First observation of $B \to \bar{D}_1(\to \bar{D}\pi^+\pi^-)\ell^+\nu_\ell$ and measurement of the $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ and $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ branching fractions with hadronic tagging at Belle

F. Meier, A. Vossen, I. Adachi, K. Adamczyk, H. Aihara, S. Al Said, D. M. Asner, H. Atmacan, T. Aushev, R. Ayad, V. Babu, S. Bahinipati, Sw. Banerjee, M. Bauer, P. Behera, K. Belous, J. Bennett, F. Bernlochner, M. Bessner, B. Bhuyan, T. Bilka, D. Biswas, A. Bobrov, D. Bodrov, G. Bonvicini, J. Borah, A. Bozek, M. Bračko, P. Branchini, T. E. Browder, A. Budano, M. Campajola, L. Cao, D. Červenkov, M.-C. Chang, P. Chang, A. Chen, B. G. Cheon, K. Chilikin, K. Cho, S.-J. Cho, S.-K. Choi, Y. Choi, S. Choudhury, D. Cinabro, S. Das, G. De Nardo, G. De Pietro, R. Dhamija, F. Di Capua, J. Dingfelder, Z. Doležal, T. V. Dong, D. Epifanov, T. Ferber, D. Ferlewicz, B. G. Fulsom, R. Garg, V. Gaur, A. Giri, P. Goldenzweig, B. Golob, E. Graziani, K. Gudkova, C. Hadjivasiliou, S. Halder, T. Hara, K. Hayasaka, H. Hayashii, M. T. Hedges, W.-S. Hou, C.-L. Hsu, K. Inami, N. Ipsita, A. Ishikawa, R. Itoh, M. Iwasaki, W. W. Jacobs, E.-J. Jang, Y. Jin, A. B. Kaliyar, K. H. Kang, T. Kawasaki, C. Kiesling, C. H. Kim, D. Y. Kim, K.-H. Kim, Y.-K. Kim, K. Kinoshita, P. Kodyš, A. Korobov, S. Korpar, E. Kovalenko, P. Križan, P. Krokovny, T. Kuhr, M. Kumar, R. Kumar, K. Kumara, Y.-J. Kwon, T. Lam, J. S. Lange, S. C. Lee, P. Lewis, C. H. Li, L. K. Li, Y. Li, Y. B. Li, L. Li Gioi, J. Libby, K. Lieret, Y.-R. Lin, D. Liventsev, T. Luo, M. Masuda, T. Matsuda, D. Matvienko, S. K. Maurya, M. Merola, F. Metzner, K. Miyabayashi, R. Mizuk, R. Mussa, I. Nakamura, M. Nakao, Z. Natkaniec, A. Natochii, L. Nayak, M. Nayak, N. K. Nisar, S. Nishida, S. Ogawa, H. Ono, P. Oskin, P. Pakhlov, G. Pakhlova, S. Pardi, H. Park, J. Park, A. Passeri, S. Patra, S. Paul, R. Pestotnik, L. E. Piilonen, T. Podobnik, E. Prencipe, M. T. Prim, A. Rostomyan, N. Rout, G. Russo, S. Sandilya, L. Santelj, V. Savinov, G. Schnell, C. Schwanda, A. J. Schwartz, Y. Seino, K. Senyo, M. E. Sevior, M. Shapkin, C. Sharma, C. P. Shen, J.-G. Shiu, F. Simon, J. B. Singh, A. Soffer, E. Solovieva, M. Starič, Z. S. Stottler, J. F. Strube, M. Sumihama, T. Sumiyoshi, M. Takizawa, U. Tamponi, K. Tanida, F. Tenchini, K. Trabelsi, M. Uchida, T. Uglov, Y. Unno, K. Uno, S. Uno, P. Urquijo, S. E. Vahsen, R. van Tonder, G. Varner, K. E. Varvell, A. Vinokurova, M.-Z. Wang, M. Watanabe, S. Watanuki, J. Wiechczynski, E. Won, X. Xu, B. D. Yabsley, W. Yan, S. B. Yang, J. H. Yin, C. Z. Yuan, L. Yuan, Y. Yusa, Z. P. Zhang, and V. Zhukova

(Belle Collaboration)

(Received 8 December 2022; accepted 23 March 2023; published 8 May 2023)

We report measurements of the ratios of branching fractions for $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ and $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ relative to $B \to \bar{D}^*\ell^+\nu_\ell$ decays with $\ell = e,\ \mu$. These results are obtained from a data sample that contains $772 \times 10^6 B\bar{B}$ pairs collected near the $\Upsilon(4S)$ resonance with the Belle detector at the KEKB asymmetric energy e^+e^- collider. Fully reconstructing both B mesons in the event, we obtain $\frac{\mathcal{B}(B^0\to\bar{D}^0\pi^-\ell^+\nu_\ell)}{\mathcal{B}(B^0\to\bar{D}^{*-}\ell^+\nu_\ell)} = (7.23 \pm 0.36 \pm 0.14)\%$, $\frac{\mathcal{B}(B^+\to D^-\pi^+\ell^+\nu_\ell)}{\mathcal{B}(B^+\to\bar{D}^{*0}\ell^+\nu_\ell)} = (6.78 \pm 0.24 \pm 0.18)\%$, $\frac{\mathcal{B}(B^0\to\bar{D}^{*0}\pi^-\ell^+\nu_\ell)}{\mathcal{B}(B^0\to\bar{D}^{*-}\ell^+\nu_\ell)} = (11.10 \pm 0.48 \pm 0.23)\%$, $\frac{\mathcal{B}(B^+\to D^{*-}\pi^+\ell^+\nu_\ell)}{\mathcal{B}(B^+\to\bar{D}^{*0}\ell^+\nu_\ell)} = (9.50 \pm 0.33 \pm 0.34)\%$, $\frac{\mathcal{B}(B^0\to D^-\pi^+\pi^-\ell^+\nu_\ell)}{\mathcal{B}(B^0\to D^{*-}\ell^+\nu_\ell)} = (2.91 \pm 0.37 \pm 0.26)\%$, $\frac{\mathcal{B}(B^+\to\bar{D}^0\pi^+\pi^-\ell^+\nu_\ell)}{\mathcal{B}(B^+\to\bar{D}^{*0}\ell^+\nu_\ell)} = (3.10 \pm 0.26 \pm 0.22)\%$, $\frac{\mathcal{B}(B^0\to D^{*-}\pi^+\pi^-\ell^+\nu_\ell)}{\mathcal{B}(B^0\to D^{*-}\ell^+\nu_\ell)} = (0.99 \pm 0.43 \pm 0.20)\%$, $\frac{\mathcal{B}(B^+\to\bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell)}{\mathcal{B}(B^+\to\bar{D}^{*0}\ell^+\nu_\ell)} = (1.25 \pm 0.27 \pm 0.15)\%$, where the uncertainties are statistical and systematic, respectively. These are the most precise measurements of these branching fraction ratios to date. The invariant mass spectra of the $D\pi$, $D^*\pi$, and $D\pi\pi$ systems are studied, and the branching fraction

products $\mathcal{B}(B^0 \to D_2^{*-}\ell^+\nu_\ell) \times \mathcal{B}(D_2^{*-} \to \bar{D}^0\pi^-) = (0.157 \pm 0.015 \pm 0.005)\%$, $\mathcal{B}(B^+ \to \bar{D}_0^{*0}\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_0^{*0} \to D^-\pi^+) = (0.054 \pm 0.022 \pm 0.005)\%$, $\mathcal{B}(B^+ \to \bar{D}_2^{*0}\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_2^{*0} \to D^-\pi^+) = (0.163 \pm 0.011 \pm 0.008)\%$, $\mathcal{B}(B^0 \to D_1^-\ell^+\nu_\ell) \times \mathcal{B}(D_1^- \to \bar{D}^{*0}\pi^-) = (0.306 \pm 0.050 \pm 0.029)\%$, $\mathcal{B}(B^0 \to D_1'^-\ell^+\nu_\ell) \times \mathcal{B}(D_1'^- \to \bar{D}^{*0}\pi^-) = (0.206 \pm 0.068 \pm 0.025)\%$, $\mathcal{B}(B^0 \to D_2^{*-}\ell^+\nu_\ell) \times \mathcal{B}(D_2^{*-} \to \bar{D}^{*0}\pi^-) = (0.051 \pm 0.040 \pm 0.010)\%$, $\mathcal{B}(B^+ \to \bar{D}_1^0\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_1^0 \to D^{*-}\pi^+) = (0.249 \pm 0.023 \pm 0.015)\%$, $\mathcal{B}(B^+ \to \bar{D}_1'^0\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_1'^0 \to D^{*-}\pi^+) = (0.138 \pm 0.036 \pm 0.009)\%$, $\mathcal{B}(B^+ \to \bar{D}_2^{*0}\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_2^{*0} \to D^{*-}\pi^+) = (0.137 \pm 0.026 \pm 0.009)\%$, $\mathcal{B}(B^0 \to D_1^-\ell^+\nu_\ell) \times \mathcal{B}(D_1^- \to D^-\pi^+\pi^-) = (0.102 \pm 0.013 \pm 0.009)\%$, $\mathcal{B}(B^+ \to \bar{D}_1^0\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_1^0 \to \bar{D}^0\pi^+\pi^-) = (0.105 \pm 0.011 \pm 0.009)\%$, are extracted. This is the first observation of the decays $B \to \bar{D}_1\ell^+\nu_\ell$ with $D_1 \to D\pi^+\pi^-$.

DOI: 10.1103/PhysRevD.107.092003

I. INTRODUCTION

Semileptonic decays of B mesons are an important tool for precision measurements of the Cabibbo-Kobayashi-Maskawa matrix elements V_{cb} and V_{ub} [1,2]. The latest determinations of $|V_{cb}|$ from inclusive semileptonic $B \to X_c\ell^+\nu_\ell$ decays, with X_c being a charmed hadronic state that is not explicitly reconstructed, differ from those using the exclusive semileptonic decays $B \to \bar{D}\ell^+\nu_\ell$ and $B \to \bar{D}^*\ell^+\nu_\ell$ by about 2.4σ [3]. The measured sum of the exclusive $B \to \bar{D}^{(*)}\ell^+\nu_\ell$, $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$, and $B^+ \to D_s^{(*)-}K^+\ell^+\nu_\ell$ rates accounts for only $85 \pm 2\%$ [3] of the inclusive rate for semileptonic B decays to charm final states.

Semileptonic decays of B mesons can also be used for other precision tests of the electroweak sector of the standard model, such as lepton flavor universality. An example is the ratio $R(D^{(*)})$ of the branching fractions $\mathcal{B}(B \to \bar{D}^{(*)}\tau^+\bar{\nu}_\tau)$ and $\mathcal{B}(B \to \bar{D}^{(*)}\ell^+\nu_\ell)$ ($\ell = e, \mu$), for which a persistent 3σ deviation between the standard model expectation [4] and the combined experimental results [4] from *BABAR* [5,6], Belle [7–9], and LHCb [10,11] has been observed. Important backgrounds in these processes are the decays $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ and $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$. The former accounts for part of the missing exclusive rate described above. The latter proceeds predominantly via $B \to \bar{D}^{**}\ell^+\nu_\ell$, $D^{**} \to D^{(*)}\pi$, where the D^{**} is an orbitally excited ($L = 1$) charmed meson. The D^{**} mass-spectrum contains two doublets of states that have light-quark total angular momenta of $j_q = \frac{1}{2}$ and $j_q = \frac{3}{2}$ [12]. The spin-0 state D_0^* can only decay to $D\pi$ and the spin-1 states D_1 and D_1' only via $D^{**} \to D^*\pi$. The spin-2 state D_2^* can decay both into $D\pi$ and $D^*\pi$. The D^{**} masses are not far from threshold. Since the $j_q = \frac{3}{2}$ states (D_1 and D_2^*) have a significant D-wave component, these states are narrow and were observed with a typical width of about 20 MeV/c^2 [13–15]. On the other hand, the states with $j_q = \frac{1}{2}$ decay mainly via S-wave and are therefore expected to be broad resonances with a width of several hundred MeV/c^2 [12,16]. The decay rate of semileptonic B decays to the $j_q = \frac{1}{2}$ states is observed to be similar to the rate to the $j_q = \frac{3}{2}$ doublet, while model calculations predict a substantially smaller rate to the $j_q = \frac{1}{2}$ doublet [17].

The decay modes with one charged pion in the final state have been measured by *BABAR* [15] and in a previous Belle analysis [18]. For the $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ channel so far only a *BABAR* result [19] with limited statistical precision is available. The results of these three measurements are listed in Table I. The current measurement improves upon the aforementioned Belle result by using a new method to reconstruct ("tag") the other B in the event, known as the Full Event Interpretation (FEI)

TABLE I. Previous results of $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ and $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ branching fraction measurements by *BABAR* [15,19] and Belle [18]. The first uncertainty is statistical, the second systematic, and the third comes from the branching fraction of the normalization mode.

Decay mode	*BABAR* (%)	Belle (%)
$B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$	$(0.43 \pm 0.08 \pm 0.03)$	$(0.405 \pm 0.036 \pm 0.041)$
$B^+ \to D^-\pi^+\ell^+\nu_\ell$	$(0.42 \pm 0.06 \pm 0.03)$	$(0.455 \pm 0.027 \pm 0.039)$
$B^0 \to \bar{D}^{*0}\pi^-\ell^+\nu_\ell$	$(0.48 \pm 0.08 \pm 0.04)$	$(0.646 \pm 0.053 \pm 0.052)$
$B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$	$(0.59 \pm 0.05 \pm 0.04)$	$(0.603 \pm 0.043 \pm 0.038)$
$B^0 \to D^-\pi^+\pi^-\ell^+\nu_\ell$	$(0.127 \pm 0.039 \pm 0.026 \pm 0.007)$	⋯
$B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$	$(0.161 \pm 0.030 \pm 0.018 \pm 0.008)$	⋯
$B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$	$(0.138 \pm 0.039 \pm 0.030 \pm 0.003)$	⋯
$B^+ \to \bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell$	$(0.080 \pm 0.040 \pm 0.023 \pm 0.003)$	⋯

algorithm [20], and by providing a result for the $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ channel as well.

II. EXPERIMENTAL APPARATUS AND DATA

The Belle detector is a large-solid-angle magnetic spectrometer. Its innermost component is a silicon vertex detector (SVD). A 50-layer central drift chamber (CDC) provides tracking and charged particle identification (PID) information using specific ionization measurements. An array of aerogel threshold Cherenkov counters (ACC), a barrel-like arrangement of time-of-flight scintillation counters (TOF) in the central part, and an electromagnetic calorimeter (ECL) comprised of CsI (Tl) crystals provide further PID information. These detector components are located inside a super-conducting solenoid coil that provides a 1.5 T magnetic field. The iron return yoke located outside of the coil is instrumented to detect K_L^0 mesons and to identify muons (KLM). The detector's z-axis is defined to be anti-parallel to the e^+ beam. More details about the detector can be found in Ref. [21].

Electron candidates are identified using the ratio between the energy deposited in the ECL and their track momentum, the ECL shower shape, the matching between the track and the ECL cluster, the energy loss in the CDC, and the number of photoelectrons in the ACC [22]. Muons are identified based on their penetration range and transverse scattering in the KLM [23]. Charged kaons and pions are identified by a combination of the energy loss in the CDC, the Cherenkov light in the ACC, and the time of flight in the TOF.

A data sample corresponding to an integrated luminosity of $L_{\rm on} = 711$ fb^{-1}, collected with the Belle detector at the KEKB asymmetric-energy e^+e^- collider [24] operating at the $\Upsilon(4S)$ resonance at $\sqrt{s} = 10.58$ GeV, is used for the measurement. The sample contains 772×10^6 $B\bar{B}$ pairs. A further data sample corresponding to an integrated luminosity of $L_{\rm off} = 89$ fb^{-1} taken slightly below the resonance, at $\sqrt{s} = 10.52$ GeV, is used for background templates. These two samples are referred to as the on-resonance and off-resonance samples, respectively.

We use a sample of simulated $B\bar{B}$ background Monte Carlo (MC) events generated with EvtGen [25]. This sample has six times more events than the Belle collision data. The full detector simulation is based on GEANT3 [26]. Final-state radiation is simulated with the PHOTOS package [27]. The $B \to \bar{D}^*\ell^+\nu_\ell$ decays are simulated using the HQET2 model [28] of EvtGen. For the $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ decay modes, dedicated MC samples of 73×10^6 events for each of five transitions (via D_0^* and D_2^* for $B \to \bar{D}\pi\ell^+\nu_\ell$, and via D_1, D_1', and D_2^* for $B \to \bar{D}^*\pi\ell^+\nu_\ell$) are generated with the ISGW2 model [29]. A signal MC sample of 50×10^6 events of $B \to \bar{D}\pi^+\pi^-\ell^+\nu_\ell$ is used for simulating the $B \to \bar{D}_1\ell^+\nu_\ell$ decay modes. A MC sample of 25×10^6 events of $B \to \bar{D}^*\pi^+\pi^-\ell^+\nu_\ell$ is used for simulating the $B \to \bar{D}_1'\ell^+\nu_\ell$ decay modes. Both are simulated with the ISGW2 model.

