end-caps, and covers polar angles from 12.4° to 155.1° . This sub-detector excels in detecting electromagnetic showers from both charged and neutral particles and provides percent-level energy resolution for electrons and photons. Additionally, the ECL provides three-momentum information for neutral particles through energy deposition (cluster) analysis. The ECL is surrounded by a superconducting solenoid which generates a magnetic field of 1.5 T. Charged tracks are bent in the x - y plane by the magnetic field and the curvature of the trajectory can be used to determine the transverse momentum of the particle. The outermost sub-detector of Belle II is the K_L^0 and muon detector (KLM). The KLM consists of 14 iron layers, each 4.7 cm thick, interspersed with 15 active layers in the barrel region and 14 (12) in the forward (backward) endcap region. The KLM covers polar angles from 20° to 155° .

Simulated samples are used to optimize the signal selections, train multivariate algorithms to suppress various sources of background, study signal and background components and validate the signal extraction. The events are simulated using dedicated generators and they are processed using a full detector simulation implemented with **GEANT4** [13]. Experimental and simulated data events are reconstructed and analyzed with the open-source **basf2** framework [14].

In the simulation, the assumed branching fractions for $B \rightarrow X_u \ell \nu$ and $B \rightarrow X_c \ell \nu$ decays are summarized in Table I. The branching fractions used for the modeling of resonant $B \rightarrow X_u \ell \nu$ decays are taken from the most up-to-date Heavy Flavor Averaging Group (HFLAV) report [5]. The inclusive $B \rightarrow X_u \ell \nu$ branching fractions are taken separately for B^0 and B^+ as quoted in the latest Particle Data Group (PDG) Review of Particle Physics [15]. The $B \rightarrow X_c \ell \nu$ branching fractions are computed from the HFLAV averages assuming isospin symmetry. Dedicated samples of signal events are produced with the **EvtGen** generator [16]. For background contributions, $\Upsilon(4S) \rightarrow B\bar{B}$ samples, which contain known semileptonic and hadronic B decays, are produced with **EvtGen** and **Pythia 8** [17].

The sample of $B \rightarrow X_u \ell \nu$ events is a mixture of non-resonant modes (such as $B \rightarrow \pi\pi\ell\nu$, $B \rightarrow \pi\rho\ell\nu$...), simulated via inclusive $B \rightarrow X_u \ell \nu$ frameworks, and resonant modes ($B \rightarrow \pi\ell\nu$, $B \rightarrow \omega\ell\nu$...). The theoretical description of resonant modes is based on functions known as *form factors*, which encapsulate the non-perturbative dynamics of the transition. Non-resonant and resonant decays are combined using the *hybrid model* initially proposed in Ref. [18]. After presenting

| Decay mode | \mathcal{B} (%) | |
|------------------------------------|---------------------|---------------------|
| | B^+ | B^0 |
| Incl. $B \rightarrow X_u \ell \nu$ | 0.192 ± 0.024 | 0.176 ± 0.022 |
| $B \rightarrow \pi\ell\nu$ | 0.0078 ± 0.0003 | 0.0150 ± 0.0006 |
| $B \rightarrow \rho\ell\nu$ | 0.0158 ± 0.0011 | 0.0294 ± 0.0021 |
| $B \rightarrow \omega\ell\nu$ | 0.0119 ± 0.0009 | - |
| $B \rightarrow \eta\ell\nu$ | 0.0035 ± 0.0004 | - |
| $B \rightarrow \eta'\ell\nu$ | 0.0024 ± 0.0007 | - |
| Incl. $B \rightarrow X_c \ell \nu$ | 11.05 ± 0.16 | 10.27 ± 0.15 |
| $B \rightarrow D\ell\nu$ | 2.27 ± 0.06 | 2.11 ± 0.05 |
| $B \rightarrow D^*\ell\nu$ | 5.27 ± 0.12 | 4.90 ± 0.11 |
| $B \rightarrow D_1\ell\nu$ | 0.64 ± 0.10 | 0.59 ± 0.10 |
| $B \rightarrow D_0^*\ell\nu$ | 0.13 ± 0.19 | 0.12 ± 0.18 |
| $B \rightarrow D_1'\ell\nu$ | 0.28 ± 0.04 | 0.26 ± 0.04 |
| $B \rightarrow D_2^*\ell\nu$ | 0.32 ± 0.03 | 0.30 ± 0.03 |
| $B \rightarrow D_s K\ell\nu$ | 0.03 ± 0.01 | - |
| $B \rightarrow D_s^* K\ell\nu$ | 0.03 ± 0.02 | - |
| $B \rightarrow D\eta\ell\nu$ | 0.90 ± 0.90 | 0.86 ± 0.86 |
| $B \rightarrow D^*\eta\ell\nu$ | 0.90 ± 0.90 | 0.86 ± 0.86 |
| $B \rightarrow D\pi\ell\nu$ | 0.07 ± 0.09 | 0.07 ± 0.08 |
| $B \rightarrow D^*\pi\ell\nu$ | 0.22 ± 0.10 | 0.20 ± 0.10 |

TABLE I. Assumed branching fractions used for simulated semileptonic B decays.

the samples of $B \rightarrow X_u \ell \nu$ decays, the hybrid model is described.

The simulation of non-resonant events is based on the De Fazio, Neubert (DFN) model [19] implemented in the **EvtGen** generator. This model describes the differential $B \rightarrow X_u \ell \nu$ rate at order $\mathcal{O}(\alpha_s)$ and at leading order in the HQE and requires two input parameters for the modeling of events, m_b and a non-perturbative parameter a used in the modeling of the shape function. In the Kagan, Neubert renormalization scheme [20] their values read: $m_b^{\text{KN}} = 4.66 \pm 0.04$ GeV and $a^{\text{KN}} = 1.3 \pm 0.5$ [21]. They are related to the difference between the B meson mass and the b quark mass and the average momentum squared of the b quark inside the B meson. The fragmentation of the hadronic system produced in non-resonant events is simulated by **Pythia 8**.

In addition, five resonant decays are included in the $B \rightarrow X_u \ell \nu$ modeling: $B \rightarrow (\pi/\rho/\omega/\eta/\eta')\ell\nu$. Their theoretical descriptions rely on specific form factor parametrizations which are also implemented in **EvtGen**. The Bourrely, Caprini, Lellouch [22] parametrization is used to model $B \rightarrow \pi\ell\nu$ decays. The form factor parameters are taken from Table 57 in Ref. [23]; $B \rightarrow \rho/\omega\ell\nu$ decays are modelled with the Bharucha, Straub, Zwicky parametrization [24] whose values

are taken from Table 4 in Ref. [25]; $B \rightarrow \eta/\eta'\ell\nu$ decays are modelled via the Isgur, Scora, Grinstein, Wise approach [26] with input parameters taken from the light-cone sum rules framework described in Ref. [27].

Resonant and non-resonant $B \rightarrow X_u\ell\nu$ contributions are combined using the hybrid approach. The resonant and non-resonant contributions are summed in three-dimensional bins of E_ℓ^B , M_X and q^2 . In each bin, the inclusive contribution ($\Delta\mathcal{B}^{\text{inc}}$) is obtained from the non-resonant component and the exclusive contribution ($\Delta\mathcal{B}^{\text{exc}}$) is obtained from the sum of exclusive components. The non-resonant part is scaled down such that the sum of non-resonant and resonant contributions matches the inclusive partial branching fraction. The non-resonant contribution is therefore scaled by weights defined as:

$$w_{ijk} = \frac{\Delta\mathcal{B}_{ijk}^{\text{inc}} - \Delta\mathcal{B}_{ijk}^{\text{exc}}}{\Delta\mathcal{B}_{ijk}^{\text{inc}}}, \quad (3)$$

where the indices i, j, k denote a three-dimensional bin. These three-dimensional bins are defined by the combinations of the following one-dimensional bins:

$$\begin{aligned} E_\ell^B &: [0, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 3.00] \text{ GeV, and} \\ q^2 &: [0, 2.50, 5.00, 7.50, 10.0, 12.5, 15.0, 20.0, 25.0] \text{ GeV}^2, \\ M_X &: [0, 1.40, 1.60, 1.80, 2.00, 2.50, 3.00, 3.50] \text{ GeV.} \end{aligned}$$

The $B \rightarrow X_c\ell\nu$ background is modelled as a sum of different types of resonant decays. Each resonance is modelled with a specific form factor parametrization implemented in `EvtGen`. The $B \rightarrow D\ell\nu$ and $B \rightarrow D^*\ell\nu$ decays are modelled using the Bernlochner, Ligeti, Papucci, Robinson, Xiong, Prim form factor model [28] with eight common parameters. The four heavy $1P$ states D_0^* , D_1' , D_1 and D_2^* are collectively referred to as D^{**} [29]. The $B \rightarrow D^{**}\ell\nu$ decays are modelled using the Bernlochner, Ligeti, Robinson (BLR) model from Ref. [30] with three parameters for the broad D_0^* and D_1' states and four parameters for the narrower D_1 and D_2^* states. The $B \rightarrow X_c\ell\nu$ modeling also includes the decays $B \rightarrow D^{(*)}\pi\pi\ell\nu$ and $B \rightarrow D_s^{(*)}K\ell\nu$. The $B \rightarrow D^{(*)}\eta\ell\nu$ modes are simulated to fill the gap between the inclusively measured $B \rightarrow X_c\ell\nu$ rate and the sum of exclusive decay rates [31]. The $D^{(*)}\pi\pi$ and $D^{(*)}\eta$ modes are collectively referred to as *gap modes* and are denoted $B \rightarrow D_{\text{Gap}}^{**}\ell\nu$. As there exists no theoretical description of these decays, they are simulated via intermediate decays to D^{**} resonances using the BLR form factor parametrization. The $D\eta$ decays are simulated via $B \rightarrow D_0^*(\rightarrow D\eta)\ell\nu$ and $D^*\eta$ decays via $B \rightarrow D_1'(\rightarrow D^*\eta)\ell\nu$. The $D^{(*)}\pi\pi$ modes are equally split between intermediate decays to D_0^* and D_1' .

Another source of background stems from continuum events which are simulated by the `KKMCee` [32]

generator (the resulting particle decays are simulated by `EvtGen`).

III. EVENT RECONSTRUCTION AND SELECTION

The inclusive $B \rightarrow X_u\ell\nu$ analysis relies on the reconstruction of the B_{tag} meson in hadronic decay channels. The rest frame of the signal B meson is determined from momentum conservation in $\Upsilon(4S) \rightarrow B\bar{B}$ decays. The lepton in the semileptonic decay is reconstructed as a single track. All detected tracks and clusters not associated with the B_{tag} or the lepton are assumed to come from the X system and are grouped into an object called the rest-of-event (ROE). The hadronic system four-momentum p_X , from which the hadronic mass M_X is derived, is computed from the four-momenta of all tracks and clusters in the ROE. All ROE clusters are assumed to be produced by photons and, since events with kaons and leptons in the ROE are rejected at a later stage, all ROE tracks are assigned a pion mass hypothesis. Particles originating from beam background can pollute the ROE and they are therefore suppressed by the selections given in Section III B 2. The neutrino cannot be detected and its momentum is therefore identified with the missing momentum of the event. Using the four momenta of the initial e^+e^- state, the B_{tag} , the lepton and X , the neutrino four-momentum can be inferred from the missing four-momentum as:

$$p_\nu = p_{\text{miss}} = p_{e^+e^-} - p_{B_{\text{tag}}} - p_\ell - p_X. \quad (4)$$

The momentum transfer squared q^2 is therefore computed as $(p_\ell + p_\nu)^2$. Finally, the missing mass squared M_{miss}^2 is calculated as $M_{\text{miss}}^2 = p_{\text{miss}}^2$.

In the following, the requirements applied on B_{tag} candidates, ROE particles and signal lepton candidates are presented. The requirements described in Sections III A and III B are also referred to as *preselections*.

A. Tag side

The B_{tag} is reconstructed using the Full Event Interpretation algorithm (FEI) [33] which reconstructs B mesons in $\mathcal{O}(10^4)$ decay chains using a series of boosted decision tree (BDT) classifiers. They are required to have an FEI output score greater than 0.01 and a beam constrained mass

$$M_{\text{bc}} = \sqrt{E_{\text{beam}}^{*2} - |p_{\text{tag}}^*|^2} > 5.27 \text{ GeV}, \quad (5)$$

where p_{tag}^* is the momentum of the B_{tag} candidate in the e^+e^- center-of-mass frame and $E_{\text{beam}}^* = \sqrt{s}/2$ is half of the center-of-mass energy of the e^+e^- collision. Furthermore, the B_{tag} candidates are required to have an energy

difference

$$\Delta E = E_{\text{tag}}^* - E_{\text{beam}}^* \quad (6)$$

in the range $-0.15 \text{ GeV} < \Delta E < 0.10 \text{ GeV}$, where E_{tag}^* is the energy of the B_{tag} candidate in the e^+e^- center-of-mass frame.

The thrust axis is defined as the axis \hat{t} along which the summed projections of the momenta of a group of particles, $\sum_{i=1}^n |\hat{t} \cdot \vec{p}_i|$, is maximized. To suppress continuum background, the absolute value of the cosine of the angle between the thrust axes of the B_{tag} and the ROE is required to be lower than 0.9. Finally, if more than one B_{tag} candidate has been selected, only the candidate with the highest FEI score is retained.

B. Signal side

1. Signal lepton

The selections applied on signal lepton candidates aim to suppress three main sources of background. The first source consists of events with secondary leptons which mostly come from D meson decays and are referred to as *secondary lepton* events. The second consists of, events in which a hadron has been misidentified as the signal lepton. These mostly correspond to pions misidentified as muons and are referred to as *fake lepton* events. The third consists of, events where the lepton originates from a J/ψ decay or from photon conversion.

