Search for lepton flavor universality violation in $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$ decays using hadronic B tagging at the Belle II experiment

Belle II実験におけるハドロニックタグ手法を用いた
$\overline{B} \to D^* \tau^- \overline{\nu}_\tau$崩壊でのレプトンフレーバー普遍性の破れの探索

Kazuki Kojima

Division of Particle and Astrophysical Science,
Graduate School of Science, Nagoya University

March, 2024

Abstract

In the standard model of particle physics (SM), lepton flavor universality is an axiom that states that the coupling constants in the interaction between leptons and gauge particles are the same regardless of the lepton flavor e, μ, or τ. This universality has been demonstrated by various experiments. On the other hand, in the semi-leptonic B decay, the ratio of the branching fractions $R(D^{(*)}) = \mathcal{B}(\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{B} \to D^{(*)}\ell^-\overline{\nu}_\ell)$, where ℓ represents an electron or muon, show a tensions with the SM expectations at a significance of 3.2σ. This deviation could be a sign of physics beyond the SM (BSM).

We search for lepton flavor universality violation by measuring $R(D^*)$ using 189 fb^{-1} data collected between 2019 and 2021 at the Belle II experiment. We find

$$R(D^*) = 0.262\ ^{+0.041}_{-0.039}\ (\text{stat.})\ ^{+0.033}_{-0.032}\ (\text{syst.}).$$

This result is consistent with the SM predictions. Therefore, no significant violation of lepton flavor universality is observed. The results also agree with the previous measurements within the uncertainty. The world average of the $R(D^{(*)})$ measurements, including our measurement, shows a slight increase in the deviation from the SM prediction from 3.2σ to 3.3σ. Finally, we discuss possible BSM contributions in $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ based on the new world average of $R(D^{(*)})$ and prospects of sensitivity on $R(D^*)$ at the Belle II experiment.

Contents

1 Introduction

1.1 The Standard Model

Elementary particles are the most fundamental particles of matter in the universe. The standard model of particle physics (SM) is a basic theory that describes the fundamental particles and forces acting between them. The model is based on the principle of gauge symmetry. It is founded on the local gauge invariance of the mathematical group $SU(3) \times SU(2) \times U(1)_Y$. This invariance is maintained by introducing gauge fields that absorb phase differences in the complex fields, denoted by $\Psi(x)$. Interactions between particles are mediated through the exchange of gauge bosons, which act as force carriers.

The SM encompasses various elementary particles as shown in Figure 1.1, distinguished by their quantum numbers. Particles are primarily categorized into three groups based on spin: fermions, gauge bosons, and the Higgs boson. The fermions have a spin of 1/2 and form all the matter in the universe. Fermions are subdivided into quarks and leptons. Gauge bosons, with a spin of one, mediate forces between fermions. The Higgs boson, a scalar boson with a spin of zero, gives mass to other particles through the Higgs mechanism.

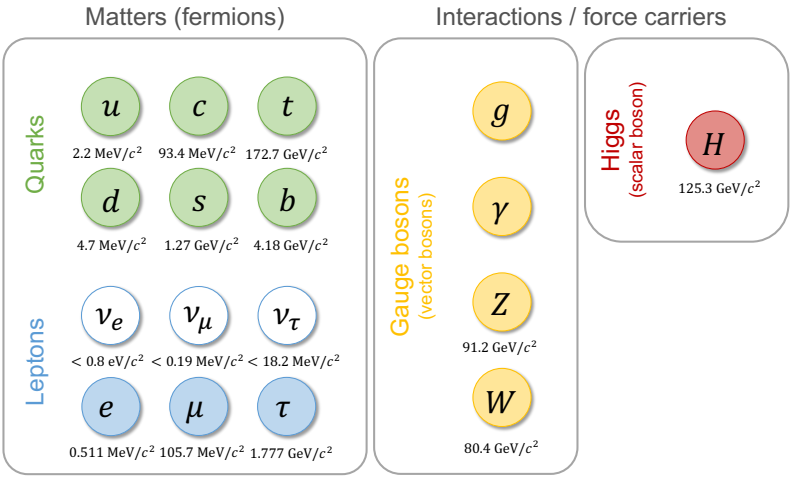


Figure 1.1: Elementray particles in SM.

The Lagrangian for the interactions of Standard Model particles with gauge fields is given by

$$\mathcal{L} = \sum_I \overline{\Psi}_I(x) i\gamma^\mu D_\mu \Psi_I(x), \tag{1.1}$$

$$D_\mu = \partial_\mu + ig_1 \frac{Y}{2} B_\mu + ig_2 \sum_{a=1}^{3} W_\mu^a \frac{\sigma_a}{2} + ig_3 \sum_{\alpha=1}^{8} G_\mu^\alpha \frac{\lambda^\alpha}{2}, \tag{1.2}$$

where I indexes the different fermion fields (quarks and leptons) in the SM, and $\overline{\Psi} = \Psi^\dagger \gamma^0$ represents the Dirac adjoint of the fermion field. The constants g_1, g_2, and g_3 are the

coupling constants for the $U(1)_Y$, $SU(2)$, and $SU(3)$ gauge fields, respectively. The charges of the fermions are listed in Table 1.1.

Table 1.1: Charge of fermions

			Charge		
			g_1	g_2	g_3
$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\begin{pmatrix} c_L \\ s_L \end{pmatrix}$	$\begin{pmatrix} t_L \\ b_L \end{pmatrix}$	$\frac{1}{6}$	1	1
u_R	c_R	t_R	$\frac{2}{3}$	0	1
d_R	s_R	b_R	$-\frac{1}{3}$	0	1
$\begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}$	$-\frac{1}{2}$	1	0
e_R	μ_R	τ_R	-1	0	0

The $SU(2) \times U(1)_Y$ symmetry is broken down to $U(1)_{\rm EM}$ gauge symmetry through spontaneous symmetry breaking. This symmetry breaking allows gauge fields to have masses. One of the $SU(2)$ gauge fields, W^3_μ, and the $U(1)_Y$ gauge field, B_μ, are recombined as follows:

$$A_\mu = W^3_\mu \sin\theta_W + B_\mu \cos\theta_W, \tag{1.3}$$

$$Z_\mu = W^3_\mu \cos\theta_W - B_\mu \sin\theta_W, \tag{1.4}$$

where $\sin\theta_W$ represents Weinberg angle, defined as $g_1/\sqrt{(g_1)^2 + (g_2)^2}$. The field A_μ corresponds to $U(1)_{\rm EM}$ gauge field and preserves gauge invariance. Physically, the gauge boson corresponds to the photon, which has a zero mass. On the other hand, three fields $W^\pm_\mu = (1/\sqrt{2})(W^1_\mu \mp iW^2_\mu)$ and Z^0_μ correspond to gauge fields that break gauge invariance among the $SU(2) \times U(1)_Y$ invariance. Their gauge bosons, known as weak bosons, acquire non-zero masses.

The transitions that change quark flavors occur with the W^+ boson through the Cabibbo-Kobayashi-Maskawa (CKM) mechanism.[1] This mechanism is mathematically represented by the CKM matrix, which is a complex unitary matrix. That describes the probability amplitudes for the transitions between quark flavors through the weak interaction. This matrix is a 3×3 matrix for three generations of quarks:

$$V_{\rm CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{td} & V_{tb} \end{pmatrix}. \tag{1.5}$$

Each element of the matrix, V_{ij}, represents the probability amplitude for a transition from an i quark to a j quark. The complex phases in the CKM matrix allows for CP violation.

[1] Charge-conjugate modes are implied throughout the thesis.

This CP violation induced in the quark sector is quantified by the Jarlskog invariant, $J = (3.08^{+0.15}_{-0.13}) \times 10^{-5}$ [1].

The SM is a powerful framework that successfully explains numerous experimental results. However, it is understood not to be the ultimate theory as it leaves several observed phenomena unanswered.

- **Non-zero neutrino masses**
  Within the currently formulated framework, neutrinos are considered only left-handed, resulting in them being massless. However, results from neutrino oscillation experiments suggest that neutrinos have mass. Therefore, the theory requires extension or modification to incorporate neutrino mass.

- **Imbalance between matter and anti-matter in the universe**
  The Big Bang should have produced equal amounts of matter and antimatter. However, observations of the universe show a clear asymmetry: there is significantly more matter than antimatter. The Standard Model lacks a robust mechanism to explain how this asymmetry arose.

- **The nature of dark matter and dark energy**
  Astrophysical and cosmological observations indicate that the universe contains far more than the matter described by the SM. That is known as dark matter. Moreover, the accelerated expansion of the universe is attributed to dark energy.

Consequently, these observations require physics beyond the SM (BSM).

1.2 Semi-leptonic B decays

The semileptonic B decays are processes in which a B meson decays into a lepton, a corresponding neutrino, and one or more hadrons. In the Standard Model (SM), the transition of a b quark to a c quark occurs at the tree level, mediated by a W boson, as shown in Figure 1.2. The effective Lagrangian is described as [2]

$$\mathcal{L}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{cb} (\bar{c}\gamma^\mu P_L b)(\bar{l}\gamma_\mu P_L \nu_l) + \text{h.c.} \tag{1.6}$$

Here the fermion field $\psi = c, b, l, \nu_l$ denotes the Dirac spinor $\psi = (\psi_L, \psi_R)$, where L and R indicates the chirality of the fermions. The lepton l is either e, μ, or τ. The Fermi constant G_F is defined as $(\sqrt{2}/8)g_W^2/m_W^2$, where g_W represents the $SU(2)$ weak coupling constant. The left- and right-handed projection operators are $P_{L,R} = (1 \mp \gamma_5)/2$. The γ_5 matrix is defined as $i\gamma^0\gamma^1\gamma^2\gamma^3 = -i\gamma_0\gamma_1\gamma_2\gamma_3$.

Since the mediator W boson is significantly heavier than the initial B meson, semileptonic B decay is effectively treated as a four Fermi interaction at the leading electroweak order. This interaction is described by the product of matrix elements for leptonic and hadronic currents [2]:

$$\mathcal{M}^{\lambda_l}_{\lambda_{D^{(*)}}}(q^2, \theta_l) = \frac{G_F}{\sqrt{2}} V_{cb} \frac{M_W^2}{M_W^2 - q^2} \sum_{\lambda_W} \eta(\lambda_W) L^{\lambda_l}_{\lambda_W}(q^2, \theta_l) H^{\lambda_{D^{(*)}}}_{\lambda W}(q^2), \tag{1.7}$$

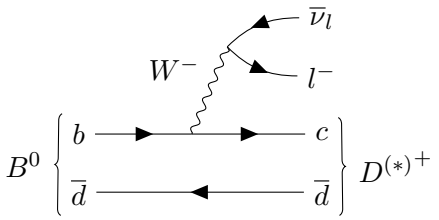


Figure 1.2: Feyman diagram of semileptonic B decays.

where

$$L^{\lambda_l}_{\lambda_W}(q^2,\theta_l) = \epsilon_\mu(q,\lambda_W)\left\langle l(p_l,\lambda_l)\overline{n u}_l(p_{\nu_l})\middle|\bar{l}\gamma^\mu(1-\gamma_5)\nu_\ell\middle|0\right\rangle \tag{1.8}$$

and

$$H^{\lambda_{D^{(*)}}}_{\lambda_W}(q^2) = \epsilon^*_\mu(q,\lambda_W)\left\langle D^{(*)}(p_{D^{(*)}},\lambda_{D^{(*)}})\middle|\bar{c}\gamma^\mu(1-\gamma_5)b\middle|\overline{B}(p_B)\right\rangle \tag{1.9}$$

represent leptonic and hadronic currents, respectively. Here $p_{l,\nu_l,D^{(*)},B}$ denotes the four-momentum of each particle. The square of momentum transfer to the lepton system, q^2, is defined as

$$q^2 = (p_l + p_\nu)^2 = (p_B - p_{D^*})^2, \tag{1.10}$$

and θ_l is the angle between the momentum of the charged lepton and B meson in the virtual W rest frame. The helicity of the W boson, λ_W, takes values $\pm$, 0, and s, where s represents the scalar state of the virtual W boson. The helicity $\lambda_W = s$ corresponds to zero. The metric factor $\eta(\lambda_W)$ is

$$\eta(\lambda_W) = \begin{cases} 1 & (\lambda_W = \pm, 0), \\ \frac{q^2 - M_W^2}{M_W^2} \overset{q^2 \ll m_W}{\approx} -1 & (\lambda_W = s), \end{cases} \tag{1.11}$$

and $\epsilon^\mu(q)$ denotes the polarization vectors of the virtual W boson.

The leptonic matrix elements $L^{\lambda_l}_{\lambda_W}(q^2,\theta_l)$ are expressed as [2]

$$L^+_\pm(q^2,\theta_l) = \pm\sqrt{2} m_l v \sin\theta_l, \tag{1.12}$$
$$L^+_0(q^2,\theta_l) = 2 m_l v \cos\theta_l, \tag{1.13}$$
$$L^+_s(q^2,\theta_l) = -2 m_l v, \tag{1.14}$$
$$L^-_\pm(q^2,\theta_l) = \sqrt{2q^2} v (1 \pm \cos\theta_l), \tag{1.15}$$
$$L^-_0(q^2,\theta_l) = -2\sqrt{q^2} v (1 \pm \cos\theta_l), \tag{1.16}$$
$$L^-_s(q^2,\theta_l) = 0, \tag{1.17}$$

where

$$v = \sqrt{1 - \frac{m_l^2}{q^2}}. \tag{1.18}$$

The interaction between a pseudoscalar meson and a vector meson is generally characterized by a vector current $V(q^2)$, and three axial currents $A_0(q^2)$, $A_1(q^2)$, and $A_2(q^2)$. The hadronic matrix elements for the vector and axial vector current operators in $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ decays are given with these currents.

Lepton flavor universality tests

In the Standard Model (SM), the couplings of vector bosons to lepton pairs are commonly shared among all generations of leptons, which is known as lepton flavor universality (LFU). This LFU is supported by numerous experimental results in tests at a wide range of energy scales involving decays of on-shell W and Z bosons, light mesons, and leptons [3–8].

The LFU tests that utilize semileptonic B decays are also performed. For the light-lepton sets, the LFU is tested with branching fractions or angular variables. The branching fractions are compared between semileptonic B decays to an electron or muon with the following ratios:

$$R(D^*)_{e/\mu} = \frac{\mathcal{B}\left(\overline{B} \to D^* e^- \overline{\nu}_\tau\right)}{\mathcal{B}\left(\overline{B} \to D^* \mu^- \overline{\nu}_\mu\right)}, \tag{1.19}$$

$$R(X)_{e/\mu} = \frac{\mathcal{B}\left(\overline{B} \to X e^- \overline{\nu}_\tau\right)}{\mathcal{B}\left(\overline{B} \to X \mu^- \overline{\nu}_\mu\right)}, \tag{1.20}$$

$$R(K)_{e/\mu} = \frac{B\left(\overline{B} \to K e^+ e^-\right)}{\mathcal{B}\left(B \to K \mu^+ \mu^-\right)}, \tag{1.21}$$

$$R(K^*)_{e/\mu} = \frac{\mathcal{B}\left(B \to K^* e^+ e^-\right)}{\mathcal{B}\left(B \to K^* \mu^+ \mu^-\right)}. \tag{1.22}$$

$$R(pK)_{e/\mu} = \frac{\mathcal{B}\left(\Lambda_b \to p K e^+ e^-\right)}{\mathcal{B}\left(\Lambda_b \to p K \mu^+ \mu^-\right)}. \tag{1.23}$$

All of the measurements on these ratios are consistent with SM predictions [9–17]. Another approach for the LFU tests is to compare angular distributions. The Belle II experiment performs the test with comprehensive sets of angular symmetry and it reveals that there is no significant LFU violation between electrons and muons [18]. Thus, the lepton flavor universality between light-lepton flavors is strongly substantiated by these measurements.

Measurements of $R(D^{(*)})$

The universality between the heaviest τ lepton and light leptons is challenged by several measurements in semileptonic B decays. The BaBar [19, 20], Belle [21–24], and LHCb experiments [25, 26] measure the ratio of branching fractions of the semileptonic B decays, defined by:

$$R(D) = \frac{\mathcal{B}\left(\overline{B} \to D \tau^- \overline{\nu}_\tau\right)}{\mathcal{B}\left(\overline{B} \to D \ell^- \overline{\nu}_\ell\right)}, \tag{1.24}$$

$$R(D^*) = \frac{\mathcal{B}\left(\overline{B} \to D^* \tau^- \overline{\nu}_\tau\right)}{\mathcal{B}\left(\overline{B} \to D^* \ell^- \overline{\nu}_\ell\right)}, \tag{1.25}$$

where ℓ represents a light lepton, e or μ. The experimental results are summrized in Table 1.2 The global average of the measurements exhibits a 3.2σ discrepancy from the SM prediction, as shown in Figure 1.3 [27]. This excess in $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays compared to $\overline{B} \to D^{(*)}\ell^-\overline{\nu}_\ell$ decays is known as the $R(D^{(*)})$ anomaly. That could indicate BSM contributions.

Experiment	τ decays	$R(D^*)$	$R(D)$	Correlation
BaBar [19, 20]	$\tau^- \to \ell^-\nu_\tau\overline{\nu}_\ell$	$0.332 \pm 0.024 \pm 0.018$	$0.440 \pm 0.058 \pm 0.042$	-0.27
Belle [21]	$\tau^- \to \ell^-\nu_\tau\overline{\nu}_\ell$	$0.293 \pm 0.038 \pm 0.015$	$0.375 \pm 0.064 \pm 0.026$	-0.49
Belle [22, 23]	$\tau^- \to \pi^-/\rho^-\nu_\tau$	$0.270 \pm 0.035^{+0.028}_{-0.025}$	—	—
Belle [24]	$\tau^- \to \ell^-\nu_\tau\overline{\nu}_\ell$	$0.283 \pm 0.018 \pm 0.014$	$0.307 \pm 0.037 \pm 0.016$	-0.51
LHCb [26]	$\tau^- \to \mu^-\nu_\tau\overline{\nu}_\mu$	$0.281 \pm 0.018 \pm 0.024$	$0.441 \pm 0.060 \pm 0.066$	-0.43
LHCb [25]	$\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$	$0.257 \pm 0.012 \pm 0.018$	—	—

Table 1.2: Summary of $R(D^{(*)})$ measurement results performed by the BaBar, Belle, and LHCb experiments. The first and second uncertainties are statistical and systematic.

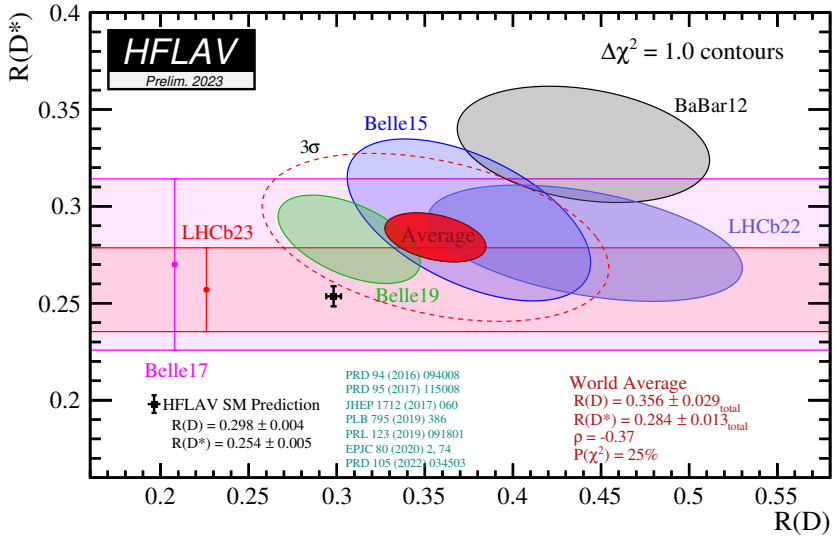


Figure 1.3: Summary of $R(D^{(*)})$ measurements and a preliminary average of $R(D)$ and $R(D^*)$ for winter in 2023 [27]. The black point with error bars indicates the SM prediction.

1.3 New physics in $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$

We considers the effective operators representing BSM contributions to semileptonic B decays as shown in Figure 1.4. The effective Hamiltonian for BSM contributions is described

as [**?**]

$$\mathcal{H}_{\text{eff}} = \frac{4G_F}{\sqrt{2}} V_{cb} \left[(1 + C_{V_L}) O_{V_L} + C_{V_R} O_{V_R} + C_{S_L} O_{S_L} + C_{S_R} O_{S_R} + C_T O_T \right], \tag{1.26}$$

where

$$O_{V_L} = (\bar{c}\gamma^\mu P_L b)(\bar{\tau}\gamma_\mu P_L \nu_\tau), \tag{1.27}$$
$$O_{V_R} = (\bar{c}\gamma^\mu P_R b)(\bar{\tau}\gamma_\mu P_L \nu_\tau), \tag{1.28}$$
$$O_{V_L} = (\bar{c} P_L b)(\bar{\tau} P_L \nu_\tau), \tag{1.29}$$
$$O_{V_R} = (\bar{c} P_R b)(\bar{\tau} P_L \nu_\tau), \tag{1.30}$$
$$O_T = (\bar{c}\sigma^{\mu\nu} P_L b)(\bar{\tau}\sigma_{\mu\nu} P_L \nu_\tau). \tag{1.31}$$

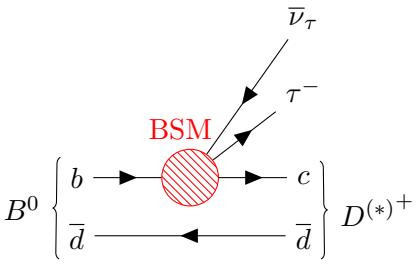


Figure 1.4: Feynman diagram of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays with effective operators of BSM.