Data-MC efficiency differences due to a variety of sources are corrected. A more detailed description can be found in Sec. VI.

III. MEASUREMENT OVERVIEW

The $B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$ branching ratio relative to $B^0 \to D^{*-}\ell^+\nu_\ell$ is measured,

$$\frac{\mathcal{B}(B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell)}{\mathcal{B}(B^0 \to D^{*-}\ell^+\nu_\ell)} = \frac{N_{\rm sig}(B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell)/\epsilon_{\rm sig}(B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell)}{N_{\rm sig}(B^0 \to D^{*-}\ell^+\nu_\ell)/\epsilon_{\rm sig}(B^0 \to D^{*-}\ell^+\nu_\ell)}, \tag{1}$$

which reduces systematic uncertainties due to data-MC differences and external branching fraction values of the charm modes as they largely cancel in the ratio. Similar expressions are used for the other measured decays. Here $N_{\rm sig}$ is the number of signal candidates and $\epsilon_{\rm sig}$ is the corresponding signal efficiency. Branching fractions of $B \to \bar{D}^{(*)}{\rm n}\pi\ell^+\nu_\ell$ (n = 1, 2) are also reported, after multiplying by the $\mathcal{B}(B^+ \to \bar{D}^{*0}\ell^+\nu_\ell) = 5.58 \pm 0.22\%$ and $\mathcal{B}(B^0 \to D^{*-}\ell^+\nu_\ell) = 4.97 \pm 0.12\%$ averages from the Particle Data Group (PDG) [3].

IV. EVENT SELECTION

The Belle data are converted into the Belle II format [30], and the particle and event reconstruction is performed within the BASF2 framework [31,32] of the Belle II experiment.

A. Common selection requirements

In each event two B meson candidates are reconstructed. One of the B meson candidates ($B_{\rm tag}$) is reconstructed with FEI. The FEI algorithm follows a hierarchical approach. Final-state particle candidates are combined to intermediate particles until the final $B_{\rm tag}$ candidates are formed. More than 100 explicit decay channels, leading to $\mathcal{O}(10\,000)$ distinct decay chains are reconstructed. For each final-state particle and for each decay channel of an intermediate particle, a multivariate classifier is trained which estimates the probability that each decay chain correctly describes the true process. In this analysis only hadronically reconstructed decay chains are considered. The $B_{\rm tag}$ meson candidates are required to have a beam-constrained mass $M_{\rm bc} = \sqrt{(E_{\rm c.m.}/c^2)^2 - (\vec{P}_{B_{\rm tag}}/c)^2} > 5.27$ GeV/c^2, and an

energy difference $\Delta E = E_{B_{\rm tag}} - E_{\rm c.m.}$ within ± 180 MeV. Here $E_{\rm c.m.}$ is half of the center-of-mass (c.m.) energy of the beams, and $\vec{P}_{B_{\rm tag}}$ and $E_{B_{\rm tag}}$ are the momentum and energy of the $B_{\rm tag}$ meson in the c.m. frame, respectively. The FEI signal probability of $B_{\rm tag}$ candidates is required to be greater than 0.5%. Distributions of $M_{\rm bc}$, ΔE, and the signal probability are shown in Fig. 1. We take into account that the composition of decay modes reconstructed by FEI differs between data and MC. The ratio between the relative abundance in each decay mode is used to correct this effect.

The other B meson candidate ($B_{\rm sig}$) is reconstructed in the decays of interest. The first selection step of the $B_{\rm sig}$ reconstruction is the requirement of at least one electron or muon candidate in the event. For both lepton types, the lepton is required to have a minimum momentum of $p > 300$ MeV/c. The lepton's point of closest approach to the KEKB interaction point (IP) is required to be within $|dz| < 2$ cm of the IP along the detector axis and within $dr < 0.5$ cm in the transverse plane.

The polar angle of muon candidate tracks is required to be within the range $45° < \theta_\mu < 145°$ to ensure that the tracks enter the KLM. Electron tracks need to be within the CDC acceptance $17° < \theta_e < 150°$. This implies that the track is within the ECL acceptance.

The likelihood ratio $\mathcal{R}_\mu = \mathcal{L}_\mu/(\mathcal{L}_\mu + \mathcal{L}_{\rm hadron})$, where $\mathcal{L}_\mu$ and $\mathcal{L}_{\rm hadron}$ are the likelihoods for muons and charged hadrons, is required to be greater than 0.9 for muon candidates. This selection has an average efficiency of 89% with a pion misidentification rate of 1.4% for muons with momenta between 1 and 3 GeV/c [23]. For electron candidates the likelihood ratio $\mathcal{R}_e$ is required to be greater than 0.8. This requirement has an average efficiency of 92% at a pion misidentification rate of 0.25% for electrons with momenta between 1 and 3 GeV/c [22].

The four-momentum of the closest photon that is within a 5° cone around an electron's momentum direction is added to that of the electron candidate to correct for bremsstrahlung. The photon's energy is required to be greater than 50, 75, and 100 MeV for the barrel ($32.2° < \theta_\gamma < 128.7°$), forward ($12.4° < \theta_\gamma < 31.4°$), and backward end cap ($130.7° < \theta_\gamma < 155.1°$) region of the ECL, respectively.

Kaons and pions are identified using the ratio $\mathcal{R}_{K/\pi} = \mathcal{L}_K/(\mathcal{L}_K + \mathcal{L}_\pi)$ between the combined ACC, TOF, and CDC likelihood for a kaon and the sum of the kaon and pion likelihoods [33]. Kaons (pions) are required to have $\mathcal{R}_{K/\pi} > 0.6$ ($\mathcal{R}_{K/\pi} < 0.4$), which has an average efficiency of 92% (93.5%). Kaon and pion candidate tracks must satisfy $dr < 2$ cm and $|dz| < 5$ cm.

Neutral kaon candidates are reconstructed from $\pi^+\pi^-$ pairs. The invariant mass of K^0_S candidates is required to be in the range 482 to 514 MeV/c^2, which is about 4σ around the nominal mass, where σ corresponds to the mass resolution. For low- ($p < 0.5$ GeV/c), medium- ($0.5 \le p \le 1.5$ GeV/c), and high-momentum ($p > 1.5$ GeV/c) K^0_S candidates, we require that the pion daughters have $dr > 0.05$, 0.03, and 0.02 cm, respectively. The angle in the transverse plane between the vector from the interaction point to the K^0_S vertex and the K^0_S flight direction is required to be less than 0.3, 0.1, and 0.03 rad for low-, medium-, and high-momentum candidates, respectively; the separation distance along the beam axis of the two pion trajectories at their point of closest approach is required to be below 0.8, 1.8, and 2.4 cm, respectively. For medium- (high-) momentum K^0_S candidates, we require the flight length in the transverse plane to be greater than 0.08 cm (0.22 cm). Finally, a mass-constrained vertex fit of the K^0_S candidate must converge.

Neutral pion candidates are reconstructed from pairs of photons, which must satisfy the same region-dependent energy requirements as the photons considered for the bremsstrahlung correction described above. The diphoton invariant mass is required to be between 120 and 150 MeV/c^2, which corresponds to about 5σ around the nominal mass. A mass-constrained fit of the two photons is required to converge. Photons are not allowed to be shared between π^0 candidates. To eliminate duplicates, all π^0 candidates of an event are sorted according to the most energetic daughter photon (and then, if needed, the second most energetic daughter). Any π^0 candidate that shares photons with one that appears earlier in this list is removed.

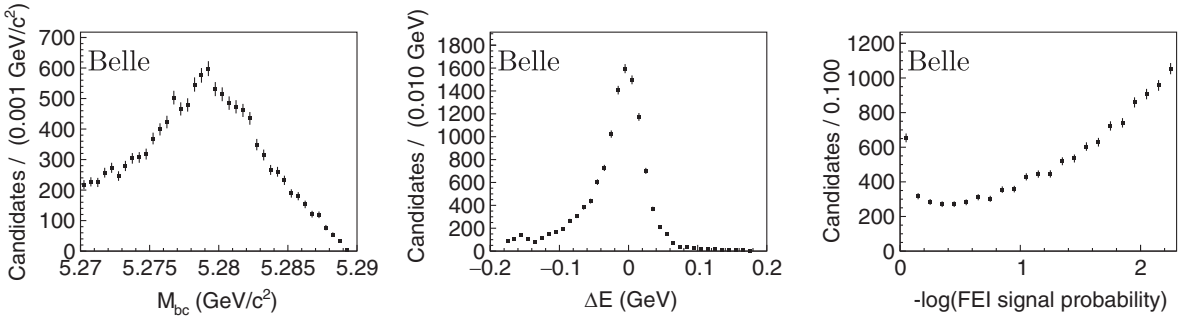

FIG. 1. Distribution of beam-constrained mass (left), energy difference (middle), and FEI signal probability (right) of charged $B_{\rm tag}$ meson candidates for the data after the full selection. The error bars show the statistical uncertainties.

Charged kaons, charged and neutral pions, and K^0_S mesons are combined to form neutral and charged D meson candidates. A total of 10 hadronic D^0 modes with the final states $K^-\pi^+$, $K^-\pi^+\pi^0$, $K^-\pi^+\pi^+\pi^-$, $K^0_S\pi^+\pi^-$, K^+K^-, $K^0_S\pi^0$, $K^0_S\pi^+\pi^-\pi^0$, $\pi^+\pi^-$, $K^-\pi^+\pi^-\pi^+\pi^0$, and $\pi^+\pi^-\pi^0$, and 9 hadronic D^+ modes with the final states $K^0_S\pi^+$, $K^0_S\pi^+\pi^-\pi^+$, $K^-\pi^+\pi^+$, $K^-K^+\pi^+$, $K^-\pi^+\pi^+\pi^0$, $K^0_S\pi^+\pi^0$, $K^0_SK^+$, $\pi^+\pi^0$, and $\pi^+\pi^-\pi^+$ are considered [34]. For D final states with at least one π^0 the D -candidate invariant mass is required to be within ± 25 MeV$/c^2$ of the nominal value [3], while the requirement for all other modes is ± 15 MeV$/c^2$, which corresponds to about 3σ. A global decay chain fit [35] is performed for all D modes except for $D^0 \to K^0_S\pi^0$. In these fits, mass constraints are applied to the D candidate as well as to K^0_S and π^0 candidates. If the fit fails, the candidate is discarded. The relative contribution and purity of the D modes is listed in Tables XIII and XIV.

Neutral D^0 meson candidates are combined with π^0 candidates to form D^{*0} candidates. The mass difference between the D^{*0} and the D^0 candidates is restricted to be between 138.9 and 145.5 MeV$/c^2$, and a global decay chain fit with mass constraints on the D^{*0}, D^0, K^0_S, and π^0 must converge. Similarly, D^{*+} meson candidates are formed from combinations of D^+ and π^0 as well as D^0 and π^+. The invariant mass of the D^{*+} candidates is allowed to deviate from the nominal mass by no more than 3 MeV$/c^2$. Again, a global decay chain fit is performed with mass constraints on the D^{*+}, D^0 or D^+, K^0_S, and π^0.

B. Specific event selection of $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ decays

By combining one $D^{(*)}$ meson candidate, one lepton candidate, and one charged pion candidate, B meson candidates are formed. The invariant mass $M(D\pi)$ is required to be below 2.8 GeV$/c^2$, as the potential D^{**} states are expected to be at lower masses. We also require $M(D\pi)$ to be above 2.05 GeV$/c^2$ to suppress $B \to \bar{D}^*\ell^+\nu_\ell$ contributions.

C. Specific event selection of $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ decays

Further B meson candidates are formed from $D^{(*)}$ meson candidates, one lepton candidate, and two oppositely-charged pion candidates. The PID requirement for the muons is tightened to $\mathcal{R}_\mu > 0.97$, which implicitly also removes all muon candidates with momenta lower than 500 MeV$/c$. To suppress the background from hadronically decaying B meson events, the missing momentum $p_{\rm miss}$ of the event is required to be greater than 200 MeV$/c$. Here $p_{\rm miss} = |\vec{p}(e^+e^-) - \vec{p}(B_{\rm tag}) - \vec{p}(D^{(*)}) - \vec{p}(\pi_1) - \vec{p}(\pi_2) - \vec{p}(\ell)|$ is the difference between the total momentum of the initial colliding beam particles and the combined momentum of all visible particles measured in the center-of-mass frame. Analogously, the missing energy $E_{\rm miss}$ is defined as the energy difference between the center-of-mass energy and the sum over the energies of the $B_{\rm sig}$ and $B_{\rm tag}$ candidates.

To suppress D^{*-} contributions to the final state in $B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$, a veto is implemented: the combined invariant mass of the neutral D meson and the pion with the opposite charge to that of the B meson is required to be above 2.05 GeV$/c^2$. The contamination from $B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$ with $D^{*-} \to \bar{D}^0\pi^-$ is reduced by 50% with this veto. However, the pions used in the reconstruction of the $B_{\rm sig}$ meson candidate can also arise from the decay of the $B_{\rm tag}$ meson. Therefore, a second veto is implemented: the invariant mass of each π^+ used in the $B_{\rm tag}$ reconstruction combined with the signal D^0 is required to be greater than 2.05 GeV$/c^2$.

The $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ modes have much more background than the $B \to \bar{D}^*\ell^+\nu_\ell$ and $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ modes. In order to increase the sensitivity of this channel, a boosted decision tree (BDT) [36] is used to further reduce the background. The following 25 input variables are used in the BDT: $E_{\rm extra}$, the unaccounted energy in the ECL; $R_1 - R_4$, the ratios of the first, second, third, and fourth to the zeroth Fox-Wolfram moments [37]; $H_0 - H_4$, the harmonic moments of zeroth to fourth order with respect to the thrust axis (Chapter 9.3 of Ref. [38]); $C_0 - C_8$, the momentum flow in nine cones of 10° around the thrust axis [39]; the sphericity and the aplanarity of the event (Chapter 9.3 of Ref. [38]); the thrust value of the event and the cosine of the polar angle of the thrust axis (Chapter 9.3 of Ref. [38]); the number of tracks used in the $B_{\rm tag}$ reconstruction; the number of neutral clusters used in the $B_{\rm tag}$ reconstruction. The BDT is trained with signal MC simulations and off-resonance data, as most of the remaining background originates from $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$) "continuum" events. The signal MC is divided into N subsamples each containing the number of expected candidates in the full Belle dataset based on the branching fraction results of the *BABAR* measurement [19]. For each subsample an individual BDT is trained using the other $N-1$ subsamples such that the size of the training sample is maximized while keeping it independent from the sample that the BDT is applied to and therefore avoiding bias. Separate BDTs are trained for the B^+ and B^0 modes. The distribution of all BDT output classifiers combined is shown in Fig. 2. The BDT output variable is required to be greater than -0.06 for $B^+ \to D^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ and greater than -0.09 for $B^0 \to D^{(*)-}\pi^+\pi^-\ell^+\nu_\ell$, which maximizes the ratio between the signal yield and its uncertainty from a fit to MC samples.

D. $\Upsilon(4S)$ selection

A total of 12 B^+ modes and 12 B^0 modes are reconstructed. Each $B_{\rm sig}$ candidate and $B_{\rm tag}$ candidate are combined to form an $\Upsilon(4S)$ candidate. In the combinations

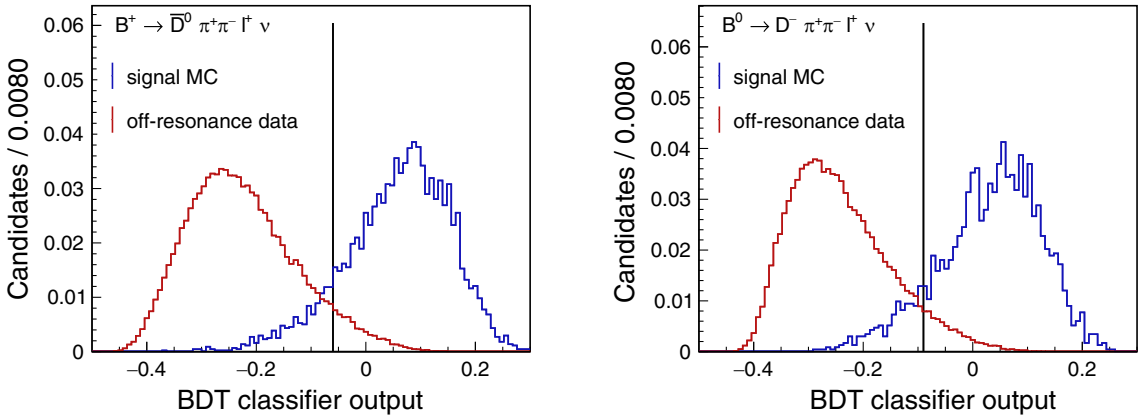

FIG. 2. Distribution of BDT output classifier for $B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$ (left) and $B^0 \to D^-\pi^+\pi^-\ell^+\nu_\ell$ (right). Candidates to the right of the vertical line are retained.

the electric charge must be conserved but the flavor of two neutral B mesons is allowed to be the same. Candidates with tracks that are not assigned to the $\Upsilon(4S)$ candidate are rejected. The average number of candidates per event after all previously described selection steps is 1.10. In events that contain more than one $\Upsilon(4S)$ candidate, a single candidate is selected, as follows. Firstly, the $B_{\rm tag}$ candidate with the highest FEI signal probability is selected. If multiple $B_{\rm sig}$ candidates remain, the D^* mode is preferred over the D mode, since otherwise an additional π^0 candidate would be left in the event. In some events, candidates are reconstructed in both the one-pion and the two-pion modes. As the distribution of $U = E_{\rm miss} - p_{\rm miss}c$ is used for signal extraction, if U is between -0.1 and 0.1 GeV for at least one candidate in both decay modes, all candidates in the event are rejected. If there are still multiple candidates, only the one with the smallest difference between the $D^{(*)}$ candidate mass and the nominal mass is retained.