To ensure that signal lepton candidates originate near the interaction point, the distances between the interaction point and the track point-of-closest-approach along the z axis (dz) and in the xy plane (dr) are required to satisfy $dr < 1 \text{ cm}$ and $|dz| < 3 \text{ cm}$. Tracks are also required to fall in the CDC acceptance by having a polar angle between 17° and 150° . In addition, they are required to have a laboratory-frame momentum greater than 300 MeV. Electrons and muons are identified using a BDT-based and a likelihood-ratio-based identification score, respectively [34]. The identification scores are calculated from information collected by all subdetectors except the PXD. The lepton identification requirements result in a sample where 45% of the events contain electron candidates and the remaining fraction contain muon candidates. About 10% of electron candidates and 35% of muon candidates are misidentified particles (mostly pions). Furthermore, the four-momentum of an electron candidate is corrected for energy loss from bremsstrahlung photons by identifying photons with an energy below 1 GeV within a 0.05 rad cone around the electron initial momentum direction. Finally, to suppress J/ψ and photon conversion leptons, each candidate is combined with oppositely charged tracks found in the ROE which are assigned the same mass hypothesis as the signal lepton candidate. Of all possible pairs, the

one with an invariant mass closest to zero or the nominal J/ψ mass is selected. Signal lepton candidates are then rejected if the invariant mass is found to be below 50 MeV or between 3.043 and 3.129 GeV for electron pairs and between 3.072 and 3.122 GeV for muon pairs. These selections are optimized to reject about 95% of photon conversion tracks and J/ψ , except for electrons below the J/ψ mass where a broad tail caused by bremsstrahlung can be observed. In this case, the threshold rejects 68% of true J/ψ . The signal lepton candidates are required to have a charge consistent with the flavor of the signal B meson, taken as the opposite of the B_{tag} charge. Even though this requirement vetoes about a fifth of neutral B signal events, it also rejects events in which a secondary lepton has been reconstructed as the signal lepton. Finally, events with multiple lepton candidates are rejected as their presence is indicative of secondary semileptonic decays of D mesons.

2. Rest-Of-Event

ROE selections are optimized to select charged particles originating near the interaction point and suppress particles that hit the edges of the detector where the response is poorly modelled. Selections applied on neutral clusters are optimized to suppress beam background and misidentified photons. A fit is performed using all tracks found in the ROE to match them to a single decay vertex. Tracks in the ROE are required to pass similar dr , dz and polar angle selections to signal lepton tracks. Duplicated tracks due to particles with low transverse momentum are suppressed by a BDT. The classifier uses the track helix parameters, the product of the track charges, the angle between the tracks and the track transverse and longitudinal momenta.

ROE neutral clusters are required to have a minimum energy of 70, 60 and 70 MeV in the forward, barrel, and backward regions of the ECL respectively and a polar angle between 17° and 150° . The barrel covers a region between 32° and 129° and there are 1° gaps between the barrel and each end cap. Photon candidates must have a reconstructed time within 70 ns of the time at which the collision occurred. Misidentified photons are suppressed by a BDT classifier trained on numerous cluster properties, such as energy, polar angle or distance from the nearest track [35].

3. Charged pions

The multiplicity of charged pions in an event is used to derive an uncertainty on the X_u hadronization modeling as described in Section VE 13. These pions are identified using a likelihood ratio score calculated from information collected by all subdetectors except the PXD and are required to have a momentum and polar angle within

the phase-space region considered in hadron identification calibration studies [36]. The likelihood requirement selects about 47% of the charged pion candidates and about 2% of pion candidates are misidentified particles.

C. Continuum

In $e^+e^- \rightarrow q\bar{q}$ events, light mesons are boosted and therefore result in two back-to-back jet-like structures, whereas B mesons are produced nearly at rest in the center-of-mass frame resulting in isotropically distributed decay products. Variables such as the two B meson thrust angles and magnitudes, the modified Fox-Wolfram moments [37, 38] and the CLEO cones [39] can be used to distinguish continuum events from B meson decays (30 variables are considered). In order to suppress the continuum background, a multi layer perceptron (MLP) is trained. The MLP is configured with two hidden layers with 256 and 128 neurons, respectively. The network uses the Swish activation function [40] and a binary cross-entropy loss function. The AdamW optimizer [41] is used to train the network, with a learning rate of 9.3×10^{-3} . The classifier is trained using a total of 14 variables: the charged track multiplicity, and a subset of 13 variables among the 30 describing the event shape as mentioned above.³ All chosen variables exhibit good agreement between simulated and observed data. A subset consisting of 5% of simulated $B \rightarrow X_u \ell \nu$ events and 20% of simulated continuum events is used for training. The sample of continuum events is corrected following the procedure described in Section IV A. The neural network is optimized using the `optuna` package [42]. The optimal selection on the classifier output score is chosen according to the maximal significance defined as $S/\sqrt{S+B}$, with S the number of $B \rightarrow X_u \ell \nu$ events and B the number of continuum events that pass the selection. The yields are counted in the region where $E_\ell^B > 1$ GeV (see Section III F for a discussion on kinematic selections). The continuum suppression selection rejects 95% of continuum events while retaining 68% of $B \rightarrow X_u \ell \nu$ events.

D. D^* reconstruction from low-momentum pions

Most decays of D^* mesons produce a D meson and low-momentum (or slow) pion, denoted as π_s , whose properties can be used to suppress $B \rightarrow D^* \ell \nu$ events. Using π_s , an inclusive reconstruction of D^* mesons in the ROE is performed. Charged π_s are identified as single tracks with a charge opposite to that of the signal lepton, and neutral π_s are reconstructed from pairs of photons.

³ The other 17 variables are dropped because they display poor agreement between experimental and simulated data in the region used to train the classifier.

The π_s are required to have a laboratory-frame momentum between 50 and 200 MeV. Duplicated tracks left by charged π_s are suppressed by a BDT as mentioned in the previous section. The D^* candidate is assumed to have a momentum vector along the π_s flight direction in the laboratory-frame and its energy is calculated from

$$E_{D^*} = \frac{M_{D^*}}{M_{D^*} - M_D} E_{\pi_s}, \quad (7)$$

where M_{D/D^*} is the known mass of the D/D^* meson [15] and E_{π_s} the energy of the π_s . Having computed the D^* four-momentum and knowing the B_{tag} and signal lepton candidate four-momenta, the missing mass of each $B \rightarrow D^* \ell \nu$ candidate can be computed as:

$$M_{\text{miss}}^2(D^* \rightarrow D\pi) = (p_{e^+e^-} - p_{B_{\text{tag}}} - p_\ell - p_{D^*})^2. \quad (8)$$

When multiple π_s candidates are reconstructed, the one with $M_{\text{miss}}^2(D^* \rightarrow D\pi)$ closest to zero is selected.

E. $B \rightarrow X_c \ell \nu$ suppression

Since most charm meson weak decays produce kaons, events where a K^\pm or K_S^0 candidate is found in the ROE are rejected. No attempt is made to reconstruct K_L^0 . Similarly to charged pions, charged kaons are selected by combining into a likelihood ratio the particle identification information from all subdetectors except the PXD. The likelihood requirement selects about 18% of all charged kaon candidates and about 9% of the selected kaon candidates are misidentified particles. They are also required to have a momentum and polar angle within the phase-space region considered in hadron identification calibration studies [36].

K_S^0 candidates are reconstructed from two oppositely charged tracks that are fit to a common production vertex. As K_S^0 are relatively long-lived particles, the ROE track selections are not applied in this case. To ensure that the momentum vector of the candidates is consistent with the vector connecting the interaction point and the fitted decay vertex, the cosine of the angle between these two vectors is required to be greater than 0.998. Finally, K_S^0 candidates are required to have a mass between 470 and 530 MeV, consistent with the known K^0 mass [15]. The K_S^0 mass resolution is about 10 MeV.

To suppress the dominant $B \rightarrow X_c \ell \nu$ background, an MLP is trained on nine kinematic variables. The trained network is similar to that used to suppress continuum background but is configured with four hidden layers with 64, 32, 128 and 128 neurons, respectively, and a learning rate of 3.3×10^{-3} . The input features exploit various differences between semileptonic decays to charmed and charmless mesons:

- the missing mass squared, M_{miss}^2 ;

- the total event charge, Q_{tot} ;
- the p -value of the ROE vertex fit, $p_{\text{vtx}}^{\text{ROE}}$;
- when a D^{*+} or D^{*0} has been inclusively reconstructed from a π_s (either neutral or charged):
 - the missing mass squared $M_{\text{miss}}^2(D^* \rightarrow D\pi_s)$ as defined in equation 8;
 - the cosine of the angle between the flight directions of the reconstructed D^* and the signal lepton in the center-of-mass frame, $\cos\theta_c(D^* \rightarrow D\pi_s)$;
 - the $\cos\theta_{BY}(D^* \rightarrow D\pi_s)$ variable defined as

$$\cos\theta_{BY} = \frac{2E_{\text{beam}}^* E_Y^* - m_B^2 - m_Y^2}{2|p_B^*||p_Y^*|},$$

where $Y = D^*\ell$; E_Y^* , $|p_Y^*|$ and m_Y are the energy, three-momentum magnitude and mass of the Y system, respectively; $|p_B^*|$ and m_B are the three-momentum magnitude and mass of the B meson, respectively. Energies and three-momenta are defined in the center-of-mass frame. This variable is expected to be between -1 and 1 if only a massless particle such as a neutrino is missing in the reconstructed event.

Since the average particle multiplicity of $B \rightarrow X_c\ell\nu$ events is expected to be larger than for $B \rightarrow X_u\ell\nu$ events, the event reconstruction quality is expected to be poorer because of missed particles or because of the limited detector resolution. The first three variables therefore help to reject poorly reconstructed $B \rightarrow X_c\ell\nu$ events which tend to populate the high M_{miss}^2 tail, have a low $p_{\text{vtx}}^{\text{ROE}}$ and a Q_{tot} different than zero. The π_s variables allow the identification of events with a D^* decay. The distributions of the nine input features are shown in Appendix A. The optimal selection on the classifier output score is chosen as the one which minimizes the $\Delta\mathcal{B}(B \rightarrow X_u\ell\nu)$ total uncertainty as obtained from the nominal fit described in Section V. The output score is shown in Figure 2 and the optimal selection value for the score is found to be 0.87. The $B \rightarrow X_c\ell\nu$ suppression requirement rejects 98% of $B \rightarrow X_c\ell\nu$ events while retaining 25% of $B \rightarrow X_u\ell\nu$ events.

F. Kinematic selections

In addition to the selections discussed so far, requirements on the three kinematic variables E_ℓ^B , M_X and q^2 can be used to further suppress background. The three sets of kinematic selections listed in Table II are considered. In the following, a selection of $E_\ell^B > 1$ GeV is applied to all samples, as it rejects a portion of phase space mainly populated by continuum and fake and secondary lepton events. Furthermore, a region where

an $M_X < 1.7$ GeV selection is applied is also explored. The mass of the lightest charm state – the D meson – being around 1.86 GeV, this selection efficiently rejects $B \rightarrow X_c\ell\nu$ decays. Finally, since $B \rightarrow X_u\ell\nu$ decays are expected to dominate at higher leptonic energies, a region with a $q^2 > 8$ GeV² selection, in addition to the E_ℓ^B and M_X selections, is also considered.

The three signal regions used to measure $\Delta\mathcal{B}(B \rightarrow X_u\ell\nu)$ are defined by applying all preselections, the continuum suppression, the $B \rightarrow X_c\ell\nu$ suppression, and one of the three sets of kinematic selections. The signal phase-space acceptance (*i.e.* the covered fraction of the accessible $B \rightarrow X_u\ell\nu$ phase-space) for the three sets of kinematic selections is given in Table II. Inclusive $B \rightarrow X_u\ell\nu$ models are most reliable in the broadest phase-space region but the additional selections on M_X and q^2 increase the signal purity and provide additional tests of the theoretical predictions.

TABLE II. $B \rightarrow X_u\ell\nu$ phase-space acceptance for the three sets of kinematic selections used to define the signal regions.

| Region label | Selections | Acceptance |
|--------------|---|------------|
| 1 | $E_\ell^B > 1.0$ GeV | 87% |
| 2 | $E_\ell^B > 1.0$ GeV $M_X < 1.7$ GeV | 57% |
| 3 | $E_\ell^B > 1.0$ GeV $M_X < 1.7$ GeV $q^2 > 8$ GeV ² | 31% |

IV. MODELING CORRECTIONS

A. Continuum modeling corrections

The modeling of continuum events is known to be poorly understood and is therefore corrected using a data-driven approach. First, the total number of expected events is corrected by comparing the number of events in 40% of the off-resonance experimental and simulated data samples. The remaining 60% are used to test and validate the BDT described below. The correction factor is extracted independently for events in which a charged or neutral B meson candidate has been reconstructed and is found to be 0.87 ± 0.02 and 0.94 ± 0.03 , respectively, where the uncertainties are statistical (uncertainties on these factors are derived as described in Section V E9).

In a second step, a BDT with a maximum tree depth of five is trained to distinguish between experimental and simulated off-resonance data using 30 variables that

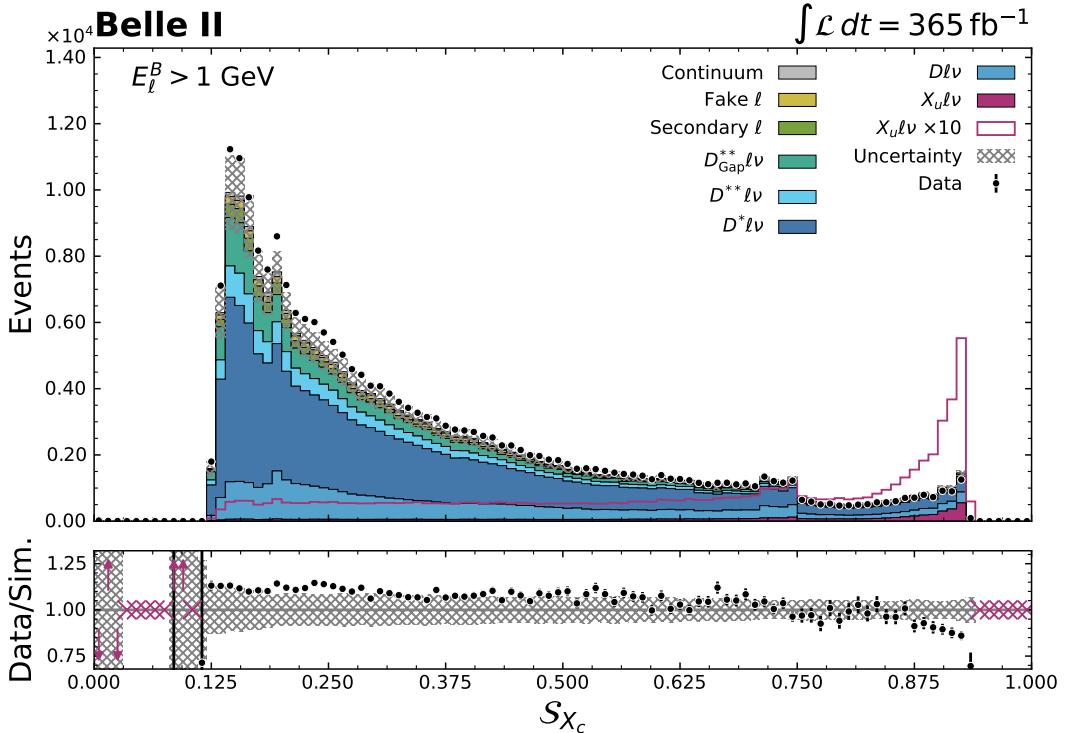


FIG. 2. $B \rightarrow X_c \ell \nu$ suppression classifier output score. Simulated templates are shown as stacked histograms and experimental data are shown as black points. The signal component is shown as a purple line with its yield multiplied by 10 to enhance visibility. The bottom panel shows the ratio between data and simulation yields in each bin. The purple arrows indicate that the point is outside of the plotting range and the purple crosses indicate the absence of simulated data in a particular bin. The step in the distribution in the range $0.70 - 0.75$ is caused by the large number of events with a $p_{\text{vtx}}^{\text{ROE}}$ close to zero (see Figure 10 in Appendix A).

describe the event shape [43] (see Section III C). In addition, the missing mass squared and the scalar sum of the transverse momenta of the signal and tag-side B final-state particles are included in the BDT input features. The BDT is trained on the off-resonance sample from which the normalization correction factors are extracted. The remainder of the sample is split into a validation sample (10% of the full sample) and a test sample (50%). The classifier output score $\mathcal{S}_{\text{cont}}$ is well calibrated and takes values between 0 and 1. The ratio

$$w_{\text{cont}} = \frac{\mathcal{S}_{\text{cont}}}{1 - \mathcal{S}_{\text{cont}}} \quad (9)$$

is an approximation to the ratio of experimental to simulated data likelihoods: it is applied as a weight to continuum events in the on-resonance simulated data sample. The continuum correction includes both the overall normalization factors and the BDT-extracted shape factors. The modeling correction is applied on the continuum sample after applying the preselections and before training the continuum suppression classifier. Its impact is illustrated in Figure 3 for the most discriminating variable in the BDT training.