From the effective Hamiltonian in Eq (1.26), the differential branching fraction of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays is obtained as [**?**]

$$\begin{aligned}
\frac{d\Gamma(\overline{B} \to D\tau^-\overline{\nu}_\tau)}{dq^2} = \frac{G_F^2 |V_{cb}|^2}{192\pi^3 m_B^3} q^2 \sqrt{\lambda_{D^*}(q^2)} \left(1 - \frac{m_\tau^2}{q^2}\right)^2 \times \Bigg\{ \\
(|1 + C_{V_L} + C_{V_R}|^2) \left[\left(1 + \frac{m_\tau^2}{2q^2}\right) H_{V,0}^2 + \frac{3}{2}\frac{m_\tau^2}{q^2} H_{V,t}^2 \right] \\
+ \frac{3}{2}|C_{S_L} + C_{S_R}|^2 H_S^2 \\
+ 8|C_T|^2 \left(1 + \frac{2m_\tau^2}{q^2}\right) H_T^2 \\
+ 3\text{Re}[(1 + C_{V_L} + C_{V_R})(C_{S_L}^* + C_{S_R}^*)] \frac{m_\tau}{\sqrt{q^2}} H_S H_{V,t} \\
- 12\text{Re}[(1 + C_{V_L} + C_{V_R}) C_T^*] \frac{m_\tau}{\sqrt{q^2}} (H_T H_{V,0}) \Bigg\},
\end{aligned} \tag{1.32}$$

and

$$\frac{d\Gamma(\overline{B}\to D^*\tau^-\overline{\nu}_\tau)}{dq^2} = \frac{G_F^2|V_{cb}|^2}{192\pi^3 m_B^3} q^2\sqrt{\lambda_{D^*}(q^2)}\left(1-\frac{m_\tau^2}{q^2}\right)^2 \times \Bigg\{$$
$$(|1+C_{V_L}|^2+|C_{V_R}|^2)\left[\left(1+\frac{m_\tau^2}{2q^2}\right)(H_{V,+}^2+H_{V,-}^2+H_{V,0}^2)+\frac{3}{2}\frac{m_\tau^2}{q^2}H_{V,t}^2\right]$$
$$-2\mathrm{Re}[(1+C_{V_L})C_{V_R}^*]\left[\left(1+\frac{m_\tau^2}{2q^2}\right)(H_{V,0}^2+2H_{V,+}H_{V,-})+\frac{3}{2}\frac{m_\tau^2}{q^2}H_{V,t}^2\right]$$
$$+\frac{3}{2}|C_{S_L}-C_{S_R}|^2H_S^2$$
$$+8|C_T|^2\left(1+\frac{2m_\tau^2}{q^2}\right)(H_{T,+}^2+H_{T,-}^2+H_{T,0}^2)$$
$$+3\mathrm{Re}[(1+C_{V_L}-C_{V_R})(C_{S_L}^*-C_{S_R}^*)]\frac{m_\tau}{\sqrt{q^2}}H_SH_{V,t}$$
$$-12\mathrm{Re}[(1+C_{V_L})C_T^*]\frac{m_\tau}{\sqrt{q^2}}(H_{T,0}H_{V,0}+H_{T,+}H_{V,+}-H_{T,-}H_{V,-})$$
$$+12\mathrm{Re}|C_{V_R}C_T^*|\frac{m_\tau}{\sqrt{q^2}}(H_{T,0}H_{V,0}+H_{T,+}H_{V,-}-H_{T,-}H_{V,+})\Bigg\}. \quad (1.33)$$

Form factor description based on Heavy quark effective theory [28] is utilized and integrated out over the variable q^2. The BSM contributions to $R(D^*)$ are expressed as a ratio over the SM prediction [29]:

$$\frac{R(D^*)}{R(D^*)_{\mathrm{SM}}} = |1+C_{V_L}|^2+|C_{V_R}|^2+0.04|C_{S_L}-C_{S_R}|^2+16.0|C_T|^2$$
$$-1.83\mathrm{Re}[(1+C_{V_L})C_{V_R}^*]-0.11\mathrm{Re}[(1+C_{V_L}-C_{V_R})(C_{S_L}^*-C_{S_R}^*)]$$
$$-5.17\mathrm{Re}[(1+C_{V_L})C_T^*]+6.60\mathrm{Re}[C_{V_R}C_T^*]. \quad (1.34)$$

1.4 This dissertation

We perform the first $R(D^*)$ measurement at the Belle II experiment. This result marks the first test of the LFU with semileptonic B decays involving τ decays, utilizing e^+e^- collision data from the Belle II detector at the SuperKEKB collider. Using a new data set from the Belle II experiment, this measurement enables a statistically independent examination of the $R(D^*)$ anomaly, augmenting the previous measurements. It contributes to the search for NP effects violating LFU in semileptonic B decays. In our analysis, we introduce a novel B reconstruction method as a B tagging technique, with selections optimized to fully exploit this approach. The analysis successfully controls the uncertainties through an understanding of the real data collected during an early data-taking period. Consequently, we deliver a timely and independent result for the high-profile $R(D^*)$ anomaly with an uncertainty comparable to the previous measurements.

I have led the establishment of an analysis framework for this $R(D^*)$ measurement with the early data set at the Belle II experiment and played a pivotal role in its development across all analysis steps. My contribution includes developing a reconstruction algorithm for $\overline{B}\to D^*\tau^-\overline{\nu}_\tau$ and $\overline{B}\to D^*\ell^-\overline{\nu}_\ell$ decays, optimizing selection criteria for those candidates, validating simulation with sideband data, formulating procedures of the signal extraction, evaluating systematic uncertainties, and deriving the final results.

This dissertation is organized as follows. Section 2 describes the SuperKEKB accelerator and the Belle II detector. Section 3 outlines our analysis strategy. Section 4 provides details of the data sets and the selection criteria for $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$ and $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ candidates. Section 5 explains the calibration procedures for simulation. Section 6 reports the validation of detector responses and background modeling in the simulation. Section 7 details the method for the signal extraction. Section 8 presents the results, while Section 9 describes the systematic uncertainties. Section 10 discusses the prospects of $R(D^*)$ measurements at Belle II and the implications for new physics parameters. Finally, the dissertation concludes in Section 11.

2 The SuperKEKB/Belle II experiment

2.1 The SuperKEKB accelerator

The SuperKEKB accelerator is an electron-positron collider located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan [30]. Figure 2.1 illustrates the SuperKEKB accelerator. The accelerator is a significant upgrade of the KEKB accelerator, designed to achieve a thirty times higher luminosity than its predecessor. It consists of an electron-positron linear injector, a positron damping ring, and two main rings with a circumference of 3 km: the high energy ring (HER) for electrons and the low energy ring (LER) for positrons. The electron and positron beams are accelerated in the linear injector to 7 GeV and 4 GeV, respectively. These beams circulate the main rings and collide at the interaction point in the center of the Belle II detector at the center-of-mass (c.m.) energy $\sqrt{s} = 10.58$ GeV, corresponding to the energy of $\Upsilon(4S)$ resonance. The SuperKEKB accelerator achieves the world's highest luminosity of 4.7×10^{35} cm^{-2}s^{-1} in 2022. The processes $e^+e^- \to B\overline{B}$ and $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$) have cross-sections of 1.1 fb^{-1} and 3.7 fb^{-1}, respectively. The $\Upsilon(4S)$ resonance decays into B meson pairs with a branching ratio of $> 96\%$ [1].

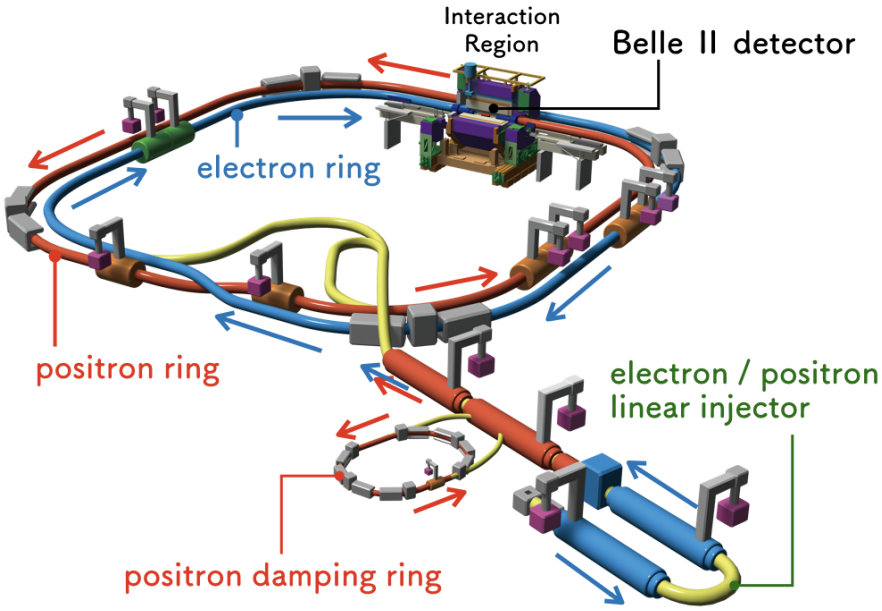


Figure 2.1: The SuperKEKB accelerator. ©KEK

2.2 The Belle II detector

The Belle II detector is a general-purpose particle detector [31]. The detector is placed at the collision point of the electron and positron beams and records events caused by the beam collisions. Figure 2.2 shows a schematic image of the Belle II detector. From the

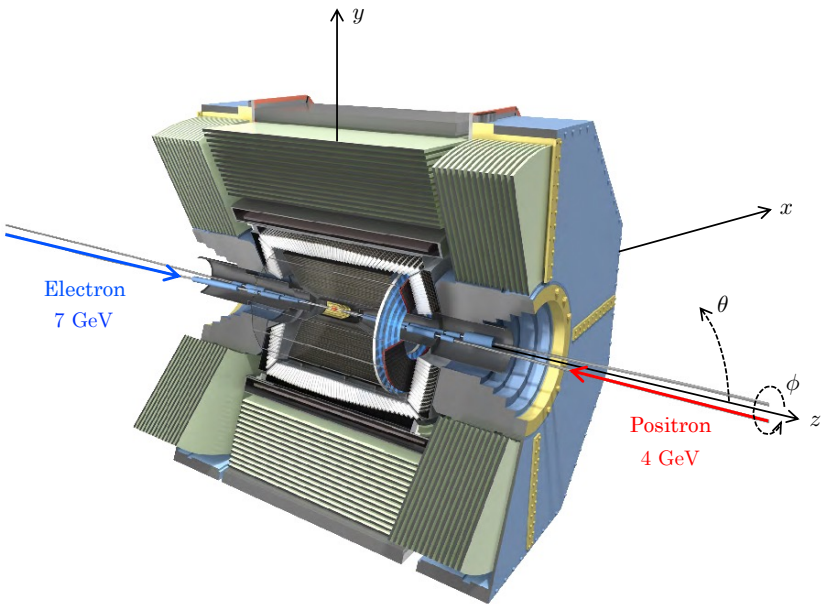


Figure 2.2: The Belle II detector. The blue and red arrows indicate the direction of the electron and positron beams.

most inner to the most outer layers, the detector consists of a vertex detector, a central drift chamber, particle identification detectors, an electromagnetic calorimeter, and a K_L^0 and muon detector. The electromagnetic calorimeter is surrounded by a superconducting solenoid magnet with an inner diameter of 3.6 m that provides a 1.5 T uniform magnetic field. The Belle II adopts a cylindrical coordinate system. The z-axis aligns with the solenoid magnetic field and points approximately parallel to the momentum direction of the electron beam. The x and y axes are oriented in the horizontal and vertical directions, respectively. The polar and azimuthal angles are denoted by θ and ϕ. The positive and negative z directions correspond to forward and backward directions, respectively.

2.2.1 Vertex detector

The vertex detector (VXD) surrounds a beryllium beam pipe with an inner diameter of 12 mm. Located at the most inner layers, this detector reconstructs decay vertices of B mesons and long-lived particles based on the trajectory of the charged particles. It comprises a pixel-type semiconductor detector (Pixel Detector, PXD) for the inner two layers and a strip-type semiconductor detector (Silicon Vertex Detector, SVD) for the outer four layers. When a charged particle traverses an n-type silicon semiconductor, it generates an electron-hole pair charge. By applying voltage, the n-type semiconductor becomes depleted. The PXD, utilizing a Depleted Field Effect Transistor (DEPFET), stores the produced electrons in the internal gate. The gate voltage of the DEPFET, which fluctuates depending on the number of stored electrons, detects the signal. Conversely, the SVD employs a double-sided silicon detector (DSSD). In this setup, electrons and holes migrate to opposite sides of the n-type and p-type silicon semiconductor along the electric

field. The electrodes on each side read out electron and hole signals from strips in the ϕ direction and strips in the z direction, respectively.

2.2.2 Central drift chamber

The central drift chamber (CDC) measures the momentum of charged particles by determining the curvature of their reconstructed tracks in a magnetic field of 1.5 T. It also identifies the types of charged particles by measuring their energy loss. Inside the chamber, charged tracks ionize the gas. The generated electrons are attracted to the sense wire and further ionize atoms near the sense wire, which causes avalanche amplification. The amplified electrons are then read out from the sense wire as a signal. The magnitude of this readout signal is used to measure the energy loss.

2.2.3 Time-of-propagation counter

The Time-of-propagation (TOP) counter, a ring-imaging Cherenkov detector, is located in the barrel region. It is composed of 16 modules, with each module consisting of a quartz radiator, a photodetector MCP-PMT, and a readout circuit. The emission angle of Cherenkov light, θ_C, is given by

$$\cos\theta_C = \frac{1}{n(\lambda)\beta}, \tag{2.1}$$

where $\beta \equiv v/c$ denotes the velocity of a charged particle, and $n(\lambda)$ is the refractive index of the radiator depending on the wavelength of light λ. The Cherenkov angle and mass of the charged particle are interconnected as

$$m = \frac{p}{\beta\gamma} = p\sqrt{n^2(\lambda)\cdot\cos^2\theta_C - 1}, \tag{2.2}$$

where $\gamma \equiv 1/\sqrt{1-\beta^2}$. Therefore, by utilizing momentum reconstructed in the inner trackers and measuring θ_C, the TOP counter identifies the type of the charged particle. It determines the Cherenkov angle by measuring the arrival time and position of the photons that propagate inside the quartz radiator.

2.2.4 Aerogel ring imaging Cherenkov detector

The Aerogel ring imaging Cherenkov (ARICH) detector is a ring-imaging Cherenkov detector situated at the forward end-cap region. It utilizes two-centimeter thick aerogel with two different refractive indices as a radiator. When a charged particle passes through the aerogel, it generates Cherenkov light, and the photons are detected by a photodetector located 20 cm away from the aerogel. This process produces a ring of Cherenkov photons on the detector, and the Cherenkov angle is reconstructed using these photons. The photodetector, a Hybrid Avalanche Photo Detector, reads out signals with an amplification factor of about 70,000 times. This amplification is achieved by accelerating photoelectrons onto the Avalanche Photodiode using high voltage and then through avalanche amplification inside the photodiode.

2.2.5 Electromagnetic calorimeter

The electromagnetic calorimeter (ECL) is a total absorption calorimeter that uses CsI(Tl) crystals to measure the energy of electrons and photons. When electrons or photons enter these crystals, they generate an electromagnetic shower. This process produces scintillation light that corresponds to the energy of the incident electrons or photons. The intensity of the scintillation light is proportional to the energy lost in the electromagnetic shower. A PIN photodiode, attached to the end of the crystal, detects this scintillation light as the photodetector.

There are discrepancies in the energy deposit to the ECL clusters and its leakage between the data and the simulation. To correct them, we apply energy bias correction to all photon candidates depending on momentum, polar angle, and azimuthal angle. The bias is evaluated using π^0 mass distributions.

2.2.6 K_L^0 and muon detector

The K_L^0 and muon detector (KLM), located outside of the solenoid magnet, identifies particles with high penetrating power, particularly K_L^0 mesons and muons. This detector features a sandwich structure consisting of 14 or 15 layers of iron and detectors for charged particles, each about 4.7 cm thick. Plastic scintillators are used in the two inner layers in the barrel region and in the forward and backward end-cap regions, while resistive plate chambers are employed in the outer layer of the barrel region. Both muons and charged hadrons produce hits in these detectors. undergo only electromagnetic interactions, while charged hadrons experience both electromagnetic and strong interactions. This leads to significant multiple scattering, enabling their discrimination through the different shapes of their showers. The K_L^0 mesons do not leave direct hits in the detectors. Instead, interact with nucleons in iron via strong interactions and emit charged particles. These particles are then detected, allowing for the identification of the K_L^0 mesons.

2.3 Reconstruction technique of B mesons

The Full Event Interpretation (FEI) algorithm [32] reconstructs one of the B mesons from $\Upsilon(4S)$ decays. It exclusively uses 68 hadronic decay channels of neutral and charged B mesons listed in Table 2.1. The reconstruction process involves a hierarchical approach with six object stages, starting from tracks, displaced vertices, and neutral clusters. In total, this method considers more than 10,000 decay cascades. This identification method is called hadronic B tagging. The algorithm utilizes `FastBDT` [33] and returns the probability of B_{tag} meson, $\mathcal{P}_{\text{FEI}}$, between 0 and 1 as the output. The probability represents a confidence level of the reconstructed B_{tag} meson. A B_{tag} candidate with a high (low) probability is more likely to be correctly (incorrectly) reconstructed B meson. The reconstruction efficiencies of the B_{tag} mesons are 0.23% and 0.30% for neutral and charged B modes at a purity of 29% [34].

Table 2.1: Reconstructed modes of hadronic B decays implemented in FEI for neutral and charged B_{tag} candidates.

Neutral B		Charged B	
ID	Decay	ID	Decay
0	$\overline{B} \to D^+\pi^-$	0	$B^- \to D^0\pi^-$
1	$\overline{B} \to D^+\pi^-\pi^0$	1	$B^- \to D^0\pi^-\pi^0$
2	$\overline{B} \to D^+\pi^-\pi^0\pi^0$	2	$B^- \to D^0\pi^-\pi^0\pi^0$
3	$\overline{B} \to D^+\pi^-\pi^+\pi^-$	3	$B^- \to D^0\pi^-\pi^+\pi^-$
4	$\overline{B} \to D^+\pi^-\pi^+\pi^-\pi^0$	4	$B^- \to D^0\pi^-\pi^+\pi^-\pi^0$
5	$\overline{B} \to D^0\pi^+\pi^-$	5	$B^- \to D^0D^-$
6	$\overline{B} \to D^+\overline{D}^0K^-$	6	$B^- \to D^0D^-\overline{K}^0_S$
7	$\overline{B} \to D^+\overline{D}^{*0}K^-$	7	$B^- \to D^{*0}D^-\overline{K}^0_S$
8	$\overline{B} \to D^{*+}\overline{D}^0K^-$	8	$B^- \to D^0\overline{D}^{*-}\overline{K}^0_S$
9	$\overline{B} \to D^{*+}\overline{D}^{*0}K^-$	9	$B^- \to D^{*0}D^{*-}\overline{K}^0_S$
10	$\overline{B} \to D^+D^-\overline{K}^0_S$	10	$B^- \to D^0\overline{D}^0K^-$
11	$\overline{B} \to D^{*+}D^-\overline{K}^0_S$	11	$B^- \to D^{*0}\overline{D}^0K^-$
12	$\overline{B} \to D^+\overline{D}^{*-}\overline{K}^0_S$	12	$B^- \to D^0\overline{D}^{*0}K^-$
13	$\overline{B} \to D^{*+}\overline{D}^{*-}\overline{K}^0_S$	13	$B^- \to D^{*0}\overline{D}^{*0}K^-$
14	$\overline{B} \to D^+D_s^-$	14	$B^- \to D^0D_s^-$
15	$\overline{B} \to D^{*+}\pi^-$	15	$B^- \to D^{*0}\pi^-$
16	$\overline{B} \to D^{*+}\pi^-\pi^0$	16	$B^- \to D^{*0}\pi^-\pi^0$
17	$\overline{B} \to D^{*+}\pi^-\pi^0\pi^0$	17	$B^- \to D^{*0}\pi^-\pi^0\pi^0$
18	$\overline{B} \to D^{*+}\pi^-\pi^+\pi^-$	18	$B^- \to D^{*0}\pi^-\pi^+\pi^-$
19	$\overline{B} \to D^{*+}\pi^-\pi^+\pi^-\pi^0$	19	$B^- \to D^{*0}\pi^-\pi^+\pi^-\pi^0$
20	$\overline{B} \to D^+D_s^{*-}$	20	$B^- \to D^0\overline{D}_s^{*-}$
21	$\overline{B} \to D^{*+}D_s^-$	21	$B^- \to D^{*0}\overline{D}_s^{*-}$
22	$\overline{B} \to D^{*+}D_s^{*-}$	22	$B^- \to D^0K^-$
23	$\overline{B} \to J/\psi\overline{K}^0_S$	23	$B^- \to D^+\pi^+\pi^-$
24	$\overline{B} \to J/\psi K^-\pi^+$	24	$B^- \to D^+\pi^+\pi^-\pi^0$
25	$\overline{B} \to J/\psi\overline{K}^0_S\pi^+\pi^-$	25	$B^- \to J/\psi K^-$
26	$\overline{B} \to \Lambda_c^+\overline{p}\pi^+\pi^-$	26	$B^- \to J/\psi K^-\pi^+\pi^-$
27	$\overline{B} \to D^0p\overline{p}$	27	$B^- \to J/\psi K^-\pi^0$
28	$\overline{B} \to D^+p\overline{p}\pi^-$	28	$B^- \to J/\psi\overline{K}^0_S\pi^-$
29	$\overline{B} \to D^{*+}p\overline{p}\pi^-$	29	$B^- \to \Lambda_c^+\overline{p}\pi^-\pi^0$
30	$\overline{B} \to D^0p\overline{p}\pi^+\pi^-$	30	$B^- \to \Lambda_c^+\overline{p}\pi^-\pi^+\pi^-$
31	$\overline{B} \to D^{*0}p\overline{p}\pi^+\pi^-$	31	$B^- \to D^0p\overline{p}\pi^-$
		32	$B^- \to D^{*0}p\overline{p}\pi^-$
		33	$B^- \to D^-p\overline{p}\pi^+\pi^-$
		34	$B^- \to D^{*-}p\overline{p}\pi^+\pi^-$
		35	$B^- \to \Lambda_c^+\overline{p}\pi^-$

3 Analysis strategy

We measure the $R(D^*)$ value by simultaneously determining the branching fractions of the signal $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$ and the normalization $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ decays. Leveraging the high reconstruction efficiency of neutral particles at Belle II, we include charged B modes as well as neutral B modes in the $R(D^*)$ measurement, under the isospin-symmetry assumption that $R(D^{*+}) = R(D^{*0})$.

We reconstruct candidates of the signal and normalization decays from the identical final states by exploiting leptonic τ decays: $\tau^- \to e^- \nu_\tau \overline{\nu}_e$ and $\tau^- \to \mu^- \nu_\tau \overline{\nu}_\mu$. This approach helps reduce experimental systematic uncertainties by the cancellation of reconstruction efficiencies for the final-state particles between the signal and normalization candidates.

Our reconstruction employs a B tagging method. Initially, one of the B mesons produced in pairs from $\Upsilon(4S)$ decays is identified and designated as a tagged B meson, or B_{tag}. This B_{tag} meson is fully reconstructed via exclusive hadronic channels using the FEI technique. The other B meson, reconstructed with a D^* candidate and a lepton candidate, is used for the signal and normalization decays and is referred to as B_{sig}. An example of the decay topology in this reconstruction is illustrated in Figure 3.1. Employing this B tagging method allows for the reconstruction of the entire event, except for missing particles like neutrinos. Thanks to the well-known initial states with the beam energies and the pair production of B mesons, the kinematics of the B_{sig} candidate is effectively constrained by the information from the partner B_{tag} candidate, even with the missing particles in the B_{sig} decays. Additionally, the increased reconstruction efficiency provided by the FEI technique is particularly beneficial in this analysis approach, where statistical uncertainty is still the predominant source of uncertainties. Comprehensive details of the reconstruction and candidate selection processes are described in Section 4.

We adopt a data-hidden analysis approach in this $R(D^*)$ analysis. The signal regions, defined by selections of the reconstructed mass difference between a D^* and its D daughter candidates, remain hidden in the data until we finalize the analysis procedures, including selections, calibration, and background estimation. We conduct calibration using data from a sideband region or a partially opened region where only the normalization decays are predominant. Both of these regions have negligible contributions from the signal decays.

In our analysis, the primary source of background candidates arises from incorrectly reconstructed D^* candidates, hereafter referred to as fake D^* candidates. This misreconstrutcion primarily results from random combinations of daughter particles and/or their cross-feed between the B_{tag} and B_{sig} decays as shown in Figure 3.2. We discriminate these candidates by examining the masses of the daughters and other daughter variables.