A wrong-sign reconstruction, e.g., $B^+ \to D^+\pi^-\ell^+\nu_\ell$ instead of $B^+ \to D^-\pi^+\ell^+\nu_\ell$, is performed to validate the background distribution.

TABLE II. Ratios between the selection efficiencies of the signal and normalization modes. The uncertainty is the MC sample statistical uncertainty.

	Electron mode	Muon mode
$B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$	1.194 ± 0.012	1.139 ± 0.010
$B^+ \to D^-\pi^+\ell^+\nu_\ell$	0.518 ± 0.004	0.482 ± 0.004
$B^0 \to \bar{D}^{*0}\pi^-\ell^+\nu_\ell$	1.156 ± 0.009	1.094 ± 0.009
$B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$	0.4040 ± 0.0026	0.3824 ± 0.0027
$B^0 \to D^-\pi^+\pi^-\ell^+\nu_\ell$	0.450 ± 0.015	0.389 ± 0.014
$B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$	0.288 ± 0.007	0.264 ± 0.007
$B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$	0.286 ± 0.016	0.270 ± 0.017
$B^+ \to \bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell$	0.220 ± 0.008	0.179 ± 0.008

For each signal MC mode, the efficiency is taken to be the fraction of correctly reconstructed candidates. A weighted average of the efficiencies based on the relative abundances of the D^{**} state reported by the PDG [3] is taken as the final efficiency value. The ratios between the efficiencies of the signal and normalization modes

$$R(B^+) = \frac{\epsilon(B^+ \to \bar{D}^{**0}\ell^+\nu_\ell)}{\epsilon(B^+ \to \bar{D}^{*0}\ell^+\nu_\ell)} \quad \text{and}$$
$$R(B^0) = \frac{\epsilon(B^0 \to D^{**-}\ell^+\nu_\ell)}{\epsilon(B^0 \to D^{*-}\ell^+\nu_\ell)} \tag{2}$$

are given in Table II.

V. EXTRACTION OF SIGNAL YIELDS

The number of signal candidates is determined with an unbinned extended maximum likelihood fit of $U = E_{\rm miss} - p_{\rm miss}c$. The probability density function (PDF) used to describe the U distribution is constructed from templates based on the MC.

A. Fit of $B \to \bar{D}^{(*)}\ell^+\nu_\ell$ sample

For the fit of the $B \to \bar{D}^{(*)}\ell^+\nu_\ell$ sample the total PDF consists of four (three) components

$$\mathcal{P}^{\bar{D}e} = N^{\bar{D}e}_{\rm sig}\mathcal{P}^{\bar{D}e}_{\rm sig} + \frac{\epsilon^{\bar{D}e}_{\rm MC}}{\epsilon^{\bar{D}e}_{\rm MC} + \epsilon^{\bar{D}^*e}_{\rm MC}} N^{\bar{D}^*e}_{\rm sig}\mathcal{P}^{\bar{D}e}_{\rm fd} + f^{\bar{D}e}_{\rm bkg}N^{\bar{D}e}_{\rm bkg}\mathcal{P}^{\bar{D}e}_{B\bar{B}} + \left(1 - f^{\bar{D}e}_{\rm bkg}\right)N^{\bar{D}e}_{\rm bkg}\mathcal{P}^{\bar{D}\ell}_{\rm cont}, \tag{3}$$

$$\mathcal{P}^{\bar{D}\mu} = N^{\bar{D}\mu}_{\rm sig}\mathcal{P}^{\rm sig}_{\bar{D}\mu} + \frac{\epsilon^{\bar{D}\mu}_{\rm MC}}{\epsilon^{\bar{D}\mu}_{\rm MC} + \epsilon^{\bar{D}^*\mu}_{\rm MC}} N^{\bar{D}^*\mu}_{\rm sig}\mathcal{P}^{\bar{D}\mu}_{\rm fd} + f^{\bar{D}\mu}_{\rm bkg}N^{\bar{D}\mu}_{\rm bkg}\mathcal{P}^{\bar{D}\mu}_{B\bar{B}} + \left(1 - f^{\bar{D}\mu}_{\rm bkg}\right)N^{\bar{D}\mu}_{\rm bkg}\mathcal{P}^{\bar{D}\ell}_{\rm cont}, \tag{4}$$

TABLE III. Fitted signal and background yields of the normalization modes $B \to \bar{D}^{(*)}\ell^+\nu_\ell$ in the full Belle data sample. The quoted uncertainties are statistical only.

	Signal		Background	
	Electron mode	Muon mode	Electron mode	Muon mode
$B^0 \to D^-\ell^+\nu_\ell$	3154 ± 67	2723 ± 60	2097 ± 60	1696 ± 52
$B^+ \to \bar{D}^0\ell^+\nu_\ell$	8974 ± 136	7752 ± 124	9840 ± 149	8548 ± 133
$B^0 \to D^{*-}\ell^+\nu_\ell$	6271 ± 102	5624 ± 91	925 ± 54	742 ± 44
$B^+ \to \bar{D}^{*0}\ell^+\nu_\ell$	19940 ± 200	18045 ± 183	523 ± 35	508 ± 34

$$\mathcal{P}^{\bar{D}^* e} = \frac{\epsilon_{\rm MC}^{\bar{D}^* e}}{\epsilon_{\rm MC}^{\bar{D} e} + \epsilon_{\rm MC}^{\bar{D}^* e}} N_{\rm sig}^{\bar{D}^* e}\mathcal{P}_{\rm sig}^{\bar{D}^* e} + f_{\rm bkg}^{\bar{D}^* e} N_{\rm bkg}^{\bar{D}^* e}\mathcal{P}_{B\bar{B}}^{\bar{D}^* e} + \left(1 - f_{\rm bkg}^{\bar{D}^* e}\right) N_{\rm bkg}^{\bar{D}^* e}\mathcal{P}_{\rm cont}^{\bar{D}^* \ell}, \quad (5)$$

$$\mathcal{P}^{\bar{D}^* \mu} = \frac{\epsilon_{\rm MC}^{\bar{D}^* \mu}}{\epsilon_{\rm MC}^{\bar{D} \mu} + \epsilon_{\rm MC}^{\bar{D}^* \mu}} N_{\rm sig}^{\bar{D}^* \mu}\mathcal{P}_{\rm sig}^{\bar{D}^* \mu} + f_{\rm bkg}^{\bar{D}^* \mu} N_{\rm bkg}^{\bar{D}^* \mu}\mathcal{P}_{B\bar{B}}^{\bar{D}^* \mu} + \left(1 - f_{\rm bkg}^{\bar{D}^* \mu}\right) N_{\rm bkg}^{\bar{D}^* \mu}\mathcal{P}_{\rm cont}^{\bar{D}^* \ell}, \quad (6)$$

where $\mathcal{P}_{\rm sig}^{\bar{D}^{(*)}\ell}$ is the signal PDF and $\mathcal{P}_{\rm fd}^{\bar{D}\ell}$, $\mathcal{P}_{B\bar{B}}^{\bar{D}^{(*)}\ell}$, and $\mathcal{P}_{\rm cont}^{\bar{D}^{(*)}\ell}$ are the PDFs describing the feeddown, $B\bar{B}$ background, and continuum background, respectively. Feeddown describes a contribution from $B \to \bar{D}^*\ell^+\nu_\ell$ that shows up in the $B \to \bar{D}\ell^+\nu_\ell$ modes if the neutral pion of a $D^{*0} \to D^0\pi^0$ or a $D^{*+} \to D^+\pi^0$ decay is missed in the reconstruction. Due to the missing π^0 it is shifted to higher values in the U distribution. Thus, this contribution can be separated and used to improve the sensitivity of the branching fraction measurement.

The fraction of the $B\bar{B}$ component among the total background, $f_{\rm bkg}$, is constrained to the values estimated in simulation. A simultaneous fit of $B \to \bar{D}\ell^+\nu_\ell$ and $B \to \bar{D}^*\ell^+\nu_\ell$ is performed, where the total $B \to \bar{D}^*\ell^+\nu_\ell$ yield $N_{\rm sig}^{\bar{D}^*\ell}$ is determined as the sum of the signal and feeddown components, which are related via their efficiencies ϵ. The templates used to construct the PDFs are created with 125 bins between -0.5 and 2 GeV. Separate PDFs are used for the electron and muon modes except for the continuum PDF, which is created from the combined sample of the two modes as their distributions are statistically compatible with each other.

The width of the signal peak in the U distribution differs between data and MC, even after all known corrections are applied. To compensate for this effect, the signal PDFs are constructed by convolving the signal-MC templates with a Gaussian whose mean and width are floating in the fit to data. Independent widths are used for the electron and muon modes. The fitted $B \to \bar{D}^{(*)}\ell^+\nu_\ell$ signal and background yields are listed in Table III and the corresponding plots are shown in Figs. 3–6. In the Appendix the fit results of the mean and width of the Gaussian are listed.

B. Fit of $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ and $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ samples

A simultaneous fit to the U distribution of 16 categories splitting the full sample according to the B flavor mode (B^0 vs B^+), the D mode (D^0/D^+ vs D^{*0}/D^{*+}), the number of pion daughters ($D\pi$ vs $D\pi\pi$), and the lepton mode (e vs μ) is performed. This allows several background sources to be constrained directly from the data, as described below. All templates are constructed with 120 bins in the range -1 to 2 GeV.

The $B^+ \to D^-\pi^+\ell^+\nu_\ell$ fit PDF consists of five components: signal, feeddown, misreconstructed $B \to \bar{D}^{**}\ell^+\nu_\ell$ background, other $B\bar{B}$, and continuum:

$$\begin{aligned}\mathcal{P}^{D^-\pi^+\ell^+} = {} & N_{\rm sig}^{D^-\pi^+\ell^+}\mathcal{P}_{\rm sig}^{D^-\pi^+\ell^+} \\ & + \frac{\epsilon_{\rm MC}^{D^-\pi^+\ell^+}}{\epsilon_{\rm MC}^{D^-\pi^+\ell^+} + \epsilon_{\rm MC}^{D^{*-}\pi^+\ell^+}} N_{\rm sig}^{D^{*-}\pi^+\ell^+}\mathcal{P}_{\rm fd}^{D^-\pi^+\ell^+} \\ & + \sum_{\rm i=1}^{14} N_{D^{**},{\rm i}}\mathcal{P}_{D^{**},{\rm i}}^{D^-\pi^+\ell^+} + N_{B\bar{B}}^{D^-\pi^+\ell^+}\mathcal{P}_{B\bar{B}}^{D^-\pi^+\ell^+} \\ & + \frac{L_{\rm on}}{L_{\rm off}}\mathcal{P}_{\rm off}^{D^-\pi^+\ell^+}. \end{aligned} \quad (7)$$

The signal template $\mathcal{P}_{\rm sig}^{D^-\pi^+\ell^+}$ is obtained from signal MC, in which the $D\pi$ is produced in D_0^{*0} decay 62% of the time, and in D_2^{*0} decay 38% of the time [3]. As in the fit of the $B \to \bar{D}^{(*)}\ell^+\nu_\ell$ sample, the signal template is convolved with a Gaussian to compensate for the different signal widths between data and MC. The feeddown component $\mathcal{P}_{\rm fd}^{D^-\pi^+\ell^+}$ comes from $B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$ decays and is taken from signal MC, in which the $D^*\pi$ final state is produced in D_1^0 decay 45% of the time, in $D_1'^0$ decay 40% of the time, and in D_2^{*0} decay the rest of the time. The $B \to \bar{D}^{**}\ell^+\nu_\ell$ background PDF $\mathcal{P}_{D^{**},{\rm i}}^{D^-\pi^+\ell^+}$ is obtained from 14 different MC samples:

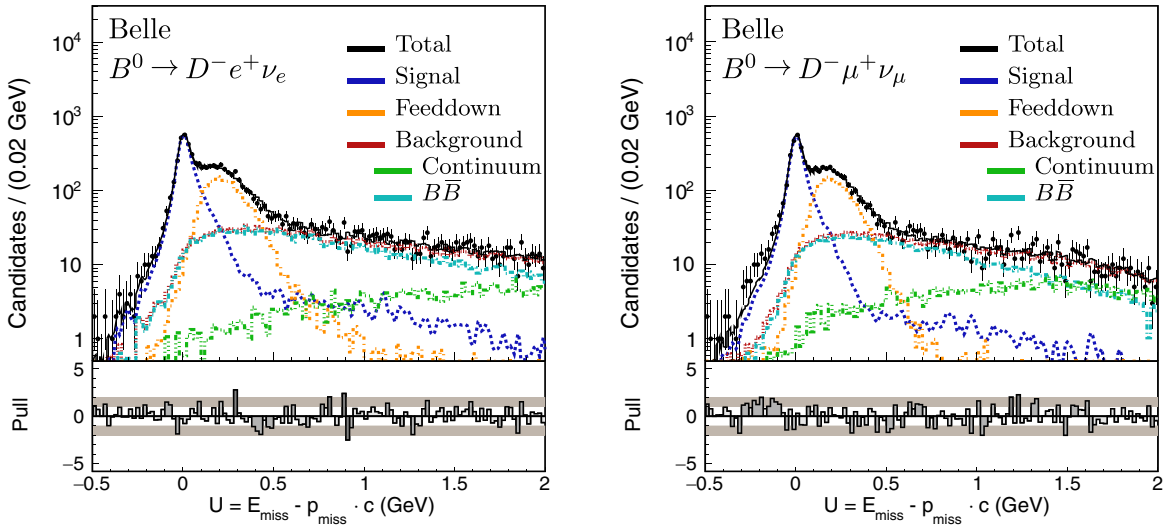

FIG. 3. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^0 \to D^- e^+ \nu_e$ (left) and $B^0 \to D^- \mu^+ \nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

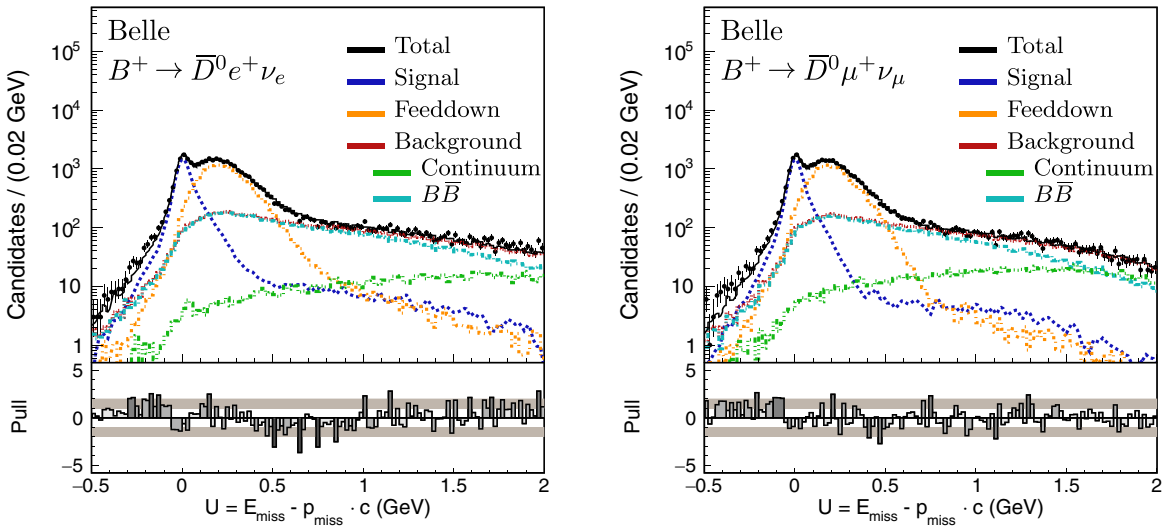

FIG. 4. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^+ \to \bar{D}^0 e^+ \nu_e$ (left) and $B^+ \to \bar{D}^0 \mu^+ \nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

(i) $B^+ \to \bar{D}^{**0} \ell^+ \nu_\ell$ with $\bar{D}^{**0} \in (\bar{D}_0^{*0}, \bar{D}_1^0, \bar{D}_1'^0, \bar{D}_2^{*0})$,
(ii) $B^+ \to \bar{D}_1^0 \ell^+ \nu_\ell$ with $\bar{D}_1^0 \to \bar{D}^0 \pi^+ \pi^-$,
(iii) $B^+ \to \bar{D}^0 \pi^+ \pi^- \ell^+ \nu_\ell$,
(iv) $B^+ \to \bar{D}^{*0} \pi^+ \pi^- \ell^+ \nu_\ell$,
(v) $B^0 \to D^{**-} \ell^+ \nu_\ell$ with $D^{**-} \in (D_0^{*-}, D_1^-, D_1'^-, D_2^{*-})$,
(vi) $B^0 \to D_1^- \ell^+ \nu_\ell$ with $D_1^- \to D^- \pi^+ \pi^-$,
(vii) $B^0 \to D^- \pi^+ \pi^- \ell^+ \nu_\ell$, and
(viii) $B^0 \to D^{*-} \pi^+ \pi^- \ell^+ \nu_\ell$.