B. $B \rightarrow X_c \ell \nu$ modeling corrections

The agreement between data and simulation is evaluated in control regions. Three control regions are defined for each kinematic selection defined in Table II by inverting the kaon veto and/or requiring a low $B \rightarrow X_c \ell \nu$ suppression classifier score (see Figure 2):

- $\text{CR}_{0,\text{low}}$: events with kaons are rejected and the $B \rightarrow X_c \ell \nu$ suppression classifier score is required to be lower than 0.60;
- $\text{CR}_{K,\text{high}}$: the presence of at least one kaon and a $B \rightarrow X_c \ell \nu$ suppression classifier score higher than 0.87 are required;
- $\text{CR}_{K,\text{low}}$: the presence of at least one kaon and a $B \rightarrow X_c \ell \nu$ suppression classifier score lower than 0.60 are required.

A comparison of the $B \rightarrow X_c \ell \nu$ composition in the three control regions and the signal region is shown in Appendix B. In addition, two validation regions, $\text{VR}_{K,\text{mid}}$ and $\text{VR}_{0,\text{mid}}$, with classifier scores between 0.60 and 0.87 are defined. In the former region, a kaon is required whereas in the latter, kaons are vetoed. The

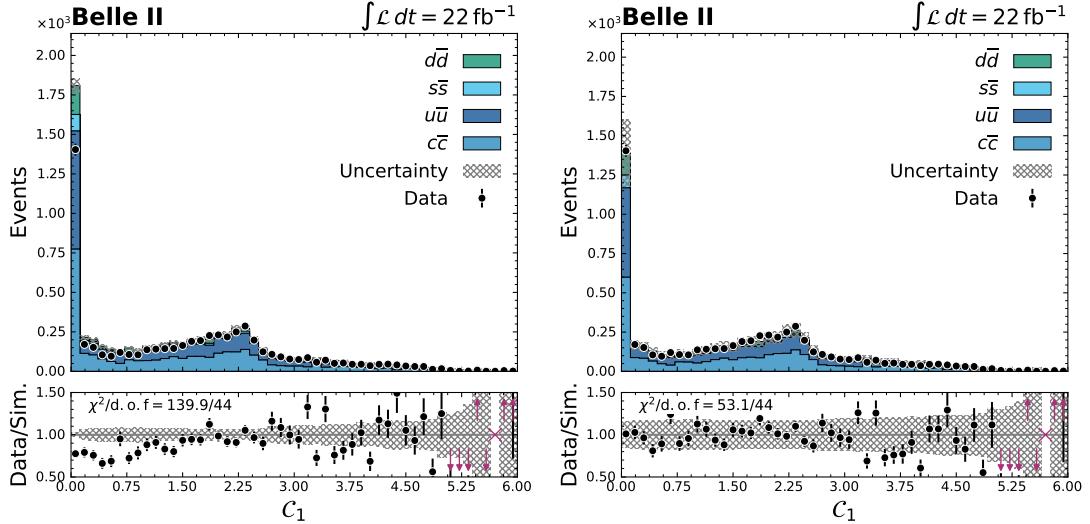


FIG. 3. Distribution of C_1 (first CLEO cone [39]) in an off-resonance sample before (left) and after (right) the continuum modeling correction is applied. The modeling uncertainties, described in Section V E 9, are included in the error band after applying the continuum calibration. The bottom panel shows the ratio between data and simulation yields in each bin. The purple arrows indicate that the point is outside of the plotting range and the purple crosses indicate the absence of simulated data in a particular bin.

subdivided selection plane is illustrated in Figure 4. In Figure 5, clear disagreements between data and simulation distributions of E_ℓ^B , M_X and q^2 can be seen.

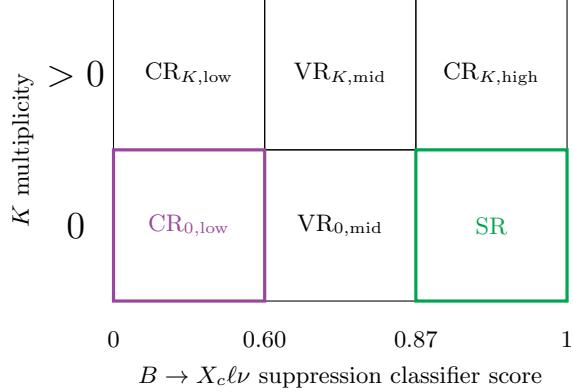


FIG. 4. The data samples after reconstruction, preselection and continuum suppression are subdivided into six regions illustrated here. The selection plane is subdivided based on the kaon multiplicity (vertical axis) and the $B \rightarrow X_c \ell \nu$ suppression classifier score (horizontal axis). The signal and control regions used for the signal extraction are highlighted in green and purple, respectively.

Background originating from continuum as well as fake and secondary leptons is known to be poorly modelled in simulation. However, continuum events are corrected following the procedure described in Section IV A and these background sources represent in total less than 10% of all events in the control

regions. The mismodeling is therefore expected to stem from $B \rightarrow X_c \ell \nu$ events. There are several known shortcomings in the simulation of $B \rightarrow X_c \ell \nu$ decays. The $B \rightarrow D^{(*)} \eta \ell \nu$ modes have never been observed and are used as a hypothesis for the decays that fill the $B \rightarrow X_c \ell \nu$ gap. Therefore, a relative uncertainty of 100% is assigned to their assumed branching fractions. This uncertainty doesn't cover the full difference between data and simulation distributions shown in Figure 5. Furthermore, most D meson decays with three or more particles are not simulated according to measured decay distributions but use simple phase-space models, which can bias the kinematics of charmed B decays. Moreover, it was observed in Belle II studies [44] that measurements can be affected by the poor modeling of charm decays involving K_L^0 mesons. The impact of the modeling of $B \rightarrow X_c \ell \nu$ events with a D meson decaying to a K_L^0 is studied in detail by scaling their contribution by arbitrary factors. Based on studies performed in the measurement of the $B^+ \rightarrow K^+ \nu \bar{\nu}$ branching fraction [44], this component is increased by 30% which only marginally improves the agreement between data and simulation. This indicates that it is unlikely that the mismodeling of $D \rightarrow K_L^0 X$ decays can explain the observed mismodeling. As no single source of mismodeling could be identified, the $B \rightarrow X_c \ell \nu$ component is corrected as a whole using the control regions defined above. The normalization and shape from simulation are corrected separately.

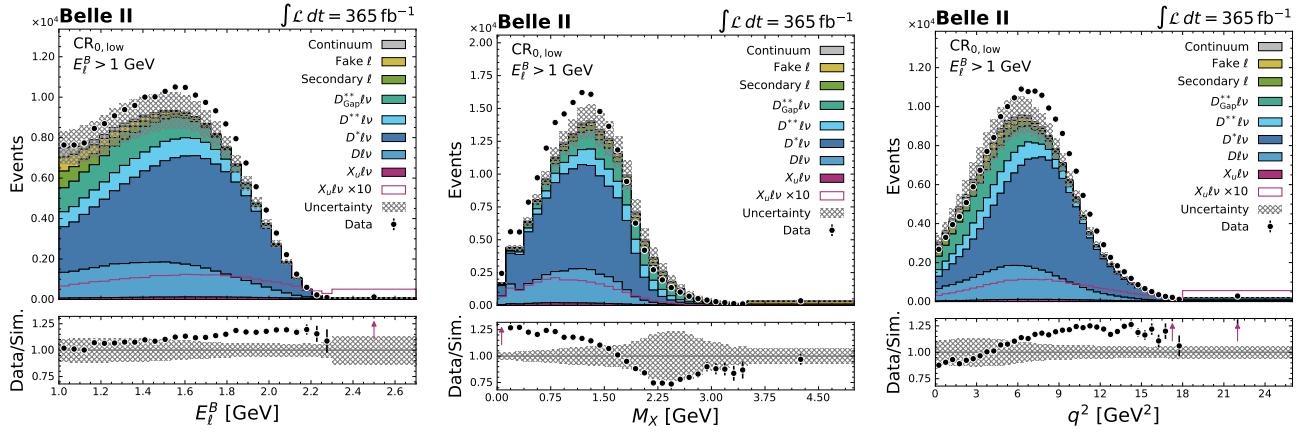


FIG. 5. E_ℓ^B , M_X and q^2 distributions in data and simulation in the $CR_{0,\text{low}}$ region defined in the text. Only the $E_\ell^B > 1$ GeV kinematic selection is applied. All systematic uncertainties as described in Section V E are included. The bottom panel shows the ratio between data and simulation yields in each bin. The purple arrows indicate that the point is outside of the plotting range.

C. $B \rightarrow X_c \ell \nu$ normalization correction

The signal-region $B \rightarrow X_c \ell \nu$ normalization is corrected before the signal extraction via a method relying on the observed normalizations in the three control regions. Since the kaon multiplicity and classifier score are only weakly correlated, it is assumed that the ratio of experimental over simulated $B \rightarrow X_c \ell \nu$ events in the signal region, r_{SR} , can be determined from the same ratio in the control regions as

$$r_{\text{SR}} = \frac{r_{0,\text{low}} \cdot r_{K,\text{high}}}{r_{K,\text{low}}}. \quad (10)$$

The $B \rightarrow X_c \ell \nu$ normalization correction factors $r_{0,\text{low}}$, $r_{K,\text{low}}$ and $r_{K,\text{high}}$ are calculated by subtracting from data and simulation all $B \rightarrow X_u \ell \nu$, continuum, as well as fake and secondary lepton simulated events. These components represent between 2% and 10% of all events in $CR_{0,\text{low}}$ and $CR_{K,\text{low}}$ and between 8% and 25% of events in $CR_{K,\text{high}}$, depending on kinematic selections. Since this computation relies on various sources of background and signal being well modelled, an uncertainty is derived on the three ratios by varying the total yields of $B \rightarrow X_u \ell \nu$, continuum and fake and secondary lepton events by 50% and computing the normalization correction factors for each variation. The calculation defined in Equation 10 is tested by replacing $CR_{K,\text{high}}$ and the signal region by $VR_{K,\text{mid}}$ and $VR_{0,\text{mid}}$ respectively. Although the signal purity is larger in $VR_{0,\text{mid}}$ (about 10–20% depending on kinematic selections) than in all other control regions, it remains lower than in the signal region (30–70%). The $B \rightarrow X_c \ell \nu$ normalization correction obtained for $VR_{0,\text{mid}}$ via the control region method (1.07, 1.15, 1.21, from the broadest set of kinematic selections to the tightest) and the factor obtained by directly taking the $B \rightarrow X_c \ell \nu$ data over simulated data yield ratio in that region (1.03, 1.16, 1.22) are

compared. The relative difference between these two factors is considered as an additional uncertainty on r_{SR} . In the signal extraction, two nuisance parameters are included to float the normalization of the $B \rightarrow X_c \ell \nu$ component in $CR_{0,\text{low}}$ and in the signal region separately. The normalization parameters are constrained by Gaussian functions with widths equal to the uncertainty on the normalization correction of each region. The values of $r_{0,\text{low}}$ and r_{SR} in the three phase-space regions considered in the signal extraction are given in Table III.

TABLE III. Values of the normalization correction factors $r_{0,\text{low}}$ and r_{SR} in the three phase-space regions considered for the signal extraction. The E_ℓ^B , M_X and q^2 kinematic selections are: $E_\ell^B > 1.0$ GeV, $M_X < 1.7$ GeV and $q^2 > 8$ GeV 2 . The total uncertainties computed as described in the text are reported.

| Kinematic selection | $CR_{0,\text{low}}$ | SR |
|----------------------|---------------------|-----------------|
| E_ℓ^B | 1.11 ± 0.05 | 1.01 ± 0.08 |
| E_ℓ^B, M_X | 1.18 ± 0.04 | 1.07 ± 0.12 |
| E_ℓ^B, M_X, q^2 | 1.24 ± 0.03 | 1.09 ± 0.21 |

The $B \rightarrow X_c \ell \nu$ shape correction is implemented directly in the signal extraction procedure by simultaneously fitting the signal region and $CR_{0,\text{low}}$ as described in Section V B.

V. SIGNAL EXTRACTION

The $B \rightarrow X_u \ell \nu$ partial branching fraction is extracted from a binned template fit with the `pyhf` package [45].