The rest of the background candidates consists of correctly reconstructed D^* candidates. These background candidates enter our signal region, as illustrated in Figure 3.3. They predominantly originate from semileptonic B decays to excited-charm mesons. These heavier 1P states in the charm meson system are collectively known as D^{**}. In these cases, the D^{**} states decay secondarily to D^* mesons accompanied by a few low-momentum pions. Often, these decays are only partially reconstructed as a B_{sig} candidate. Such partial

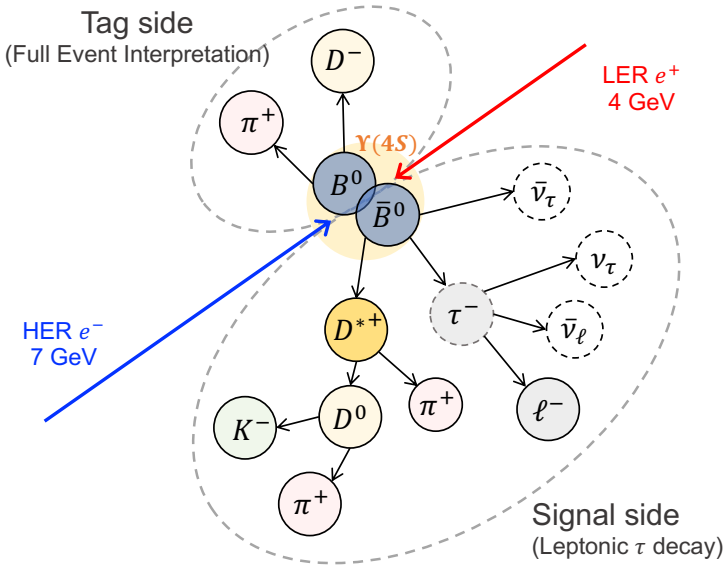
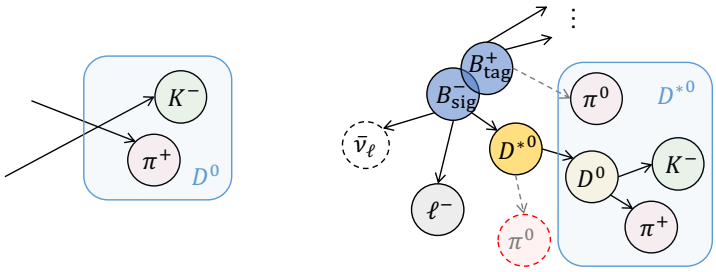


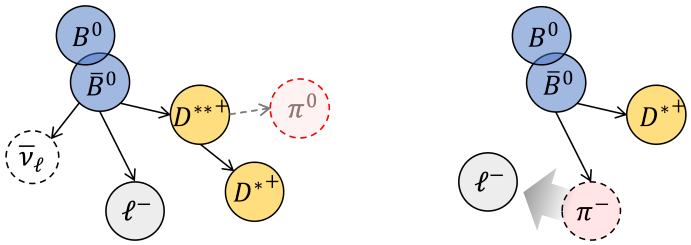
Figure 3.1: Schematic view of topology in the reconstruction of a $B\overline{B}$ candidate.



(a) (b)

Figure 3.2: Schematic view of background events with a fake D^* candidates in the reconstruction. Fake D^* candidates are formed by (a) random combination of daughter particles and/or (b) cross-feed of daughter particles between the $B_{\rm tag}$ and $B_{\rm sig}$ decays. Here a π^0 candidate from the $B_{\rm sig}$ decay chain is misreconstructed and that from the $B_{\rm tag}$ decay is used for the D^* reconstruction in the cross feed.

reconstructions lead to a missing mass in the event, which can mimic the larger missing mass typical of τ decays. To suppress these background candidates, we apply selections for particles not involved in the reconstruction of either the $B_{\rm tag}$ or $B_{\rm sig}$ candidates. These selections ensure that no extra reconstructed particles are associated with the $B_{\rm sig}$ candidate. The misidentification of hadron daughters as lepton candidates also inadvertently includes background candidates from hadronic B decays.



(a) (b)

Figure 3.3: Schematic view of background candidates with a correctly reconstructed D^* candidates in the reconstruction. (a) Failures to reconstruct a low-momentum π^0 candidate from a D^{**} state often induce the background candidates from $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ decays. (b) The misidentification of a hadron daughter as a lepton candidate allows to include the background candiadtes from hadronic B decays.

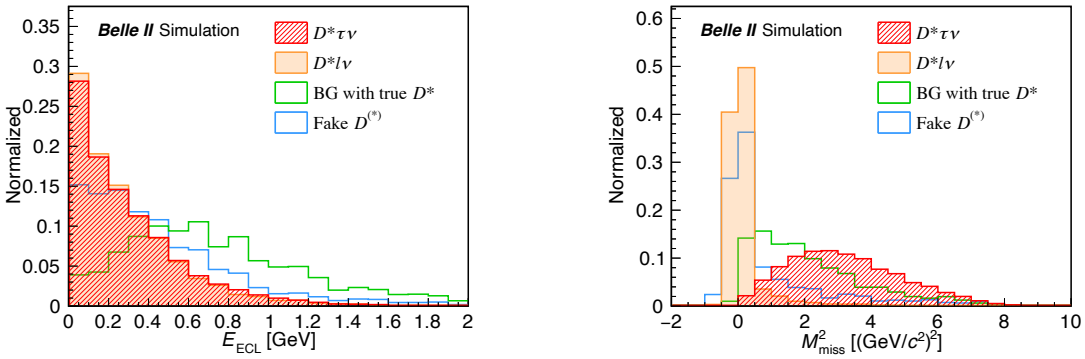
We establish a reasonable background estimation in simulation for both types of background sources: those with incorrectly and correctly reconstructed D^* candidates. We validate the modeling of these background decays in the simulation by comparing it with data from sideband regions where each background source predominantly contributes. For the background candidates with fake D^* candidates, we calibrate them using data-simulation ratios of their yields. Furthermore, the background yields are controlled in the signal extraction, implementing constraints based on these data-simulation ratios. Through these data-driven approaches, we demonstrate that both background sources are well-modeled in our simulation.

We distinguish between correctly reconstructed signal and normalization candidates, as well as background candidates, by utilizing two key variables: the extra energy detected by the ECL detector aside from the tag-side and signal-side B mesons, so-called $E_{\rm ECL}$, and the missing mass squared, $M^2_{\rm miss}$. These variables take advantage of both B mesons from $\Upsilon(4S)$ decays, so they are unique to analyses using the B-tagging technique. Figure 3.4 illustrates distributions of $E_{\rm ECL}$ and $M^2_{\rm miss}$.

Extra energy in ECL $E_{\rm ECL}$

The extra energy in ECL mainly discriminates $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ and $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$ candidates from background candidates. It is calculated as the sum of the cluster energy of selected photon candidates that are not involved in the reconstruction of the $B_{\rm tag}$ or $B_{\rm sig}$ candidates. Correctly reconstructed candidates typically show a peak at zero, whereas background candidates often exhibit higher values due to additional clusters from missing particles.

Missing mass squared $M^2_{\rm miss}$



(a) (b)

Figure 3.4: Distributions of (a) $E_{\rm ECL}$ and (b) $M^2_{\rm miss}$ in simulation. The candidates are reconstructed with $\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell$ and $D^{*+} \to D^0\pi^+$ modes. The red and orange histograms are correctly reconstructed $\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau$ and $\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell$ candidates, respectively. The green (light-blue) histograms are background candidates with a correctly and incorrectly reconstructed D^* candidate. The vertical axes are normalized in an arbitrary unit.

The missing mass squared is defined by Eq. (3.1) and is instrumental in separating $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ candidates from $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$ and background candidates.

$$M^2_{\rm miss} = \left(\frac{E^*_{e^+e^-}}{2} - E_{D^*} - E_\ell\right)^2 - \left(-\vec{p}^{\,*}_{B_{\rm tag}} - \vec{p}^{\,*}_{D^*} - \vec{p}^{\,*}_\ell\right)^2. \tag{3.1}$$

Here, $E^*_{e^+e^-}$ is the c.m. collision energy of the e^+e^- beam, $E^*_{D^*}$ and E^*_ℓ are the reconstructed energy of D^* and lepton candidates, $\vec{p}^{\,*}_{B_{\rm tag}}$, $\vec{p}^{\,*}_{B_{D^*}}$, and $\vec{p}^{\,*}_\ell$ are the reconstructed momentum of $B_{\rm tag}$, D^*, and lepton candidates in the c.m. frame, respectively. Correctly reconstructed $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$ candidates typically peak at zero, corresponding to the mass of a neutrino as all final-state particles except for one neutrino are reconstructed. In contrast, $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ candidates exhibit larger values due to the emission of multiple neutrinos from the leptonic τ decays.

Utilizing these variables in two dimensions, we can effectively discriminate between signal, normalization, and background candidates, as demonstrated in Figure 3.5. We ascertain the yields of these candidates and determine the ratio $R(D^*)$ from a fit to data. This fitting process employs these variables as fitting observables within two-dimensional probability density functions (PDFs).

The most significant source of systematic uncertainty in this measurement arises from the shapes of the fitting observables in simulation. Corrections are required due to discrepancies observed between data and simulation. We implement data-driven correction procedures and evaluate their uncertainties.

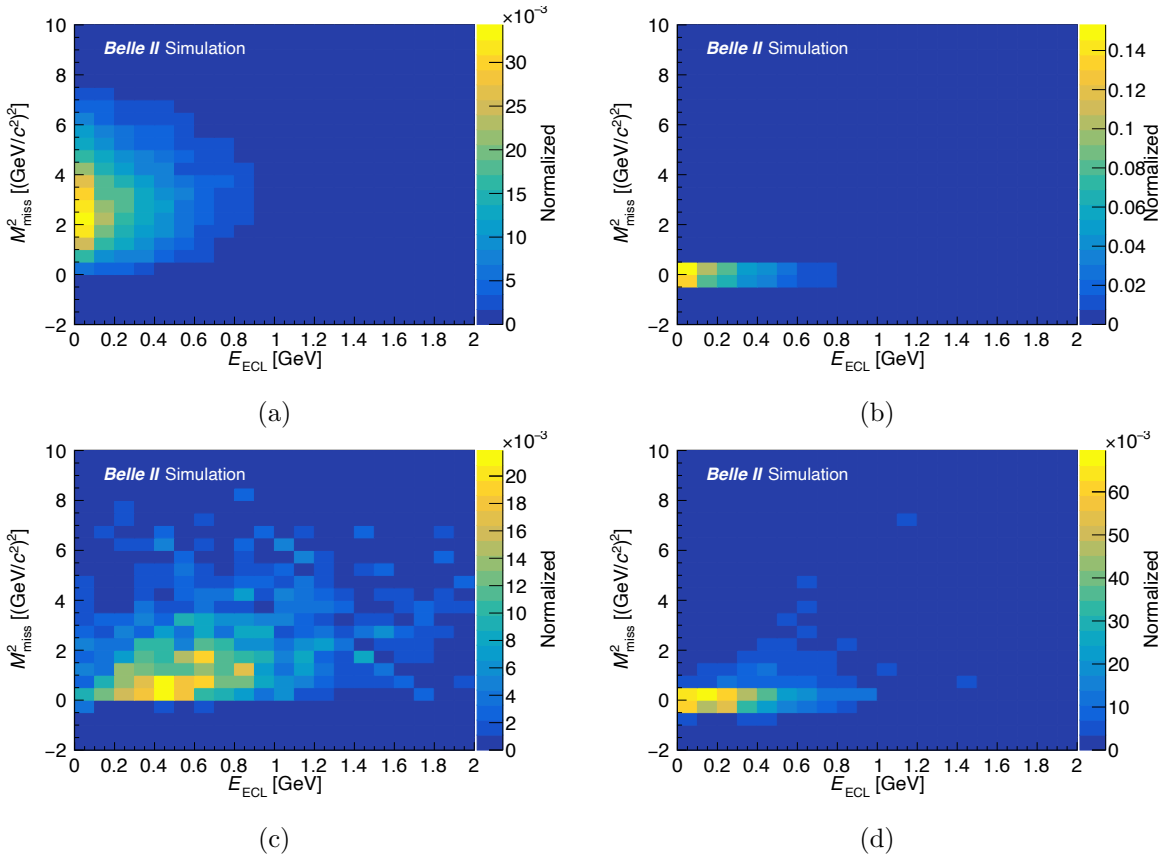


Figure 3.5: Two-dimensional distributions of E_{ECL} and M^2_{miss} in simulation. The candidates are reconstructed with $\overline{B} \to D^{*+}\ell^-\overline{\nu}_\ell$ and $D^{*+} \to D^0\pi^+$ modes. The signal (a), normalization candidates (b), and background candidates with a true D^* candidate (c) and a fake D^* candidate (d) are shown, respectively. The z-axes are normalized in an arbitrary unit.

4 Signal reconstruction and selection

4.1 Data set

4.1.1 Data

We use (189.26 ± 0.01) fb^{-1} of the e^+e^- collision data at a center-of-mass energy of 10.58 GeV collected from 2019 to 2021. The total number of $B\overline{B}$ events is $(198.0 \pm 3.0) \times 10^6$.

4.1.2 Monte Carlo (MC) samples

Table 4.1: Branching fractions of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ in the signal MC samples and equivalent luminosity of the generated samples.

Decay	Branching fraction	Equivalent luminosity [fb^{-1}]
$\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau$	0.0125416	40239.5
$\overline{B}^0 \to D^{+}\tau^-\overline{\nu}_\tau$	0.0063687	42066.0
$B^- \to D^{*0}\tau^-\overline{\nu}_\tau$	0.0134976	35427.2
$B^- \to D^{0}\tau^-\overline{\nu}_\tau$	0.0068527	36854.6

Monte Carlo (MC) simulation samples are used to develop the signal selection criteria, examine the leading background processes, and determine the final fit model. The decay chains in the MC samples are simulated using the EvtGen package [35] and the detector response is modeled with the Geant4 framework [36]. Electromagnetic final-state radiation is simulated using the PHOTOS package [37, 38]. All MC samples are generated under the nominal beam-induced background conditions estimated at an instantaneous luminosity of 2.8×10^{35} cm^{-2}s^{-1} in the simulation.

Signal $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ MC

We prepare 800 million events for the signal decays from $B^0\overline{B}^0$ and B^+B^- pairs, respectively. In the simulation, one of the B mesons decays exclusively to $D\tau^-\overline{\nu}_\tau$ and $D^*\tau^-\overline{\nu}_\tau$ with a ratio of these branching fractions. The other B meson decays inclusively. The τ lepton from $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays inclusively. Branching fractions of the signal decays are also summarized in Table 4.1.

Generic MC

We use MC samples that consist of $e^+e^- \to B\overline{B}$ events where each B meson decays inclusively, and continuum events where $e^+e^- \to q\overline{q}$ $(q = u, d, s, c)$. These samples have event compositions based on production cross-sections and decay branching fractions. The decay rates of B meson decays for which no measurements exist are modeled by PYTHIA [39]. The continuum events are simulated with the KKMC package [40, 41] using PYTHIA [39] for hadronization. We have MC samples of 900 fb^{-1} $B^0\overline{B}^0$ and B^+B^- events, and 1,000 fb^{-1} $q\overline{q}$ events. Another 100 fb^{-1} sample of $B\overline{B}$ events is used for the FEI training.

Background $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ MC

We prepare dedicated simulation samples for the $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ decays with $D^{**} \to D^{(*)}\pi\pi$ or $D^{**} \to D^{(*)}\eta$, which are not included in the generic MC samples. One B mesons decays to $D^{**}\ell^-\overline{\nu}_\ell$ or $D^{**}\tau^-\overline{\nu}_\tau$ and the other B meson decays inclusively. The τ lepton from $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ decays inclusively. 10(3) million events of $B^0\overline{B}^0$ and B^+B^- pairs each are generated for $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ ($\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$) decays.

Entries from both signal and background MC samples are scaled to the luminosity of the collision data. For the dedicated MC samples with the exclusive B decays, the equivalent luminosity, denoted as $\mathcal{L}_{\text{equiv}}(B \to X)$, is used for the scaling. The equivalent luminosity is determined using the formula:

$$\mathcal{L}_{\text{equiv}}(B \to X) = \frac{N_{B\overline{B}}}{\sigma_{B\overline{B}} \cdot 2\mathcal{B}(B \to X)}, \tag{4.1}$$

where $N_{B\overline{B}}$ represents the number of $B\overline{B}$ pairs generated in each dedicated MC sample, $\sigma_{B\overline{B}}$ is the cross section of $B\overline{B}$ pairs at the $\Upsilon(4S)$ resonance in Table 4.2, and $\mathcal{B}(B \to X)$ denotes the branching fractions of $B \to X$ decays assumed in the simulation. Consequently, entries from these dedicated MC samples are combined, taking into account the reconstruction efficiencies.

Table 4.2: Cross sections of $B\overline{B}$ events in the generic MC.

Event type	Cross section [nb]
$B^0\overline{B}^0$	0.510
B^+B^-	0.540

In our simulation, we have updated the branching fractions to reflect the latest values from the HFLAV group [27] and the PDG [1]. We adopt isospin-averaged branching fractions for $\overline{B} \to X_c\ell^-\overline{\nu}_\ell$ decays, where X_c represents a charm meson. Table 4.3 details the branching fractions for $B \to D^*\ell^-\overline{\nu}_\ell$, $B \to D^{**}\ell^-\overline{\nu}_\ell$, and $B \to D^{**}\tau^-\overline{\nu}_\tau$. The branching fractions for hadronic B decays and τ decays have also been updated with the values listed in Table 4.4 and Table 4.5, respectively.

Additionally, our simulation incorporates parameterizations of the hadronic form factors based on HQET in order to model semileptonic B decays. Both the signal and normalization decays utilize the Bernlochner-Ligeti-Papucci-Robinson-Xiong-Prim (BLPRXP) form factor parameterization scheme, as detailed in Ref. [42]. For $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ decays, we have adopted the Bernlochner-Ligeti-Robinson (BLR) parameterization [43, 44]. The parameters of these form factors have been updated to their latest values by a software package for form factor reweighting, called Helicity Amplitude Module for Matrix Element Reweighting (HAMMER) [45].

Table 4.3: Isospin-averaged branching fractions of semileptonic B decay chains in the simulation. The second column shows subsequent decay modes of resonant D^{**} states. The branching fractions used for the evaluation of the systematic uncertainty due to non-resonant $\overline{B} \to D^{(*)}\pi\ell^-\overline{\nu}_\ell$ are in parentheses.

Decay		Branching fraction $[10^{-3}]$	
		$\mathcal{B}(B^0)$	$\mathcal{B}(B^+)$
$\overline{B} \to D^*\ell^-\overline{\nu}_\ell$		50.3 ± 1.1	54.1 ± 1.1
$\overline{B} \to D_1\ell^-\overline{\nu}_\ell$	$D_1 \to D\pi\pi/D^*\pi$	6.16 ± 1.01	6.63 ± 1.09
$\overline{B} \to D_0^*\ell^-\overline{\nu}_\ell$	$D_0^* \to D\pi$	3.90 ± 0.70	4.20 ± 0.75
	$D_0^* \to D\pi\pi$	0.29 ± 0.29	0.31 ± 0.31
	$D_0^* \to D^*\pi\pi$	1.00 ± 1.00	1.08 ± 1.08
	$D_0^* \to D\eta$	4.09 ± 4.09	3.77 ± 3.77
$\overline{B} \to D_1'\ell^-\overline{\nu}_\ell$	$D_1' \to D^*\pi$	3.90 ± 0.84	4.20 ± 0.90
	$D_1' \to D\pi\pi$	0.29 ± 0.29	0.31 ± 0.31
	$D_1' \to D^*\pi\pi$	1.00 ± 1.00	1.08 ± 1.08
	$D_1' \to D^*\eta$	4.09 ± 4.09	3.77 ± 3.77
$\overline{B} \to D_2^*\ell^-\overline{\nu}_\ell$	$D_2^* \to D^{(*)}\pi$	2.73 ± 0.30	2.93 ± 0.32
$\overline{B} \to D_sK^+\ell^-\overline{\nu}_\ell$		—	0.30 ± 0.14
$\overline{B} \to D_s^*K^+\ell^-\overline{\nu}_\ell$		—	0.29 ± 0.19
$\overline{B} \to D\pi\ell^-\overline{\nu}_\ell$		$0 \quad (0.3 \pm 0.9)$	$0 \quad (0.3 \pm 0.9)$
$\overline{B} \to D^*\pi\ell^-\overline{\nu}_\ell$		$0\ (-1.1 \pm 1.1)$	$0\ (-1.1 \pm 1.1)$
$\overline{B} \to D_1\tau^-\overline{\nu}_\tau$	$D_1 \to D\pi\pi/D^*\pi$	0.52 ± 0.52	0.56 ± 0.56
$\overline{B} \to D_0^*\tau^-\overline{\nu}_\tau$	$D_0^* \to D\pi$	0.33 ± 0.33	0.36 ± 0.36
	$D_0^* \to D^*\pi\pi$	0.17 ± 0.17	0.18 ± 0.18
	$D_0^* \to D\eta$	0.35 ± 0.35	0.32 ± 0.32
$\overline{B} \to D_1'\tau^-\overline{\nu}_\tau$	$D_1' \to D^*\pi$	0.33 ± 0.33	0.36 ± 0.36
	$D_1' \to D\pi\pi$	0.05 ± 0.05	0.05 ± 0.05
	$D_1' \to D^*\eta$	0.35 ± 0.35	0.32 ± 0.32
$\overline{B} \to D_2^*\tau^-\overline{\nu}_\tau$	$D_2^* \to D^{(*)}\pi$	0.23 ± 0.23	0.25 ± 0.25

Table 4.4: Branching fractions of hadronic B decays in the simulation, taken from the PDG [1].

Decay	Branching fraction $[10^{-3}]$	
	$\mathcal{B}(B^0)$	$\mathcal{B}(B^+)$
$\overline{B} \to D^* D_s^-$	8.0 ± 1.1	8.2 ± 1.7
$\overline{B} \to D^* D_s^{*-}$	17.7 ± 1.4	17.1 ± 2.4
$\overline{B} \to D^* {D_{s0}^*}^-$	1.5 ± 0.6	9.0 ± 7.0
$\overline{B} \to D^* {D'_{s1}}^-$	0.83 ± 0.28	0.94 ± 0.42
$\overline{B} \to D^* D^- K^0$	6.4 ± 0.5	2.1 ± 0.5
$\overline{B} \to D^* D^{*-} K^0$	8.1 ± 0.7	9.2 ± 1.2
$\overline{B} \to D^* \pi^-$	2.74 ± 0.13	4.90 ± 0.17
$\overline{B} \to D^* \rho^-$	6.8 ± 0.9	9.8 ± 1.7
$\overline{B} \to D^* \pi^- \pi^0$	8.2 ± 0.5	—
$\overline{B} \to D^* a_1^-$	13.0 ± 0.27	19 ± 5
$\overline{B} \to D^* \pi^+ \pi^- \pi^- \pi^0$	17.6 ± 2.7	18 ± 4
$\overline{B} \to D^* \pi^+ \pi^- \pi^+ \pi^- \pi^-$	—	2.6 ± 0.4

Table 4.5: Branching fractions of leptonic τ decays in the simulation.

Decay	Branching fraction [%]
$\tau^- \to e^- \nu_\tau \overline{\nu}_e$	17.82 ± 0.04
$\tau^- \to \mu^- \nu_\tau \overline{\nu}_\mu$	17.39 ± 0.04

4.2 Categorization of reconstructed events

In our simulation, we classify the reconstructed candidates into four main event categories as follows:

1. **Signal events**
   This category includes correctly reconstructed $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$ events. When $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$ candidates accompany a misidentified lepton candidate from a τ decay which passes the requirement of the lepton identification, they are also considered as the signal events.

2. **Normalization events**
   Correctly reconstructed $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ events are grouped under this category.

3. **Background events with a correctly reconstructed D^* candidate**
   In these events, a D^* candidate is reconstructed correctly, but the decay of the B_{sig} meson is incorrectly reconstructed. They are further divided into three subcategories:

(a) $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$
These semileptonic B decays $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ are the primary contributors in this background events. Missing daughter particles, like π^0, in the decays of excited charm meson can result in a signal-like event topology with a higher missing mass. $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ events are also included in this subcategory.

(b) Hadronic B decays
The background events from hadronic B decays, predominantly $\overline{B} \to D^*\overline{D}_s^{(*)}$, $\overline{B} \to D^*\overline{D}^{(*)}K$, and $\overline{B} \to D^*n\pi$ decays, are categorized in this subcategory.

(c) "Other" background events
This subcategory encompasses miscellaneous minor background contributions, such as:

- $B^0 \leftrightarrow B^+$ cross-feeds in semileptonic B decays, where the charge of $B_{\rm tag}$ is incorrectly assigned due to mis-reconstruction involving missing charged particles.
- Cross-feed events within $B\overline{B}$, where daughters of $B_{\rm tag}$ and $B_{\rm sig}$ are interchanged, with correct reconstructions of D^* and ℓ, but one of these particles originating from $B_{\rm tag}$.
- Events from the continuum containing a correctly reconstructed D^* candidate.