These events constitute background due to misreconstructed signal candidates, swapping of final state particles between the $B_{\rm sig}$ and $B_{\rm tag}$ candidates, or events with $D^{**} \to D^{(*)} \pi^0$. The composition of the different D^{**} states is set to the world averages of these modes [3]. The yields of the $B \to \bar{D}^{**} \ell^+ \nu_\ell$ background components, $N_{D^{**},{\rm i}}$, are calculated as the product of the terms listed in Table IV. The other $B\bar{B}$ background is taken from a generic $b \to c$ MC sample with six times the luminosity of data. Off-resonance data is used to model the continuum PDF $\mathcal{P}_{\rm off}^{D^- \pi^+ \ell^+}$. The yield of the continuum contribution in the fit is constrained via the ratio of the on- and off-resonance luminosities. The ratio is allowed to float in the fit within a Gaussian constraint with a width of 1%. This accounts for the uncertainty in the determination of the luminosity ratio.

The fit model for $B^0 \to \bar{D}^0 \pi^- \ell^+ \nu_\ell$ is constructed similarly:

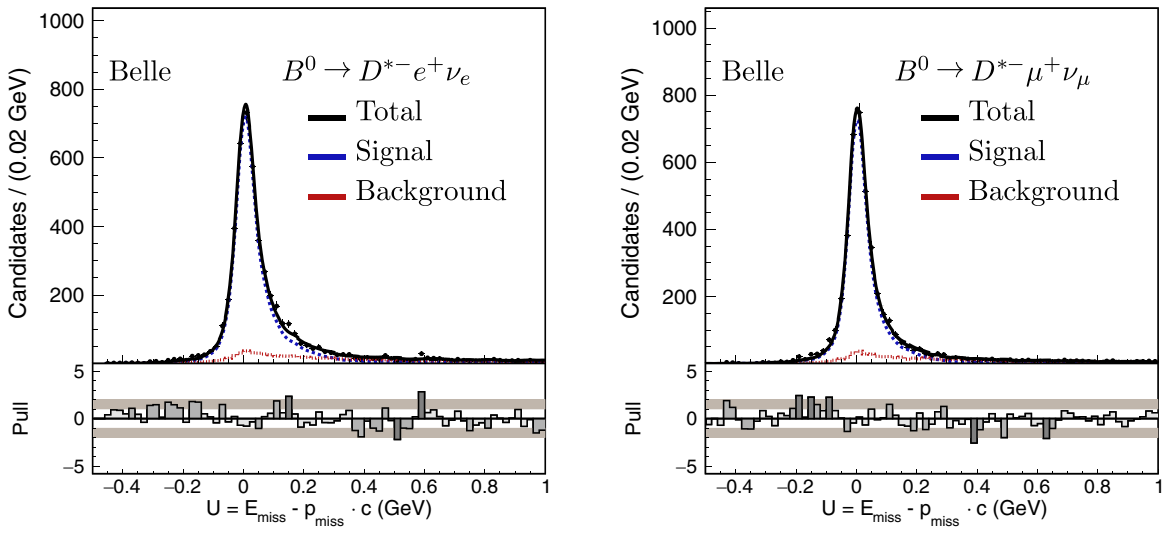

FIG. 5. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^0 \to D^{*-}e^+\nu_e$ (left) and $B^0 \to D^{*-}\mu^+\nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown.

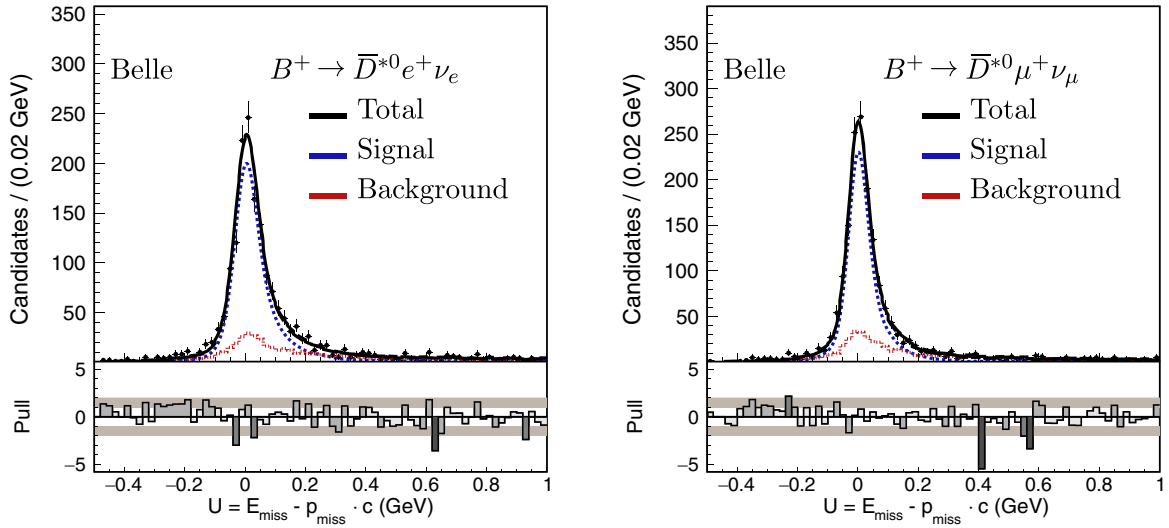

FIG. 6. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^+ \to \bar{D}^{*0}e^+\nu_e$ (left) and $B^+ \to \bar{D}^{*0}\mu^+\nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown.

$$\mathcal{P}^{\bar{D}^0\pi^-\ell^+} = N_{\rm sig}^{\bar{D}^0\pi^-\ell^+}\mathcal{P}_{\rm sig}^{\bar{D}^0\pi^-\ell^+} + \frac{\epsilon_{\rm MC}^{\bar{D}^0\pi^-\ell^+}}{\epsilon_{\rm MC}^{\bar{D}^0\pi^-\ell^+} + \epsilon_{\rm MC}^{\bar{D}^{*0}\pi^-\ell^+}} N_{\rm sig}^{\bar{D}^{*0}\pi^-\ell^+}\mathcal{P}_{\rm fd}^{\bar{D}^0\pi^-\ell^+} + \sum_{\rm i=1}^{14} N_{D^{**},{\rm i}}\mathcal{P}_{D^{**},{\rm i}}^{\bar{D}^0\pi^-\ell^+} + N_{B\bar{B}}^{\bar{D}^0\pi^-\ell^+}\mathcal{P}_{B\bar{B}}^{\bar{D}^0\pi^-\ell^+} + \frac{L_{\rm on}}{L_{\rm off}}\mathcal{P}_{\rm off}^{\bar{D}^0\pi^-\ell^+}$$
$$+ \frac{\epsilon(B^0 \to D^{*-}\ell^+\nu_\ell \text{ reconstructed as } B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell)}{\epsilon(B^0 \to D^{*-}\ell^+\nu_\ell \text{ reconstructed as } B^0 \to D^{*-}\ell^+\nu_\ell)} N_{\rm sig}^{D^{*-}\ell^+}\mathcal{P}_{D^{*-}\ell^+}^{\bar{D}^0\pi^-\ell^+}. \quad (8)$$

The signal composition is 71% $B^0 \to D_0^{*-}\ell^+\nu_\ell$ and 29% $B^0 \to D_2^{*-}\ell^+\nu_\ell$ decays [3]. Again, the signal template is convolved with a Gaussian. The feeddown is produced in $B^0 \to D_1^-\ell^+\nu_\ell$ decay 42.5% of the time, in $B^0 \to D_1'^-\ell^+\nu_\ell$ decay 47% of the time, and in $B^0 \to D_2^{*-}\ell^+\nu_\ell$ decay 10.5% of the time. Compared to the B^+ mode an additional sixth component is added to account for misreconstructed $B^0 \to D^{*-}\ell^+\nu_\ell$ candidates that survive the D^* veto. The yield of this component is fixed by the product of the $B^0 \to D^{*-}\ell^+\nu_\ell$ signal yield $N_{\rm sig}^{D^{*-}\ell^+}$ from the fit described in Sec. VA and the ratio of efficiencies of the $B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$ and $B^0 \to D^{*-}\ell^+\nu_\ell$ selections.

TABLE IV. The $B \to \bar{D}^{**}\ell^+\nu_\ell$ background yields are the product of the scaling factor, the corresponding signal yield, and the efficiency ratio. The scaling factor accounts for the assumed relative abundance of the D^{**} state, e.g., that for $B^+ \to D^-\pi^+\ell^+\nu_\ell$ in 62% of the time the decay proceeds via an intermediate $\bar{D}_0^{*0}$ resonance.

Component	Scaling factor	Signal yield	Efficiency ratio
$B^+ \to \bar{D}_0^{*0}\ell^+\nu_\ell$	0.62	$N_{\rm sig}^{D^-\pi^+\ell^+}$	$\frac{N(B^+\to\bar{D}_0^{*0}\ell^+\nu_\ell \text{ with error in reconstruction as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^+\to\bar{D}_0^{*0}\ell^+\nu_\ell \text{ correctly reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}$
$B^+ \to \bar{D}_1^{0}\ell^+\nu_\ell$	0.45	$N_{\rm sig}^{D^{*-}\pi^+\ell^+}$	$\frac{N(B^+\to\bar{D}_1^{0}\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^+\to\bar{D}_1^{0}\ell^+\nu_\ell \text{ correctly reconstructed as } B^+\to D^{*-}\pi^+\ell^+\nu_\ell)}$
$B^+ \to \bar{D}_1'^{0}\ell^+\nu_\ell$	0.4	$N_{\rm sig}^{D^{*-}\pi^+\ell^+}$	$\frac{N(B^+\to\bar{D}_1'^{0}\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^+\to\bar{D}_1'^{0}\ell^+\nu_\ell \text{ correctly reconstructed as } B^+\to D^{*-}\pi^+\ell^+\nu_\ell)}$
$B^+ \to \bar{D}_2^{*0}\ell^+\nu_\ell$	0.38	$N_{\rm sig}^{D^-\pi^+\ell^+}$	$\frac{N(B^+\to\bar{D}_2^{*0}\ell^+\nu_\ell \text{ with error in reconstruction as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^+\to\bar{D}_2^{*0}\ell^+\nu_\ell \text{ correctly reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}$
$B^+ \to \bar{D}_1^{0}\ell^+\nu_\ell$ $(D_1^0 \to D^0\pi^+\pi^-)$	0.55	$N_{\rm sig}^{D^0\pi^+\pi^-\ell^+}$	$\frac{N(B^+\to\bar{D}_1^{0}\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^+\to\bar{D}_1^{0}\ell^+\nu_\ell \text{ correctly reconstructed as } B^+\to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell)}$
$B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$	0.45	$N_{\rm sig}^{D^0\pi^+\pi^-\ell^+}$	$\frac{N(B^+\to\bar{D}^0\pi^+\pi^-\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^+\to\bar{D}^0\pi^+\pi^-\ell^+\nu_\ell \text{ correctly reconstructed as } B^+\to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell)}$
$B^+ \to \bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell$	1	$N_{\rm sig}^{D^{*0}\pi^+\pi^-\ell^+}$	$\frac{N(B^+\to\bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^+\to\bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell \text{ correctly reconstructed as } B^+\to \bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell)}$
$B^0 \to D_0^{*-}\ell^+\nu_\ell$	0.71	$N_{\rm sig}^{D^0\pi^-\ell^+}$	$\frac{N(B^0\to D_0^{*-}\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^0\to D_0^{*-}\ell^+\nu_\ell \text{ correctly reconstructed as } B^0\to \bar{D}^0\pi^-\ell^+\nu_\ell)}$
$B^0 \to D_1^{-}\ell^+\nu_\ell$	0.425	$N_{\rm sig}^{D^{*0}\pi^-\ell^+}$	$\frac{N(B^0\to D_1^{-}\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^0\to D_1^{-}\ell^+\nu_\ell \text{ correctly reconstructed as } B^0\to \bar{D}^{*0}\pi^-\ell^+\nu_\ell)}$
$B^0 \to D_1'^{-}\ell^+\nu_\ell$	0.47	$N_{\rm sig}^{D^{*0}\pi^-\ell^+}$	$\frac{N(B^0\to D_1'^{-}\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^0\to D_1'^{-}\ell^+\nu_\ell \text{ correctly reconstructed as } B^0\to \bar{D}^{*0}\pi^-\ell^+\nu_\ell)}$
$B^0 \to D_2^{*-}\ell^+\nu_\ell$	0.29	$N_{\rm sig}^{D^0\pi^-\ell^+}$	$\frac{N(B^0\to D_2^{*-}\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^0\to D_2^{*-}\ell^+\nu_\ell \text{ correctly reconstructed as } B^0\to \bar{D}^0\pi^-\ell^+\nu_\ell)}$
$B^0 \to D_1^{-}\ell^+\nu_\ell$ $(D_1^- \to D^-\pi^+\pi^-)$	0.55	$N_{\rm sig}^{D^-\pi^+\pi^-\ell^+}$	$\frac{N(B^0\to D_1^{-}\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^0\to D_1^{-}\ell^+\nu_\ell \text{ correctly reconstructed as } B^0\to D^-\pi^+\pi^-\ell^+\nu_\ell)}$
$B^0 \to D^-\pi^+\pi^-\ell^+\nu_\ell$	0.45	$N_{\rm sig}^{D^-\pi^+\pi^-\ell^+}$	$\frac{N(B^0\to D^-\pi^+\pi^-\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^0\to D^-\pi^+\pi^-\ell^+\nu_\ell \text{ correctly reconstructed as } B^0\to D^-\pi^+\pi^-\ell^+\nu_\ell)}$
$B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$	1	$N_{\rm sig}^{D^{*-}\pi^+\pi^-\ell^+}$	$\frac{N(B^0\to D^{*-}\pi^+\pi^-\ell^+\nu_\ell \text{ reconstructed as } B^+\to D^-\pi^+\ell^+\nu_\ell)}{N(B^0\to D^{*-}\pi^+\pi^-\ell^+\nu_\ell \text{ correctly reconstructed as } B^0\to D^{*-}\pi^+\pi^-\ell^+\nu_\ell)}$

The $B \to \bar{D}^*\pi\ell^+\nu_\ell$ fit models consist of only four components as there is no feeddown. The strategy for modeling background from $B \to \bar{D}^{**}\ell^+\nu_\ell$ is the same as for $B \to \bar{D}\pi\ell^+\nu_\ell$. The signal PDF template is obtained from signal MC, in which the $D^*\pi$ final state is produced in D_1 decay, D_1' decay, and D_2^* decay at the same proportions as the feeddown components in $B \to \bar{D}\pi\ell^+\nu_\ell$ described above. For the $B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$ signal a convolution with a Gaussian is performed, but not for the $B^0 \to \bar{D}^{*0}\pi^-\ell^+\nu_\ell$ signal. The fitted mean and width of the Gaussian are listed in Table XII.