A. Statistical model

The statistical model relies on the simultaneous measurement of event counts n in two independent regions⁴ – a signal region and a control region – using binned distributions of a given variable. These distributions are modelled as the sum of three different components (*i.e.* samples or templates, specified in Section V C). Within the statistical model, a set of free and constrained parameters, η and χ respectively, parametrize the variations of each template yield. The impact of each nuisance parameter χ is limited by a Gaussian constraint term c_χ . The constraints are defined using a set of auxiliary measurements \mathbf{a} (one measurement a_χ for each parameter), which correspond to component yield variations caused by various systematic effects. Thus, the statistical model f , in its most general form, can be written as

$$f(\mathbf{n}, \mathbf{a} | \eta, \chi) = \prod_c \prod_b \text{Pois}(n_{cb} | \nu_{cb}(\eta, \chi)) \prod_{\chi \in \chi} c_\chi(a_\chi | \chi), \quad (11)$$

where $\mathbf{n} \equiv \{n_{cb}\}$ are the observed event yields. The products run over all regions c , bins b and constrained parameters χ ; and ν_{cb} are the expected yields.

$$\begin{aligned} \nu_{cb} &= \sum_s \nu_{scb}(\eta, \chi) \\ &= \sum_s \left(\prod_{\kappa \in \kappa} \kappa_{scb}(\eta, \chi) \right) \\ &\times \left(\nu_{scb}^0(\eta, \chi) + \sum_{\Delta \in \Delta} \Delta_{scb}(\eta, \chi) \right), \end{aligned} \quad (12)$$

where the sum runs over all templates s , κ_{scb} and Δ_{scb} are multiplicative and additive modifiers respectively and ν_{scb}^0 is the nominal yield in a given bin. The different types of modifiers implemented in the fit are discussed in the following subsections.

B. Fitted regions and $B \rightarrow X_c \ell \nu$ shape correction

As detailed in Section IV B, $B \rightarrow X_c \ell \nu$ decays appear to be poorly modelled. Normalization correction factors are calculated and applied before the fit whereas the correction of the $B \rightarrow X_c \ell \nu$ component shape is implemented directly into the fit. In order to constrain the shape of the $B \rightarrow X_c \ell \nu$ background in the signal region, distributions are fitted simultaneously in the signal region and $\text{CR}_{0,\text{low}}$, the latter being orthogonal to the signal region and enriched in $B \rightarrow X_c \ell \nu$ events. This method not only corrects the $B \rightarrow X_c \ell \nu$ shape in the signal region, it also properly accounts for all correlations between correction factors

and other fit parameters to determine a $B \rightarrow X_u \ell \nu$ branching fraction value that incorporates uncertainties related to the $B \rightarrow X_c \ell \nu$ background corrections. Using any control region to correct the shape of the $B \rightarrow X_c \ell \nu$ component relies on the assumption that the disagreement between data and simulation is similar in the signal region and the chosen control region. This assumption is supported by the fact that the ratio of $B \rightarrow X_c \ell \nu$ data over simulated data yields follows a similar trend in relevant kinematic variables in all three control regions as illustrated in Figure 6. To choose the control region used for correcting the signal region mismodeling, two considerations must be taken into account: the similarity in composition and shape of the $B \rightarrow X_c \ell \nu$ component in the control and signal region and the sample size relative to the signal region. Based on these two considerations, $\text{CR}_{0,\text{low}}$ is chosen for the fit.

The partial branching fraction of $B \rightarrow X_u \ell \nu$ is measured in three different phase-space regions defined by applying the kinematic selections of Table II in the signal region. Each set of selections is also applied in the control region and therefore characterizes a separate fit setup. The differences between these setups are highlighted in the following sections. The setups are referred to as fits 1, 2 and 3 from the broadest phase-space region to the tightest (see Table II).

C. Fit templates

The fit is performed using the following templates:

- Signal ($B \rightarrow X_u \ell \nu$): this template includes all signal events whose generated kinematics fall in the chosen phase-space region (*signal-in*). Signal decays are modelled via the DFN framework [19]. The template is assigned an unconstrained normalization factor in the statistical model whose postfit value can be directly translated into a measurement of the branching fraction. This factor is fixed to its value from simulation in $\text{CR}_{0,\text{low}}$ where its contribution is negligible. An additional template (*signal-out*) is defined to cover the small fraction of signal events which pass the reconstruction-level selections given in Table II but whose generated kinematic parameters fall outside the target signal region (based on the generator-level selections). The resolution of the E_ℓ^B distribution is extremely good, allowing the $B \rightarrow X_u \ell \nu$ -out yield to be neglected in fit 1. In fits 2 and 3, this component is kept as it represents 2% and 5% of all signal events in their respective signal regions. The template is assigned the same unconstrained normalization as the signal-in template but nuisance parameters can vary its normalization and shape independently.
- $B \rightarrow X_c \ell \nu$ background: this template represents the largest source of background. It is composed

⁴ Referred to as *channels* in the `pyhf` documentation.

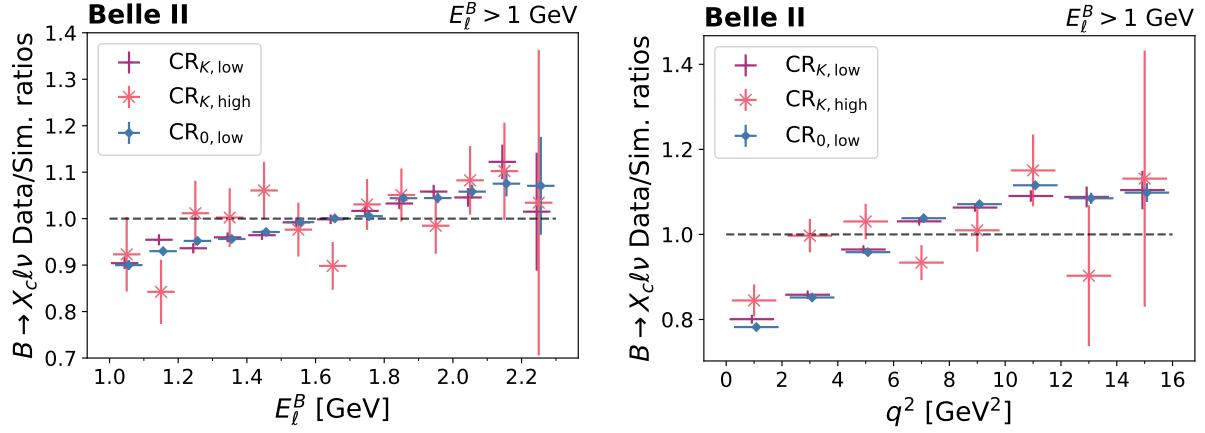


FIG. 6. The normalized ratios of $B \rightarrow X_c \ell \nu$ yields in data to simulation are shown for the three control regions $CR_{K,\text{low}}$ (purple), $CR_{K,\text{high}}$ (pink) and $CR_{0,\text{low}}$ (blue) in E_ℓ^B (top) and q^2 (bottom). Only the $E_\ell^B > 1$ GeV kinematic selection is applied. The error bars include the statistical uncertainty of the experimental and simulated data samples.

of $B \rightarrow D \ell \nu$, $B \rightarrow D^* \ell \nu$, $B \rightarrow D^{**} \ell \nu$ and $B \rightarrow D_{\text{Gap}}^{**} \ell \nu$ decays. To capture the shape of this template in $CR_{0,\text{low}}$, one free-floating normalization factor per bin, common to the signal and control-region distributions, is used. The normalization of this template is corrected prior to fitting, following the procedure described in Section IV C. A constrained normalization parameter is assigned to the $B \rightarrow X_c \ell \nu$ template in each region to float the overall normalization within the uncertainties given in Table III.

- *Other backgrounds:* all other backgrounds are summed up in a single template. They include continuum events as well as fake and secondary lepton events. After applying the continuum suppression classifier, the yields of these components are relatively small compared to semileptonic event yields. In fit 1, this template is assigned a normalization modifier which is constrained by a Gaussian function with a width of 50% of the total yields to conservatively vary this component. In fits 2 and 3, the contribution from other backgrounds is negligible and this modifier is therefore not added.

D. Fitted variables

Different variables are tested for the fit. To combine information from different shapes, two and three-dimensional fits can be performed. In this analysis, the two-dimensional $E_\ell^B:q^2$ variable is fitted in fits 1 and 2 by splitting E_ℓ^B into four bins and q^2 into three bins. When applying both the tight M_X and q^2 selections, only E_ℓ^B spans a range large enough to extract the signal branching fraction, and thus only E_ℓ^B , split into eight bins, is used in fit 3. The binning for each variable is chosen to

mitigate model dependence. Since inclusive $B \rightarrow X_u \ell \nu$ models are known to poorly describe the shape of kinematic distributions in the high E_ℓ^B and q^2 region, a single bin is defined for this region. The variable and binning for each fit are given in Table IV.

TABLE IV. Variable and binning for the three fit configurations described in the text.

| Variable | Binning |
|------------------------|--|
| Fit 1 $E_\ell^B : q^2$ | $[1.0, 1.3, 1.6, 1.9, 2.7] \text{ GeV} \times [0, 4, 8, 26] \text{ GeV}^2$ |
| Fit 2 $E_\ell^B : q^2$ | $[1.0, 1.3, 1.6, 1.9, 2.7] \text{ GeV} \times [0, 4, 8, 26] \text{ GeV}^2$ |
| Fit 3 E_ℓ^B | $[1.0, 1.4, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.7] \text{ GeV}$ |

E. Systematic uncertainties

Sources of systematic uncertainty are directly implemented in the fit as nuisance parameters. Each nuisance parameter can affect either the shape or the normalization of the template(s) to which it is assigned, or both. An overview of systematic uncertainties considered is given in the following. Two sources of uncertainty (number of $\Upsilon(4S)$ decays and X_u hadronization modeling) which are not considered in the fit are also described.

For the uncertainties related to the D decay branching fractions, the continuum calibration, the charged particle identification and the hadronic tagging, the number of parameters is truncated to avoid implementing a large number of nuisance parameters with a negligible impact. To do so, the principal components of each covariance matrix are taken such that at least 99% of the total variance is covered. This method reduces the total number of parameters in each fit from about 1,000 to

about 150.

1. $B \rightarrow X_u \ell \nu$ modeling

The resonant and non-resonant branching fractions (given in Table I) are varied within their respective uncertainties therefore yielding one nuisance parameter per decay type. The modeling of $B \rightarrow X_u \ell \nu$ resonant modes relies on assumed form factor models which depend on a set of parameters extracted from lattice QCD and dedicated measurements. The form factor parameter variations are calculated using the **eFFORT** [46] package. The parameter values, uncertainties and correlation matrices are taken from the references given for each model in Section II. Two nuisance parameters are added to represent the uncertainties in $m_b^{\text{KN}} = 4.66 \pm 0.04$ GeV and $a^{\text{KN}} = 1.3 \pm 0.5$, including their covariance. In addition, to factor in differences between inclusive $B \rightarrow X_u \ell \nu$ models, a set of events is simulated via the Bosch, Lange, Neubert, Paz (BLNP) framework [10, 47]. The difference between the nominal fitted distribution and the BLNP distribution is added as a symmetric uncertainty. Moreover, applying a kaon veto to suppress charm background in the signal region also rejects $B \rightarrow X_u \ell \nu$ events where the X_u system hadronizes as a pair of kaons. The branching fraction of $B \rightarrow K\bar{K}(X)\ell \nu$ decays (where X is a charmless hadronic system) is not known and an uncertainty related to this contribution is therefore extracted by varying the production of s quarks relative to u or d production, γ_s . The two values measured by the TASSO [48] ($\gamma_s = 0.35 \pm 0.05$) and JADE [49] ($\gamma_s = 0.27 \pm 0.06$) collaborations are averaged and the uncertainty is chosen to cover the full spread of 1σ variations from the two values, giving $\gamma_s = 0.30 \pm 0.09$ (which corresponds to a branching fraction of roughly 0.02%). Three sets of events with γ_s equal to 0.21, 0.30 and 0.39 are produced. The sample simulated with γ_s equal to its central value is used as nominal and the variations with respect to the other two distributions are added as systematic uncertainties. For each variation related to the $B \rightarrow X_u \ell \nu$ modeling, the hybrid weights are recalculated (see Equation 3).

2. $B \rightarrow X_c \ell \nu$ modeling

All $B \rightarrow X_c \ell \nu$ and all D meson decay branching fractions are varied within their respective uncertainties (see Table I). The uncertainties of the unmeasured $B \rightarrow D^{(*)}\eta\ell\nu$ branching fractions are chosen to be equal to 100%. The branching fraction uncertainty of $B \rightarrow D_0^*\ell\nu$ decays is increased by about a factor of six to partially cover the large difference between the two measurements entering the HFLAV average calculation [50, 51]. The number of parameters related to the branching fractions of D meson decays is truncated from 227 to one. For $B \rightarrow X_c \ell \nu$ decays, the form factor parameter variations are computed using the **Hammer** [52] package. The pa-

rameter values, uncertainties and correlation matrices are taken from the references given for each model in Section II.

3. f^{+-}/f^{00}

The ratio of production fractions of charged and neutral pairs of B mesons in $\Upsilon(4S)$ decays is taken from Ref. [5]: $f^{+-}/f^{00} = 1.052 \pm 0.031$. A nuisance parameter is added in the fit to vary the relative fraction of B^+ and B^0 decays accordingly.

4. π_s efficiency

Efficiency correction factors for π_s originating from D^* decays are extracted separately for charged and neutral pions from studies of $B^0 \rightarrow D^{*-}(\rightarrow \bar{D}^0\pi_s^-)\pi^+$ and $B^+ \rightarrow \bar{D}^{*0}(\rightarrow \bar{D}^0\pi_s^0)\pi^+$ events. The correction factors are extracted in three laboratory-frame momentum bins between 50 and 200 MeV. Separately for charged and neutral π_s , covariance matrices are derived to encode the uncertainties versus momentum. This results in a total of six nuisance parameters.

5. Tracking efficiency

Track finding efficiency differences between data and simulation for tracks with a momentum higher than 200 MeV are studied in $e^+e^- \rightarrow \tau^+\tau^-$ events with one τ decaying to three charged tracks. A correction factor consistent with unity with an associated uncertainty of 0.24% per track is found. Therefore, the efficiency of each reconstructed track is varied up and down by 0.24% in simulated events and the difference between the resulting up and down variations of each template is added as an uncertainty.

6. Charged particle identification

The efficiencies and misidentification rates of e^\pm , μ^\pm and K^\pm are extracted from independent studies in bins of the track charge, laboratory-frame momentum and polar angle and applied to simulated events [34, 36]. The correction factors, which combine efficiency and misidentification corrections, are typically in the range 0.95 to 1.02 for kaons and 0.90 to 1.03 for leptons. Each correction factor has associated statistical and systematic uncertainties. A total of 200 variations built from random samples are drawn from the probability density function for lepton and hadron identification factors separately. The number of parameters implemented in the fit is truncated from 400 to six in each fit.

7. K_S^0 reconstruction efficiency

The reconstruction efficiency of K_S^0 as a function of the flight distance from the interaction point is studied in $D^{*+} \rightarrow D^0(\rightarrow K_S^0\pi^+\pi^-)\pi^+$ decays. An efficiency correction factor of 0.55% per cm of flight distance is extracted. The correction factor is applied to each K_S^0 candidate to reweight simulated events and the full difference between the nominal and the corrected template is added as a symmetric uncertainty.