The first two components are predominant in these "other" background events.

4. **Background events with a wrongly reconstructed D^* candidate**
This category is a dominant source of background in this measurement. In these events, $B_{\rm sig}$ decays are misreconstructed with a wrong D^* candidate. The yields of these events are adjusted and constrained using data from the side-band regions.

4.3 Reconstruction

4.3.1 Reconstruction of tag-side B meson

Candidate $B_{\rm tag}$ mesons are reconstructed from tracks and photon clusters passing the basic selections using tracking parameters and kinematics in Table 4.6. All tracks used to reconstruct the tag-side B mesons must have absolute impact parameters, $|d_0|$ and $|z_0|$, less than 0.5 cm and 2.0 cm, respectively. Here, d_0 and z_0 are signed distances in the r-ϕ plane and z coordinate, respectively, to the point of closest approach to the nominal beam interaction point. The transverse momentum, $p_{\rm T}$, is above 0.1 GeV/c. Photon clusters are selected with energy above 0.1 GeV in the CDC acceptance, $-0.8660 < \cos\theta < 0.9563$.

Events must contain at least five tracks and at least three energy deposits in the ECL, where the latter includes electromagnetic showers associated with charged tracks. Total visible energy, defined as the sum of all the measured energy of the tracks and neutral clusters, which are electromagnetic showers not associated with charged tracks, needs to be larger than 4.0 GeV in an event. This selection can exclude events from two-photon processes. We also require the ratio of the second to zeroth Fox-Wolfram moment [46], R_2, to be less than 0.4. This selection can suppress $q\bar{q}$ events.

Table 4.6: Requirements for the final state particles in the tag-side B reconstruction.

Particles	Selections
Track	$\lvert d_0\rvert < 0.5$ cm, $\lvert z_0\rvert < 2$ cm, $p_{\rm T} > 0.1$ GeV/c
Photons	$E > 0.1$ GeV, $-0.8660 < \cos\theta < 0.9563$

We apply selections using $M_{\rm bc}$ and ΔE to $B_{\rm tag}$ candidates. $M_{\rm bc}$ is the beam-constrained mass, defined by

$$M_{\rm bc} = \sqrt{\left(\frac{E^*_{e^+e^-}}{2}\right)^2 - \left|\vec{p}^{\,*}_{B_{\rm tag}} c\right|^2}/c^2, \tag{4.2}$$

where $E^*_{e^+e^-}$ is the measured collision energy in the c.m. frame. ΔE is the energy difference from the half of the measured collision energy, $\Delta E = E^*_{B_{\rm tag}} - E^*_{e^+e^-}/2$, where $E^*_{B_{\rm tag}}$ is the c.m. energy of the $B_{\rm tag}$ candidate. The correctly reconstructed B mesons peak around the nominal B meson mass and zero in the $M_{\rm bc}$ and ΔE distributions, respectively. We set selections for the $B_{\rm tag}$ candidates as $M_{\rm bc} > 5.27$ GeV/c^2 and $-0.15 < \Delta E < 0.10$ GeV. Finally, we select the $B_{\rm tag}$ candidates with $\mathcal{P}_{\rm FEI} > 0.01$. The efficiencies of the tag-side B mesons are approximately 0.23% and 0.30% for B^0 and B^+ candidates at a purity of 29%, respectively [34].

4.3.2 Reconstruction of signal-side B meson

Selections applied to final state particles used for the signal-side B reconstruction are summarized in Table 4.7 and the details are described as follows.

All tracks for signal-side B mesons are reconstructed with a selection of impact parameters, $dr < 2$ cm and $\lvert dz\rvert < 4$ cm. Here, dr and dz are signed distances in the r-ϕ plane and z coordinate respectively to the point of closest approach to the measured beam interaction point. The transverse momentum, $p_{\rm T}$, is required to be above 0.1 GeV/c with an exception for low-momentum π^+ mesons arising from D^{*+} decays. Candidates of π^+ (K^+) mesons are selected with $\mathcal{P}_{\pi(K)} = \mathcal{L}_{\pi(K)}/(\mathcal{L}_\pi + \mathcal{L}_K) > 0.1$, where $\mathcal{L}_{\pi(K)}$ represents the likelihood of π^+ (K^+) calculated from SVD, CDC, TOP, ARICH, ECL, and KLM hits. We require that the number of hits associated with a track in the CDC exceeds 20 to ensure adequate dE/dx information for accurate identification. For electron and muon candidates, identification is based on likelihood ratios: $\mathcal{P}_{e(\mu)} = \mathcal{L}_{e(\mu)}/(\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d) > 0.9$. These likelihood values are constructed for electron, muon, pion, kaon, proton, and deuteron with CDC, (TOP,) ARICH, ECL, and KLM information. For low-momentum π^+ mesons from D^{*+} decays, referred to as $\pi^+_{\rm slow}$, we implement a lower momentum threshold of $p > 0.05$ GeV/c, and the π^+ identification criterion is not applied in these cases.

All photon candidates must satisfy the following criteria: the number of cluster hits in ECL, $N_{\rm hit,ECL}$, should be greater than 1.5, and the polar angle must fall within the CDC and ECL acceptance ranges, which are $0.8536 < \cos\theta < 0.9563$, $-0.6252 < \cos\theta < 0.8462$, or $-0.8660 < \cos\theta < -0.6521$. Additionally, the difference in cluster timing from the event

Table 4.7: Requirements for the final state particles in the signal-side B reconstruction.

Particles	Selections
π^+ (K^+)	$dr < 2$ cm, $\lvert dz\rvert < 4$ cm, $p_{\rm T} > 0.1$ GeV/c, $\mathcal{P}_{\pi(K)} > 0.1$, $N_{\rm hit,CDC} > 20$
$\pi^+_{\rm slow}$	$dr < 2$ cm, $\lvert dz\rvert < 4$ cm, $p > 0.05$ GeV/c
e^- (μ^-)	$dr < 2$ cm, $\lvert dz\rvert < 4$ cm, $p_{\rm T} > 0.1$ GeV/c, $\mathcal{P}_{e(\mu)} > 0.9$
γ (except for $\pi^0_{\rm slow}$)	$N_{\rm hit,ECL} > 1.5$, $\lvert\Delta T_\gamma\rvert < 200$ ns, $\begin{cases} E_\gamma > 0.080 \text{ GeV} & (0.8536 < \cos\theta < 0.9563), \\ E_\gamma > 0.030 \text{ GeV} & (-0.6252 < \cos\theta < 0.8462), \\ E_\gamma > 0.060 \text{ GeV} & (-0.8660 < \cos\theta < -0.6521) \end{cases}$
γ for $\pi^0_{\rm slow}$	$N_{\rm hit,ECL} > 1.5$, $\lvert\Delta T_\gamma\rvert < 200$ ns, $\begin{cases} E_\gamma > 0.025 \text{ GeV} & (0.8536 < \cos\theta < 0.9563), \\ E_\gamma > 0.025 \text{ GeV} & (-0.6252 < \cos\theta < 0.8462), \\ E_\gamma > 0.040 \text{ GeV} & (-0.8660 < \cos\theta < -0.6521) \end{cases}$
π^0	$0.120 < M_{\pi^0} < 0.145$ GeV/c^2
$\pi^0_{\rm slow}$	$0.105 < M_{\pi^0} < 0.150$ GeV/c^2
K^0_S	`ksSelector`: $P_{K^0_S} > 0.90$, $P_{\Lambda^0\text{-veto}} > 0.11$

time t_0, denoted as ΔT_γ, should be less than 200 ns. This event time is measured by SVD and ECL.

Neutral pion π^0 candidates are reconstructed by combining two photon candidates for the decay $\pi^0 \to \gamma\gamma$. These candidates are required to have a mass between 0.120 and 0.145 GeV/c^2. The π^0 mesons originating from D^* decays, denoted as $\pi^0_{\rm slow}$, are characterized by low momentum, and we relax the photon energy selection criteria and broaden the π^0 mass range for those candidates as follows. The minimum energy threshold for photon candidates is set at 0.025 (0.040) GeV for the forward endcap and barrel (backward endcap) regions, respectively. We accept $\pi^0_{\rm slow}$ candidates with π^0 mass between 0.105 and 0.150 GeV/c^2.

Candidates for K^0_S mesons are selected by employing a specialized multi-variate-analysis tool, `ksSelector`, developed by the Belle II Collaboration. A `FastBDT` classifier [33] is optimized to reconstruct K^0_S candidates from $\pi^+\pi^-$. The classifier utilizes kinematics of the K^0_S and its daughter pions, the flight length of K^0_S, and the number of hits in PXD and SVD as input variables. The classifier yields two K^0_S probabilities, referred to as $P_{K^0_S}$ and $P_{\Lambda^0\text{-veto}}$, discriminating correct K^0_S candidates from $\pi^+\pi^-$ combinatorial background and Λ^0 baryon decays to $p^+\pi^-$, respectively. The criteria $P_{K^0_S} > 0.90$ and $P_{\Lambda^0\text{-veto}} > 0.11$ are imposed. These selection criteria achieve approximately 90% efficiency.

Table 4.8: List of reconstructed D^0 and D^+ decay modes.

Decay	Branching fraction
$D^0 \to K^-\pi^+\pi^0$	$(14.4 \pm 0.6)\%$
$D^0 \to K^-\pi^+\pi^-\pi^+$	$(8.23 \pm 0.14)\%$
$D^0 \to K^0_S\pi^+\pi^-\pi^0$	$(5.2 \pm 0.6)\%$
$D^0 \to K^-\pi^+$	$(3.947 \pm 0.030)\%$
$D^0 \to K^0_S\pi^+\pi^-$	$(2.80 \pm 0.18)\%$
$D^0 \to K^0_S\pi^0$	$(1.240 \pm 0.020)\%$
$D^0 \to K^-K^+$	$(0.408 \pm 0.006)\%$
$D^0 \to \pi^-\pi^+$	$(0.1454 \pm 0.0024)\%$
$D^+ \to K^-\pi^+\pi^+$	$(9.38 \pm 0.16)\%$
$D^+ \to K^0_S\pi^+$	$(1.56 \pm 0.03)\%$
$D^+ \to K^-K^+\pi^+$	$(0.968 \pm 0.018)\%$

We reconstruct D mesons from the decay modes listed in Table 4.8. The D candidates must have an invariant mass M_D between 1.80 and 1.90 GeV/c^2. Candidate D^* mesons are reconstructed through $D^{*+} \to D^0\pi^+$, $D^{*+} \to D^+\pi^0$, and $D^{*0} \to D^0\pi^0$ decays. The mass difference between reconstructed D^* and D candidates, defined by $\Delta M_{D^*} = M_{D^*} - M_D$, are required to be within $0.100 < \Delta M_{D^*} < 0.190$ GeV/c^2.

The signal-side B mesons are reconstructed by a D^* and a lepton candidate. We form the $\Upsilon(4S)$ candidate by combining the tag-side and signal-side B mesons in the three sets of charges: $(B^0_{\rm tag}, \bar{B}^0_{\rm sig})$, $(B^0_{\rm tag}, B^0_{\rm sig})$, and $(B^+_{\rm tag}, B^-_{\rm sig})$.

Particles in the ROE are reconstructed using selections listed in Table 4.9. We require no additional tracks in the ROE, which are searched with selections of $dr < 5$ cm, $|dz| < 10$ cm, $p_{\rm T} > 0.1$ GeV/c, and $N_{\rm hit,CDC} > 0$. Additional photons are accepted in the ROE and are summed as the extra energy in ECL, $E_{\rm ECL}$. π^0 candidates are also reconstructed from the $\pi^0 \to \gamma\gamma$ decay. For those daughter photons, we require that the ratio of energies in the central crystal, E_1, to 3×3 crystals around the central crystal, E_9, is larger than 0.4 in addition to the selection criteria used for π^0 candidates from D decays. Since the two photons are back-to-back in the π^0 rest frame, we require daughter angles: $\cos\Delta\phi_{\gamma\gamma} > 0.5403$ and $\cos\Delta\psi_{\gamma\gamma} > 0.6216$, where $\Delta\phi_{\gamma\gamma}$ and $\Delta\psi_{\gamma\gamma}$ are differences of ϕ and three-dimensional angles between two photons of π^0 daughters in the laboratory frame, respectively.

4.4 Selection optimization

After the primary reconstruction of the $\Upsilon(4S)$, additional cuts are applied to further optimize the selection. For the optimization, we maximize the figure of merit (FOM) defined as a statistical significance of $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ candidates in Eq. (4.3).

$$\text{FOM} = \frac{N_{\rm sig}}{\sqrt{N_{\rm sig} + N_{\rm bkg}}}, \tag{4.3}$$

Table 4.9: Requirements for the final state particles in the Rest of Event.

Particles	Selections
Track	$\lvert d_0\rvert < 5$ cm, $\lvert z_0\rvert < 10$ cm, $p_{\rm T} > 0.1$ GeV/c $N_{\rm hit,CDC} > 0$
γ (except for $\pi^0_{\rm ROE}$)	$N_{\rm hit,ECL} > 1.5$, $\lvert\Delta T_\gamma\rvert < 200$ ns, $d_{{\rm track}-\gamma} > 20$ cm, $\begin{cases} E_\gamma > 0.080 \text{ GeV} & (0.8536 < \cos\theta < 0.9563), \\ E_\gamma > 0.030 \text{ GeV} & (-0.6252 < \cos\theta < 0.8462), \\ E_\gamma > 0.060 \text{ GeV} & (-0.8660 < \cos\theta < -0.6521) \end{cases}$
γ for $\pi^0_{\rm ROE}$	$N_{\rm hit,ECL} > 1.5$, $\lvert\Delta T_\gamma\rvert < 200$ ns, $E_1/E_9 > 0.4$, $\begin{cases} E_\gamma > 0.120 \text{ GeV} & (0.8536 < \cos\theta < 0.9563), \\ E_\gamma > 0.030 \text{ GeV} & (-0.6252 < \cos\theta < 0.8462), \\ E_\gamma > 0.080 \text{ GeV} & (-0.8660 < \cos\theta < -0.6521) \end{cases}$
$\pi^0_{\rm ROE}$	$0.112 < M_{\pi^0} < 0.142$ GeV/c^2 $\cos\Delta\phi_{\gamma\gamma} > 0.5403$, $\cos\Delta\psi_{\gamma\gamma} > 0.6216$

where $N_{\rm sig}$ ($N_{\rm bkg}$) denotes the number of properly reconstructed (misreconstructed) $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ candidates in the region of $M^2_{\rm miss} > 0.5$ GeV/c^2. $N_{\rm bkg}$ includes the number of correctly reconstructed $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$ candidates.

The $E_{\rm ECL}$ quantity and the square of momentum transfer to the lepton system $\ell^-\overline{\nu}_\ell$ in the signal B decays, q^2, are required to be less than 2.0 GeV and larger than 4.0 (GeV/c)^2, respectively. The following selection criteria are optimized for candidates at this signal region after these requirements.

Two selections are applied to reduce fake D^* candidates, which are the largest background contribution in this analysis: Each event must contain less than 21 clusters, all of which satisfy selections for the π^0; the positive χ probability of the second vertex fit on the signal B meson is required to allow only B candidates with a successful fit.

4.4.1 γ selections

Additional requirements are applied to the photons used in the reconstruction of the D or D^* candidates. These requirements are optimized using simulation according to the FOM. The selections for photons consist of two variables: the distance between the cluster and the nearest track hitting the ECL, $\Delta_{\rm TC}$, and output of a multi-variate analysis [47] which employs the first eleven Zernike moments [48, 49] determined for each cluster, Z_0. The latter variable is originally developed to distinguish between clusters belonging to real photons and those that result from K^0_Ls. The selection is determined using a quadratic function:

$$\left(\frac{\Delta_{\rm TC}}{X \text{ cm}}\right)^2 + \left(\frac{Z_0}{Y}\right)^2 > 1.0. \tag{4.4}$$

Table 4.10: Parameter values of the cut function Eq. 4.4 for γ selections. $\gamma_{\text{high(low)}}$ denotes either photons of π^0 daughters with higher(lower) energy.

Candidates		cluster region	Parameter values	
			X [cm]	Y
π^0 daughters in D decays	γ_{high}	Forward	85	0.65
		Barrel	85	0.60
		Backward	85	0.55
	γ_{low}	Forward	85	0.45
		Barrel	85	0.45
		Backward	40	0.65
π^0 daughters in D^* decays	γ_{high}	Forward	70	0.45
		Barrel	70	0.30
		Backward	60	0.45
	γ_{low}	Forward	40	0.70
		Barrel	40	0.40
		Backward	85	0.20
D^{*0} daughter in $D^* \to D^0\gamma$		Forward	85	0.55
		Barrel	85	0.40
		Backward	50	0.55

The optimal values of X and Y are chosen depending on cluster regions or decay mothers as listed in Table 4.10. For daughters from a π^0 candidate, different criteria are adopted between γ candidates with the higher and lower energy.

4.4.2 π^0 selections

The π^0 candidates from $D^{(*)}$ should pass a selection using the difference of reconstructed mass from the known π^0 mass [1], $|M_{\pi^0} - M_{\pi^0}^{\text{PDG}}|$. It is required to be within $1.2\sigma_{M_{\pi^0}}$ for π^0 candidates from D mesons and $1.5\sigma_{M_{\pi^0}}$ for π^0_{slow} candidates from D^* mesons, respectively. The resolution of reconstructed π^0 mass $\sigma_{M_{\pi^0}}$ is determined by fitting a double-sided single Gaussian on correctly reconstructed π^0 candidates in the simulation. Table 4.11 lists the asymmetric $\sigma_{M_{\pi^0}}$ values given by the fits. The above $1.2\sigma_{M_{\pi^0}}$ and $1.5\sigma_{M_{\pi^0}}$ range of the mass difference provide the highest FOM in simulation.The resulting π^0 mass selections are between 122.4 and 143.0 MeV/c^2 and between 118.3 and 147.3 MeV/c^2, respectively. The mass range corresponds to an efficiency of 94.0% (87.5%) for correctly reconstructed π^0 (π^0_{slow}) candidates.

Furthermore, the selection for the π^0 candidates of the D daughter is also optimized using the energy asymmetry between the two-photon daughters, which is defined as

$$A_{E_\gamma} = \frac{E_{\gamma_{\text{high}}} - E_{\gamma_{\text{low}}}}{E_{\gamma_{\text{high}}} + E_{\gamma_{\text{low}}}}, \tag{4.5}$$

Table 4.11: Fitted values of asymmetric σs for π^0 candidates in $D^{(*)}$ decays. $\sigma_{L(H)}$ denotes a Gaussian sigma for the lower and higher side from $M_{\pi^0}^{\rm PDG}$.

Candidates	σ_L [GeV/c^2]	σ_H [GeV/c^2]
π^0 in D decays	$(8.39 \pm 0.03) \times 10^{-3}$	$(5.37 \pm 0.07) \times 10^{-3}$
$\pi^0_{\rm slow}$ in D^* decays	$(13.9 \pm 0.1) \times 10^{-3}$	$(10.0 \pm 0.3) \times 10^{-3}$

where $E_{\gamma_{\rm high(low)}}$ represents energy of the photon with the higher(lower) energy. The selection $A_{E_\gamma} < 0.65$ allows for a higher FOM without significant efficiency loss for signal and normalization decays.

4.4.3 K^0_S selections

Candidate K^0_S mesons are reconstructed in their $K^0_S \to \pi^-\pi^+$ decays. We employ a FastBDT classifier [33] to discriminate K^0_S candidates from $\pi^+\pi^-$ combinatorial background and Λ^0 baryon decays to $p\pi^-$. The classifier returns two K^0_S probabilities, referred to as $P_{K^0_S}$ and $P_{\Lambda^0\text{-veto}}$, respectively, utilizing the kinematic properties of the K^0_S and its daughter pions, the flight length of the K^0_S, and the number of hits in PXD and SVD as input variables. We require $P_{K^0_S} > 0.90$ and $P_{\Lambda^0\text{-veto}} > 0.11$. In addition, the reconstructed K^0_S invariant mass is required to be between 0.4768 and 0.5146 GeV/c^2 and its flight length from the IP less than 5.0 cm. The selection criteria for K^0_S candidates have an efficiency of approximately 90%.

4.4.4 $D^{(*)}$ selections

Mode-dependent mass cuts in terms of σ are applied on M_D and ΔM_{D^*} for each decay chain involving a D^* or D candidate as shown in Table 4.12. Here σ refers to the width of the M_D or ΔM_{D^*} distribution for each decay mode. We require the masses of π^0 and K^0_S from D mesons to be the known values [1] for reconstruction of M_{D^*} while we do not apply any mass constraints in the calculation of M_D. These cuts have been chosen such that the highest FOM can be achieved. A limit of 5σ is set as the maximum value when searching the optimal FOM. The size of σ for each M_D and ΔM_{D^*} is determined asymmetrically by fitting a single double-sided Gaussian function on these mass distributions in simulation. The mean of the Gaussian is fixed at the known mass (difference) value [1]. The left- and right-hand sigmas of the Gaussian function, represented as $\sigma_{\rm L}$ and $\sigma_{\rm R}$, respectively, are obtained as listed in Table 4.13.

4.5 Best candidate selection

There are multiple $\Upsilon(4S)$ candidates from an event after the selection. The best candidate in each event is identified by selecting only one candidate through the following five selections:

1. **$B_{\rm tag}$ probability selection**
   Candidates with the highest $B_{\rm tag}$ probability are kept, while the other candidates are discarded.

Table 4.12: Selected mass ranges on M_D and ΔM_{D^*} distributions.

D^* decays	D decays	M_D	ΔM_{D^*}
$D^{*+} \to D^0\pi^+$	$D^0 \to K^-\pi^+\pi^0$	2.9σ	4.4σ
	$D^0 \to K^-\pi^+\pi^-\pi^+$	3.1σ	5.0σ
	$D^0 \to K^0_S\pi^+\pi^-\pi^0$	3.3σ	4.7σ
	$D^0 \to K^-\pi^+$	5.0σ	5.0σ
	$D^0 \to K^0_S\pi^+\pi^-$	5.0σ	5.0σ
	$D^0 \to K^0_S\pi^0$	2.4σ	5.0σ
	$D^0 \to K^-K^+$	2.2σ	5.0σ
	$D^0 \to \pi^+\pi^-$	3.1σ	2.9σ
$D^{*+} \to D^+\pi^0$	$D^+ \to K^-\pi^+\pi^+$	2.8σ	2.8σ
	$D^+ \to K^0_S\pi^+\pi^0$	2.8σ	2.0σ
	$D^+ \to K^-\pi^+\pi^+\pi^0$	1.8σ	2.2σ
	$D^+ \to K^0_S\pi^+\pi^-\pi^+$	2.1σ	3.3σ
	$D^+ \to K^0_S\pi^+$	5.0σ	2.6σ
	$D^+ \to K^-K^+\pi^+$	3.5σ	1.0σ
	$D^+ \to K^0_SK^+$	1.4σ	2.0σ
$D^{*0} \to D^0\pi^0$	$D^0 \to K^-\pi^+\pi^0$	2.2σ	1.8σ
	$D^0 \to K^-\pi^+\pi^-\pi^+$	3.0σ	2.2σ
	$D^0 \to K^0_S\pi^+\pi^-\pi^0$	1.9σ	1.6σ
	$D^0 \to K^-\pi^+$	5.0σ	2.3σ
	$D^0 \to K^0_S\pi^+\pi^-$	4.0σ	1.6σ
	$D^0 \to K^0_S\pi^0$	1.9σ	1.9σ
	$D^0 \to K^-K^+$	1.9σ	3.0σ
	$D^0 \to \pi^+\pi^-$	1.4σ	1.5σ
$D^{*0} \to D^0\gamma$	$D^0 \to K^-\pi^+\pi^0$	1.8σ	1.2σ
	$D^0 \to K^-\pi^+\pi^-\pi^+$	2.4σ	1.2σ
	$D^0 \to K^0_S\pi^+\pi^-\pi^0$	1.6σ	1.1σ
	$D^0 \to K^-\pi^+$	3.7σ	1.1σ
	$D^0 \to K^0_S\pi^+\pi^-$	2.8σ	1.2σ
	$D^0 \to K^0_S\pi^0$	1.8σ	1.1σ
	$D^0 \to K^-K^+$	1.6σ	1.6σ
	$D^0 \to \pi^+\pi^-$	1.4σ	0.7σ

Table 4.13: M_D and ΔM_{D^*} resolutions.