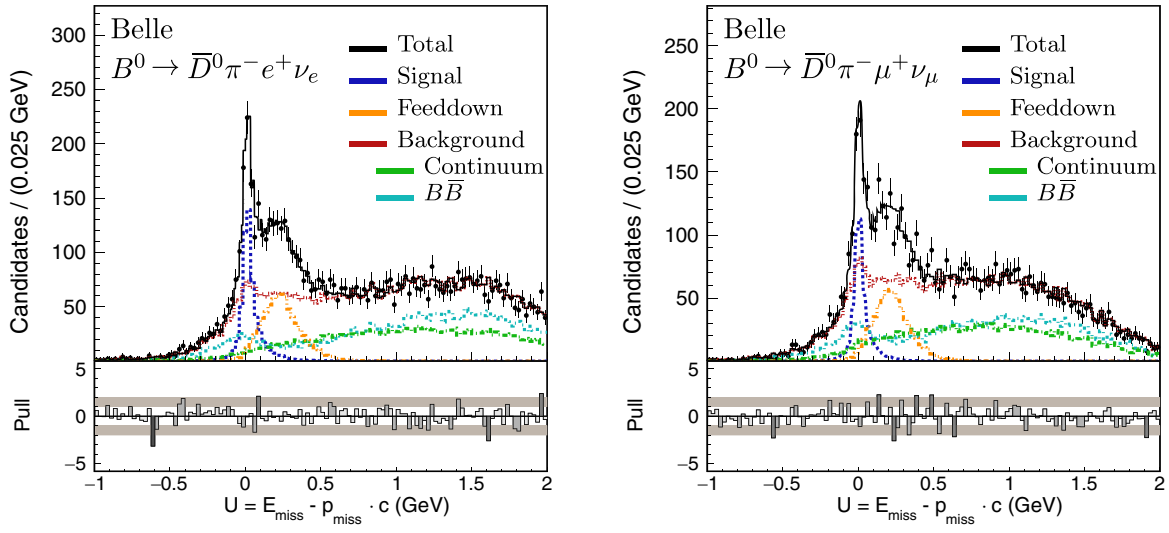

FIG. 7. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^0 \to \bar{D}^0\pi^- e^+\nu_e$ (left) and $B^0 \to \bar{D}^0\pi^-\mu^+\nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

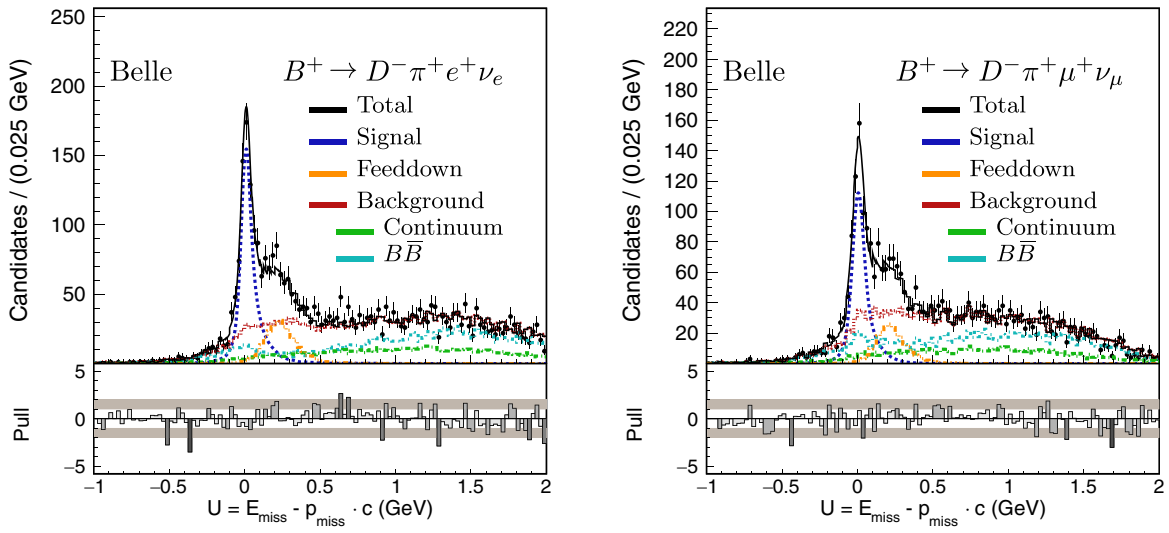

FIG. 8. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^+ \to D^-\pi^+e^+\nu_e$ (left) and $B^+ \to D^-\pi^+\mu^+\nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

For $B \to \bar{D}\pi^+\pi^-\ell^+\nu_\ell$ the fit model contains four components (signal, feeddown, other $B\bar{B}$, continuum), while for $B \to \bar{D}^*\pi^+\pi^-\ell^+\nu_\ell$ only three components are needed as there is no feeddown. Following the findings of the *BABAR* measurement [19] the signal is assumed to proceed via a D_1 resonance for the $D\pi\pi$ modes and via a D_1' resonance for the $D^*\pi\pi$ modes. The $B \to \bar{D}^*\pi^+\pi^-\ell^+\nu_\ell$ templates are constructed with 30 bins in the range −0.5 to 1 GeV.

The plots of the data and fit results are shown in Figs. 7–14. The signal and background yields are summarized in Table V.

VI. SYSTEMATIC UNCERTAINTIES

The systematic uncertainties mainly arise from the fit modeling, the uncertainty on the branching fraction values of the normalization mode $B \to \bar{D}^*\ell^+\nu_\ell$ and the charm modes, and the hadron PID. For the two-pion modes there are additional sizable systematic uncertainties from the BDT and from the limited size of the MC sample used to calculate the signal efficiency of the selection. The various considered sources of systematic uncertainties are described below. Their numerical values are summarized in Tables VI and VII.

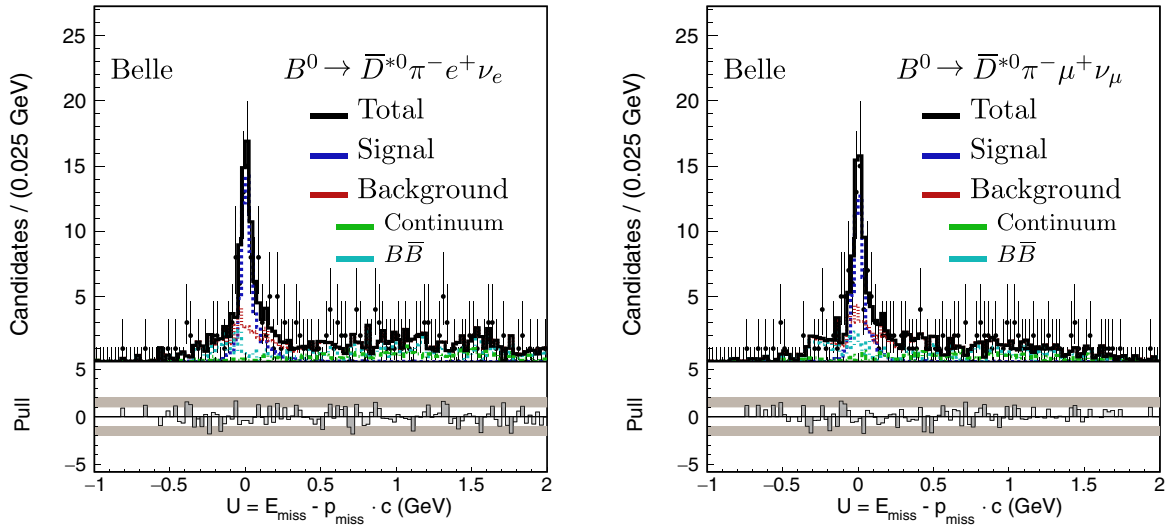

FIG. 9. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^0 \to \bar{D}^{*0}\pi^-e^+\nu_e$ (left) and $B^0 \to \bar{D}^{*0}\pi^-\mu^+\nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

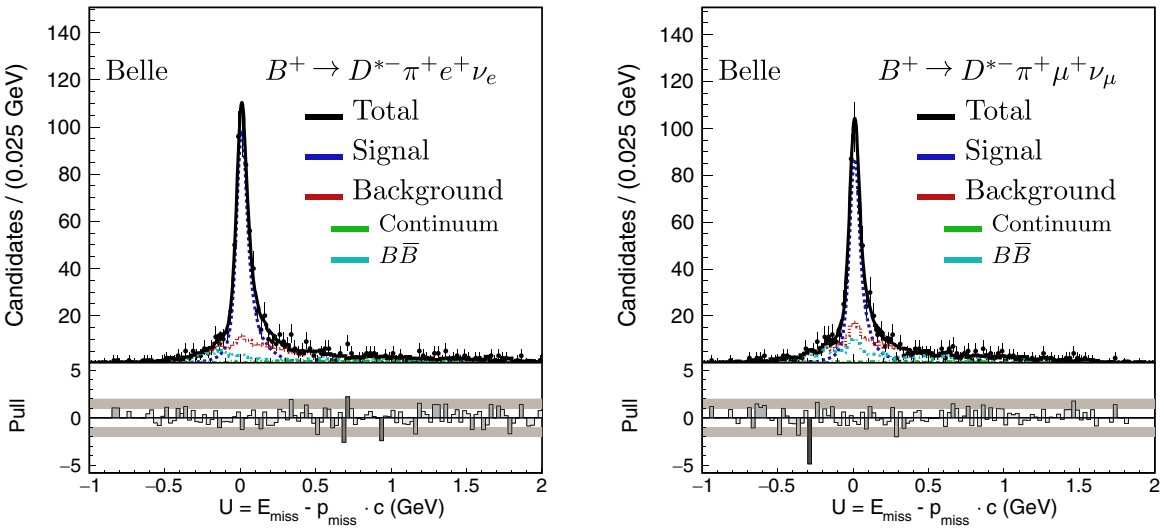

FIG. 10. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^+ \to D^{*-}\pi^+ e^+ \nu_e$ (left) and $B^+ \to D^{*-}\pi^+ \mu^+ \nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

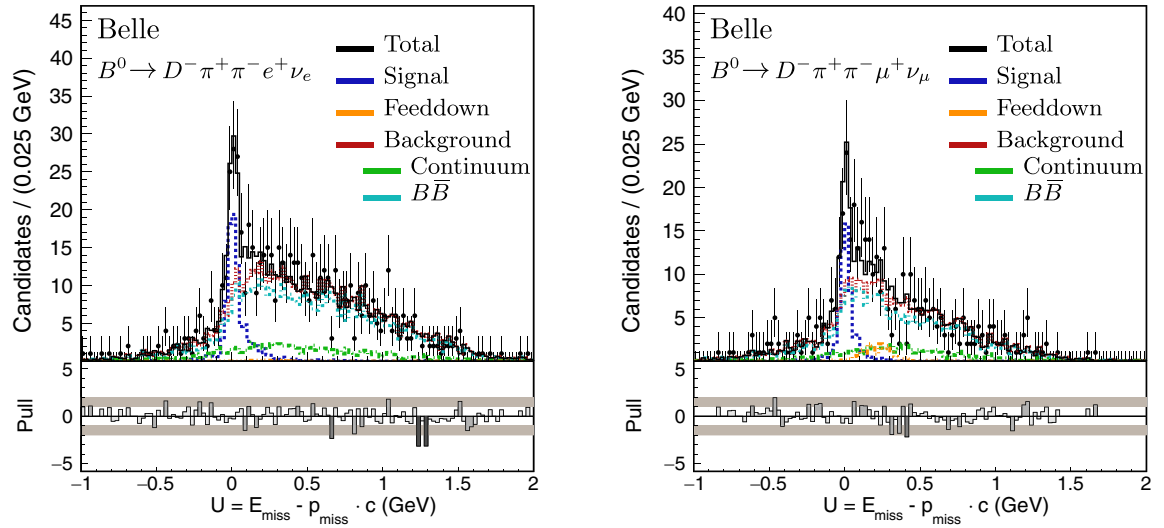

FIG. 11. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^0 \to D^-\pi^+\pi^- e^+ \nu_e$ (left) and $B^0 \to D^-\pi^+\pi^- \mu^+ \nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

A. MC statistics fit model

To account for the finite size of the MC samples used to produce the PDF templates, alternative fit PDFs are created by varying the bin contents of each PDF template according to a Poisson distribution. This is done 1000 times, and after each variation the fit to the collision data is performed with the new set of templates. It is checked that the pull distributions are unbiased, where the pull is defined as the difference between the yields using the varied fit PDF and the nominal yields divided by the statistical uncertainty of the new yields. The spread of the new signal yields (about 1% for the one-pion modes, 5%–20% for the two-pion modes) is used as an estimate for the systematic uncertainty.

B. MC statistics signal efficiency

The uncertainty on the calculated signal efficiency ratios in Table II due to the finite size of the MC samples is propagated to the branching fractions and ratios, and assigned as systematic uncertainty.

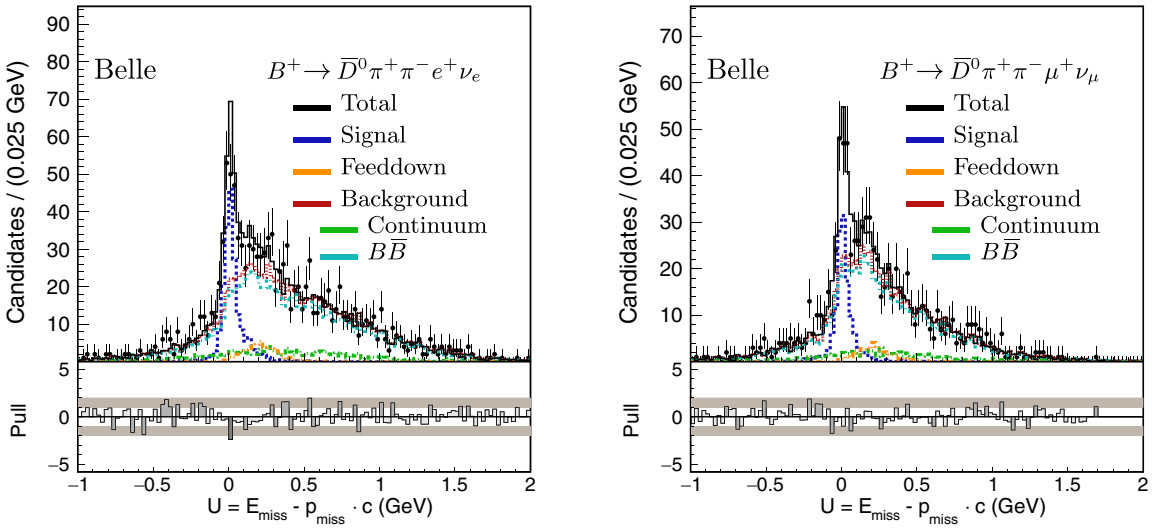

FIG. 12. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^+ \to \bar{D}^0\pi^+\pi^-e^+\nu_e$ (left) and $B^+ \to \bar{D}^0\pi^+\pi^-\mu^+\nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

C. Charm branching ratios

To estimate the uncertainty due to the uncertainties on the branching ratios of the charm decays, we sample each charm branching ratio 10 000 times from a Gaussian distribution with mean and width that equal to the PDG central value and uncertainty [3]. It is assumed that the branching fractions for different D modes are independent. For each sampled set of D branching fractions, the new sum of branching fractions is calculated for the signal and normalization channels. The reconstruction efficiency is taken into account via the relative abundance of the modes. The ratio of the sums is calculated and the spread of the resulting distribution assigned as systematic uncertainty.

D. Signal $B \to \bar{D}^{**}\ell^+\nu_\ell$ composition

The signal PDF U shapes slightly vary for different intermediate D^{**} states. Therefore, the overall U shape depends on the D^{**} composition. To estimate the signal branching-fraction uncertainty due to the uncertainties in the D^{**} composition, we generate the U distribution using the template of one D^{**} state and then fit with the nominal signal template described in Sec. V B whose composition is

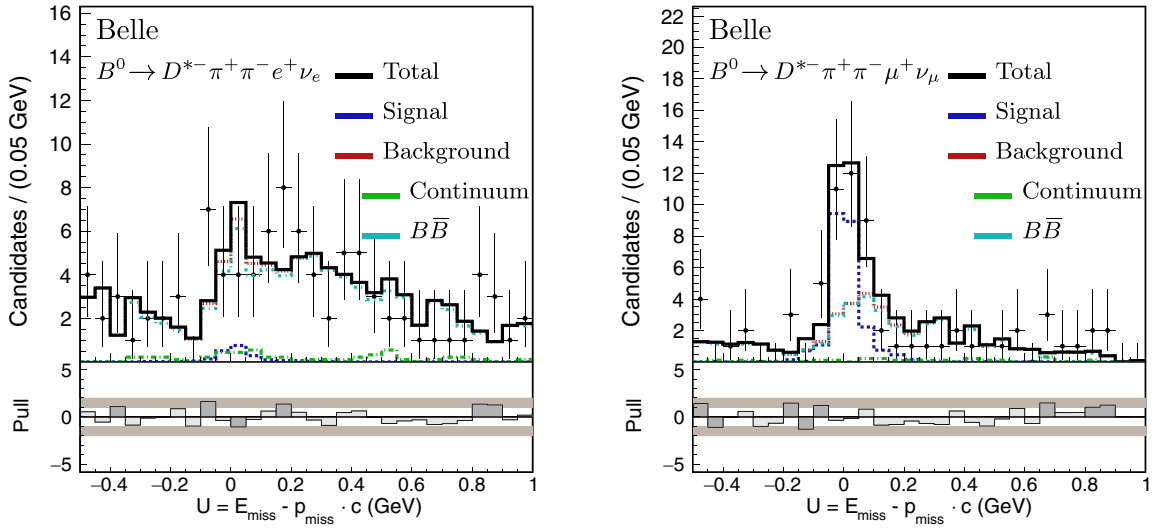

FIG. 13. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^0 \to D^{*-}\pi^+\pi^-e^+\nu_e$ (left) and $B^0 \to D^{*-}\pi^+\pi^-\mu^+\nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

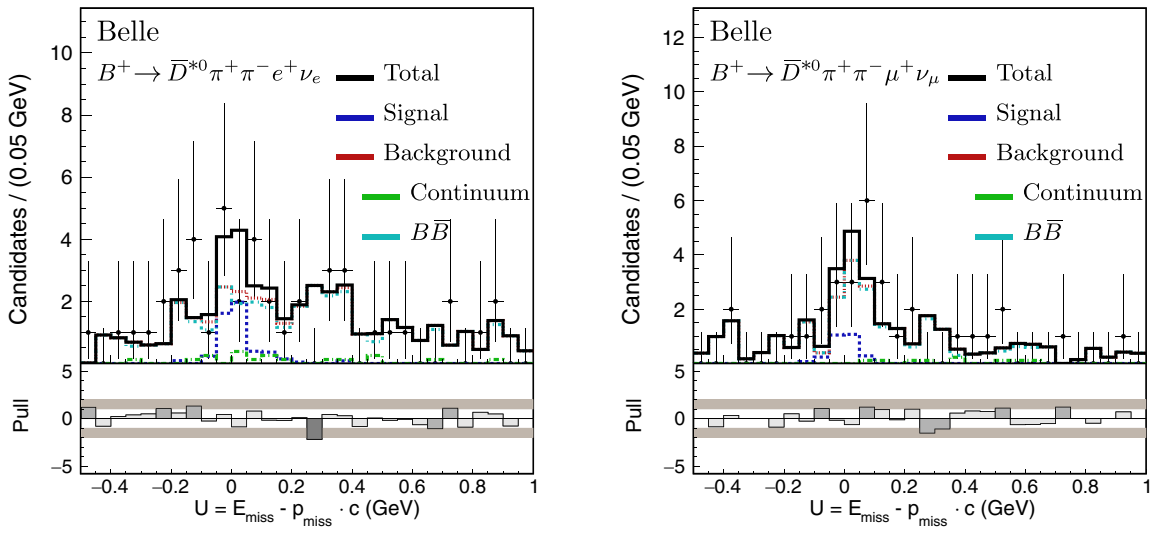

FIG. 14. Distribution of $E_{\rm miss} - p_{\rm miss}c$ of $B^+ \to \bar{D}^{*0}\pi^+\pi^- e^+\nu_e$ (left) and $B^+ \to \bar{D}^{*0}\pi^+\pi^-\mu^+\nu_\mu$ (right) for the data. The MC shapes, normalized according to the result of the fit, are also shown. The component labeled background is the sum of the continuum and $B\bar{B}$ components.

taken from Ref. [3]. The largest average difference between the generated and fitted signal yields among the tested D^{**} scenarios is assigned as a systematic uncertainty.