8. Hadronic tagging

Hadronic tagging efficiency correction factors are extracted from independent studies where the signal side B meson is reconstructed in $B \rightarrow X\ell\nu$ and $B \rightarrow D^{(*)}\pi$ decays. The correction factors are extracted in 12 (11) separate tag-side B^+ (B^0) channels covering all decay modes used by the FEI algorithm and applied to simulated events. The number of parameters included in the fit is truncated from 23 to two.

9. Continuum calibration

Simulated continuum events are corrected using the data-driven approach described in Section III C. The full difference between the nominal distribution and the distribution corrected by the normalization factors is added as an uncertainty. Uncertainties on the continuum component shape also arise because of the limited size of the samples used to extract the correction factors. A total of 200 bootstrapped variations of the off-resonance experimental and simulated data samples are trained separately yielding as many sets of correction factors. The number of parameters implemented in the fit is truncated from 200 to five.

10. Finite size of simulated data samples

To account for the limited size of simulated samples, one constrained nuisance parameter per fit bin is added.

11. Fit template normalizations

In fit 1, a normalization parameter is assigned to the *other backgrounds* template, which contributes to the total signal branching fraction uncertainty. The nuisance parameter is constrained by a Gaussian function with a width of 50% of the total component yields to vary the component normalization within a conservative range. In fits 2 and 3, the yields of non-semileptonic events being negligible, this parameter is not added but other nuisance parameters related to this template can vary both

its normalization and shape. Furthermore, the control region and signal region $B \rightarrow X_c\ell\nu$ template normalizations are corrected before the fit and they are allowed to float separately within the correction factor uncertainties given in Table III.

12. Number of $\Upsilon(4S)$ decays

To extract a $B \rightarrow X_u\ell\nu$ branching fraction from the fit result, the output signal normalization is multiplied by the assumed input branching fraction. Since charged and neutral $B \rightarrow X_u\ell\nu$ decays are simulated separately but the measurement is performed simultaneously on all decays, the combined inclusive $B \rightarrow X_u\ell\nu$ input branching fraction is computed from the number of simulated signal events N_{X_u} and the number of collected $\Upsilon(4S)$ events such that $\mathcal{B}(B \rightarrow X_u\ell\nu) = N_{X_u}/(2N_{\Upsilon(4S)})$. The small number of $\Upsilon(4S)$ decays to final states without open- b mesons has been neglected. The number of $\Upsilon(4S)$ events collected by Belle II has been measured to be $N_{\Upsilon(4S)} = (387 \pm 6) \times 10^6$ and its uncertainty is therefore added to the total uncertainty of the measured branching fraction.

13. X_u hadronization modeling

An uncertainty on the fragmentation modeling of the X_u system in non-resonant signal events is derived using the signal region π^\pm multiplicity. It was checked beforehand in control regions that this variable is reasonably well modelled in $B \rightarrow X_c\ell\nu$ decays. In each of the three signal regions, the fit results are projected onto the π^\pm multiplicity distribution and it is assumed that the remaining discrepancy between data and simulation is entirely due to the imperfect X_u hadronization modeling. For each kinematic region, three scale factors are extracted in bins of the generator-level π^\pm multiplicity of non-resonant $B \rightarrow X_u\ell\nu$ events (0, 1, 2 and more π^\pm) such that the simulated postfit π^\pm multiplicity distribution matches perfectly the data distribution. The scale factors are applied to the pre-fit signal region distributions and the three fits are repeated. The difference between the measured $B \rightarrow X_u\ell\nu$ branching fraction and the branching fraction extracted from the rescaled distribution is taken as an uncertainty. Part of the uncertainty could originate from mismodeling of backgrounds in the π^\pm multiplicity distribution. However, because of the relatively large signal purity in the three kinematic regions, the full difference is conservatively attributed to the $B \rightarrow X_u\ell\nu$ modelling.

F. Signal extraction validation

Before fitting signal region data, the signal extraction is validated using the regions where a kaon is found. A

simulated $B \rightarrow X_u \ell \nu$ component with known normalization is injected into the $\text{CR}_{K,\text{high}}$ experimental and simulated data and the $\text{CR}_{K,\text{high}}$ and $\text{CR}_{K,\text{low}}$ regions are used as input to the fit. This allows the normalization of the signal injected into experimental data to be controlled while keeping the simulated data normalization fixed. The $B \rightarrow X_c \ell \nu$ normalization in the two regions is corrected before the fit, as is done in the actual signal fits.

The first test performed with this setup is to vary the normalization of the signal component injected into data in the range $0.85 - 1.15$ in steps of 0.05 while keeping the normalization of the signal in simulated data fixed. In this test, when the measured signal normalization is compared to the input, a slight bias is observed. The bias arises because of differences in shape and composition of the $B \rightarrow X_c \ell \nu$ component in the control and signal regions used in the fit. Because of these differences, the $B \rightarrow X_c \ell \nu$ shape correction extracted from the control region doesn't adjust perfectly the $B \rightarrow X_c \ell \nu$ shape in the signal region, thus forcing the fit to pull on other fit parameters, including the free-floating $B \rightarrow X_u \ell \nu$ normalization, to match the simulated templates to data. A correction for this bias is determined by using the response obtained in the validation setup and results in a $1 - 3\%$ correction for fits 1, 2 and 3 depending on the measured signal normalization.

Since this test relies on a single specific data configuration, an uncertainty is extracted by producing a large number of randomly sampled data sets. Each sampled data set is created by varying the $\text{CR}_{K,\text{high}}$ data yields within their statistical uncertainties while keeping the injected signal yields and the $\text{CR}_{K,\text{low}}$ data fixed. A set of 1,000 sampled data sets is produced for each injected signal normalization and each sampled data set is fitted. The standard deviation of the resulting seven output signal normalization Gaussian distributions is nearly constant and the largest of the seven values for each set of kinematic selections is added as an uncertainty on the measured $B \rightarrow X_u \ell \nu$ branching fraction. This bias correction results in a 2.6% , 2.3% and 1.8% relative uncertainty for fits 1, 2 and 3 respectively.

Since the correction and uncertainty are derived from the $\text{CR}_{K,\text{high}}$ and $\text{CR}_{K,\text{low}}$ regions instead of $\text{CR}_{0,\text{low}}$ and the signal region, an additional uncertainty related to the differences in $B \rightarrow X_c \ell \nu$ composition in each region is extracted. The $B \rightarrow X_c \ell \nu$ composition in the signal and control regions for the three sets of kinematic selections is shown in Appendix B, Table X. The injected signal fit setup is repeated by fixing the normalization of the injected signal to unity and by rescaling the $B \rightarrow X_c \ell \nu$ composition of $\text{CR}_{K,\text{high}}$ ($\text{CR}_{K,\text{low}}$) such that it matches the composition of the signal region ($\text{CR}_{0,\text{low}}$). The difference in measured normalization with and without rescaling the composition is taken as an uncertainty, adding a 1.4% , 5.7% and 0.2%

contribution to the uncertainty of fits 1, 2 and 3, respectively. The difference in $B \rightarrow X_c \ell \nu$ composition between $\text{CR}_{K,\text{high}}$ and the signal region in the phase-space region considered for fit 2 being relatively large (see Table X in Appendix B), the composition uncertainty extracted for this fit is larger than in the other fits. Besides, in fit 3, the E_ℓ^B distribution is not particularly sensitive to the $B \rightarrow X_c \ell \nu$ composition and this uncertainty is therefore relatively low. The $B \rightarrow X_c \ell \nu$ composition uncertainty is subleading in fits 1 and 3 but is one of the dominant sources in fit 2.

G. Prefit distributions

The three fitted variables in $\text{CR}_{0,\text{low}}$ and the signal region are shown in Figure 7. While the agreement between data and simulation appears to be relatively good in signal enriched regions, this is not the case in regions of the phase space where $B \rightarrow X_c \ell \nu$ events dominate. The shape mismodeling of this component is expected to be corrected by simultaneously fitting each distribution in the signal region and $\text{CR}_{0,\text{low}}$. However, the normalization of the charm background in the signal regions appears to be overestimated after applying the normalization correction factors (*cf.* Table III). To evaluate the impact of the overestimated $B \rightarrow X_c \ell \nu$ normalizations, the injected signal setup described in Section V F is used. The injected signal normalization is fixed to unity and the $B \rightarrow X_c \ell \nu$ template in $\text{CR}_{K,\text{high}}$ is scaled up by an *overestimation factor* in the range $[1.05, 1.10, 1.15, 1.20]$. After correcting for the bias described in Section V F, a slight bias increasing linearly with the $B \rightarrow X_c \ell \nu$ overestimation factor is observed. In the three signal regions considered, it is not possible to extract precisely the true $B \rightarrow X_c \ell \nu$ normalization without making assumptions about the $B \rightarrow X_u \ell \nu$ normalization, which is extracted from the fit. By checking the ratio of data to simulated data yields, it is estimated that the $B \rightarrow X_c \ell \nu$ normalization is overestimated by a factor roughly in the range $1.10 - 1.15$ in all three regions. A flat correction factor is therefore extracted assuming that the overestimation factor is equal to 1.15. The entire difference between the measured branching fraction and the corrected one is added as an uncertainty. The $B \rightarrow X_c \ell \nu$ normalization bias correction is tested with different correction factors assuming various overestimation factors in the range $1.00 - 1.30$. In all cases the assigned uncertainty is large enough to cover the true injected signal normalization after applying the bias correction.

VI. RESULTS

A. Partial branching fractions

The postfit distributions in each of the three signal regions are shown in Figure 8. The excellent post-fit

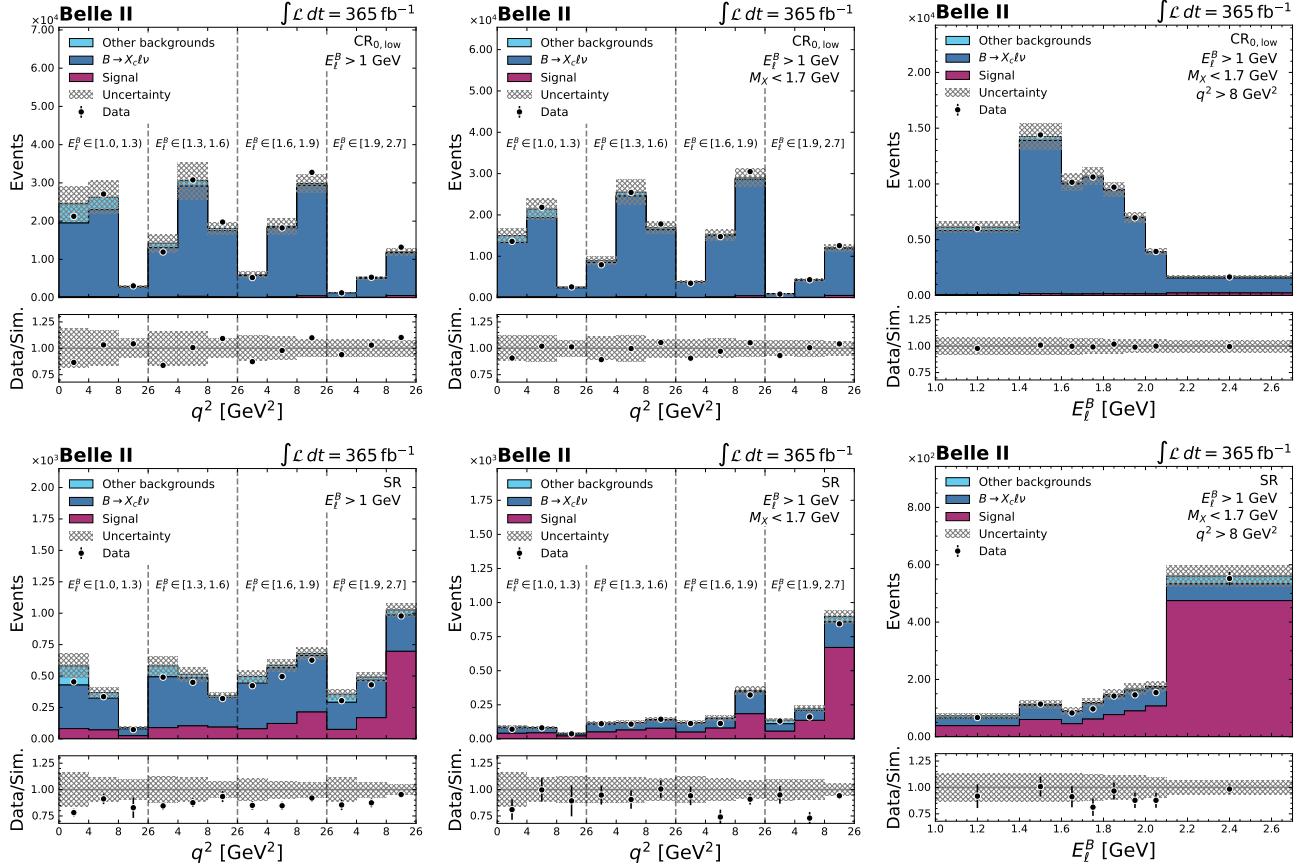


FIG. 7. The $E_\ell^B:q^2$ and E_ℓ^B distributions in data and simulation before the fit in $CR_{0,\text{low}}$ (top) and the signal region (bottom) considered for fit 1 (left), 2 (middle) and 3 (right). The $E_\ell^B:q^2$ distribution is shown in bins of E_ℓ^B (in GeV) and q^2 as indicated in the figures. The $B \rightarrow X_c \ell \nu$ normalization correction factors have been applied (see Table III). The signal-in and out components are merged in the *Signal* template. The bottom panel shows the ratio between data and simulation yields in each bin.

agreement is confirmed by the p -values obtained for each fit: 0.79, 0.20 and 0.84 for fits 1, 2 and 3, respectively. The event yields as extracted from fits 1, 2 and 3 are given in Table V. In the kinematic regions considered for fits 1 and 2, the signal yields are similar to those reported in the latest Belle measurement [31] while the background yields are approximately 30–45% lower, despite using a data set that is about half the size of the 711 fb^{-1} data set collected by the Belle detector. The higher signal reconstruction efficiency can be explained by the improvements in the FEI performance compared to the Full Reconstruction algorithm used in Belle [53], most notably the inclusion of more B decay chains. Furthermore, in this measurement, the continuum and $B \rightarrow X_c \ell \nu$ backgrounds are suppressed by relying on two neural networks which are optimized separately whereas a single BDT was used for that purpose in the Belle measurement. Finally, the increased reconstruction efficiency for π_s helps to reject a larger fraction of $B \rightarrow D^* \ell \nu$ events.