$D^{(*)}$ decays	$\sigma_{\rm L}$ [MeV/c^2]	$\sigma_{\rm H}$ [MeV/c^2]
$D^0 \to K^-\pi^+\pi^0$	16.9	10.6
$D^0 \to K^-\pi^+\pi^-\pi^+$	10.3	7.44
$D^0 \to K^0_S\pi^+\pi^-\pi^0$	12.7	8.35
$D^0 \to K^-\pi^+$	9.71	6.32
$D^0 \to K^0_S\pi^+\pi^-$	7.30	4.86
$D^0 \to K^0_S\pi^0$	23.2	13.1
$D^0 \to K^-K^+$	6.51	5.61
$D^0 \to \pi^+\pi^-$	5.89	5.03
$D^+ \to K^-\pi^+\pi^+$	6.61	5.72
$D^+ \to K^0_S\pi^+\pi^0$	15.5	9.97
$D^+ \to K^-\pi^+\pi^+\pi^0$	13.4	9.11
$D^+ \to K^0_S\pi^+\pi^-\pi^+$	5.48	4.58
$D^+ \to K^0_S\pi^+$	5.15	4.37
$D^+ \to K^-K^+\pi^+$	4.80	4.33
$D^+ \to K^0_SK^+$	3.53	3.57
$D^{*+} \to D^0\pi^+$	0.625	0.763
$D^{*+} \to D^+\pi^0$	1.04	1.29
$D^{*0} \to D^0\pi^0$	1.18	1.46
$D^{*0} \to D^0\gamma$	9.94	7.02

2. D^* mode selection

Only candidates of the $D^{*+} \to D^0\pi^+$ mode exclusively retained when the neutral $B_{\rm sig}$ candidates are reconstructed through both $D^{*+} \to D^0\pi^+$ and $D^{*+} \to D^+\pi^0$ modes. This selection is based on their cross-feed rates between $D^{*+} \to D^0\pi^+$ and $D^{*+} \to D^+\pi^0$.

3. D^* candidate selection

- For the $D^{*+} \to D^0\pi^+$ mode, candidates with the highest χ^2 probability from the second vertex fit are selected.
- For the other D^* modes, candidates with the lowest $\chi^2(M_{\pi^0})$ for $\pi^0_{\rm slow}$ mass from D^* decays are chosen. This selection uses

$$\chi^2(M_{\pi^0}) \equiv \frac{(M_{\pi^0} - M^{\rm PDG}_{\pi^0})^2}{\sigma^2_{M_{\pi^0}}}, \tag{4.6}$$

where $M^{\rm PDG}_{\pi^0}$ is the known π^0 mass [1], and $\sigma_{M_{\pi^0}}$ is the resolution of M_{π^0}. The value of $\sigma_{M_{\pi^0}}$ is determined by fitting a single double-sided Gaussian function to the M_{π^0} distribution of correctly reconstructed π^0 candidates in simulation.

Table 4.14: Expected yields for each event categories and D^* modes at 189.26 fb^{-1} in simulation after all the selections. The SM $R(D^*)$ value of 0.254 is assumed for the yields of $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$. The uncertainties are only statistical.

Event category	Expected yields in MC		
	$D^{*+} \to D^0\pi^+$	$D^{*+} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
1. Signal events	47.9 ± 0.4	7.3 ± 0.1	47.6 ± 0.4
2. Normalization events	1041.0 ± 11.2	133.2 ± 4.3	927.2 ± 10.7
3. Background events with a correct D^*			
(a) $\overline{B} \to D^{**}\ell^-\overline{\nu}$	61.6 ± 2.2	9.0 ± 0.9	46.0 ± 2.0
(b) Hadronic B decay	38.8 ± 2.1	4.5 ± 0.8	21.5 ± 1.6
(c) "Other" background events	12.4 ± 1.1	1.9 ± 0.5	6.7 ± 0.9
4. Background events with a fake D^*	160.7 ± 5.2	27.6 ± 1.7	251.5 ± 4.8

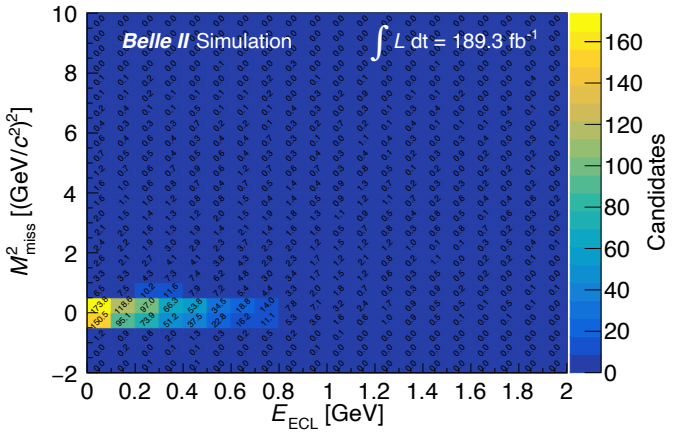
4. ***D* branching fraction selection**
   Preference is given only to candidates reconstructed from a single D mode with the highest branching fractions.

5. **Random selection**
   For any remaining candidates, a random selection process is applied.

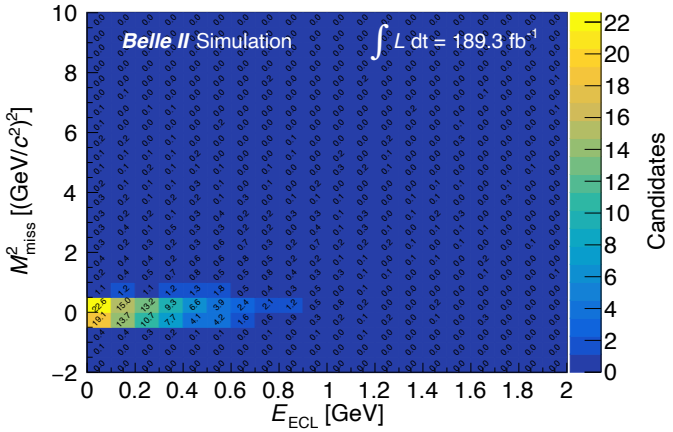
Through this approach, we systematically select the most promising candidate in each event.

4.6 Expected yields in simulation

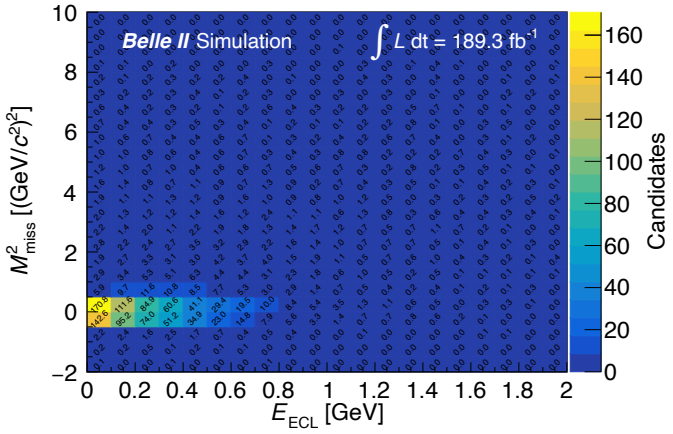
Table 4.14 summarizes the yields of candidates that pass all selections predicted in simulation at 189.26 fb^{-1}. Figure 4.1 displays the expected two-dimensional total distributions of $E_{\rm ECL}$ and $M^2_{\rm miss}$. Figures 4.2 to 4.4 present one-dimensional projections of these expected MC distributions to $E_{\rm ECL}$ or $M^2_{\rm miss}$. Calibration procedures outlined in Section 5 have been applied to these simulated results. For the expected number of $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ decays, we have adopted a $R(D^*)$ value of 0.254. Tables 4.15 and 4.16 provide detailed breakdowns of yields in the categories of signal events and "other" background events.



(a)



(b)



(c)

Figure 4.1: Expected two-dimensional distributions of E_{ECL} and M^2_{miss} from (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ in simulation. The figures in the plots are the number of entries in a bin.

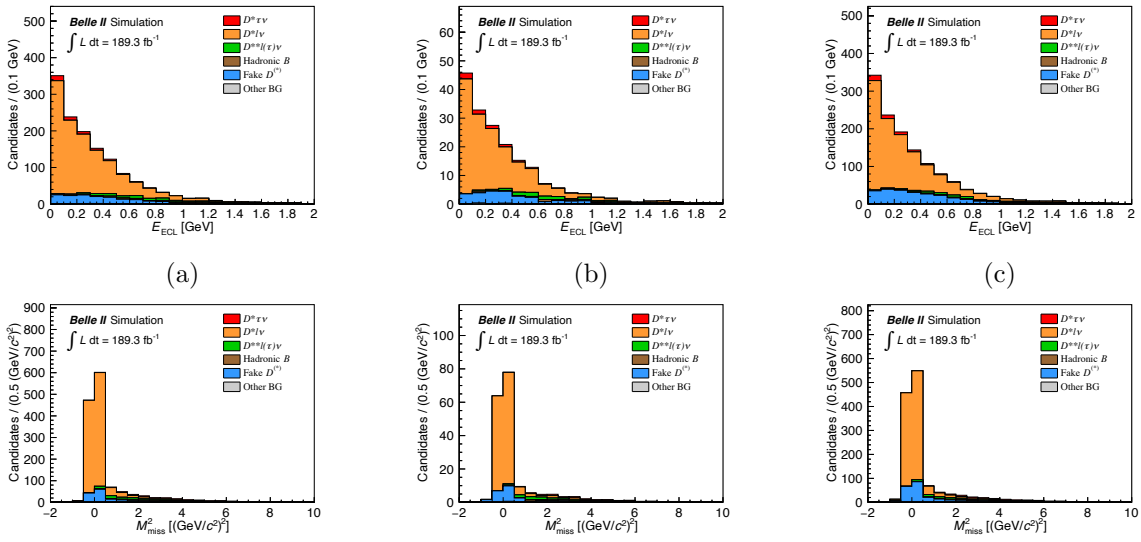


Figure 4.2: Expected distributions of (a)–(c) E_{ECL} and (d)–(f) M^2_{miss} in the $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ (right) modes of simulation.

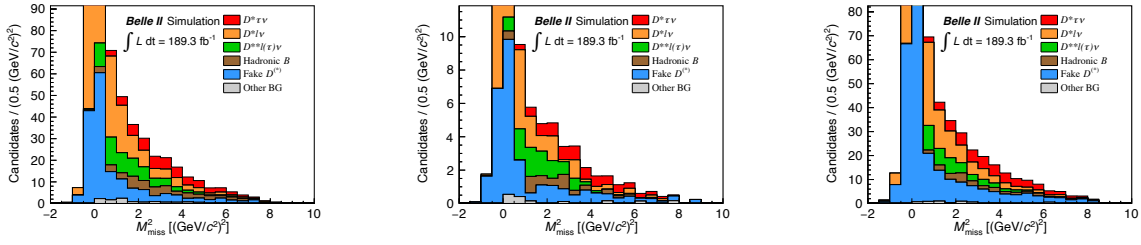


Figure 4.3: Distributions of M^2_{miss} with zoomed vertical axes in the (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ modes of the simulation.

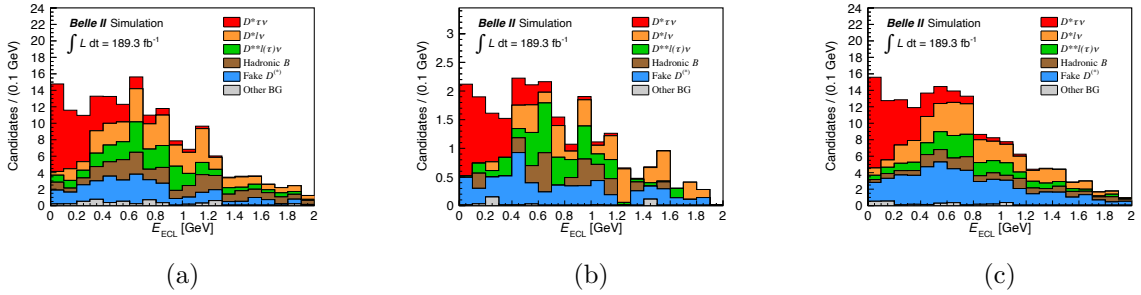


Figure 4.4: Distributions of E_{ECL} for candidates that pass a signal-enhanced selection of $1.5 < M^2_{\mathrm{miss}} < 6.0$ $(\mathrm{GeV}/c^2)^2$ in the (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ modes of simulation.

Table 4.15: Breakdown of expected yields in the category of the signal events at 189.26 fb^{-1} in simulation. The SM $R(D^*)$ value of 0.254 is assumed for yields of $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$. The uncertainties are only statistical.

	$D^{*+} \to D^0 \pi^+$	$D^{*+} \to D^+ \pi^0$	$D^{*0} \to D^0 \pi^0$
Correctly reconstructed signal events	44.9 ± 0.3	6.9 ± 0.1	44.6 ± 0.4
Signal events with a misidentified lepton	3.0 ± 0.1	0.5 ± 0.0	3.1 ± 0.1
Total	47.9 ± 0.4	7.3 ± 0.1	47.6 ± 0.4

Table 4.16: Breakdown of expected yields in the category of the "other" background events at 189.26 fb^{-1} in the simulation. The uncertainties are only statistical.

	$D^{*+} \to D^0 \pi^+$	$D^{*+} \to D^+ \pi^0$	$D^{*0} \to D^0 \pi^0$
$B^0 \leftrightarrow B^+$ cross feed	8.8 ± 0.9	1.6 ± 0.4	4.3 ± 0.7
"Other" $B\overline{B}$ background events	2.2 ± 0.4	0.3 ± 0.1	1.5 ± 0.3
Continuum events	1.5 ± 0.5	0.0 ± 0.0	0.9 ± 0.3
Total	12.4 ± 1.1	1.9 ± 0.5	6.7 ± 0.9

5 Calibration for simulation

This section summarizes the calibration for simulation. The calibration procedures from Sections **??** adopt correction factors officially provided by the Belle II collaboration. Calibration procedures from Sections **??** are developed specifically for this $R(D^*)$ measurement.

5.1 B_{tag} reconstruction efficiency correction

The hadronic FEI algorithm is trained using simulation. Since there are imperfections in the simulation, it induces differences in the performance of the FEI algorithm between the data and the simulation. To rectify these discrepancies, we correct the reconstruction efficiencies of the tag-side B mesons of hadronic decays using inclusive semi-leptonic B decays $\bar{B} \to X\ell^-\bar{\nu}_\ell$ on the signal side [34]. By comparing the yields between data and simulation, the correction factors are derived individually for neutral and charged B mesons, and also separated by a lepton flavor of the semileptonic B decay final state, an electron or muon. The obtained values are listed in Table 5.1. These correction factors are applied to all simulation events. The correction factors of $\overline{B} \to Xe^-\overline{\nu}_e$ and $\overline{B} \to X\mu^-\overline{\nu}_\mu$ are used for B_{sig} candidates reconstructed with an electron and muon candidate, respectively.

Table 5.1: Correction factors on the efficiency of the tag-side B mesons reconstructed through the FEI algorithm.

	Correction factor	
	$\overline{B} \to Xe^-\overline{\nu}_e$	$\overline{B} \to X\mu^-\overline{\nu}_\mu$
Neutral B_{tag}	0.710 ± 0.023	0.673 ± 0.025
Charged B_{tag}	0.686 ± 0.021	0.650 ± 0.024

5.2 Tracking correction

5.2.1 Tracking efficiency correction

The efficiency of tracks with momentum ranging from 0.2 to 3.5 GeV/c are assesed using $e^-e^+ \to \tau^-\tau^+$ events, where one τ decays leptonically and the other τ decays hadronically into three charged pions. These τ-pair events comprise four tracks in total. Out of them, a lepton candidate and two pion candidates are reconstructed so that the sum of the charges is ± 1. The efficiency of track reconstruction is then evaluated using the remaining track, which has a charge of ∓ 1. All the efficiencies are observed to be consistent between data and simulation as functions of transverse momentum, polar angle, and azimuthal angle, respectively. As a result, we do not implement any corrections but assign a systematic uncertainty of 0.3% to the reconstruction efficiencies for tracks above 0.2 GeV/c. This uncertainty is predominantly due to the charge asymmetry in the efficiencies.

For tracks with low momentum, below 0.2 GeV/c, the efficiencies are validated using $\overline{B} \to D^{*+}\pi^-$ decays. In these decays, D^{*+} candidates are reconstructed through $D^0\pi^+_{\text{slow}}$

Table 5.2: Correction factors for efficiencies of low-momentum tracks.

Momentum [GeV/c]	Correction factor
[0.05, 0.12]	0.909 ± 0.038
[0.12, 0.16]	1.033 ± 0.034
[0.16, 0.20]	0.971 ± 0.034
[0.20, 0.32]	1.000 (normalization)

decay, with $\pi^+_{\rm slow}$ typically being low-momentum. The $\pi^+_{\rm slow}$ candidates within the momentum range of 0.05 to 0.32 GeV/c are selected for this validation. The D^0 candidates are reconstructed from three decay modes: $K^-\pi^+$, $K^-\pi^+\pi^-\pi^+$, and $K^0_S\pi^+\pi^-$. The ΔE distributions are fitted to count reconstructed B candidates across four momentum regions of $\pi^+_{\rm slow}$. The correction factors for low-momentum tracking efficiency are calculated as the yield ratio of the data to the simulation and are normalized to the ratio at the highest $\pi^+_{\rm slow}$ momentum bin, from 0.20 to 0.32 GeV/c. This approach presupposes an absolute efficiency of unity in the highest momentum bin based on the measurement above 0.2 GeV/c using the $e^-e^+ \to \tau^-\tau^+$ events. The resulting relative efficiencies, serving as correction factors, are listed in Table 5.2. These factors are applied as weights to the low-momentum tracks in the simulation.

5.2.2 Momentum scale correction

We have identified imperfections in the magnetic field map utilized for data processing, leading to shifts in the momentum of reconstructed tracks. These momentum shifts are quantified using control channels involving a decay chain of $D^{*+} \to D^0\pi^+$ and $D^0 \to K^-\pi^+$. For this evaluation, we have tagged D^0 candidates using a D^{*+} candidate that satisfies ΔM_{D^*} between ± 1.5 MeV/c^2 from the nominal mass difference between D^{*+} and D^0 in the PDG [1]. To align the reconstructed mass peak of the D^0 candidates with the PDG mass, we apply a common global scale factor to the reconstructed momenta of all D^0 daughter particles. This scale factor is determined to be 0.99971 and it is cross-verified using the decay channels, $D^0 \to K^-\pi^+\pi^-\pi^+$, $D^+ \to K^-\pi^+\pi^+$, $J/\psi \to \mu^+\mu^-$, and $K^0_S \to \pi^+\pi^-$. Consequently, we adjust the magnitude of momenta for all tracks in the data by applying this common global scale factor.

5.3 Efficiency correction for particle reconstruction

5.3.1 Lepton identification efficiency

In order to correct the efficiencies and fake rates of lepton identification between data and MC, we use correction tables that describe dependencies on momentum and polar angle.

5.3.2 $K^\pm$ and $\pi^\pm$ identification efficiency

The pion and kaon ID efficiencies and the difference between data and MC are corrected based on correction tables. The correction is applied depending on the momentum and polar angle of the tracks.

Table 5.3: Correction factors on π^0 reconstruction efficiency.

Momentum range [GeV/c]	Correction factor	
	$\pi^0_{\rm slow}$ from D^* decays	π^0 from D decays
[0.05, 0.2]	1.000 ± 0.074	—
[0.2, 0.4]	0.978 ± 0.046	0.960 ± 0.061
[0.4, 0.6]	0.985 ± 0.041	0.985 ± 0.042
[0.6, 0.8]	1.030 ± 0.052	1.032 ± 0.053
[0.8, 1.0]	1.038 ± 0.045	1.033 ± 0.042
[1.0, 1.5]	1.056 ± 0.050	1.052 ± 0.047
[1.5, 2.0]	1.050 ± 0.050	1.060 ± 0.047
[2.0, 3.0]	1.032 ± 0.060	1.023 ± 0.048

5.3.3 π^0 reconstruction efficiency

We adopt correction factors for efficiencies of π^0 candidates depending on their momentum, as listed in Table 5.3. Different factors are applied to $\pi^0_{\rm slow}$ candidates from D^* decays and π^0 candidates from D decays.

5.4 Branching fraction correction of D decays

There are inconsistencies between the branching fractions of D decays implemented in the Belle II simulation and the corresponding world average reported by PDG [1]. To align the branching fractions with the average, we have derived correction factors by comparing the generated number of events in the simulation of 100 fb^{-1}, denoted as $N_{\rm gen}$, with the expected number of events based on the PDG values, $N_{\rm PDG}$. This calibration approach allows for the correction of branching fractions for D decays via intermediate states, such as ρ and K^*, inclusively. The correction factors are calculated as a ratio of $N_{\rm PDG}$ to $N_{\rm gen}$. The events are counted in a decay chain of $\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell$, followed by the decays $D^{*+} \to D^0\pi^+$ and $D^{*+} \to D^+\pi^0$ for the neutral and charged D modes, respectively. The values of $N_{\rm PDG}$ for each D^0 and D^+ decay are computed using Eq. (5.1) and Eq. (5.2), respectively. In these calculations, the branching fractions of D decay, $\mathcal{B}(D^{0,+} \to X)$, refer to the PDG, while the other branching fractions, $\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell$, $D^{*+} \to D^+\pi^0$, and $D^{*+} \to D^0\pi^+$, are taken from the values used in our simulation.

$$N_{\rm PDG}(D^0 \to X) = \begin{cases} 2N_{B\overline{B}} \cdot 2\mathcal{B}(\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell)^{\rm gen} \cdot \mathcal{B}(D^{*+} \to D^0\pi^+)^{\rm gen} \cdot \mathcal{B}(D^0 \to X), & (5.1) \\ 2N_{B\overline{B}} \cdot 2\mathcal{B}(\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell)^{\rm gen} \cdot \mathcal{B}(D^{*+} \to D^+\pi^0)^{\rm gen} \cdot \mathcal{B}(D^+ \to X). & (5.2) \end{cases}$$

Here, $\mathcal{B}^{\rm gen}$ represents the branching fractions in the simulation, and $N_{B\overline{B}}$ denotes the number of $B\overline{B}$ events. The counted numbers and the derived correction factors for each of the D decays are listed in Table 5.4. These factors are applied to simulation events that include a correctly reconstructed D candidate.

Table 5.4: Correction factors for branching fractions of D decays.