Also the signal efficiency differs depending on the assumed intermediate D^{**} state. Therefore, the integrated signal efficiency depends on the relative abundance of the D^{**} states. The uncertainty on the world average of the D^{**} composition is propagated to the signal efficiency and assigned as a systematic uncertainty.

E. Lepton PID

By using $\gamma\gamma \to \ell^+\ell^-$ processes, lepton PID efficiency factors in kinematic ranges of the momentum and polar angle have been calculated (Chapter 5.4 of Ref. [38]), which correct for the difference between the selection efficiency in data and MC. The systematic uncertainties on the PID efficiency factors account for the method itself and for a possible effect from a hadronic environment, which is determined using inclusive $B \to J/\psi X$ decays. To propagate the uncertainties to the branching fractions we sample lepton correction factors for each kinematic bin using a Gaussian around the nominal value with a width corresponding to the uncertainty of the correction factor. The average correction factor over all truth-matched signal events as well as the average correction factor over all truth-matched candidates of the normalization channels are calculated. The spread of the distribution of the ratio of the two means is taken as the systematic uncertainty due to lepton identification. This procedure is performed separately for each of the D^{**} states, and the largest uncertainty per B and $D^{(*)}$ mode among all $B \to \bar{D}^{**}\ell^+\nu_\ell$ modes is assigned as the systematic uncertainty.

TABLE V. Fitted $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ and $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ signal and background yields in the full Belle data sample. The quoted uncertainties are statistical only.

	Signal		Background	
	Electron mode	Muon mode	Electron mode	Muon mode
$B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$	570 ± 35	433 ± 34	2641 ± 80	2190 ± 78
$B^+ \to D^-\pi^+\ell^+\nu_\ell$	721 ± 32	569 ± 31	1329 ± 53	1302 ± 54
$B^0 \to \bar{D}^{*0}\pi^-\ell^+\nu_\ell$	798 ± 44	690 ± 43	77 ± 12	64 ± 11
$B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$	787 ± 35	634 ± 34	172 ± 19	242 ± 23
$B^0 \to D^-\pi^+\pi^-\ell^+\nu_\ell$	88 ± 14	58 ± 12	452 ± 26	271 ± 21
$B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$	196 ± 20	132 ± 18	852 ± 37	603 ± 33
$B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$	3 ± 10	41 ± 10	86 ± 11	41 ± 8
$B^+ \to \bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell$	57 ± 15	38 ± 14	37 ± 7	26 ± 6

TABLE VI. Relative systematic uncertainties (in %) in the determination of the $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ branching fractions.

	$B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$	$B^+ \to D^-\pi^+\ell^+\nu_\ell$	$B^0 \to \bar{D}^{*0}\pi^-\ell^+\nu_\ell$	$B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$
MC statistics: Fit model	1.1	0.7	1.1	0.7
MC statistics: Efficiency	0.6	0.5	0.7	0.7
Charm branching ratios	1.0	1.4	1.1	1.2
Signal D^{**} composition	0.8	0.8	0.9	1.1
Lepton PID	0.1	0.1	0.1	0.1
Charged hadron PID	0.3	1.5	0.2	2.0
Tracking efficiency	0.3	0.6	0.6	0.6
π^0 efficiency	0.1	0.4	0.3	0.2
$B \to \bar{D}^*\ell^+\nu_\ell/B \to \bar{D}^{**}\ell^+\nu_\ell$ form factors	0.5	0.4	0.3	0.5
Sum	1.9	2.6	2.1	3.5
$\mathcal{B}(B \to \bar{D}^*\ell^+\nu_\ell)$	2.4	3.9	2.4	3.9
Sum including $\mathcal{B}(B \to \bar{D}^*\ell^+\nu_\ell)$	3.1	4.7	3.2	5.3

TABLE VII. Relative systematic uncertainties (in %) in the determination of the $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ branching fractions.

	$B^0 \to D^-\pi^+\pi^-\ell^+\nu_\ell$	$B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$	$B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$	$B^+ \to \bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell$
MC statistics: Fit model	7.0	4.5	17.1	10.6
MC statistics: Efficiency	2.4	1.8	8.1	3.7
Charm branching ratios	1.4	0.9	1.0	0.9
BDT	3.9	4.0	5.5	4.2
Lepton PID	0.3	0.3	0.6	0.4
Charged hadron PID	2.4	3.6	1.6	1.6
Tracking efficiency	0.5	0.7	0.6	0.7
π^0 efficiency	0.4	0.2	0.4	0.3
$B \to \bar{D}^*\ell^+\nu_\ell$ form factors	0.1	0.1	0.1	0.1
Sum	8.9	7.3	19.8	12.2
$\mathcal{B}(B \to \bar{D}^*\ell^+\nu_\ell)$	2.4	3.9	2.4	3.9
Sum including $\mathcal{B}(B \to \bar{D}^*\ell^+\nu_\ell)$	9.2	8.3	20.0	12.8

F. Charged hadron PID

Similar to the study for the lepton PID, correction factors for the hadron PID selection requirements are sampled in bins of the momentum and polar angle to evaluate the systematic uncertainty on the branching fraction due to the uncertainties in the determination of the correction factors using inclusive D^* samples (Chapter 5.4 of Ref. [38]). The average correction factors of the signal and normalization samples are calculated, then divided by each other, and the spread of the resulting distribution of ratios is interpreted as the systematic uncertainty for the hadron PID. Similar to the lepton PID described above, the largest value over the possible D^{**} states is assigned as the final systematic uncertainty.

G. Tracking efficiency

For each signal and normalization mode the average track multiplicity over the various D modes is determined in simulation. The difference between the signal and normalization mode average track multiplicity is multiplied by 0.35% (Chapter 15.1.1.2 of Ref. [38]) and the result is taken as systematic uncertainty due to tracking efficiency differences between data and MC. For low-momentum tracks ($p_T < 200$ MeV/c) an additional tracking-related systematic uncertainty is calculated. Using a $B^0 \to D^{*-}\pi^+$ sample the slow pion efficiency is determined in six momentum bins for data and MC (Chapter 15.1.1.2 of Ref. [38]). The relative uncertainty of the ratio between the data and MC efficiencies is taken as systematic uncertainty due to low-momentum tracking. The two tracking-related systematic uncertainties are added in quadrature.

H. π^0 efficiency

The π^0 efficiency differs between data and MC. The effect is corrected in the calculation of the signal efficiency and the uncertainty on the ratio between the data and MC

efficiency of about 2.4% (Chapter 15.1.4 of Ref. [38]) is propagated to the systematic uncertainty of the branching fraction measurement. First, the average π^0 multiplicity for each signal and normalization mode is determined and the difference between the signal and normalization values is calculated. This difference is multiplied by the aforementioned uncertainty to obtain the systematic uncertainty due to the π^0 efficiency data-MC ratio.

I. $B \to \bar{D}^* \ell^+ \nu_\ell$ and $B \to \bar{D}^{**} \ell^+ \nu_\ell$ form factors

The $B \to \bar{D}^{**} \ell^+ \nu_\ell$ MC samples are generated with the ISGW2 model [29]. A more accurate description can be achieved with the LLSW model [16]. To estimate the systematic uncertainty due to using the ISGW2 model two-dimensional form factor weights in $\omega = \frac{m_B^2 + m_{D^{**}}^2 - q^2}{2 m_B m_{D^{**}}}$, with the masses of the B meson m_B and the D^{**} system $m_{D^{**}}$, and the four-momentum transfer squared to the lepton-neutrino system q^2, and the cosine of the angle between the charged lepton and the D meson $\cos\theta_l$ are determined. These weights are calculated separately for decays via D_0^*, D_1, D_1', and D_2^* mesons. The U distribution is generated using the nominal ISGW2-based templates and fit with signal and feeddown templates that are reweighted with the form factor weights described above. The average difference between the fitted and generated yields over 1000 iterations of generating and fitting is calculated and divided by the generated yield (f_{sig}).

Similarly, the simulation of the $B \to \bar{D}^{(*)} \ell^+ \nu_\ell$ modes is based on heavy quark effective theory (HQET) [28]. A reweighting in the momentum transfer and the momentum of the charged lepton is applied to account for outdated values of the CLN [40] form factor parameters ρ^2, R_1, and R_2. The U distribution is generated with the nominal HQET-based templates and fit with the reweighted templates. The difference between the fitted and generated yields divided by the generated yield is calculated (f_{norm}).

The difference of the ratio $f_{\text{sig}}/f_{\text{norm}}$ from unity is taken as the systematic uncertainty due to the form factors.

J. $\mathcal{B}$ $(B \to \bar{D}^* \ell^+ \nu_\ell)$

The PDG average of the branching ratio of the normalization mode $B^+ \to \bar{D}^{*0} \ell^+ \nu_\ell$ is $\mathcal{B}(B^+ \to \bar{D}^{*0} \ell^+ \nu_\ell) = (5.58 \pm 0.22) \times 10^{-2}$ [3], introducing a systematic uncertainty of 3.9%. The corresponding PDG average for the B^0 mode is $\mathcal{B}(B^0 \to D^{*-} \ell^+ \nu_\ell) = (4.97 \pm 0.12) \times 10^{-2}$ [3], which introduces a systematic uncertainty of 2.4%.

K. BDT

The BDT to suppress continuum background in $B \to \bar{D}^{(*)} \pi^+ \pi^- \ell^+ \nu_\ell$ is trained with signal MC and off-resonance data. Differences in the input variable distributions between the signal simulation and signal events in real data might introduce a bias in the calculation of the signal efficiency. To estimate the associated uncertainty, the BDT output is calculated for the cross-check and normalization modes $B \to \bar{D} \ell^+ \nu_\ell$ and $B \to \bar{D}^* \ell^+ \nu_\ell$. The same requirement on the BDT output as for the $D\pi\pi$ signal-candidate selection is applied for these $B \to \bar{D}^{(*)} \ell^+ \nu_\ell$ modes and the fit to the $B \to \bar{D}^{(*)} \ell^+ \nu_\ell$ sample described in Sec. V A is performed. The ratio between the $B \to \bar{D}^{(*)} \ell^+ \nu_\ell$ yield of this fit and the yield obtained without the BDT requirement is considered a data-based efficiency of the BDT requirement. This efficiency is compared with the signal MC efficiency of the $B \to \bar{D}^{(*)} \ell^+ \nu_\ell$ samples. The largest relative difference between the data- and MC-based efficiencies among the $B \to \bar{D} \ell^+ \nu_\ell$ and $B \to \bar{D}^* \ell^+ \nu_\ell$ values is taken as the BDT-related systematic uncertainty. This procedure assumes that the BDT, which uses variables of the B_{tag} meson reconstruction and event-shape variables, is mostly independent of the B_{sig} meson reconstruction.

VII. BRANCHING FRACTION RESULTS

The weighted average branching fraction ratios are calculated based on the total uncertainties. The calculation takes into account that some component uncertainties are correlated between the electron and muon mode.

TABLE VIII. Branching fraction ratio results and ratios between electron and muon decay modes with statistical and systematic uncertainties. The denominator for the branching fraction ratios is $B^0 \to D^{*-} \ell^+ \nu_\ell$ for the B^0 modes and $B^+ \to \bar{D}^{*0} \ell^+ \nu_\ell$ for the B^+ modes.

Decay mode	Branching fraction ratio [%]	e/μ ratio
$B^0 \to \bar{D}^0 \pi^- \ell^+ \nu_\ell$	$7.23 \pm 0.36(\text{stat}) \pm 0.14(\text{syst})$	$1.13 \pm 0.12(\text{stat})$
$B^+ \to D^- \pi^+ \ell^+ \nu_\ell$	$6.78 \pm 0.24(\text{stat}) \pm 0.18(\text{syst})$	$1.07 \pm 0.08(\text{stat})$
$B^0 \to \bar{D}^{*0} \pi^- \ell^+ \nu_\ell$	$11.10 \pm 0.48(\text{stat}) \pm 0.23(\text{syst})$	$0.98 \pm 0.08(\text{stat})$
$B^+ \to D^{*-} \pi^+ \ell^+ \nu_\ell$	$9.50 \pm 0.33(\text{stat}) \pm 0.34(\text{syst})$	$1.06 \pm 0.08(\text{stat})$
$B^0 \to D^- \pi^+ \pi^- \ell^+ \nu_\ell$	$2.91 \pm 0.37(\text{stat}) \pm 0.26(\text{syst})$	$1.18 \pm 0.31(\text{stat})$
$B^+ \to \bar{D}^0 \pi^+ \pi^- \ell^+ \nu_\ell$	$3.10 \pm 0.26(\text{stat}) \pm 0.22(\text{syst})$	$1.23 \pm 0.21(\text{stat})$
$B^0 \to D^{*-} \pi^+ \pi^- \ell^+ \nu_\ell$	$0.99 \pm 0.43(\text{stat}) \pm 0.20(\text{syst})$	$0.06 \pm 0.21(\text{stat})$
$B^+ \to \bar{D}^{*0} \pi^+ \pi^- \ell^+ \nu_\ell$	$1.25 \pm 0.27(\text{stat}) \pm 0.15(\text{syst})$	$1.1 \pm 0.5(\text{stat})$

The results and the ratios between the electron and muon mode branching fractions are listed in Table VIII.

The results are the most precise determinations of these branching fraction ratios to date (except for $B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$). All values are compatible with the previous world averages. The electron and muon values are compatible with each other within one standard deviation apart from those for $B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$. The p-value of the hypothesis that the latter are compatible is 0.5%.[1]

The branching fraction ratios are converted into absolute branching fractions by multiplying them with the branching fraction of $B \to \bar{D}^*\ell^+\nu_\ell$. The results are listed in Table IX.

TABLE IX. Branching fraction results with statistical and systematic uncertainties.