The inclusive $B \rightarrow X_u \ell \nu$ partial branching fractions

TABLE V. Signal, background and data yields in the signal region as extracted from each fit. The error given for $B \rightarrow X_u \ell \nu$ and background yields is the total uncertainty as obtained from each fit. The three sets of kinematic selections used in each fit are given in Table II.

| Fit | $B \rightarrow X_u \ell \nu$ in $B \rightarrow X_u \ell \nu$ out | Backgrounds | Data |
|-------|--|-------------|------------------------------|
| Fit 1 | 1615 ± 122 | – | 3782 ± 128 5383 ± 73 |
| Fit 2 | 1240 ± 105 | 33 ± 5 | 965 ± 95 2236 ± 47 |
| Fit 3 | 876 ± 68 | 42 ± 4 | 439 ± 62 1355 ± 37 |

obtained from fits 1, 2 and 3 are

$$\Delta \mathcal{B}(B \rightarrow X_u \ell \nu)_1 = (1.54 \pm 0.08 \pm 0.12) \times 10^{-3}, \quad (13)$$

$$\Delta \mathcal{B}(B \rightarrow X_u \ell \nu)_2 = (0.95 \pm 0.05 \pm 0.10) \times 10^{-3}, \quad (14)$$

$$\Delta \mathcal{B}(B \rightarrow X_u \ell \nu)_3 = (0.55 \pm 0.03 \pm 0.05) \times 10^{-3}, \quad (15)$$

where the uncertainties are statistical and systematic, respectively. With a relative uncertainty of 9.6%, the measurement of the branching fraction over the

phase-space region defined by the selection $E_\ell^B > 1$ GeV (fit 1) is competitive with the most precise measurement performed by the BaBar collaboration (8.1% relative uncertainty) using an inclusive tagging method [54]. It is more precise than similar measurements performed by the Belle (11.4%) and BaBar (14.5%) collaborations using a hadronic tagging method [31, 55]. With relative uncertainties of 11.9% and 10.7%, the measurements of the branching fraction over the two tighter phase-space regions are competitive with the most precise determinations from these two experiments. When extrapolated to the full phase-space region, the three measured $B \rightarrow X_u \ell \nu$ branching fractions are $(1.78 \pm 0.17) \times 10^{-3}$, $(1.67 \pm 0.20) \times 10^{-3}$ and $(1.76 \pm 0.19) \times 10^{-3}$ which can be compared to the latest average value listed in the PDG Review of Particle Physics [15], $(1.92 \pm 0.21) \times 10^{-3}$.⁵ The three values are correlated. Assuming a systematic correlation of 100% and estimating the statistical correlation from bootstrapping, it was checked that the values agree with each other within one standard deviation. The correlation matrix of the statistical uncertainties is given in Table VI. The statistical correlation is similar between each pair of fits because there is a large overlap in signal events and because, in all three fits, the measurement is sensitive to the region with best signal purity *i.e.* the region with high E_ℓ^B and q^2 .

TABLE VI. Correlations between statistical uncertainties of the branching fractions measured in fits 1, 2 and 3.

| | Fit 1 | Fit 2 | Fit 3 |
|-------|-------|-------|-------|
| Fit 1 | 1 | 0.86 | 0.81 |
| Fit 2 | 0.86 | 1 | 0.83 |
| Fit 3 | 0.81 | 0.83 | 1 |

The total uncertainty breakdown in terms of individual sources of uncertainty is shown in Table VII and the numerical impact of uncertainties related to the bias correction and the sample composition on the measured partial branching fractions is summarized in Table VIII. These individual uncertainties are intended to provide insight into the sensitivity of the measurement to inputs and assumptions. However, it is not possible to account for correlations between fit parameters when calculating the individual uncertainties and the numbers quoted are therefore approximate. Hence, the sum in quadrature of all individual sources of uncertainty does not necessarily match the total uncertainty obtained from the fit. For

all three measured branching fractions, the DFN model parameter uncertainty represents one of the largest sources of uncertainty. It is larger in fit 3 than in the other two fits because the $B \rightarrow X_u \ell \nu$ acceptance is the smallest among the three kinematic regions making the dependence on the model stronger. One of the leading uncertainties of the branching fractions measured with fits 2 and 3 is the hadronic system fragmentation modeling uncertainty. The result obtained from fit 2 is significantly impacted by the $B \rightarrow X_c \ell \nu$ composition uncertainty described in Section V F because the relative fractions of $B \rightarrow X_c \ell \nu$ sub-components are considerably different between the signal region and $CR_{K,\text{high}}$ when applying the E_ℓ^B and M_X selections (see Table X).

B. $|V_{ub}|$ determination

The value of $|V_{ub}|$ can be determined from the measured partial branching fractions using Equation 2. We use for the value of the B meson lifetime, the average of the B^+ and B^0 lifetimes, $\tau_B = 1.578 \pm 0.003$ ps [5]. Following Ref. [5], three different theoretical frameworks are used to extract $|V_{ub}|$:

- the BLNP model [10, 47] with input values in the shape-function (SF) renormalization scheme [56, 57]: $m_b^{\text{SF}} = 4.600 \pm 0.022$ GeV and $\mu_\pi^{2\text{SF}} = 0.184^{+0.051}_{-0.062}$ GeV 2 ;
- the Dressed Gluon Exponentiation (DGE) model [11] with input values in the $\overline{\text{MS}}$ scheme: $m_b^{\overline{\text{MS}}} = 4.206 \pm 0.040$ GeV;
- the Gambino, Giordano, Ossola and Uraltsev (GGOU) model [9] with input values in the kinetic scheme [58]: $m_b^{\text{kin}} = 4.573 \pm 0.012$ GeV and $\mu_\pi^{2\text{kin}} = 0.454 \pm 0.043$ GeV 2 .

The values of $|V_{ub}|$ obtained from each model and each signal extraction fit using Equation 2 are summarized in Table IX. Theoretical uncertainties, *e.g.*, those due to the assumed value of the b quark mass, impact both the determination of the branching fraction and the calculation of $|V_{ub}|$. However, the correlations between these quantities are difficult to determine due to the different parameterizations of non-perturbative effects used in the different theoretical calculations of $|V_{ub}|$ and in the DFN model used in our simulation. Therefore, similarly to what is done in the HFLAV averaging calculation, correlations between the branching fraction and partial decay rate theoretical uncertainties are ignored when extracting $|V_{ub}|$.

VII. CONCLUSIONS

A measurement of the partial branching fraction of inclusive charmless semileptonic decays in three phase-space regions is reported. The three regions are defined

⁵ The uncertainty on the value quoted by HFLAV is inflated based on theoretical errors. The branching fractions we quote are simply scaled by the phase-space acceptances given in Table II. They lack theoretical uncertainties related to this procedure and are therefore simply quoted here as a point of comparison.

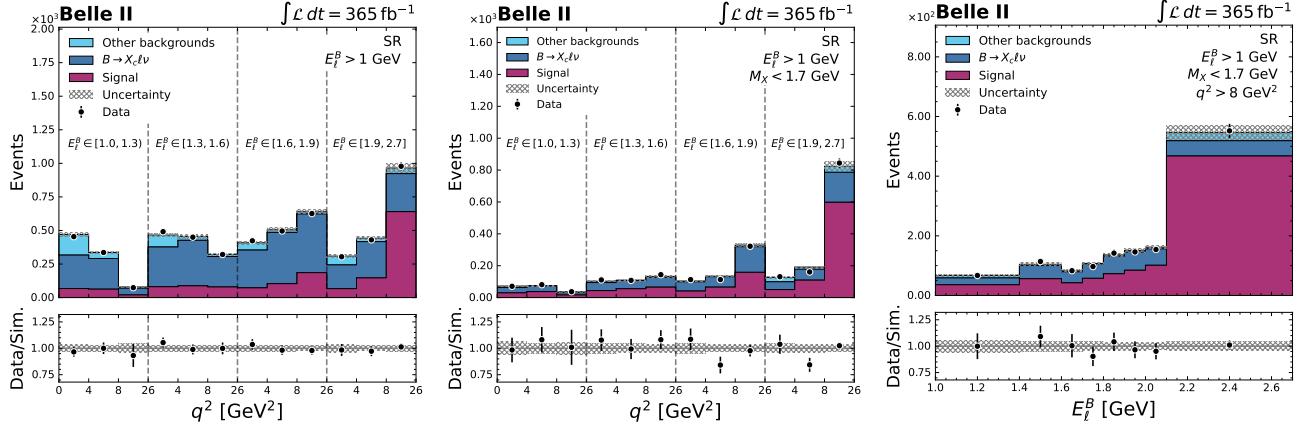


FIG. 8. Postfit signal region distributions for fit 1 (left), 2 (middle) and 3 (right). The $E_\ell^B:q^2$ variable is flattened in bins of E_ℓ^B (in GeV) and q^2 . The control region postfit distributions are not shown as the data/simulation agreement is almost perfect. The signal-in and out components are merged in the *Signal* template. The bottom panel shows the ratio between data and simulation yields in each bin.

by selections on the kinematic properties of $B \rightarrow X_u \ell \nu$ decays and cover, respectively, 87%, 57% and 31% of the full phase space. These kinematic selections, in addition to the use of neural networks, help to suppress various sources of background and in particular the contribution of CKM-favored $B \rightarrow X_c \ell \nu$ decays. The partner B meson from the $\Upsilon(4S)$ decay is reconstructed to constrain the kinematics of the signal-side decays and precisely determine several variables which play a key role in the measurement such as q^2 and the variables used to suppress the $B \rightarrow X_c \ell \nu$ background. The reconstruction and selection strategy developed for this analysis leads to a relatively higher signal efficiency compared to previous similar measurements such as Ref. [31]. The broadest phase-space region, where inclusive $B \rightarrow X_u \ell \nu$ theoretical predictions are most reliable, is defined by a selection $E_\ell^B > 1$ GeV. The partial branching fraction in this region is measured from a fit to the two-dimensional variable $E_\ell^B:q^2$ and the obtained value is

$$\Delta\mathcal{B}(B \rightarrow X_u \ell \nu) = (1.54 \pm 0.08 \pm 0.12) \times 10^{-3}, \quad (16)$$

where the uncertainties are statistical and systematic, respectively. From this branching fraction, the value of $|V_{ub}|$ obtained using a partial decay rate predicted by the GGOU framework is

$$|V_{ub}| = (4.01 \pm 0.11 \pm 0.16^{+0.07}_{-0.08}) \times 10^{-3}, \quad (17)$$

where the uncertainties are statistical, systematic and theoretical, respectively. The GGOU framework is used to allow a consistent comparison with the inclusive $|V_{ub}|$ average reported by HFLAV, with which our result is compatible [5]. The measured value also agrees within uncertainties with the average value obtained from measurements using $B \rightarrow \pi \ell \nu$ decays but it exceeds the HFLAV exclusive average. The values of $|V_{ub}|$ extracted from the three phase-space regions are compared to the

latest inclusive and exclusive HFLAV averages in Figure 9.

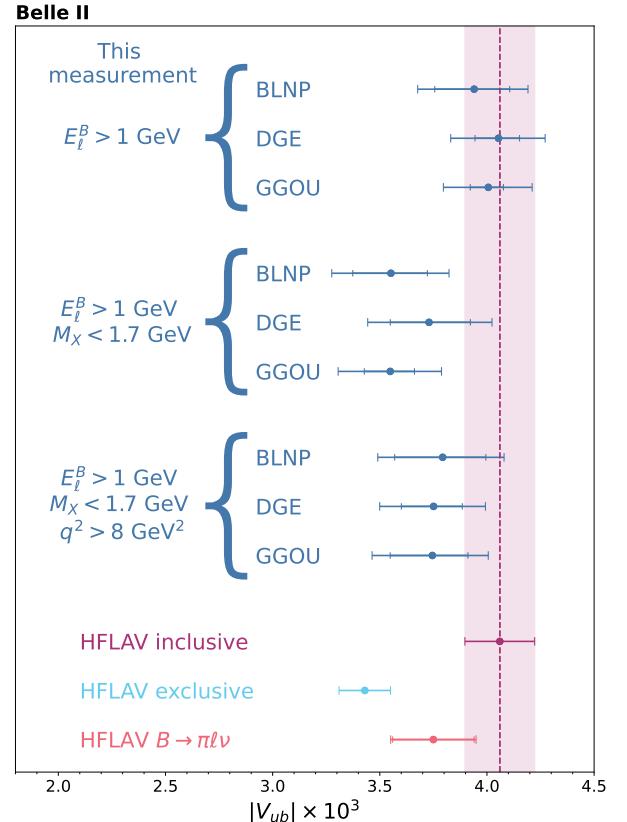


FIG. 9. Comparison between the three values of $|V_{ub}|$ obtained from fits 1, 2 and 3 (blue), the inclusive (purple band), exclusive (cyan) and $B \rightarrow \pi \ell \nu$ (pink) averages quoted in the latest HFLAV report [5]. The outer error bars represent the total uncertainty and, where shown, the inner error bars represent the contribution of the theoretical uncertainty.