Decay	N_{gen} [10^3]	N_{PDG} [10^3]	Correction factor
$D^0 \to K^-\pi^+\pi^0$	1,016	1,086	0.936
$D^0 \to K^-\pi^+\pi^-\pi^+$	580.8	690.4	0.841
$D^0 \to K^0_S\pi^+\pi^-\pi^0$	253.6	438.4	0.578
$D^0 \to K^-\pi^+$	278.8	279.6	0.997
$D^0 \to K^0_S\pi^+\pi^-$	136.5	147.5	0.926
$D^0 \to K^0_S\pi^0$	60.47	64.20	0.942
$D^0 \to K^-K^+$	28.80	28.88	0.997
$D^0 \to \pi^+\pi^-$	10.27	10.22	1.005
$D^+ \to K^-\pi^+\pi^+$	300.2	297.4	1.009
$D^+ \to K^0_S\pi^+$	34.54	36.96	0.935
$D^+ \to K^-K^+\pi^+$	30.98	31.50	0.984

5.5 Calibration of the M_D and ΔM_{D^*} selection range

The mass resolution of $D^{(*)}$ mesons is known to be different between data and simulation, which is attributable to imperfections in the simulation of the detector responses. To address this, we scale the range of M_D and ΔM_{D^*} by a correction factor, r_{range}, for each $D^{(*)}$ mode. This eventually calibrates the signal yields of $D^{(*)}$ candidates in the data. The correction factors are determined through fits on M_D and ΔM_{D^*} distributions. For this calibration, we prepare samples with all selections except for the M_D or ΔM_{D^*} selections.

Table 5.5 and Table 5.6 detail the functions employed to model the signal and background distributions for the individual D and D^* modes. The shape parameters of these functions for signal (background) are determined from simulations using correctly reconstructed (fake) D^* candidates. To define the signal shapes, we use bins with percentile from 0.5% to 99.5% in the M_D and ΔM_{D^*} distributions. It minimizes the shape distortion from the distribution tails. In these signal models, the mean of either the triple Gaussian or the Crystal Ball function is anchored to the known masses [1].

The fake D^* candidates exhibit a notable peaking structure in the ΔM_{D^*} distributions. Therefore, we combine components of the peaking and the smooth backgrounds. The shape of the smooth background component is initially fixed using only the ΔM_{D^*} sideband regions defined in Table 5.7 for each D^* mode. The modeling function for this component is given by the threshold function in Eq. (5.3):

$$f(\Delta M_{D^*}|M_\pi^{\text{PDG}}, A, B, C) = \left(1 - \exp\left(-\frac{\Delta M_{D^*} - M_\pi^{\text{PDG}}}{A}\right)\right) \times \left(\frac{\Delta M_{D^*}}{M_\pi^{\text{PDG}}}\right)^B + C\left(\frac{\Delta M_{D^*}}{M_\pi^{\text{PDG}}} - 1\right). \quad (5.3)$$

Here M_π^{PDG} represents the known mass of a charged (neutral) pion for $D^* \to D^0\pi^+$ ($D^{*+} \to D^+\pi^0$ and $D^{*0} \to D^0\pi^0$) [1], while A, B, and C are shape parameters refined

Table 5.5: Modeling functions of the M_D distributions. A crystal ball function is allowed to have asymmetric tails.

Decays	Signal	Background
D decays with π^0	Triple Gaussian	1st Chebychev
D decays without π^0	Crystal ball + double Gaussian	1st Chebychev

Table 5.6: Modeling functions on ΔM_{D^*}

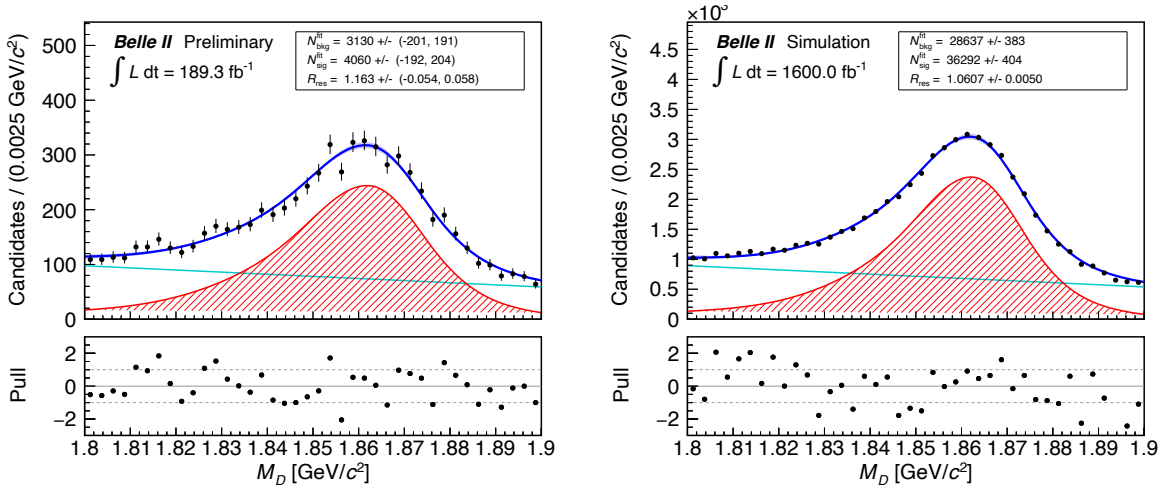
D^* decays	Signal	Background	
		Peaking	Smooth
$D^* \to D^0\pi^+$	Triple Gaussian	Double Gaussian	Threshold function
$D^* \to D^+\pi^0$	Triple Gaussian	Double Gaussian	Threshold function
$D^* \to D^0\pi^0$	Triple Gaussian	Double Gaussian	Threshold function

Table 5.7: ΔM_{D^*} sideband regions.

Decay	ΔM_{D^*} sideband region [GeV/c^2]
$D^{*+} \to D^0\pi^+$	$0.140 < \Delta M_{D^*} < 0.141$, $0.155 < \Delta M_{D^*} < 0.170$
$D^{*+} \to D^+\pi^0$	$0.135 < \Delta M_{D^*} < 0.137$, $0.150 < \Delta M_{D^*} < 0.170$
$D^{*0} \to D^0\pi^0$	$0.140 < \Delta M_{D^*} < 0.141$, $0.150 < \Delta M_{D^*} < 0.170$

using the fake D^* candidates in simulation. Subsequently, a shape of the peaking background component is fitted to fake D^* candidates on top of the pre-fixed shape of the smooth background component across the entire ΔM_{D^*} range. For this peaking background component, the mean of one of a double Gaussian is fixed to ΔM_{D^*} in the PDG. The observed peaking backgrounds are found to be mainly induced by the signal and normalization events. These peaking backgrounds contain a correctly reconstructed charged particle that is recognized as a fake candidate due to a software failure in its identification to the corresponding generated particle or π^0 candidates reconstructed with a photon from beam-induced backgrounds or hadronic showers. The uncertainty of this treatment for the peaking backgrounds is assigned as a systematic uncertainty described in Section 9.

The sum of the signal and background components are fitted on total M_D or ΔM_{D^*} distributions to deduce a global resolution factor, denoted as $R_{\rm res}^{\rm data}$ and $R_{\rm res}^{\rm MC}$ for data and simulation, respectively. The global resolution factor is multiplied commonly to all σs of Gaussian functions for both the signal and peaking background components. It thus adjusts the mass resolution of their peak shapes as determined by the fits. Figures 5.1 and 5.2 illustrate the fit results of the M_D and ΔM_{D^*} distributions for $D^0 \to K^-\pi^+\pi^0$ and $D^{*0} \to D^0\pi^+$ modes, respectively. Figure 5.3 compares those fit results between the data and the simulation at the same normalization. The correction factors are computed as



(a) (b)

Figure 5.1: Fit results of the M_D distributions for $D^0 \to K^-\pi^+\pi^0$ in (a) the data and (b) the simulation. The blue lines are the total fit functions as the sum of the signal and background components shown by the red and light blue lines.

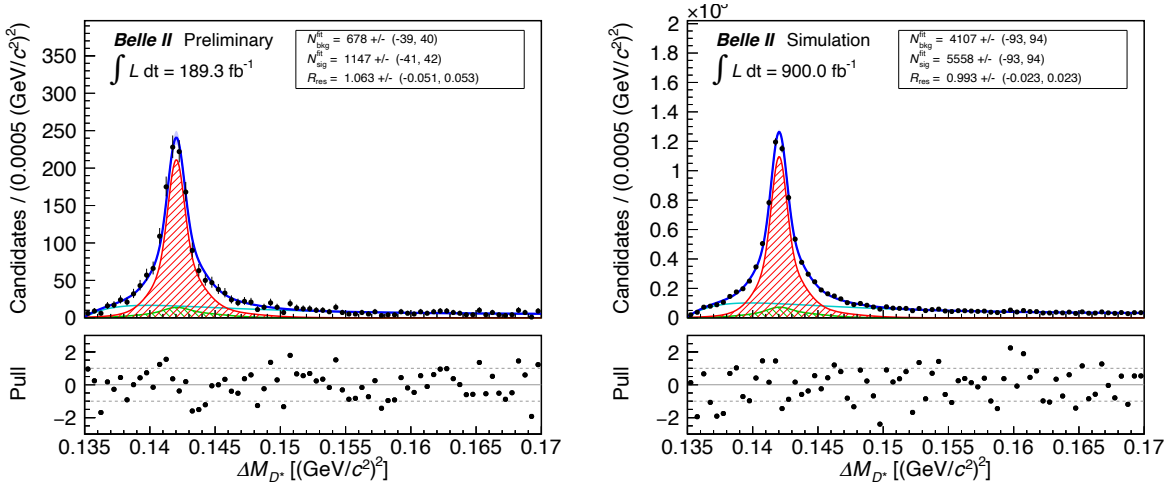
the ratio of these global resolution factors, defined as $r_{\text{range}} \equiv R_{\text{res}}^{\text{data}}/R_{\text{res}}^{\text{MC}}$. The resulting factors for M_D and ΔM_{D^*} are shown in Figures 5.4 and 5.5, respectively. The r_{range} ratios are then employed to adjust the selection windows for the signal region in the data, labeled as $\Delta m_{\text{data}}^{\text{SR}}$. This calibration is achieved by scaling the signal window range determined in the simulation, $\Delta m_{\text{MC}}^{\text{SR}}$, through Eq. (5.4).

$$\Delta m_{\text{data}}^{\text{SR}} = \Delta m_{\text{MC}}^{\text{SR}} \cdot r_{\text{range}} \tag{5.4}$$

5.6 Fake D^* yield correction

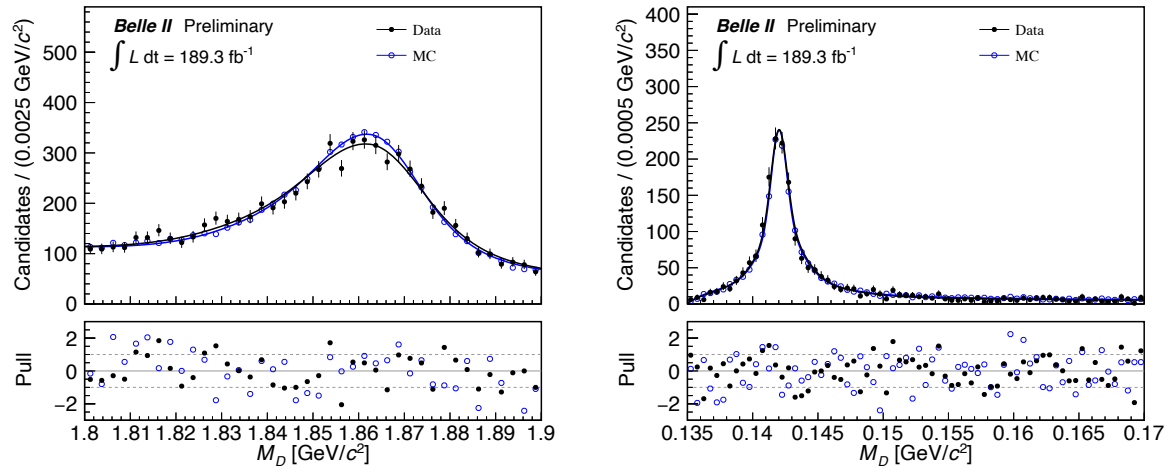
Following the calibration of the selection range of M_D and ΔM_{D^*}, we calibrate the yields of fake D^* components. In the sideband regions of ΔM_{D^*} specified in Table 5.7, which are dominated by the fake D^* candidates. The fake D^* yields are extracted by fitting the ΔM_{D^*} distributions. The fitting incorporates the same modeling functions for the smooth background component in Eq. (5.3). These function shapes are determined with simulation.

Fit results for $D^{*0} \to D^0\pi^0$ in the ΔM_{D^*} sideband regions are illustrated in Figure 5.6. All the fit results are shown in Appendix 11. The correction factors are computed as the data-simulation ratios of the yields, as presented in Figure 5.7. These correction factors are applied to constrain the fake D^* yield in the signal region for the $R(D^*)$ fit. Specifically for the $D^{*0} \to D^0\pi^0$ mode, we observe a slight dependence of the correction factor on M_{miss}^2. To accommodate the dependency, the calibration are segmented into three M_{miss}^2 intervals: $[-2, 1)$, $[1, 5)$, and $[5, 10]$ $(\text{GeV}/c^2)^2$.



(a) (b)

Figure 5.2: Fit results of the ΔM_{D^*} distributions for $D^{*0} \to D^0\pi^0$ in (a) the data and (b) the simulation. The blue lines are the total fit functions as the sum of the signal, peaking background, and smooth background components shown by the red, green, and light blue lines.



(a) (b)

Figure 5.3: Comparison of fit results of (a) the M_D and (b) ΔM_{D^*} distributions for $D^0 \to K^-\pi^+\pi^0$ and $D^{*0} \to D^0\pi^0$ between the data (black points) and the simulation (blue open points), respectively. The number of entries from the simulation is normalized to that of the data.

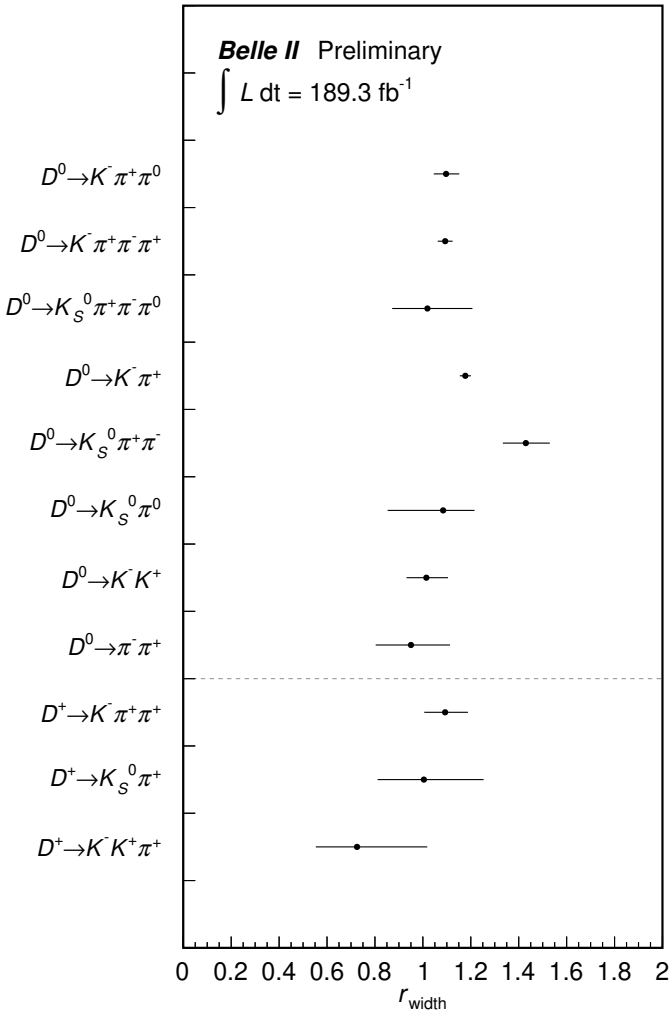


Figure 5.4: Correction factors r_{range} on the selection windows of the M_D signal region.

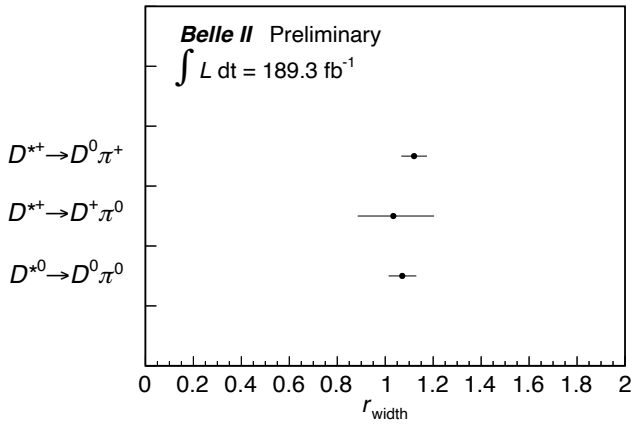


Figure 5.5: Correction factors r_{range} on the selection windows of the ΔM_{D^*} signal region.

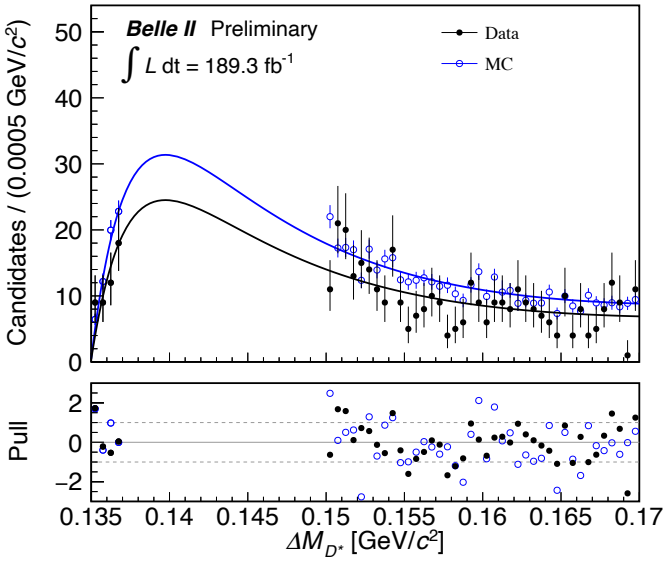


Figure 5.6: Fit results of the ΔM_{D^*} distributions in the ΔM_{D^*} sideband regions. The black (blue open) points are the data (the simulation). The solid lines show respective fitted threshold functions. The entries of the simulation are scaled to the luminosity of the data.

5.7 Shape correction for fit variables

The region $M^2_{\text{miss}} < 1$ (GeV/c^2)^2 is partially opened to calibrate the simulation, involving corrections to distributions of the final fit variables. This specific region is primarily populated by the normalization events and fake D^* background events from $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$. Two corrections are made with this region: background photon corrections in the ROE as E_{ECL} correction, and M^2_{miss} resolution correction.

5.7.1 E_{ECL} correction for background clusters

Significant discrepancies are observed in the E_{ECL} distributions between the data and the simulation, particularly in the $D^{*+} \to D^0 \pi^+$ mode, as illustrated in Figure 5.8. Differences are noted in the distributions of energy and multiplicity of photons in the ROE between the data and the simulation. The discrepancy is more significant in the low-energy regions, where photons induced by beam background and hadronic split-off showers predominantly distribute. We apply background photon corrections by shifting the energy of each of these clusters in two steps:

1. Beam-induced background photons
   The E_{ECL} distributions are analyzed by dividing the data set into two distinct data-taking periods before and after noon of April 16th, 2021 (JST), as shown in Figure 5.9. A more pronounced discrepancy is found in the earlier run period, where the condition of beam-induced background is less severe. These discrepancies are thought to arise from variations in the beam background across different run periods, which are not

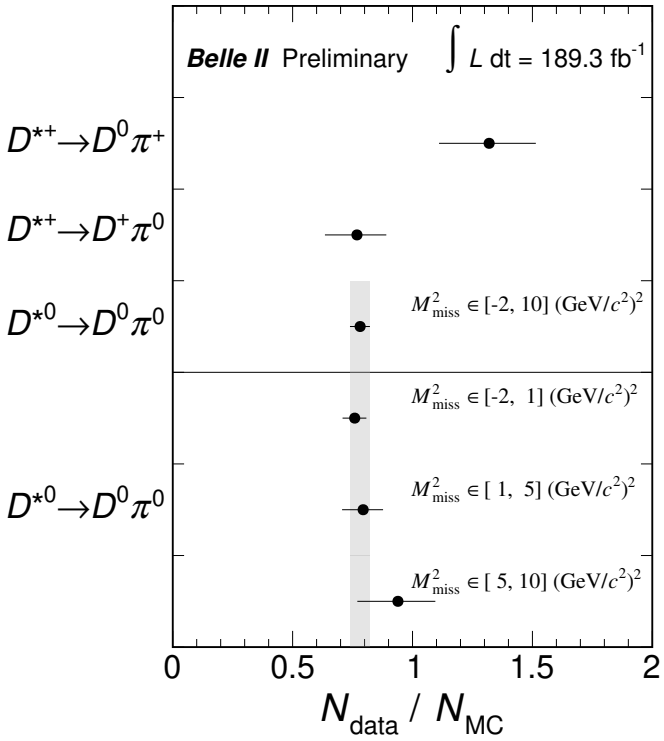
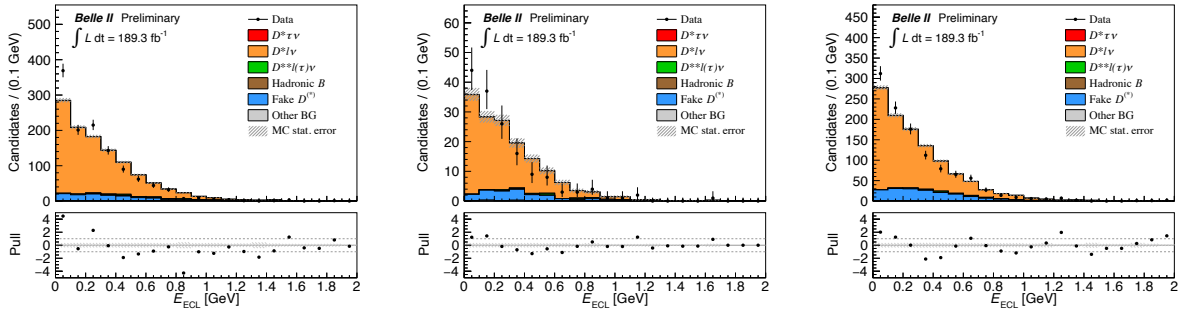


Figure 5.7: Correction factors for fake D^* yields. The correction factors evaluated in the full $M^2_{\rm miss}$ region are shown in the top panel. The $M^2_{\rm miss}$-dependent factors for $D^{*0} \to D^0\pi^0$ are in the bottom panel. The gray band is 1σ region of the overall factor for $D^{*0} \to D^0\pi^9$.

accounted for in our simulation. The remaining discrepancies are attributed to the modeling of photons from the hadronic split-off showers in the simulation.