Decay mode	Branching fraction [%]
$B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$	$0.360 \pm 0.018(\text{stat}) \pm 0.011(\text{syst})$
$B^+ \to D^-\pi^+\ell^+\nu_\ell$	$0.378 \pm 0.013(\text{stat}) \pm 0.018(\text{syst})$
$B^0 \to \bar{D}^{*0}\pi^-\ell^+\nu_\ell$	$0.551 \pm 0.024(\text{stat}) \pm 0.017(\text{syst})$
$B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$	$0.530 \pm 0.019(\text{stat}) \pm 0.028(\text{syst})$
$B^0 \to D^-\pi^+\pi^-\ell^+\nu_\ell$	$0.145 \pm 0.018(\text{stat}) \pm 0.013(\text{syst})$
$B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$	$0.173 \pm 0.014(\text{stat}) \pm 0.014(\text{syst})$
$B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$	$0.049 \pm 0.021(\text{stat}) \pm 0.010(\text{syst})$
$B^+ \to \bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell$	$0.070 \pm 0.015(\text{stat}) \pm 0.009(\text{syst})$

VIII. EXCLUSIVE $B \to \bar{D}^{**}\ell^+\nu_\ell$ BRANCHING FRACTIONS

Using the sPlot technique [41] with the implementation of Ref. [42], signal weights are assigned to each event based on the fit to the U distribution. This allows the background contribution to the $m(D\pi)$, $m(D^*\pi)$, and $m(D\pi\pi)$ distributions to be statistically subtracted, and the signal-only distribution to be studied. For all fit components only very small values ($< 5\%$) for the correlations between the U distribution and the invariant masses are found, which is a requirement for the validity of the sPlot technique. We perform weighted unbinned maximum likelihood fits to the invariant mass distributions. The uncertainty calculation is based on Ref. [43].

For the $B \to \bar{D}\pi\ell^+\nu_\ell$ modes the PDG reports decays via the D_0^* and D_2^* resonances. These two contributions are parametrized with Breit-Wigner functions that are convolved with a Gaussian distribution. The width of the Gaussian is fixed from simulations to 3.4 MeV/c^2. The peak position and width of the D_0^* and D_2^* resonances are allowed to float in the fit. However, they are constrained within Gaussian distributions using their world averages and corresponding uncertainties [3]. In a second fit the peak positions and widths are fixed to the results from the first fit. The difference in the statistical uncertainties between the two fits is used to single out the uncertainty introduced by the Gaussian constraint. It is interpreted as a systematic uncertainty. The weighted $m(D\pi)$ distribution (see Fig. 15) shows that a third component must be added to the fit model. Here, we choose an exponential distribution. The yields, which are listed in Table X, are converted into branching fractions using Eq. (1). The statistical uncertainty is extracted directly from the fit, while the systematic uncertainty is the sum in quadrature of the relative uncertainties of the inclusive branching fractions reported in Table IX and the uncertainties introduced by the limited knowledge of the D^{**} peak positions and width described above. In the fit to the $m(\bar{D}^0\pi^-)$ distribution the yield of the D_0^{*-} component is compatible with zero. Therefore, instead of calculating a branching fraction, an upper limit at 90% confidence level (CL) is set. We create 2000 new data samples by bootstrapping [44] the original data (randomly selecting events, each with its corresponding weight, while allowing repetition of the events). The $\bar{D}^0\pi^-$ mass fit is performed for each sample. The 90% CL upper limit on the yield is the value that is higher than that found in 90% of the samples in which a positive D_0^{*-} yield is obtained. This yield is then converted into the upper limit.

Three D^{**} resonances are known for the $D^*\pi$ final state, D_1, D_1', and D_2^*. The three components are parametrized with Breit-Wigner functions convolved with a Gaussian. The shape parameters of the two narrow resonances D_1 and D_2^* are constrained within Gaussian distributions to their world averages [3], while the peak position and width of the broad D_1' resonance is fixed to its world average. Instead of fitting the $m(D^*\pi)$ mass directly the invariant mass of the D^* is subtracted. This allows us to conveniently incorporate the feeddown component as well. By subtracting the invariant mass of the D meson from $m(D\pi)$ the peaks align. We perform the fit in the range 0.2 to 0.8 GeV/c^2. The data and the overlaid fit projections are shown in Fig. 16. The yields of the three components and the resulting branching fractions are listed in Table XI. The systematic uncertainty is dominated by the shape uncertainties. It is determined by fitting twice, once with the shape parameters floating and once fixed.

The weighted unbinned maximum likelihood fit to the $m(D\pi\pi)$ distribution is performed in the range 2.15 to 5 GeV/c^2 (see Fig. 17). Initially, the fit model consists of a single Gaussian and a first-order polynomial. The fitted peak position and width are compatible with the D_1 resonance for the B^0 and B^+ modes. Therefore, the Gaussian component is interpreted as $B^0 \to D_1^-\ell^+\nu_\ell$ with $D_1^- \to D^-\pi^+\pi^-$ and $B^+ \to \bar{D}_1^0\ell^+\nu_\ell$ with $\bar{D}_1^0 \to \bar{D}^0\pi^+\pi^-$, respectively. The peaking component is replaced with a Breit-Wigner function convolved with a Gaussian.

[1]The deviation of the ratios from unity cannot naively be interpreted in terms of standard deviations.

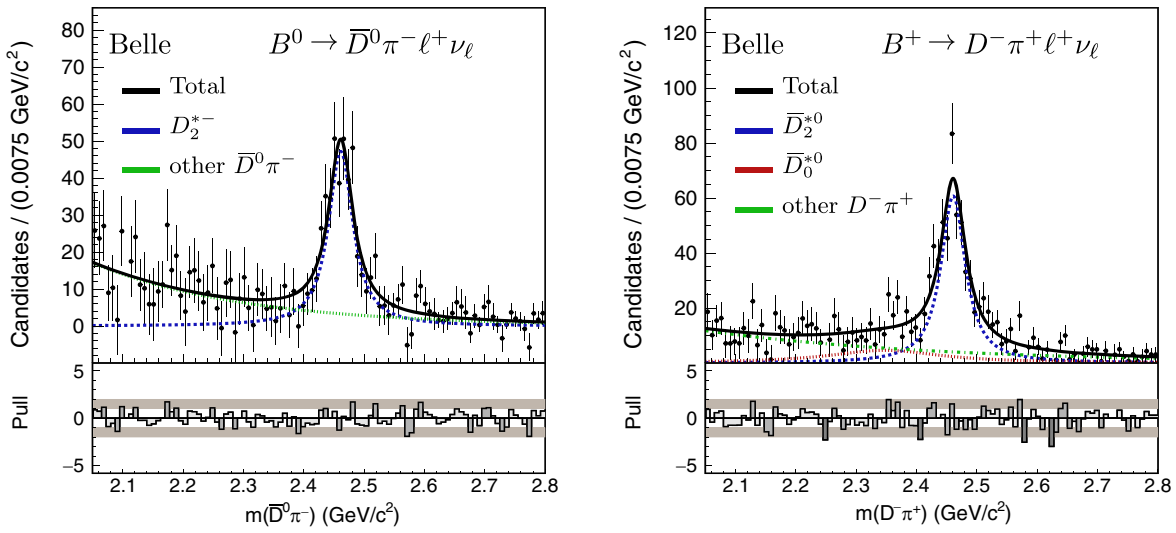

FIG. 15. Invariant $m(D\pi)$ mass distribution of $B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$ (left) and $B^+ \to D^-\pi^+\ell^+\nu_\ell$ (right) reconstruction after applying signal weights determined from a fit of the U distribution using the sPlot technique.

The shape parameters of the Breit-Wigner are set to the PDG values, but allowed to float within a Gaussian constraint. We find 103 ± 13 events for the B^0 mode and 197 ± 20 events for the B^+ mode. By comparing the log-likelihood with a fit, in which the D_1 yield is fixed to zero, the statistical significance is determined to be 17.3 for the B^0 mode and 25.1 for the B^+ mode. The remaining signal events (42 ± 13 events in the B^0 mode and 131 ± 20 events in the B^+ mode), which are parametrized with the polynomial, can either be a nonresonant decay process or a decay via a very broad resonance, such as the D_0^* or D_1'. However, with our statistical power we can only state that there must be at least one additional process besides the decay via the D_1 resonance, but cannot characterize it further. The D_1 yields are converted into the following branching fractions:

$$\mathcal{B}(B^0 \to D_1^-\ell^+\nu_\ell) \times \mathcal{B}(D_1^- \to D^-\pi^+\pi^-) = (0.102 \pm 0.013(\text{stat}) \pm 0.009(\text{syst}))\% \quad (9)$$

TABLE X. Fitted D^{**} yields, statistical significances, and branching fractions for the $D\pi$ final state. The statistical significance is calculated as $\mathcal{S} = \sqrt{2\Delta\mathcal{L}}$, where $\Delta\mathcal{L}$ is the difference between the log-likelihood value of the nominal fit and of a fit with the signal yield fixed to zero.

	Yield	$\mathcal{S}$	Branching fraction [%]
$B^0 \to D_0^{*-}\ell^+\nu_\ell$ with $D_0^{*-} \to \bar{D}^0\pi^-$	⋯	⋯	< 0.044 at 90% CL
$B^0 \to D_2^{*-}\ell^+\nu_\ell$ with $D_2^{*-} \to \bar{D}^0\pi^-$	457 ± 45	25.2	$0.157 \pm 0.015(\text{stat}) \pm 0.005(\text{syst})$
other $B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$	547 ± 45	not applicable	⋯
$B^+ \to \bar{D}_0^{*0}\ell^+\nu_\ell$ with $\bar{D}_0^{*0} \to D^-\pi^+$	180 ± 72	3.9	$0.054 \pm 0.022(\text{stat}) \pm 0.005(\text{syst})$
$B^+ \to \bar{D}_2^{*0}\ell^+\nu_\ell$ with $\bar{D}_2^{*0} \to D^-\pi^+$	590 ± 39	24.9	$0.163 \pm 0.011(\text{stat}) \pm 0.008(\text{syst})$
other $B^+ \to D^-\pi^+\ell^+\nu_\ell$	520 ± 70	not applicable	⋯

TABLE XI. Fitted D^{**} yields, statistical significances, and branching fractions for the $D^*\pi$ final state.

	Yield	$\mathcal{S}$	Branching fraction [%]
$B^0 \to D_1^-\ell^+\nu_\ell$ with $D_1^- \to \bar{D}^{*0}\pi^-$	866 ± 142	25.3	$0.306 \pm 0.050(\text{stat}) \pm 0.029(\text{syst})$
$B^0 \to D_1'^-\ell^+\nu_\ell$ with $D_1'^- \to \bar{D}^{*0}\pi^-$	523 ± 173	17.3	$0.206 \pm 0.068(\text{stat}) \pm 0.025(\text{syst})$
$B^0 \to D_2^{*-}\ell^+\nu_\ell$ with $D_2^{*-} \to \bar{D}^{*0}\pi^-$	145 ± 114	4.4	$0.051 \pm 0.040(\text{stat}) \pm 0.010(\text{syst})$
$B^+ \to \bar{D}_1^0\ell^+\nu_\ell$ with $\bar{D}_1^0 \to D^{*-}\pi^+$	698 ± 65	24.2	$0.249 \pm 0.023(\text{stat}) \pm 0.015(\text{syst})$
$B^+ \to \bar{D}_1'^0\ell^+\nu_\ell$ with $\bar{D}_1'^0 \to D^{*-}\pi^+$	353 ± 93	13.3	$0.138 \pm 0.036(\text{stat}) \pm 0.009(\text{syst})$
$B^+ \to \bar{D}_2^{*0}\ell^+\nu_\ell$ with $\bar{D}_2^{*0} \to D^{*-}\pi^+$	382 ± 74	11.8	$0.137 \pm 0.026(\text{stat}) \pm 0.009(\text{syst})$

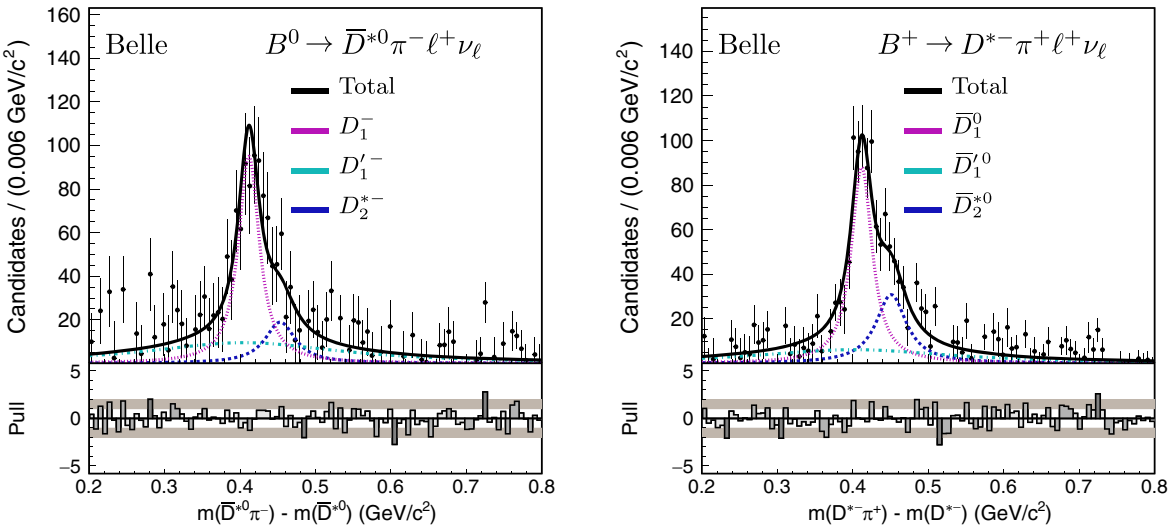

FIG. 16. Distribution of the mass difference $m(D^*\pi) - m(D^*)$ of $B^0 \to \bar{D}^{*0}\pi^-\ell^+\nu_\ell$ (left) and $B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$ (right) reconstruction after applying signal weights determined from a fit of the U distribution using the sPlot technique.

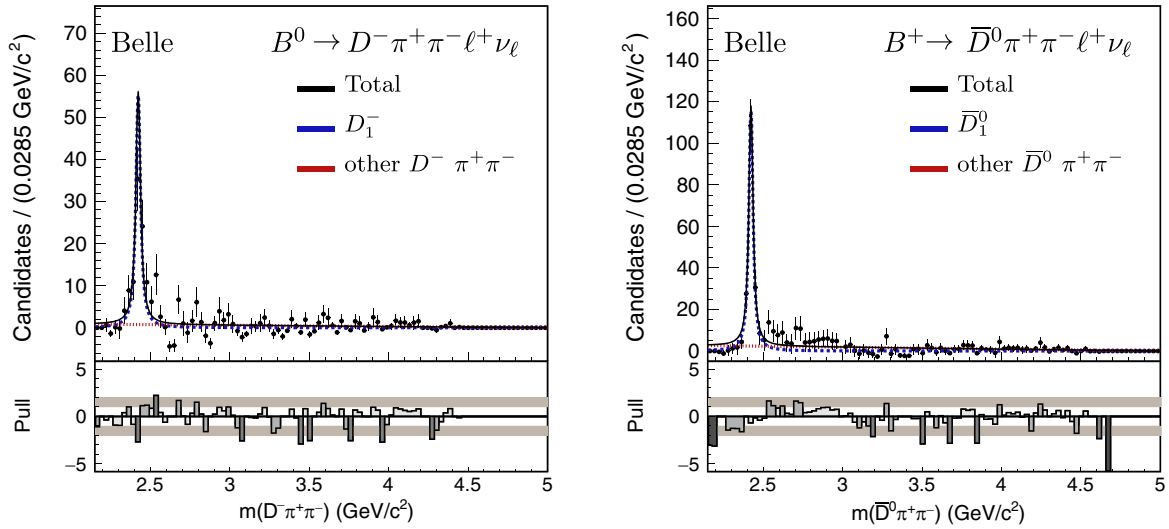

FIG. 17. Invariant $m(D\pi\pi)$ mass distribution of $B^0 \to D^-\pi^+\pi^-\ell^+\nu_\ell$ (left) and $B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$ (right) reconstruction after applying signal weights determined from a fit of the U distribution using the sPlot technique. A fit function consisting of a Gaussian and a first-order polynomial is overlaid.

$$\mathcal{B}(B^+ \to \bar{D}_1^0\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_1^0 \to \bar{D}^0\pi^+\pi^-) = (0.105 \pm 0.011(\text{stat}) \pm 0.009(\text{syst}))\% \quad (10)$$

This is the first observation of these decay modes.

IX. CONCLUSION

In conclusion, using hadronic tagging, we have measured the $B \to \bar{D}^{(*)}\pi\ell^+\nu_\ell$ and $B \to \bar{D}^{(*)}\pi^+\pi^-\ell^+\nu_\ell$ branching fractions, achieving the highest precision to date (except for $B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$). These results were obtained from a data sample that contains $772 \times 10^6 B\bar{B}$ pairs collected near the $\Upsilon(4S)$ resonance with the Belle detector at the KEKB asymmetric energy e^+e^- collider. All values are compatible with the previous world averages, but tend to be slightly lower. Therefore, the gap between the sum of the exclusive semileptonic B decays to charm final states and the inclusive rate has slightly increased and increased in significance. Furthermore, the mass spectra of the hadronic final state particles were studied after statistically subtracting the background contributions.