TABLE VII. Breakdown of the systematic uncertainties of the partial branching fraction obtained from each fit.

| Uncertainty source | Relative uncertainty (%) | | |
|--|--------------------------|-------|-------|
| | Fit 1 | Fit 2 | Fit 3 |
| DFN parameters | 4.4 | 4.5 | 5.7 |
| DFN \rightarrow BLNP | 0.2 | 0.8 | 1.3 |
| γs | 1.7 | 2.1 | 2.1 |
| $B \rightarrow \pi \ell \nu$ form factors | 0.3 | 0.3 | 0.3 |
| $B \rightarrow \rho \ell \nu$ form factors | 0.3 | 0.3 | 0.2 |
| $B \rightarrow \omega \ell \nu$ form factors | 0.1 | 0.1 | 0.1 |
| $B \rightarrow \eta/\eta' \ell \nu$ form factors | < 0.1 | < 0.1 | < 0.1 |
| $B^\pm \rightarrow X_u \ell \nu$ branching fractions | 0.9 | 0.6 | 0.5 |
| $B^0 \rightarrow X_u \ell \nu$ branching fractions | 0.6 | 0.5 | 0.5 |
| $B \rightarrow D_{\text{Broad}}^{**}$ form factors | 0.5 | 0.1 | 0.2 |
| $B \rightarrow D_{\text{Narrow}}^{**}$ form factors | 0.1 | < 0.1 | < 0.1 |
| $B \rightarrow D/D^* \ell \nu$ form factors | < 0.1 | < 0.1 | < 0.1 |
| $B^\pm \rightarrow X_c \ell \nu$ branching fractions | 0.7 | 0.5 | 0.2 |
| $B^0 \rightarrow X_c \ell \nu$ branching fractions | 0.6 | 0.2 | 0.1 |
| D decay branching fractions | 0.1 | 0.3 | 0.1 |
| SR $X_c \ell \nu$ normalization | 1.6 | 3.5 | 3.4 |
| CR $X_c \ell \nu$ normalization | 0.9 | 1.1 | 0.4 |
| <i>Other backgrounds</i> normalization | 0.3 | N/A | N/A |
| X_u fragmentation | 0.3 | 4.4 | 3.9 |
| $N_{\Upsilon(4S)}$ | 1.4 | 1.4 | 1.4 |
| FEI | 1.3 | 1.3 | 1.4 |
| Slow pion efficiency | 0.4 | 0.2 | 0.3 |
| ℓ identification | 0.7 | 0.7 | 0.6 |
| $f^{\pm/00}$ | 0.6 | 0.7 | 0.6 |
| Continuum calibration | 0.2 | 0.2 | 0.2 |
| Tracking | 0.3 | 0.3 | 0.3 |
| K_S^0 efficiency | 0.1 | 0.1 | < 0.1 |
| K^\pm ID | < 0.1 | < 0.1 | < 0.1 |
| Simulated data statistics | 1.1 | 1.1 | 0.8 |

ACKNOWLEDGMENTS

The authors would like to thank Keri Vos for the useful discussions related to the theoretical description of semileptonic B decays. This work, based on data collected using the Belle II detector, which was built and commissioned prior to March 2019, was supported by Higher Education and Science Committee of the Republic of Armenia Grant No. 23LCG-1C011; Australian Research Council and Research Grants No. DP200101792, No. DP210101900, No. DP210102831, No. DE220100462, No. LE210100098, and No. LE230100085; Aus-

trian Federal Ministry of Education, Science and Research, Austrian Science Fund (FWF) Grants DOI: 10.55776/P34529, DOI: 10.55776/J4731, DOI: 10.55776/J4625, DOI: 10.55776/M3153, and DOI: 10.55776/PAT1836324, and Horizon 2020 ERC Starting Grant No. 947006 “InterLeptons”; Natural Sciences and Engineering Research Council of Canada, Digital Research Alliance of Canada, and Canada Foundation for Innovation; National Key R&D Program of China under Contract No. 2024YFA1610503, and No. 2024YFA1610504 National Natural Science Foundation of China and Research Grants No. 11575017,

TABLE VIII. Breakdown of uncertainties added following the bias corrections described in Sections V F and V G. In addition, we show the systematic, statistical and total uncertainties of the corrected branching fractions.

| | Relative uncertainty (%) | | |
|--|--------------------------|-------|-------|
| | Fit 1 | Fit 2 | Fit 3 |
| Fit bias | 2.6 | 2.3 | 1.8 |
| Composition uncertainty | 1.3 | 5.7 | 0.2 |
| $B \rightarrow X_c \ell \nu$ overestimation correction | 2.6 | 0.4 | 1.3 |
| Total systematic | 7.9 | 10.5 | 9.7 |
| Statistical | 5.4 | 5.6 | 4.5 |
| Total | 9.6 | 11.9 | 10.7 |

TABLE IX. Theoretical $B \rightarrow X_u \ell \nu$ decay rates $\Delta\tilde{\Gamma}$ (excluding the $|V_{ub}|^2$ term) and $|V_{ub}|$ obtained from three different theoretical predictions in three separate phase-space regions. The values of $\Delta\tilde{\Gamma}$ are used to compute $|V_{ub}|$ following Equation 2. The uncertainties on $|V_{ub}|$ are statistical, systematic and theoretical, respectively.

| Inclusive $B \rightarrow X_u \ell \nu$ model | Fit 1 | Fit 2 | Fit 3 | |
|--|-------|--|--|--|
| $\Delta\tilde{\Gamma}$ (ps $^{-1}$) | BLNP | $62.9^{+5.3}_{-5.9}$ | $47.5^{+4.5}_{-4.8}$ | $24.4^{+2.6}_{-2.9}$ |
| | DGE | $59.4^{+2.9}_{-3.2}$ | $43.1^{+4.5}_{-4.2}$ | $25.0^{+1.8}_{-2.0}$ |
| | GGOU | $60.8^{+2.2}_{-2.6}$ | $47.6^{+3.0}_{-3.2}$ | $25.0^{+2.2}_{-2.6}$ |
| $ V_{ub} \times 10^3$ | BLNP | $3.94 \pm 0.11 \pm 0.16^{+0.17}_{-0.18}$ | $3.55 \pm 0.10 \pm 0.19^{+0.17}_{-0.18}$ | $3.79 \pm 0.09 \pm 0.18^{+0.20}_{-0.22}$ |
| | DGE | $4.05 \pm 0.11 \pm 0.16^{+0.10}_{-0.11}$ | $3.73 \pm 0.10 \pm 0.20^{+0.19}_{-0.18}$ | $3.75 \pm 0.08 \pm 0.18^{+0.13}_{-0.15}$ |
| | GGOU | $4.01 \pm 0.11 \pm 0.16^{+0.07}_{-0.08}$ | $3.55 \pm 0.10 \pm 0.19^{+0.11}_{-0.12}$ | $3.75 \pm 0.08 \pm 0.18^{+0.17}_{-0.20}$ |

No. 11761141009, No. 11705209, No. 11975076, No. 12135005, No. 12150004, No. 12161141008, No. 12405099, No. 12475093, and No. 12175041, and Shandong Provincial Natural Science Foundation Project ZR2022JQ02; the Czech Science Foundation Grant No. 22-18469S, Regional funds of EU/MEYS: OPJAK FORTE CZ.02.01.01/00/22_008/0004632 and Charles University Grant Agency project No. 246122; European Research Council, Seventh Framework PIEF-GA-2013-622527, Horizon 2020 ERC-Advanced Grants No. 267104 and No. 884719, Horizon 2020 ERC-Consolidator Grant No. 819127, Horizon 2020 Marie Skłodowska-Curie Grant Agreement No. 700525 “NIOBE” and No. 101026516, and Horizon 2020 Marie Skłodowska-Curie RISE project JENNIFER2 Grant Agreement No. 822070 (European grants); L’Institut National de Physique Nucléaire et de Physique des Particules (IN2P3) du CNRS and L’Agence Nationale de la Recherche (ANR) under Grant No. ANR-23-CE31-0018 (France); BMFTR, DFG, HGF, MPG, and AvH Foundation (Germany); Department of Atomic Energy

under Project Identification No. RTI 4002, Department of Science and Technology, and UPES SEED funding programs No. UPES/R&D-SEED-INFRA/17052023/01 and No. UPES/R&D-SOE/20062022/06 (India); Israel Science Foundation Grant No. 2476/17, U.S.-Israel Binational Science Foundation Grant No. 2016113, and Israel Ministry of Science Grant No. 3-16543; Istituto Nazionale di Fisica Nucleare and the Research Grants BELLE2, and the ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU; Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research Grants No. 16H03968, No. 16H03993, No. 16H06492, No. 16K05323, No. 17H01133, No. 17H05405, No. 18K03621, No. 18H03710, No. 18H05226, No. 19H00682, No. 20H05850, No. 20H05858, No. 22H00144, No. 22K14056, No. 22K21347, No. 23H05433, No. 26220706, and No. 26400255, and the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan; National Research Foun-

dation (NRF) of Korea Grants No. 2021R1-F1A-1064008, No. 2022R1-A2C-1003993, No. 2022R1-A2C-1092335, No. RS-2016-NR017151, No. RS-2018-NR031074, No. RS-2021-NR060129, No. RS-2023-00208693, No. RS-2024-00354342 and No. RS-2025-02219521, Radiation Science Research Institute, Foreign Large-Size Research Facility Application Supporting project, the Global Science Experimental Data Hub Center, the Korea Institute of Science and Technology Information (K25L2M2C3) and KREONET/GLORIAD; Universiti Malaya RU grant, Akademi Sains Malaysia, and Ministry of Education Malaysia; Frontiers of Science Program Contracts No. FOINS-296, No. CB-221329, No. CB-236394, No. CB-254409, and No. CB-180023, and SEP-CINVESTAV Research Grant No. 237 (Mexico); the Polish Ministry of Science and Higher Education and the National Science Center; the Ministry of Science and Higher Education of the Russian Federation and the HSE University Basic Research Program, Moscow; University of Tabuk Research Grants No. S-0256-1438 and No. S-0280-1439 (Saudi Arabia), and Researchers Supporting Project number (RSPD2025R873), King Saud University, Riyadh, Saudi Arabia; Slovenian Research Agency and Research Grants No. J1-50010 and No. P1-0135; Ikerbasque, Basque Foundation for Science, State Agency for Research of the Spanish Ministry of Science and Innovation through Grant No. PID2022-136510NB-C33, Spain, Agencia Estatal de Investigacion, Spain Grant No. RYC2020-029875-I and Generalitat Valenciana, Spain Grant No. CIDEVENT/2018/020; The Knut and Alice Wallenberg Foundation (Sweden), Contracts No. 2021.0174, No. 2021.0299, and No. 2023.0315; National Science and Technology Council, and Ministry of Education (Taiwan); Thailand Center of Excellence in Physics; TUBITAK ULAKBIM (Turkey); National Research Foundation of Ukraine, Project No. 2020.02/0257, and Ministry of Education and Science of Ukraine; the U.S. National Science Foundation and Research Grants No. PHY-1913789 and No. PHY-2111604, and the U.S. Department of Energy and Research Awards No. DE-AC06-76RLO1830, No. DE-SC0007983, No. DE-SC0009824, No. DE-SC0009973, No. DE-SC0010007, No. DE-SC0010073, No. DE-SC0010118, No. DE-SC0010504, No. DE-SC0011784, No. DE-SC0012704, No. DE-SC0019230, No. DE-SC0021274, No. DE-SC0021616, No. DE-SC0022350, No. DE-SC0023470; and the Vietnam Academy of Science and Technology (VAST) under Grants No. NVCC.05.02/25-25 and No. DL0000.05/26-27.

These acknowledgements are not to be interpreted as an endorsement of any statement made by any of our institutes, funding agencies, governments, or their representatives.

We thank the SuperKEKB team for delivering high-luminosity collisions; the KEK cryogenics group for the efficient operation of the detector solenoid magnet and IBBelle on site; the KEK Computer Research Center for on-site computing support; the NII for SINET6 net-

work support; and the raw-data centers hosted by BNL, DESY, GridKa, IN2P3, INFN, and the University of Victoria.

Appendix A: $B \rightarrow X_c \ell \nu$ suppression classifier input features

The shapes of the nine variables used to train the $B \rightarrow X_c \ell \nu$ suppression classifier (see Section III E) are compared for $B \rightarrow X_u \ell \nu$ and $B \rightarrow X_c \ell \nu$ events in Figure 10. The performance of the classifier appears to be mostly driven by the event missing mass, the quality of ROE vertex fit and the missing mass in $D^* \rightarrow D\pi^0$ decays.

Appendix B: $B \rightarrow X_c \ell \nu$ composition

The composition of the $B \rightarrow X_c \ell \nu$ component is given in Table X for three different sets of kinematic selections and for the four control and signal regions defined in the main text. The compositions are extracted from simulated data.

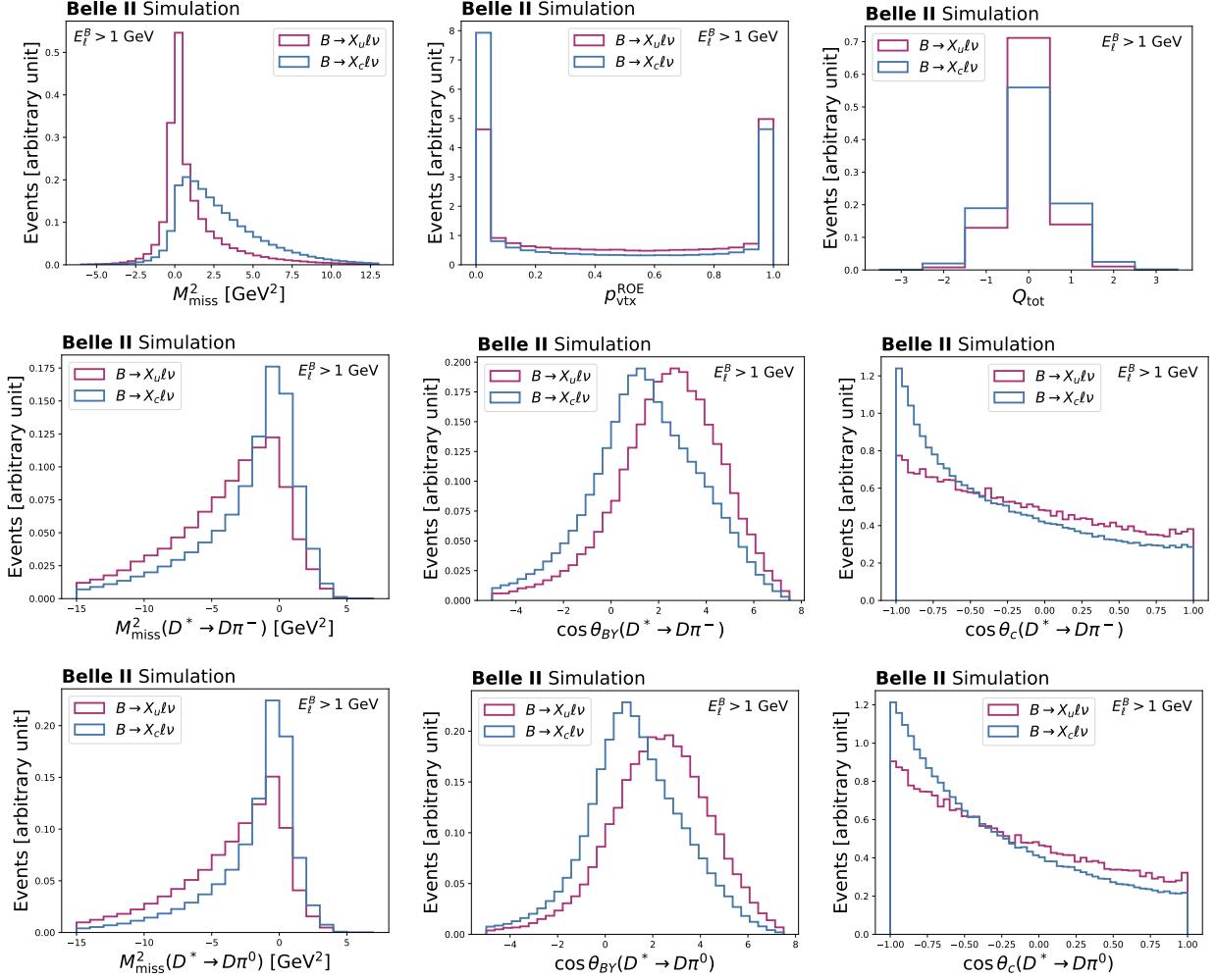


FIG. 10. The shapes of the nine $B \rightarrow X_c \ell \nu$ suppression classifier input features are compared for $B \rightarrow X_u \ell \nu$ (purple) and $B \rightarrow X_c \ell \nu$ (blue) events in the preselected region on which the continuum suppression and kaon veto have been applied. The default values chosen for $p_{\text{vtx}}^{\text{ROE}}$ when the vertex fit fails and for the six $D^* \rightarrow D\pi$ variables when no low momentum π is found in the event are not shown. These values are $-1, -20 \text{ GeV}^2, -10$ and 1 for $p_{\text{vtx}}^{\text{ROE}}, M_{\text{miss}}^2, \cos \theta_{BY}$ and $\cos \theta_c$, respectively.