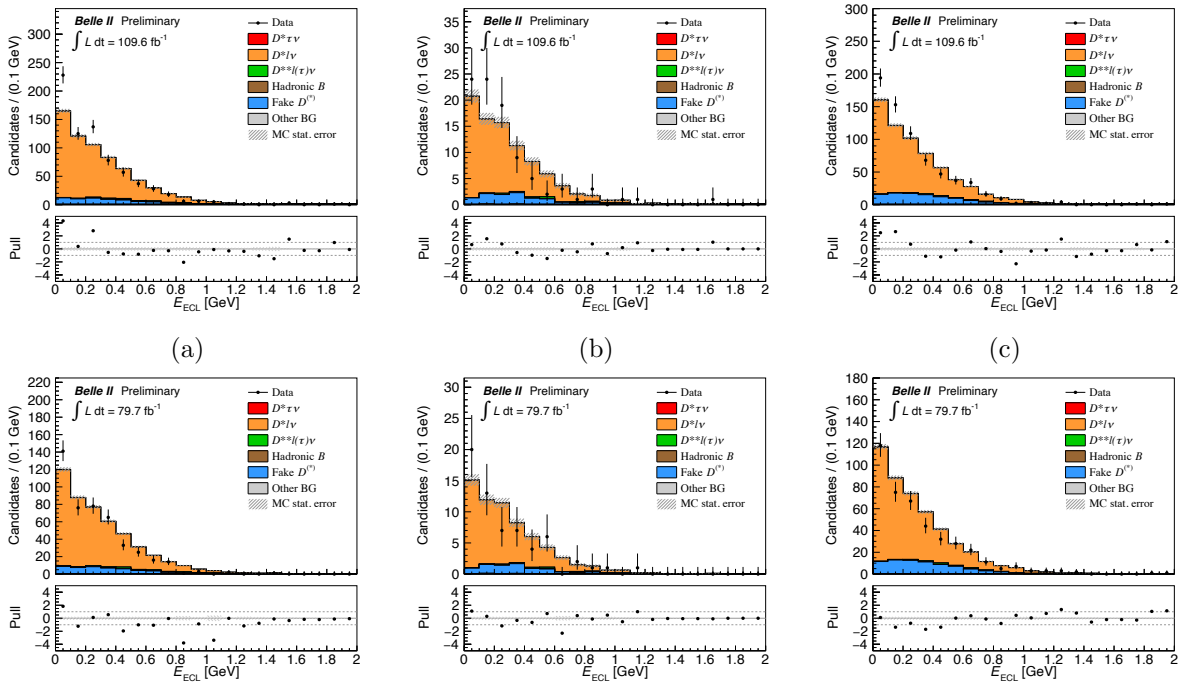
2. Photons produced by hadronic interactions in ECL
   To address the remaining discrepancies, the energy of each photon cluster from hadronic split-off showers is reduced by a certain amount. If the reduced energy falls below the polar-angle-dependent energy selection criteria for photons in the ROE, the clusters is discarded. We vary the energy shift for each cluster from 0.0 MeV to -30 MeV. The energy shift alters the $E_{\rm ECL}$ distribution, as demonstrated in Figure 5.10. The optimal energy shift is determined based on the χ^2 of the $E_{\rm ECL}$ distributions. The χ^2 and $\Delta\chi^2$ values in different energy shifts are presented in Figure 5.11. We adopt -15 MeV for a nominal energy shift, around which the summed χ^2 value of three D^* modes is minimized.

Other potential sources contributing to the discrepancy in the $E_{\rm ECL}$ distributions are investigated, but none of them provide a reasonable explanation. Details are discussed in Appendix 11.



(a) (b) (c)

Figure 5.8: Comparison of the E_{ECL} distributions between data and simulation in (a) the $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ modes. The black points are the data, and the histogram is the simulation. No corrections are applied to clusters in the ROE. The bottom panel shows the pull values of the data to the simulation. Shaded regions correspond to the statistical uncertainty of the simulation.



(a) (b) (c)

(d) (e) (f)

Figure 5.9: Comparison of the E_{ECL} distributions between data and simulation in (a)–(c) the first and (d)–(f) the second half run periods. The $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ modes (right) are shown. The black points are the data, and the histogram is the simulation. No corrections are applied to clusters in the ROE. The bottom panel shows the pull values of the data to the simulation. Shaded regions correspond to the statistical uncertainty of the simulation.

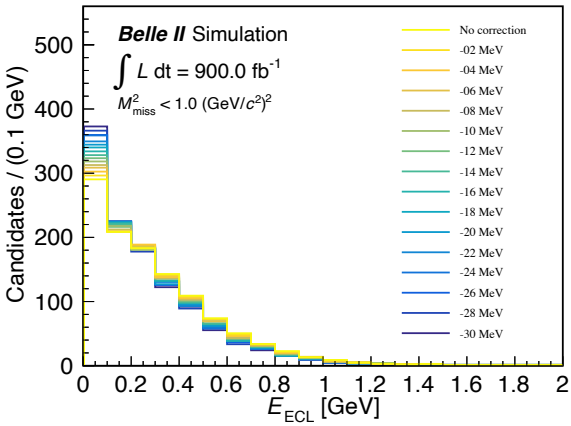
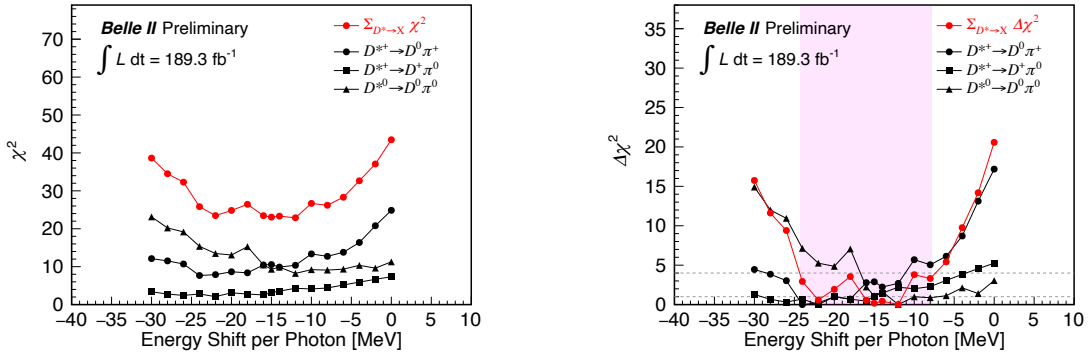


Figure 5.10: Comparison of the E_{ECL} distributions before and after the corrections for clusters from hadronic split-off showers for $D^{*+} \to D^0\pi^+$ in simulation. The lightest yellow histogram shows the distribution before the correction. The other histograms show those after the correction with the energy shift scanned from -2 MeV to -30 MeV in 2 MeV steps, corresponding to light yellow to dark blue of histogram colors. Here, the correction for photons of beam-induced background is applied to all histograms.



(a) (b)

Figure 5.11: χ^2 (a) and $\Delta\chi^2$ values of the E_{ECL} distributions (b) with energy shift scan in the E_{ECL} correction. The χ^2 and $\Delta\chi^2$ are calculated between 0.0 to 0.8 GeV in E_{ECL}. The shaded region in the $\Delta\chi^2$ plot represents the uncertainty range for the fake photon correction.

5.7.2 M_{miss} resoution correction

Figure 5.13 displays the M^2_{miss} distributions of the neutral and charged B modes in the region $|M^2_{\mathrm{mass}}| < 0.5$ (GeV/c^2)^2. The peak widths show slight differences between the data and the simulation. These differences in resolution primarily result from the beam energy spread in the data, which is not simulated in the simulation where the beam collision energy is fixed at the nominal value of 10.58 GeV in the c.m. frame. We evaluate the M^2_{miss} resolution by fitting the peak of $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$ decays in the M^2_{miss} distributions.

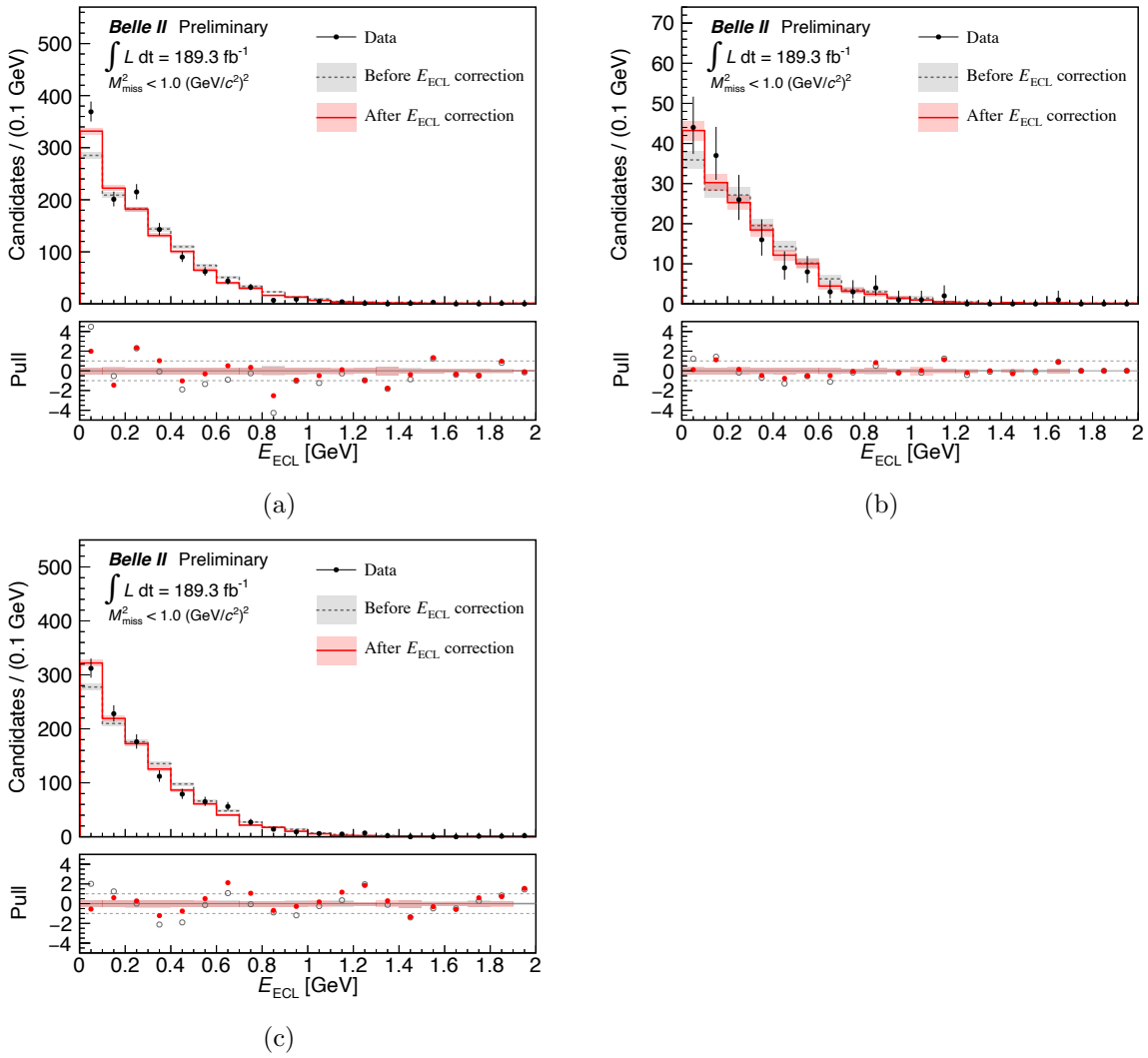
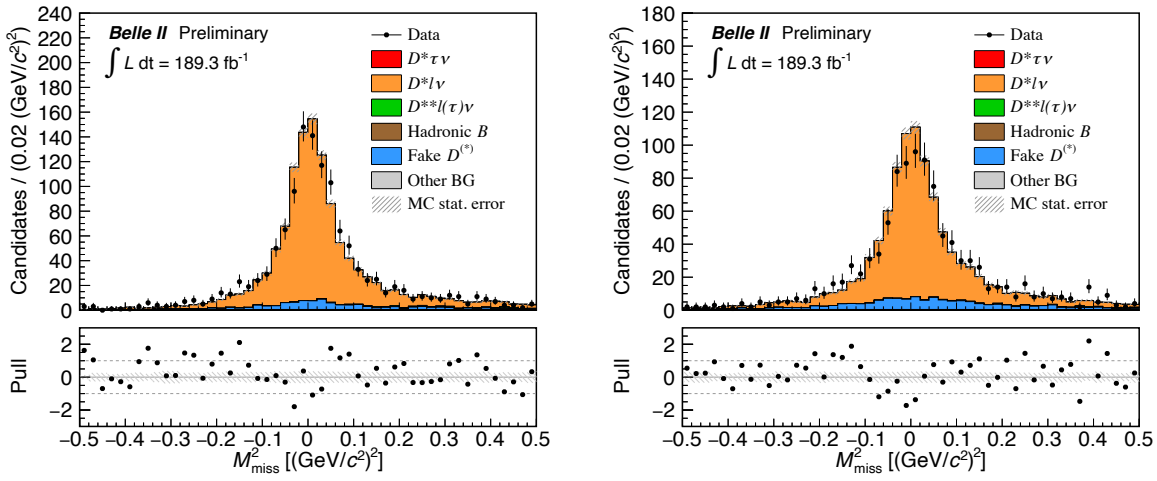


Figure 5.12: Comparison of the E_{ECL} distributions between data and simulation for (a) $D^{*+} \to D^0\pi^+$, (a) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$. The black points are the data. The red (gray) histograms are the distributions after (before) the background cluster corrections in the simulation. The energy shift uses the nominal value of -15 MeV. The bottom panel shows the pull values of the data to the simulation. The rectangle-shaded regions illustrate the statistical uncertainty of the simulation.

Triple and double Gaussian functions are used to model the signal $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$ and background components in the fit, respectively. Their shapes are determined using the simulation. The mean of one of the multiple Gaussian functions is fixed to 0.0 $(\mathrm{GeV}/c^2)^2$, representing the expected M^2_{miss} value, equivalent to a squared mass of a neutrino. The signal shape determination uses only the candidates in the central region from 0.5% to 99.5% of the cumulative ratio in $|M^2_{\mathrm{mass}}| < 0.5$ $(\mathrm{GeV}/c^2)^2$. Figures 5.14 and 5.15 illustrate the determined shapes for the signal and background components, respectively. A global resolution factor, referred to as $R_{\mathrm{res}}^{\mathrm{data(MC)}}$, is floated in the fit for the data (the simulation)



(a) (b)

Figure 5.13: Comparison of the M^2_{mass} distributions between the data and the simulation for (a) the neutral and (a) charged B modes in the region $|M^2_{\text{mass}}| < 0.5$ $(\text{GeV}/c^2)^2$.

Table 5.8: Smearing factors for the M^2_{miss} resolution.

Decay	Smearing factor $[(\text{GeV}/\text{c}^2)^2]$
Neutral B	$0.061^{+0.015}_{-0.017}$
Charged B	$0.060^{+0.028}_{-0.054}$

so that it scales σ of the Gaussian functions. These global resolution factors are common to both components of signal and background components in the fit. The factors are determined for each of the neutral and charged B modes. Figures 5.16 and 5.17 show the fit results in the data and the simulation, respectively, while Figure 5.18 compares these fit results. Figure 5.19 displays the ratios of the M^2_{miss} resolution factors, which are calculated as $R_{\text{res}} \equiv R^{\text{data}}_{\text{res}}/R^{\text{MC}}_{\text{res}}$. The resolution in the original simulation, represented as $\sigma^{\text{MC}}_{M^2_{\text{miss}}}$, is defined as an average resolution weighted by yield compositions of the Gaussian function in the triple Gaussian function of the signal $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ component. The M^2_{miss} in simulation is smeared event-by-event by smearing factors, $\Delta\sigma_{M^2_{\text{miss}}}$, which are obtained from Eq. (5.5) and are listed in Table 5.8.

$$\Delta\sigma_{M^2_{\text{miss}}} = \sigma^{\text{MC}}_{M^2_{\text{miss}}} \cdot \sqrt{R^2_{\text{res}} - 1}. \tag{5.5}$$

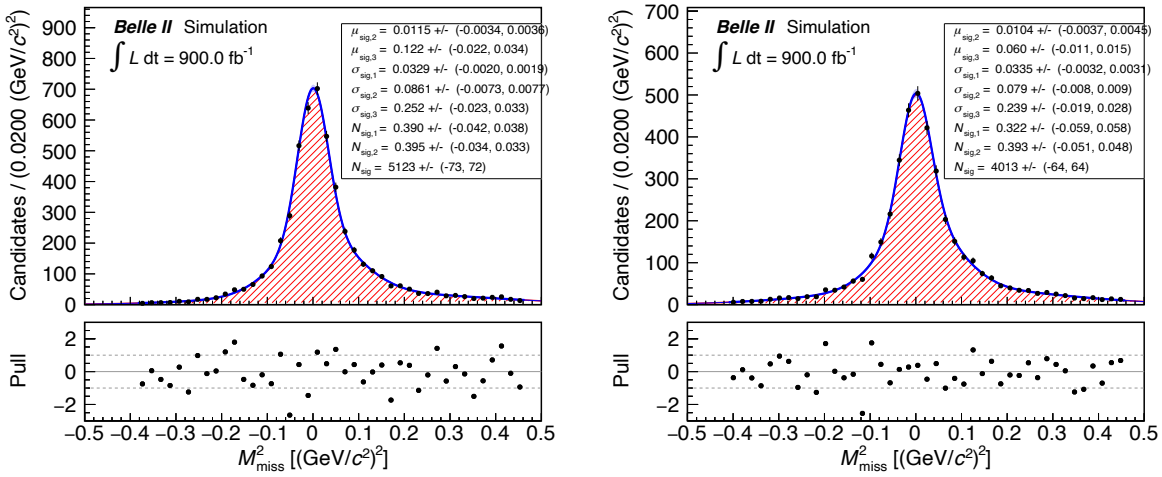


Figure 5.14: Fit results of the M^2_{miss} peaks with correctly reconstructed candidates of $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ decays in the region $|M^2_{\text{miss}}| < 0.5$ $(\text{GeV}/c^2)^2$ for (a) the neutral and (b) charged B modes in simulation, respectively.

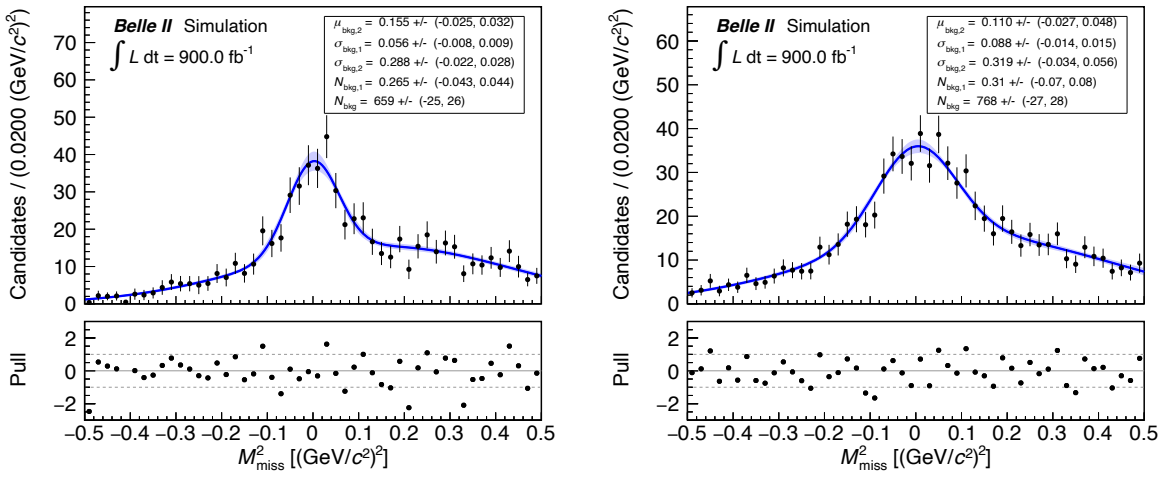
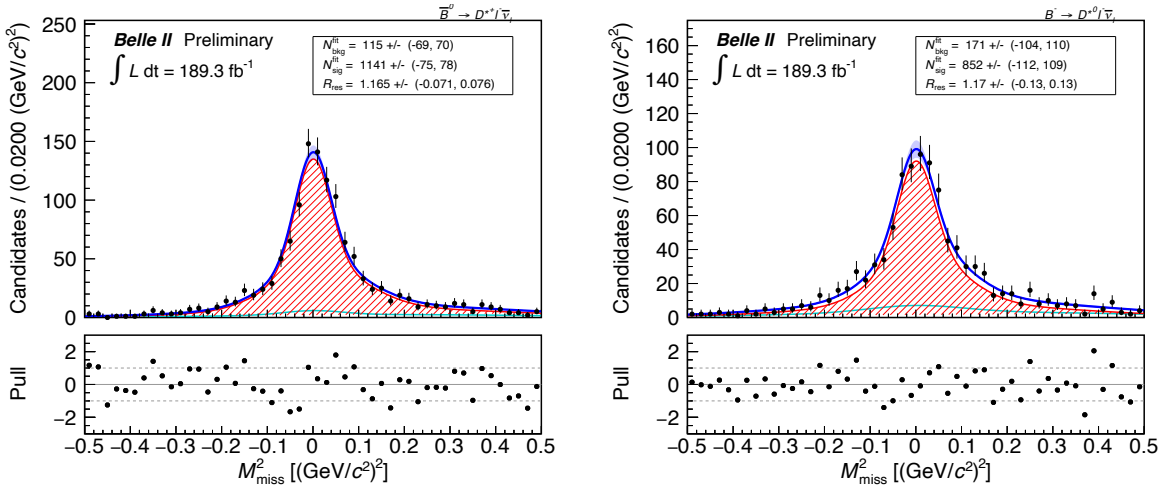
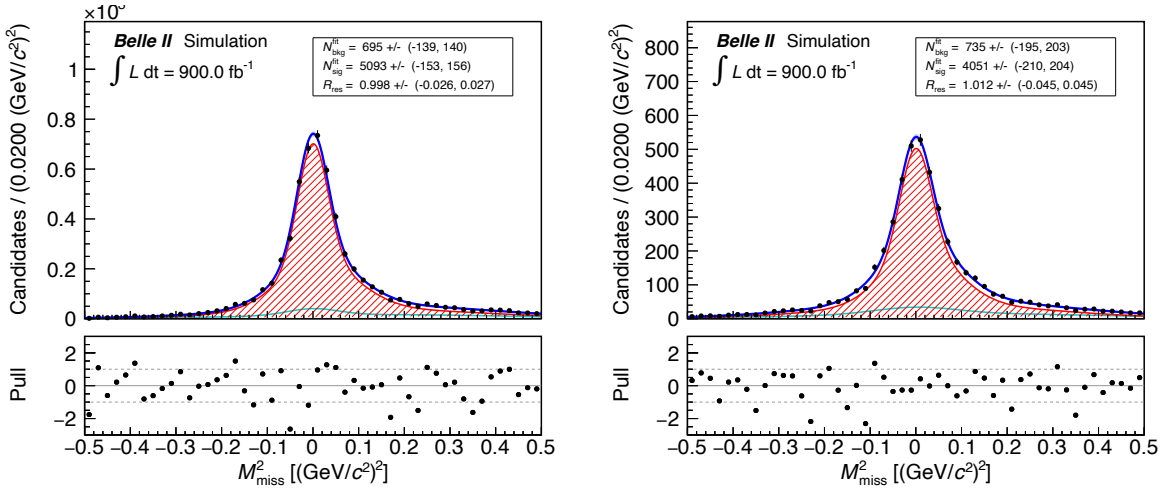


Figure 5.15: Fit results of the M^2_{miss} peaks with incorrectly reconstructed candidates of $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ decays in the region $|M^2_{\text{miss}}| < 0.5$ $(\text{GeV}/c^2)^2$ for (a) the neutral and (b) charged B modes in simulation, respectively.



(a) (b)

Figure 5.16: Fit results of the M^2_{miss} peaks in the region $|M^2_{\text{miss}}| < 0.5$ $(\text{GeV}/c^2)^2$ for (a) the neutral and (b) charged B modes in data, respectively.



(a) (b)

Figure 5.17: Fit results of the M^2_{miss} peaks in the region $|M^2_{\text{miss}}| < 0.5$ $(\text{GeV}/c^2)^2$ for (a) the neutral and (b) charged B modes in simulation, respectively.