The results for the $B \to \bar{D}\pi\ell^+\nu_\ell$ decays via the D_2^* resonance are compatible with the world averages. They constitute the most precise measurements of these branching fractions to date. The value for $\mathcal{B}(B^+ \to \bar{D}_0^{*0}\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_0^{*0} \to D^-\pi^+)$ is significantly smaller than previous measurements. This applies even more so to the B^0 mode, where no contribution could be found in this analysis. Instead, an unknown new $D\pi$ signal component with a significant contribution is found, whose mass distribution follows a falling exponential function. The results for the $B \to \bar{D}^*\pi\ell^+\nu_\ell$ decays via the narrower D_1 and D_2^* resonances are compatible with previous measurements and the world averages. For the decay via the wider D_1' resonance the branching fractions are measured to be 35% (50%) lower than the world average in the B^0 (B^+) mode. Additionally, we have performed the first observations of $B \to \bar{D}_1\ell^+\nu_\ell$ with $D_1 \to D\pi^+\pi^-$.

ACKNOWLEDGMENTS

This work, based on data collected using the Belle detector, which was operated until June 2010, was supported by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, the Japan Society for the Promotion of Science (JSPS), and the Tau-Lepton Physics Research Center of Nagoya University; the Australian Research Council including Grants No. DP180102629, No. DP170102389, No. DP170102204, No. DE220100462, No. DP150103061, No. FT130100303; Austrian Federal Ministry of Education, Science and Research (FWF) and FWF Austrian Science Fund No. P 31361-N36; the National Natural Science Foundation of China under Contracts No. 11675166, No. 11705209, No. 11975076, No. 12135005, No. 12175041, No. 12161141008; Key Research Program of Frontier Sciences, Chinese Academy of Sciences (CAS), Grant No. QYZDJ-SSW-SLH011; the Ministry of Education, Youth and Sports of the Czech Republic under Contract No. LTT17020; the Czech Science Foundation Grant No. 22-18469S; Horizon 2020 ERC Advanced Grant No. 884719 and ERC Starting Grant No. 947006 "InterLeptons" (European Union); the Carl Zeiss Foundation, the Deutsche Forschungsgemeinschaft, the Excellence Cluster Universe, and the VolkswagenStiftung; the Department of Atomic Energy (Project Identification No. RTI 4002) and the Department of Science and Technology of India; BSF and ISF (Israel); the Istituto Nazionale di Fisica Nucleare of Italy; National Research Foundation (NRF) of Korea Grants No. 2016R1D1A1B02012900, No. 2018R1A2B3003643, No. 2018R1A6A1A06024970, No. RS202200197659, No. 2019R1I1A3A01058933, No. 2021R1A6A1A03043957, No. 2021R1F1A1060423, No. 2021R1F1A1064008, No. 2022R1A2C1003993; Radiation Science Research Institute, Foreign Large-size Research Facility Application Supporting project, the Global Science Experimental Data Hub Center of the Korea Institute of Science and Technology Information and KREONET/Global Ring Network for Advanced Application Development (GLORIAD); the Polish Ministry of Science and Higher Education and the National Science Center; the Ministry of Science and Higher Education of the Russian Federation, Agreement No. 14.W03.31.0026, and the HSE University Basic Research Program, Moscow; University of Tabuk research Grants No. S-1440-0321, No. S-0256-1438, and No. S-0280-1439 (Saudi Arabia); the Slovenian Research Agency Grants No. J1-9124 and No. P1-0135; Ikerbasque, Basque Foundation for Science, Spain; the Swiss National Science Foundation; the Ministry of Education and the Ministry of Science and Technology of Taiwan; and the United States Department of Energy and the National Science Foundation. 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 KEKB group for the excellent operation of the accelerator; the KEK cryogenics group for the efficient operation of the solenoid; and the KEK computer group and the Pacific Northwest National Laboratory (PNNL) Environmental Molecular Sciences Laboratory (EMSL) computing group for strong computing support; and the National Institute of Informatics, and Science Information NETwork 6 (SINET6) for valuable network support.

APPENDIX: ADDITIONAL FIT RESULTS

TABLE XII. Fitted mean and width of Gaussian used to smear signal U templates.

	Mean [MeV]		σ [MeV]	
	Electron mode	Muon mode	Electron mode	Muon mode
$B^0 \to D^-\ell^+\nu_\ell$	2.7 ± 1.0	1.4 ± 0.9	7.7 ± 2.1	9.7 ± 2.1
$B^+ \to \bar{D}^0\ell^+\nu_\ell$	0.7 ± 0.7	-0.3 ± 0.7	10.7 ± 1.8	9.7 ± 2.5
$B^0 \to D^{*-}\ell^+\nu_\ell$	1.3 ± 0.8	1.4 ± 0.8	9.3 ± 3.3	10.5 ± 1.7
$B^+ \to \bar{D}^{*0}\ell^+\nu_\ell$	-0.4 ± 1.8	-0.1 ± 1.5	22.5 ± 3.2	16.8 ± 2.8
$B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$	9.9 ± 0.8	2.4 ± 2.9	0.5 ± 0.5	4 ± 6
$B^+ \to D^-\pi^+\ell^+\nu_\ell$	5.4 ± 2.4	3.4 ± 3.3	13 ± 5	23 ± 6
$B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$	5.1 ± 2.8	7.1 ± 3.0	15 ± 5	15 ± 5

TABLE XIII. Relative contribution (R) and purity (p) of D^0 decay modes in the data (both in %).

	$B^0 \to \bar{D}^0\pi^-\ell^+\nu_\ell$		$B^+ \to \bar{D}^0\pi^+\pi^-\ell^+\nu_\ell$		$B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$		$B^0 \to \bar{D}^{*0}\pi^-\ell^+\nu_\ell$		$B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$	$B^+ \to \bar{D}^{*0}\pi^+\pi^-\ell^+\nu_\ell$
	R	p	R	p	R	p	R	p	R	R
$D^0 \to K^-\pi^+$	13.5	15	19.3	19	24.7	67	17.8	62	35.0	24.0
$D^0 \to K^-\pi^+\pi^0$	28.1	6	36.4	16	28.0	68	22.3	29	27.9	28.0
$D^0 \to K^-\pi^+\pi^+\pi^-$	26.7	9	35.5	14	26.3	55	30.9	23	23.6	38.7
$D^0 \to K^0_S\pi^+\pi^-$	7.8	7	6.2	14	6.3	66	6.1	22	4.3	8.0
$D^0 \to K^-K^+$	2.4	9	2.7	21	2.5	71	2.1	55	2.1	1.3
$D^0 \to K^0_S\pi^0$	0.9	21	…	…	…	…	…	…	…	…
$D^0 \to K^0_S\pi^+\pi^-\pi^0$	14.7	4	…	…	4.3	39	…	…	…	…
$D^0 \to \pi^+\pi^-$	5.9	4	…	…	1.4	31	…	…	…	…
$D^0 \to K^-\pi^+\pi^-\pi^+\pi^0$	…	…	…	…	6.7	34	…	…	…	…
$D^0 \to \pi^+\pi^-\pi^0$	…	…	…	…	…	…	20.7	0	7.1	…

TABLE XIV. Relative contribution (R) and purity (p) of D^+ decay modes in the data (both in %).

	$B^+ \to D^-\pi^+\ell^+\nu_\ell$		$B^0 \to D^-\pi^+\pi^-\ell^+\nu_\ell$		$B^+ \to D^{*-}\pi^+\ell^+\nu_\ell$		$B^0 \to D^{*-}\pi^+\pi^-\ell^+\nu_\ell$
	R	p	R	p	R	p	R
$D^+ \to K^0_S\pi^+$	5.9	29	5.4	14	3.8	39	4.8
$D^+ \to K^0_S\pi^+\pi^-\pi^+$	13.1	7	12.9	10	…	…	4.8
$D^+ \to K^-\pi^+\pi^+$	52.5	25	38.8	27	40.4	48	66.7
$D^+ \to K^-K^+\pi^+$	11.7	12	13.2	7	7.2	32	14.3
$D^+ \to K^-\pi^+\pi^+\pi^0$	…	…	…	…	28.4	6	…
$D^+ \to K^0_S\pi^+\pi^0$	15.3	12	17.1	5	5.3	68	9.5
$D^+ \to K^0_SK^+$	1.5	30	2.1	20	…	…	…
$D^+ \to \pi^+\pi^0$	…	…	10.4	0	…	…	…
$D^+ \to \pi^+\pi^-\pi^+$	…	…	…	…	14.9	0	…

TABLE XV. Assumptions on D^{**} modeling [3].

Decay mode	Resonance	Branching fraction $[10^{-3}]$	Relative abundance
$B^0 \to \bar{D}^0 \pi^- \ell^+ \nu_\ell$	D_0^{*-}	3.0 ± 1.2	0.71
	D_2^{*-}	1.21 ± 0.33	0.29
$B^+ \to D^- \pi^+ \ell^+ \nu_\ell$	D_0^{*0}	2.5 ± 0.5	0.62
	D_2^{*0}	1.53 ± 0.16	0.38
$B^0 \to \bar{D}^{*0} \pi^- \ell^+ \nu_\ell$	D_1^-	2.80 ± 0.28	0.425
	$D_1'^-$	3.1 ± 0.9	0.47
	D_2^{*-}	0.68 ± 0.12	0.105
$B^+ \to D^{*-} \pi^+ \ell^+ \nu_\ell$	D_1^0	3.03 ± 0.20	0.45
	$D_1'^0$	2.7 ± 0.6	0.40
	D_2^{*0}	1.01 ± 0.24	0.15

[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. L. Workman *et al.* (Particle Data Group), Review of particle physics, Prog. Theor. Exp. Phys. **2022**, 083C01 (2022).

[4] Y. Amhis *et al.*, Averages of *b*-hadron, *c*-hadron, and τ-lepton properties as of 2021, Phys. Rev. D **107**, 052008 (2023).

[5] J. Lees *et al.* (*BABAR* Collaboration), Measurement of an excess of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ decays and implications for charged Higgs bosons, Phys. Rev. D **88**, 072012 (2013).

[6] J. Lees *et al.* (*BABAR* Collaboration), Evidence for an Excess of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ Decays, Phys. Rev. Lett. **109**, 101802 (2012).

[7] M. Huschle *et al.* (Belle Collaboration), Measurement of the branching ratio of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)} \ell^- \bar{\nu}_\ell$ decays with hadronic tagging at Belle, Phys. Rev. D **92**, 072014 (2015).

[8] S. Hirose *et al.* (Belle Collaboration), Measurement of the τ Lepton Polarization and $R(D^*)$ in the Decay $B \to D^* \tau^- \bar{\nu}_\tau$, Phys. Rev. Lett. **118**, 211801 (2017).

[9] G. Caria *et al.* (Belle Collaboration), Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a Semileptonic Tagging Method, Phys. Rev. Lett. **124**, 161803 (2020).

[10] R. Aaij *et al.* (LHCb Collaboration), Measurement of the Ratio of Branching Fractions $\mathcal{B}(\bar{B}^0 \to D^{*+} \tau^- \bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+} \mu^- \bar{\nu}_\mu)$, Phys. Rev. Lett. **115**, 111803 (2015); **115**, 159901(E) (2015).

[11] R. Aaij *et al.* (LHCb Collaboration), Test of lepton flavor universality by the measurement of the $B^0 \to D^{*-} \tau^+ \nu_\tau$ branching fraction using three-prong τ decays, Phys. Rev. D **97**, 072013 (2018).

[12] N. Isgur and M. B. Wise, Spectroscopy with Heavy Quark Symmetry, Phys. Rev. Lett. **66**, 1130 (1991).

[13] D. Liventsev *et al.* (Belle Collaboration), Study of $B \to \bar{D}^{**} \ell^+ \nu_\ell$ with full reconstruction tagging, Phys. Rev. D **77**, 091503 (2008).

[14] B. Aubert *et al.* (*BABAR* Collaboration), Measurement of Semileptonic B Decays into Orbitally-Excited Charmed Mesons, Phys. Rev. Lett. **103**, 051803 (2009).

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

[16] A. K. Leibovich, Z. Ligeti, I. W. Stewart, and M. B. Wise, Semileptonic *B* decays to excited charmed mesons, Phys. Rev. D **57**, 308 (1998).

[17] A. Le Yaouanc, J. P. Leroy, and P. Roudeau, A model for NL and SL decays by $\bar{B}^0 \to D^{**}$ transitions with $\mathcal{B}(j = 1/2) \ll \mathcal{B}(j = 3/2)$ using the *LLSW* scheme, Phys. Rev. D **105**, 013004 (2022).

[18] A. Vossen *et al.* (Belle Collaboration), Measurement of the branching fraction of $B \to \bar{D}^{(*)} \pi \ell^+ \nu_\ell$ at Belle using hadronic tagging in fully reconstructed events, Phys. Rev. D **98**, 012005 (2018).

[19] J. Lees *et al.* (*BABAR* Collaboration), Observation of $\bar{B} \to D^{(*)} \pi^+ \pi^- \ell^- \bar{\nu}$ Decays in $e^+ e^-$ Collisions at the $\Upsilon(4S)$ Resonance, Phys. Rev. Lett. **116**, 041801 (2016).

[20] T. Keck, F. Abudinén, F. Bernlochner *et al.*, The Full Event Interpretation, Comput. Softw. Big Sci. **3**, 6 (2019).

[21] A. Abashian *et al.* (Belle Collaboration), The Belle detector, Nucl. Instrum. Methods Phys. Res., Sect. A **479**, 117 (2002).

[22] K. Hanagaki, H. Kakuno, H. Ikeda, T. Iijima, and T. Tsukamoto, Electron identification in Belle, Nucl. Instrum. Methods Phys. Res., Sect. A **485**, 490 (2002).

[23] A. Abashian *et al.* (Belle Collaboration), Muon identification in the Belle experiment at KEKB, Nucl. Instrum. Methods Phys. Res., Sect. A **491**, 69 (2002).

[24] S. Kurokawa and E. Kikutani, Overview of the KEKB accelerators, Nucl. Instrum. Methods Phys. Res., Sect. A **499**, 1 (2003).

[25] D. J. Lange, The EvtGen particle decay simulation package, Nucl. Instrum. Methods Phys. Res., Sect. A **462**, 152 (2001).
[26] R. Brun, F. Bruyant, M. Maire, A. McPherson, and P. Zanarini, GEANT 3: User's guide, Report No. CERN-DD-EE-84-01, CERN, 1987.
[27] P. Golonka and Z. Was, PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays, Eur. Phys. J. C **45**, 97 (2006).
[28] K. Abe *et al.* (Belle Collaboration), Measurement of B ($\bar{B}^0 \to D^+\ell^-\bar{\nu}$) and determination of $|V_{cb}|$, Phys. Lett. B **526**, 258 (2002).
[29] D. Scora and N. Isgur, Semileptonic meson decays in the quark model: An update, Phys. Rev. D **52**, 2783 (1995).
[30] M. Gelb, T. Keck, M. Prim, H. Atmacan, J. Gemmler, R. Itoh, B. Kronenbitter, T. Kuhr, M. Lubej, F. Metzner, C. Park, S. Park, C. Pulvermacher, M. Ritter, and A. Zupanc, B2BII: Data Conversion from Belle to Belle II, Comput. Softw. Big Sci. **2**, 9 (2018).
[31] T. Kuhr, C. Pulvermacher, M. Ritter *et al.*, The Belle II Core Software, Comput. Softw. Big Sci. **3**, 1 (2019).
[32] Belle II Collaboration, Belle II Analysis Software Framework (BASF2), 10.5281/zenodo.5574115.
[33] E. Nakano, Belle PID, Nucl. Instrum. Methods Phys. Res., Sect. A **494**, 402 (2002).
[34] Throughout this paper, the inclusion of the charge conjugate mode decay is implied unless otherwise stated.
[35] J. F. Krohn *et al.* (Belle-II Analysis Software Group), Global decay chain vertex fitting at Belle II, Nucl. Instrum. Methods Phys. Res., Sect. A **976**, 164269 (2020).
[36] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, Boosted decision trees as an alternative to artificial neural networks for particle identification, Nucl. Instrum. Methods Phys. Res., Sect. A **543**, 577 (2005).
[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] Ed. A. J. Bevan, B. Golob, Th. Mannel, S. Prell, and B. D. Yabsley, The Physics of the B Factories, Eur. Phys. J. C **74**, 3026 (2014).
[39] D. M. Asner *et al.* (CLEO Collaboration), Search for exclusive charmless hadronic B decays, Phys. Rev. D **53**, 1039 (1996).
[40] I. Caprini, L. Lellouch, and M. Neubert, Dispersive bounds on the shape of $B \to \bar{D}^*\ell^+\nu_\ell$ form-factors, Nucl. Phys. **B530**, 153 (1998).
[41] M. Pivk and F. R. Le Diberder, ${}_sPlot$: A statistical tool to unfold data distributions, Nucl. Instrum. Methods Phys. Res., Sect. A **555**, 356 (2005).
[42] H. Dembinski, M. Kenzie, C. Langenbruch, and M. Schmelling, Custom Orthogonal Weight functions (COWs) for event classification, Nucl. Instrum. Methods Phys. Res., Sect. A **1040**, 167270 (2022).
[43] C. Langenbruch, Parameter uncertainties in weighted unbinned maximum likelihood fits, Eur. Phys. J. C **82**, 393 (2022).
[44] B. Efron, Bootstrap methods: Another look at the jackknife, Ann. Stat. **7**, 1 (1979).