TABLE X. $B \rightarrow X_c \ell \nu$ composition in terms of the four types of subcomponents in each control and signal region for the three sets of kinematic selections.

| | CR _{K,low} | CR _{K,high} | CR _{0,low} | SR |
|--|---------------------|----------------------|---------------------|-------|
| $E_\ell^B > 1.0 \text{ GeV}$ | | | | |
| $B \rightarrow D \ell \nu$ | 17.3% | 40.4% | 19.1% | 37.3% |
| $B \rightarrow D^* \ell \nu$ | 61.2% | 47.3% | 58.7% | 49.2% |
| $B \rightarrow D^{**} \ell \nu$ | 9.7% | 6.2% | 10.0% | 6.9% |
| $B \rightarrow D_{\text{Gap}}^{**} \ell \nu$ | 11.8% | 6.1% | 12.2% | 6.6% |
| $E_\ell^B > 1.0 \text{ GeV}, M_X < 1.7 \text{ GeV}$ | | | | |
| $B \rightarrow D \ell \nu$ | 22.0% | 64.3% | 21.5% | 47.7% |
| $B \rightarrow D^* \ell \nu$ | 65.3% | 33.6% | 61.9% | 47.3% |
| $B \rightarrow D^{**} \ell \nu$ | 6.4% | 1.5% | 8.0% | 2.7% |
| $B \rightarrow D_{\text{Gap}}^{**} \ell \nu$ | 6.3% | 0.6% | 8.6% | 2.3% |
| $E_\ell^B > 1.0 \text{ GeV}, M_X < 1.7 \text{ GeV}, q^2 > 8 \text{ GeV}^2$ | | | | |
| $B \rightarrow D \ell \nu$ | 14.6% | 35.9% | 16.5% | 32.1% |
| $B \rightarrow D^* \ell \nu$ | 79.9% | 61.4% | 75.2% | 63.2% |
| $B \rightarrow D^{**} \ell \nu$ | 3.5% | 2.0% | 5.1% | 3.2% |
| $B \rightarrow D_{\text{Gap}}^{**} \ell \nu$ | 1.9% | 0.8% | 3.3% | 1.5% |

[1] N. Cabibbo, Unitary Symmetry and Leptonic Decays, *Phys. Rev. Lett.* **10**, 531 (1963).

[2] M. Kobayashi and T. Maskawa, CP Violation in the Renormalizable Theory of Weak Interaction, *Prog. Theor. Phys.* **49**, 652 (1973).

[3] R. Aaij *et al.*, Measurement of the branching fraction of the $B^0 \rightarrow D_s^+ \pi^-$ decay, *Eur. Phys. J. C* **81**, 314 (2021).

[4] L. Aggarwal *et al.*, Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond (2022), arXiv:2207.06307 [hep-ex].

[5] S. Banerjee *et al.* (Heavy Flavor Averaging Group), Averages of b -hadron, c -hadron, and τ -lepton properties as of 2023 (2024), arXiv:2411.18639 [hep-ex].

[6] T. Mannel, Operator product expansion for inclusive semileptonic decays in heavy quark effective field theory, *Nuclear Physics B* **413**, 396 (1994).

[7] J. Chay, H. Georgi, and B. Grinstein, Lepton energy distributions in heavy meson decays from QCD, *Physics Letters B* **247**, 399 (1990).

[8] B. Blok, M. Shifman, and D.-X. Zhang, Illustrative example of how quark-hadron duality might work, *Phys. Rev. D* **57**, 2691 (1998).

[9] P. Gambino, P. Giordano, G. Ossola, and N. Uraltsev, Inclusive semileptonic B decays and the determination of $|V_{ub}|$, *J. High Energy Phys.* **10** (2007), 058.

[10] B. O. Lange, M. Neubert, and G. Paz, Theory of charmless inclusive B decays and the extraction of V_{ub} , *Phys. Rev. D* **72**, 073006 (2005).

[11] J. R. Andersen and E. Gardi, Inclusive spectra in charmless semileptonic B decays by dressed gluon exponentiation, *J. High Energy Phys.* **01** (2006), 097.

[12] K. Akai, K. Furukawa, and H. Koiso, SuperKEKB Collider, *Nucl. Instrum. Meth. A* **907**, 188 (2018).

[13] S. Agostinelli *et al.*, Geant4—a simulation toolkit, *Nucl. Instrum. Meth. A* **506**, 250 (2003).

[14] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun (Belle-II Framework Software Group), The Belle II Core Software, *Comput. Softw. Big Sci.* **3**, 1 (2018).

[15] S. Navas *et al.* (Particle Data Group), Review of particle physics, *Phys. Rev. D* **110**, 030001 (2024), updated results and plots available at: <https://pdglive.lbl.gov/Viewer.action>.

[16] D. J. Lange, The EvtGen particle decay simulation package, *Nucl. Instrum. Meth. A* **462**, 152 (2001), *Beauty 2000*, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[17] C. Bierlich *et al.*, A Comprehensive Guide to the Physics and Usage of PYTHIA 8.3, *SciPost Phys. Codebases* **8**, 10.21468/SciPostPhysCodeb.8 (2022).

[18] C. Ramirez, J. F. Donoghue, and G. Burdman, Semileptonic $b \rightarrow u$ decay, *Phys. Rev. D* **41**, 1496 (1990).

[19] F. De Fazio and M. Neubert, $B \rightarrow X_u \ell \bar{\nu}_\ell$ decay distributions to order α_s , *J. High Energy Phys.* **06** (1999), 017.

[20] A. L. Kagan and M. Neubert, QCD anatomy of $B \rightarrow X_s \gamma$ decays, *Eur. Phys. J. C* **7**, 5 (1999).

[21] O. L. Buchmüller and H. U. Flächer, Fit to moments of inclusive $B \rightarrow X_c \ell \bar{\nu}$ and $B \rightarrow X_s \gamma$ decay distributions using heavy quark expansions in the kinetic scheme, *Phys. Rev. D* **73**, 073008 (2006).

[22] C. Bourrely, I. Caprini, and L. Lellouch, Model-independent description of $B \rightarrow \pi \ell \nu$ decays and a determination of $|V_{ub}|$, *Phys. Rev. D* **79**, 013008 (2009).

[23] Y. Aoki *et al.* (Flavour Lattice Averaging Group), FLAG Review 2021, *Eur. Phys. J. C* **82**, 869 (2022).

[24] A. Bharucha, D. M. Straub, and R. Zwicky, $B \rightarrow V \ell^+ \ell^-$ in the Standard Model from light-cone sum rules, *J. High Energy Phys.* **8** (2016), 98.

[25] F. U. Bernlochner, M. T. Prim, and D. J. Robinson, $B \rightarrow \rho \ell \bar{\nu}$ and $\omega \ell \bar{\nu}$ in and beyond the Standard Model: Improved predictions and $|V_{ub}|$, *Phys. Rev. D* **104**, 034032 (2021).

[26] D. Scora and N. Isgur, Semileptonic meson decays in the quark model: An update, *Phys. Rev. D* **52**, 2783 (1995).

[27] G. Duplancic and B. Melic, Form factors of $B, B_s \rightarrow \eta^{(\prime)}$ and $D, D_s \rightarrow \eta^{(\prime)}$ transitions from QCD light-cone sum rules, *J. High Energy Phys.* **11** (2015), 138.

[28] F. U. Bernlochner, Z. Ligeti, M. Papucci, M. T. Prim, D. J. Robinson, and C. Xiong, Constrained second-order power corrections in HQET: $R(D^{(*)})$, $|V_{cb}|$, and new physics, *Phys. Rev. D* **106**, 096015 (2022).

[29] F. U. Bernlochner, Z. Ligeti, and S. Turczyk, A proposal to solve some puzzles in semileptonic B decays, *Phys. Rev. D* **85**, 094033 (2012).

[30] F. U. Bernlochner, Z. Ligeti, and D. J. Robinson, Model independent analysis of semileptonic B decays to D^{**} for arbitrary new physics, *Phys. Rev. D* **97**, 075011 (2018).

[31] L. Cao *et al.* (Belle Collaboration), Measurements of Partial Branching Fractions of Inclusive $B \rightarrow X_u \ell^+ \nu_\ell$ Decays with Hadronic Tagging, *Phys. Rev. D* **104**, 012008 (2021).

[32] S. Jadach, B. Ward, Z. Was, S. Yost, and A. Siodmok, Multi-photon Monte Carlo event generator KKMCEe for lepton and quark pair production in lepton colliders, *Computer Physics Communications* **283**, 108556 (2023).

[33] T. Keck *et al.*, The full event interpretation, Computing and Software for Big Science **3**, 6 (2019).

[34] M. Milesi, J. Tan, and P. Urquijo, Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers, *EPJ Web Conf.* **245**, 06023 (2020).

[35] P. Cheema, Suppressing Beam Background and Fake Photons at Belle II using Machine Learning, *EPJ Web Conf.* **295**, 09035 (2024).

[36] I. Adachi *et al.* (Belle II Collaboration), Charged-hadron identification at Belle II, *Eur. Phys. J. C* **85**, 1237 (2025).

[37] G. C. Fox and S. Wolfram, Observables for the Analysis of Event Shapes in $e^+ e^-$ Annihilation and Other Processes, *Phys. Rev. Lett.* **41**, 1581 (1978).

[38] S. Lee *et al.* (Belle Collaboration), Evidence for $B^0 \rightarrow \pi^0 \pi^0$, *Phys. Rev. Lett.* **91**, 261801 (2003).

[39] D. Asner *et al.* (CLEO Collaboration), Search for exclusive charmless hadronic B decays, *Phys. Rev. D* **53**, 1039 (1996).

[40] P. Ramachandran, B. Zoph, and Q. V. Le, Searching for Activation Functions (2017), arXiv:1710.05941.

[41] I. Loshchilov and F. Hutter, Decoupled Weight Decay Regularization, in *7th Int. Conf. Learn. Represent. ICLR 2019 New Orleans USA May 6-9 2019* (2019).

[42] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, Optuna: A Next-generation Hyperparameter Optimiza-

tion Framework (2019), arXiv:1907.10902.

[43] D. Martschei, M. Feindt, S. Honc, and J. Wagner-Kuhr, Advanced Event Reweighting Using Multivariate Analysis, *J. Phys. Conf. Ser.* **368**, 012028 (2012).

[44] I. Adachi *et al.* (Belle II Collaboration), Evidence for $B^+ \rightarrow K^+ \nu \bar{\nu}$ decays, *Phys. Rev. D* **109**, 112006 (2024).

[45] L. Heinrich, M. Feickert, G. Stark, and K. Cranmer, pyhf: pure-Python implementation of HistFactory statistical models, *Journal of Open Source Software* **6**, 2823 (2021).

[46] eFFORT, <https://github.com/b2-hive/eFFORT> (2022).

[47] S. Bosch, B. Lange, M. Neubert, and G. Paz, Factorization and shape-function effects in inclusive B -meson decays, *Nuclear Physics B* **699**, 335 (2004).

[48] M. Althoff *et al.* (TASSO), A Detailed Study of Strange Particle Production in e^+e^- Annihilation at High-energy, *Z. Phys. C* **27**, 27 (1985).

[49] W. Bartel *et al.* (JADE Collaboration), Charged particle and neutral kaon production in e^+e^- annihilation at PETRA, *Z. Phys. C* **20**, 187 (1983).

[50] B. Aubert *et al.* (BABAR Collaboration), Measurement of the Branching Fractions of $\bar{B} \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ Decays in Events Tagged by a Fully Reconstructed B Meson, *Phys. Rev. Lett.* **101**, 261802 (2008).

[51] F. Meier *et al.* (Belle Collaboration), First observation of $B \rightarrow \bar{D}_1(\rightarrow \bar{D} \pi^+ \pi^-) \ell^+ \nu_\ell$ and measurement of the $B \rightarrow \bar{D}^(*) \pi \ell^+ \nu_\ell$ and $B \rightarrow \bar{D}^(*) \pi^+ \pi^- \ell^+ \nu_\ell$ branching fractions with hadronic tagging at Belle, *Phys. Rev. D* **107**, 092003 (2023).

[52] F. U. Bernlochner, S. Duell, Z. Ligeti, M. Papucci, and D. J. Robinson, Das ist der HAMMER: Consistent new physics interpretations of semileptonic decays, *Eur. Phys. J. C* **80**, 883 (2020).

[53] M. Feindt, F. Keller, M. Kreps, T. Kuhr, S. Neubauer, D. Zander, and A. Zupanc, A hierarchical neurobayes-based algorithm for full reconstruction of B mesons at B factories, *Nucl. Instrum. Meth. A* **654**, 432 (2011).

[54] J. Lees *et al.* (BaBar Collaboration), Measurement of the inclusive electron spectrum from B meson decays and determination of $|V_{ub}|$, *Phys. Rev. D* **95**, 072001 (2017).

[55] J. Lees *et al.* (BaBar Collaboration), Study of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays in $B\bar{B}$ events tagged by a fully reconstructed B -meson decay and determination of $|V_{ub}|$, *Phys. Rev. D* **86**, 032004 (2012).

[56] M. Neubert, Two-loop relations for heavy-quark parameters in the shape-function scheme, *Physics Letters B* **612**, 13 (2005).

[57] M. Neubert, Advanced predictions for moments of the $\bar{B} \rightarrow X_s \gamma$ photon spectrum, *Phys. Rev. D* **72**, 074025 (2005).

[58] I. Bigi, M. Shifman, N. Uraltsev, and A. Vainshtein, High power n of m_b in b -flavored widths and $n = 5 \rightarrow \infty$ limit, *Phys. Rev. D* **56**, 4017 (1997).