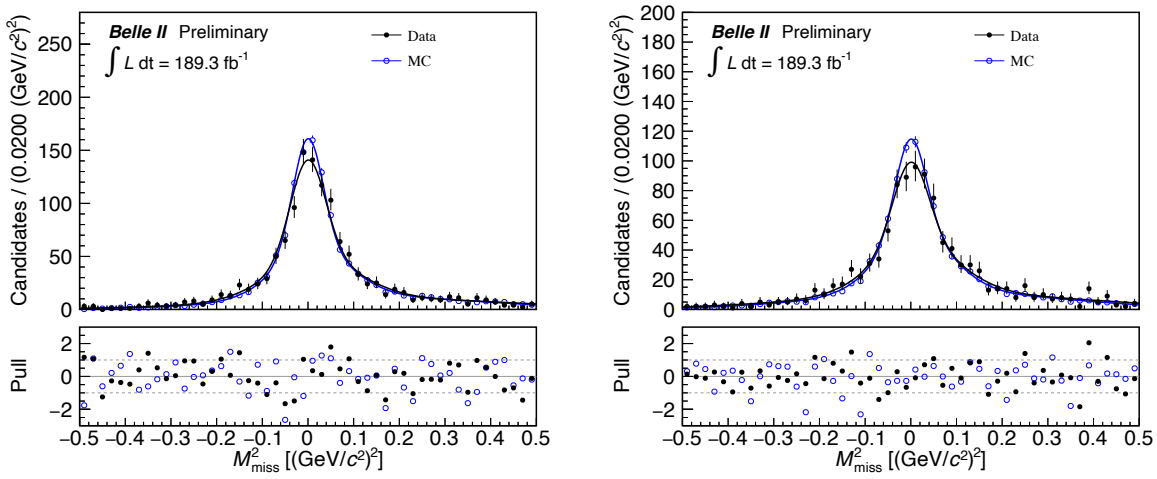


Figure 5.18: Comparison of fit results of the M^2_{miss} peaks between data (black points) and simulation (blue open points) for (a) the neutral and (b) charged B modes, respectively. The number of entries from the simulation is normalized to that of the data.

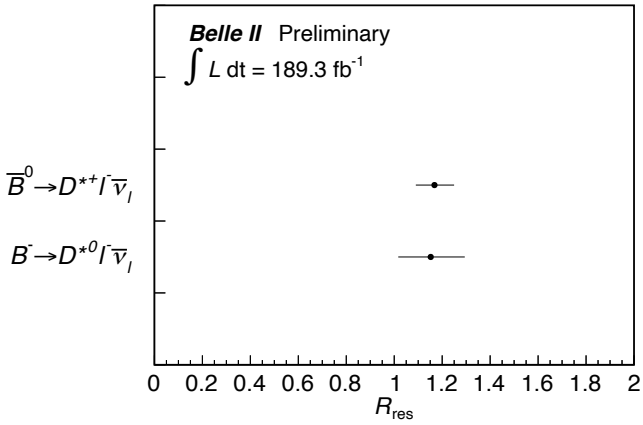


Figure 5.19: Ratios of the M^2_{miss} resolution factors in data to simulation, R_{res}.

6 Validation

We validate the simulation, where the corrections in Section 5 are applied, using sideband data. We employ four sideband regions for this validation: q^2, $N_{\pi^0_{\rm ROE}}$, $M_{\rm bc}$, and ΔM_{D^*}. Agreement with the data for observables listed in Table 6.1 is evaluated. These sideband data allow us to evaluate the detector responses and the background modeling in simulation. The goodness of the fit is estimated with p-values through χ^2 tests in one-dimensional distributions. The p-values are calculated based on a method to compare weighted and unweighted histograms in Ref. [50]. The p-values generally hold more than 5%. Thus, we find sufficient agreement between data and simulation. Detailed discussions for each sideband data are presented in Sections 6.1 to 6.4. Except where specifically mentioned, the simulation is normalized by the luminosity in this section.

Furthermore, we validate the reconstruction efficiencies and distribution shapes of kinematic observables using the data in a range of $M^2_{\rm miss} < 1$ (GeV/c^2)^2, where the signal events are suppressed, and the normalization events are predominant. All distributions of the kinematic variables demonstrate good agreement between the data and the simulation according to the χ^2 tests. The results of this validation are detailed in Section 6.5.

Table 6.1: Test observables for the validation of the simulation with sideband data and data in the range $M^2_{\rm miss} < 1$ (GeV/c^2)^2. $E'_{\rm ECL}$ is defined by $E_{\rm ECL} - \sum E_{\pi^0_{\rm ROE}}$.

Region	Test observables	Major process
q^2 sideband region	$E_{\rm ECL}$, $M^2_{\rm miss}$	$\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ events
$N_{\pi^0_{\rm ROE}}$ sideband region	$E_{\rm ECL}$, $M^2_{\rm miss}$, $E'_{\rm ECL}$	$\overline{B} \to D^{**} \ell^- \overline{\nu}_\ell$ events
$M_{\rm bc}$ sideband region	M_{bc}, $E_{\rm ECL}$, $M^2_{\rm miss}$	Background events associated with incorrectly reconstructed $B_{\rm tag}$
ΔM_{D^*} sideband region	$E_{\rm ECL}$, $M^2_{\rm miss}$	Fake D^* events
$M^2_{\rm miss} < 1$ (GeV/c^2)^2	q^2, $\lvert\vec{p}^{\,*}_{D^*}\rvert$, $\lvert\vec{p}^{\,*}_{\ell}\rvert$, $\lvert\vec{p}_{\pi^+_{\rm slow}}\rvert$, $\lvert\vec{p}_{\pi^0_{\rm slow}}\rvert$, $\lvert\vec{p}_{K^0_S}\rvert$	$\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ events

6.1 Validation in q^2 side-band region

The q^2 sideband region is defined by $q^2 < 3.5$ (GeV/c)^2, where $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ decays are dominant. The distribution shapes of the fitting observables, $E_{\rm ECL}$ and $M^2_{\rm miss}$, are validated. Figure 6.1 compares the distribution shapes of the simulation with those of the data for $E_{\rm ECL}$ and $M^2_{\rm miss}$ in the q^2 sideband region. The yields of the simulation are scaled to match those of the data. Their χ^2 values and corresponding p-values are evaluated in the ranges $0.0 \leqq E_{\rm ECL} \leqq 0.8$ GeV and $-1.0 \leqq M^2_{\rm miss} \leqq 1.0$ (GeV/c^2)^2 and listed in Table 6.2. All the p-values exceed 5%, demonstrating good agreement between the data and the simulation.

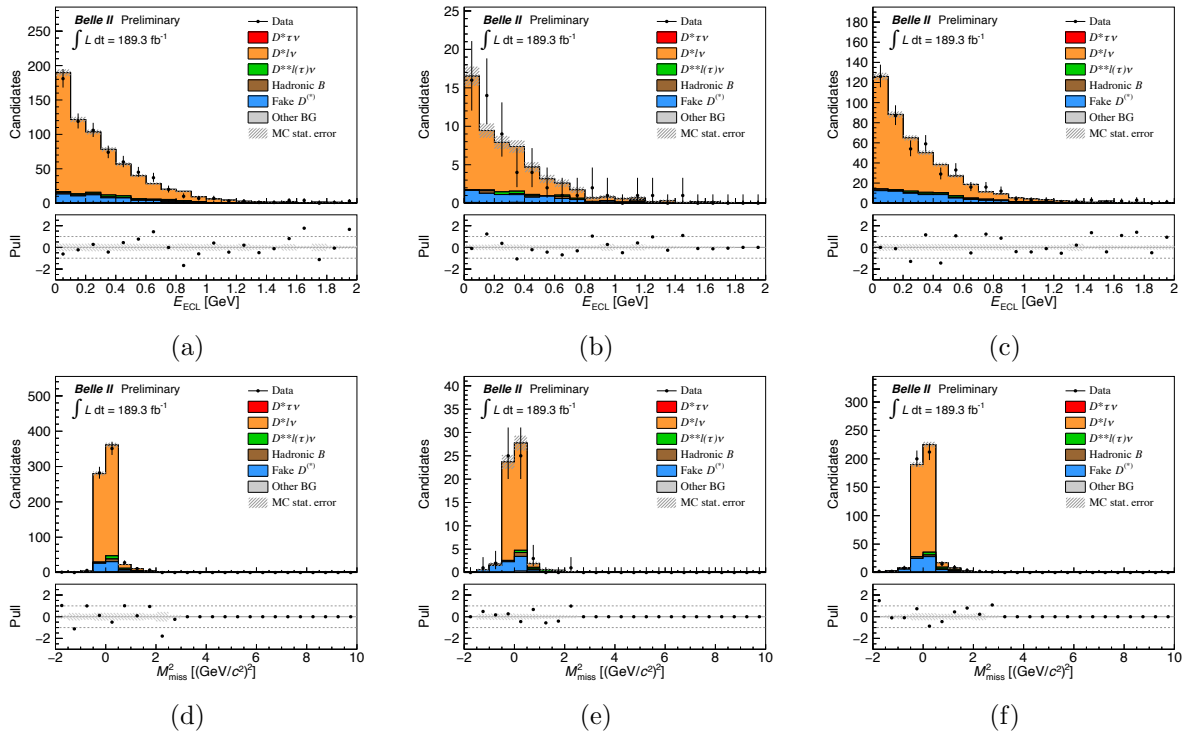


Figure 6.1: Comparison of (a)–(c) E_{ECL} and (d)–(f) M^2_{miss} distributions between data and simulation in the q^2 sideband region. $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ modes (right) are shown. The yields of the simulation are normalized to those of the data.

Table 6.2: Results of statistical tests for E_{ECL} and M^2_{miss} distributions in the q^2 sideband region. The simulation is compared with the sideband data in ranges of $[0.0, 0.8]$ GeV and $[-1.0, 1.0]$ $(\mathrm{GeV}/c^2)^2$, respectively. Only statistical uncertainties of the data and the simulation are considered in the evaluation.

Observable	NDF	χ^2 (p-value [%])		
		$D^{*+} \to D^0\pi^+$	$D^{*0} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
E_{ECL}	7	3.68 (81.6)	5.40 (61.1)	8.62 (28.1)
M_{miss^2}	3	2.63 (45.3)	0.893 (82.7)	1.40 (70.5)

6.2 Validation in $N_{\pi^0_{\rm ROE}}$ side-band region

The $N_{\pi^0_{\rm ROE}}$ sideband region is used to validate background events from $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ decays. Especially, since semileptonic B decays to some of the D^{**} excited states have not yet been observed, data-driven validation is crucial. The $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ background decays are enhanced by reconstructing an additional π^0 candidate produced in $D^{**} \to D^* n\pi$ decays. Therefore, at least one $\pi^0_{\rm ROE}$ is required in addition to the reconstructed $B\overline{B}$ pair in the $N_{\pi^0_{\rm ROE}}$ sideband region, or $N_{\pi^0_{\rm ROE}} \geqq 1$. Figure 6.2 compares the simulation with the data for $E_{\rm ECL}$ and $M^2_{\rm miss}$ in the $N_{\pi^0_{\rm ROE}}$ sideband region. $E'_{\rm ECL}$ is also utilized, defined as $E_{\rm ECL}$ minus the energy sum of $\pi^0_{\rm ROE}$ candidates, as shown in Figure 6.3. Their χ^2 values and corresponding p-values are evaluated in the ranges $0.0 \leqq E_{\rm ECL} \leqq 2.0$ GeV, $1.0 \leqq M^2_{\rm miss} \leqq 5.0$ $(\mathrm{GeV}/c^2)^2$, and $0.0 \leqq E'_{\rm ECL} \leqq 0.8$ GeV, and listed in Table 6.3. The $E_{\rm ECL}$ distributions of the $D^{*+} \to D^+\pi^0$ and $D^{*0} \to D^0\pi^0$ modes, and the $M^2_{\rm miss}$ distribution for the $D^{*+} \to D^+\pi^0$ mode, yield low p-values. The other p-values indicate good agreement between data and simulation, exceeding 5%.

For $D^{*+} \to D^+\pi^0$, the χ^2 values of the $E_{\rm ECL}$ and $M^2_{\rm miss}$ distribution are abnormally large, which result in small p-values. Nevertheless, when the χ^2 tests are conducted using histograms with a broader bin widths, as shown in Figure 6.4, the χ^2 values become 4.71 for the five bins in the range $0.0 \leqq E_{\rm ECL} \leqq 2.0$ GeV and 0.138 for the two bins in the range of $1.0 \leqq M^2_{\rm miss} \leqq 7.0$ $(\mathrm{GeV}/c^2)^2$. Consequently, the p-value increases significantly to 31.9% and 70.1%. Therefore, low p-values could be attributed to the small number of entries in the bins.

In the $D^{*0} \to D^0\pi^0$ mode, the p-value obtained from the χ^2 test with a broader bin width recovers only up to 3.2% in Figure 6.5. However, it is observed that the systematic uncertainty associated with the fake photon corrections can adequately explain the discrepancy seen in this mode. The energy shifts for the fake photon corrections are varied from -30 to 0.0 MeV in the simulation, and the resulting χ^2 values for the $E_{\rm ECL}$ distributions are then evaluated. The results of these variations, presented in Figure 6.6, show that the uncertainty range for the fake photon corrections, specifically -24 to -8.0 MeV in energy shifts, covers p-values above 5%. This indicates that the systematic uncertainty sufficiently accounts for the observed discrepancy. Consequently, this ensures the adequacy of the simulation in modeling the shape of the $E_{\rm ECL}$ distribution for $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ events in this D^* mode.

The $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ decays dominate the region of $1.0 < M^2_{\rm miss} < 5.0$ $(\mathrm{GeV}/c^2)^2$. The χ^2 values and corresponding p-values are additionally evaluated in this $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$-enhanced $M^2_{\rm miss}$ regions with the $N_{\pi^0_{\rm ROE}}$ sideband data. Figure 6.7 compares the simulation with the data for $E_{\rm ECL}$ and $E'_{\rm ECL}$ in the $N_{\pi^0_{\rm ROE}}$ sideband region. They are evaluated in ranges $0.0 \leqq E_{\rm ECL} \leqq 2.0$ GeV and $0.0 \leqq E'_{\rm ECL} \leqq 0.8$ GeV, and listed in Table 6.4. All the p-values exceed 5% to show good agreement between the data and the simulation

6.3 Validation in $M_{\rm bc}$ side-band region

The background events, such as fake D^*, $B^0 \leftrightarrow B^+$ cross-feed, and "other" $B\overline{B}$ background events, often originate from incorrectly reconstructed $B_{\rm tag}$ candidates. These

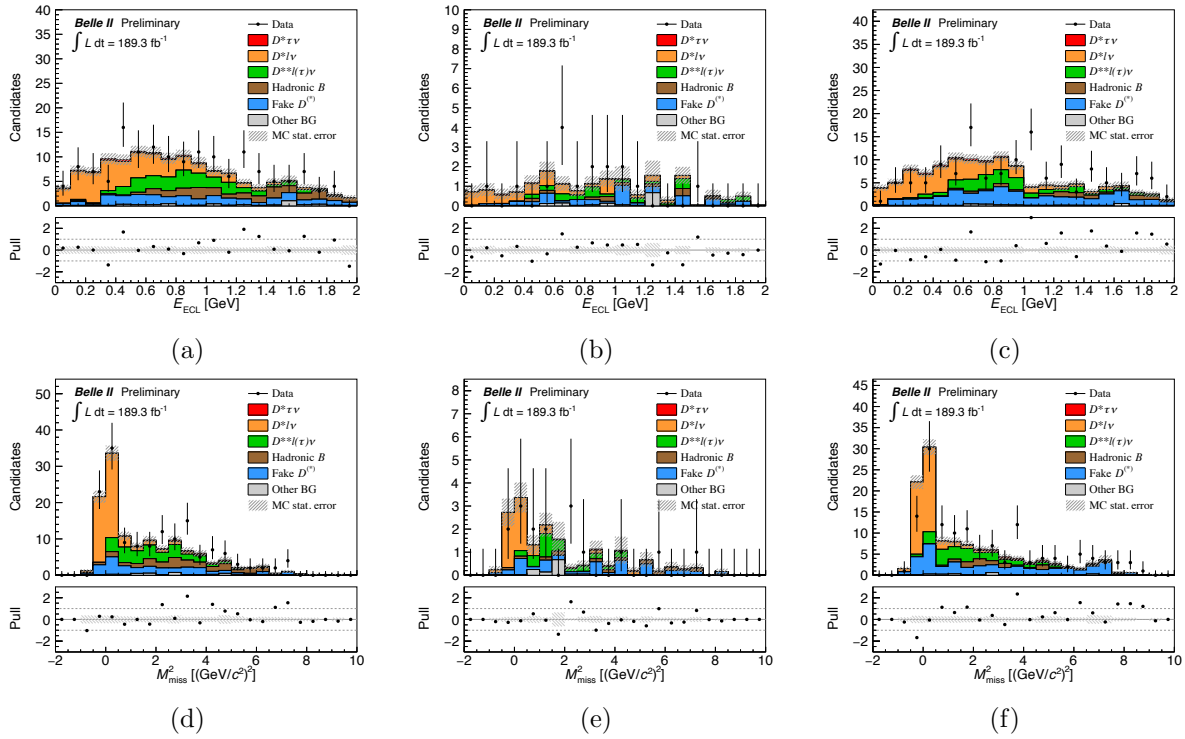


Figure 6.2: Comparison of (a)–(c) E_{ECL} and (d)–(f) M^2_{miss} distributions between data and simulation in the $N_{\pi^0_{\mathrm{ROE}}}$ sideband region. $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ modes (right) are shown.

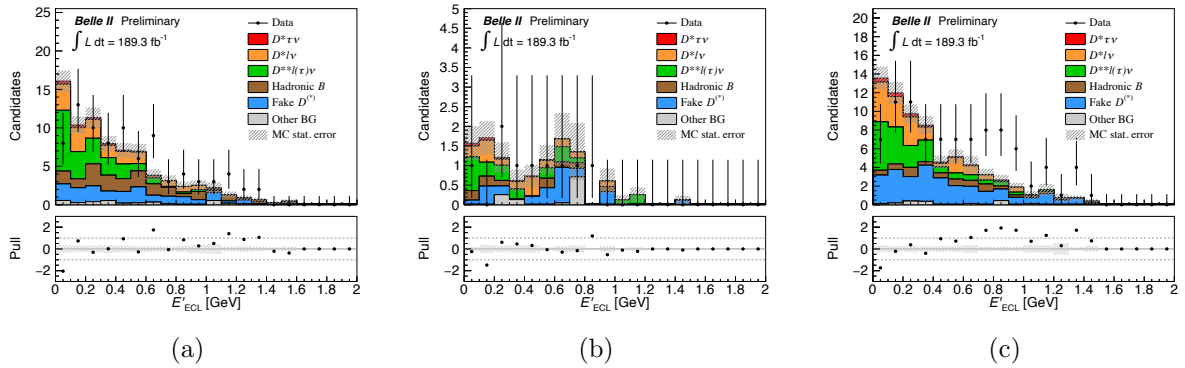
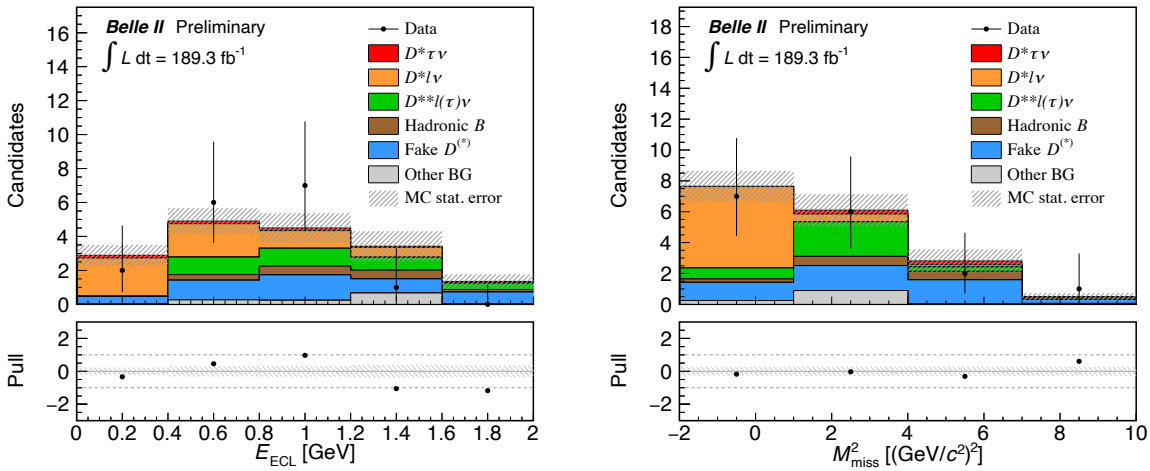


Figure 6.3: Comparison of E_{ECL} distributions between data and simulation in the $N_{\pi^0_{\mathrm{ROE}}}$ sideband region. (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ modes are shown.

Table 6.3: Results of statistical tests for E_{ECL}, M^2_{miss}, and E'_{ECL} distributions in the $N_{\pi^0_{\mathrm{ROE}}}$ sideband region. The simulation is compared with the sideband data in ranges of $[0.0, 2.0]$ GeV, $[1.0, 5.0]$ $(\mathrm{GeV}/c^2)^2$, and $[0.0, 0.8]$ GeV, respectively. Only statistical uncertainties of the data and the simulation are considered in the evaluation.

Observable	NDF	χ^2 (p-value [%])		
		$D^{*+} \to D^0\pi^+$	$D^{*0} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
E_{ECL}	19	17.3 (56.7)	106 (0.322×10^{-12})	48.1 (0.0251)
M_{miss^2}	7	8.77 (26.9)	18.9 (0.840)	9.25 (23.5)
E'_{ECL}	7	11.1 (13.6)	3.13 (87.3)	11.3 (12.7)



(a) (b)

Figure 6.4: Comparison of (a) E_{ECL} and (b) M^2_{miss} distributions between data and simulation for $D^{*+} \to D^+\pi^0$ mode with a broader bin width in the $N_{\pi^0_{\mathrm{ROE}}}$ sideband region.

misreconstructions typically result from either exchanging daughters between the signal-side and tag-side B mesons or from missing a few particles from a B_{tag} candidate. The fractions of these incorrect B_{tag} candidates are validated using the data. The M_{bc} sideband samples are prepared by inverting the requirement for M_{bc} of B_{tag} candidates to $5.24\ \mathrm{GeV}/c^2 < M_{\mathrm{bc,tag}} < 5.27\ \mathrm{GeV}/c^2$, focusing on the region where correctly reconstructed B_{tag} are suppressed. The M_{bc} distributions in this sideband region are displayed in Figure 6.8. Figure 6.9 compares the simulation with the data for E_{ECL} and M^2_{miss} in the M_{bc} sideband region. Their χ^2 values and corresponding p-values are evaluated in ranges $5.24 \leqq M_{\mathrm{bc,tag}} \leqq 5.27\ \mathrm{GeV}/c^2$, $0.0 \leqq E_{\mathrm{ECL}} \leqq 0.8$ GeV, and $-1.0 \leqq M_{\mathrm{miss}^2} \leqq 8.0\ (\mathrm{GeV}/c^2)^2$, and listed in Table 6.5. All the p-values exceed 5%, reflecting good agree-

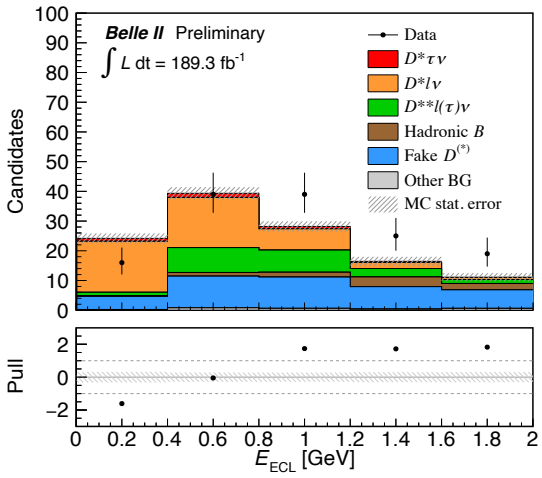


Figure 6.5: Comparison of E_{ECL} distributions between data and simulation for $D^{*0} \to D^0\pi^0$ mode with a broader bin width in $N_{\pi^0_{\mathrm{ROE}}}$ sideband region.

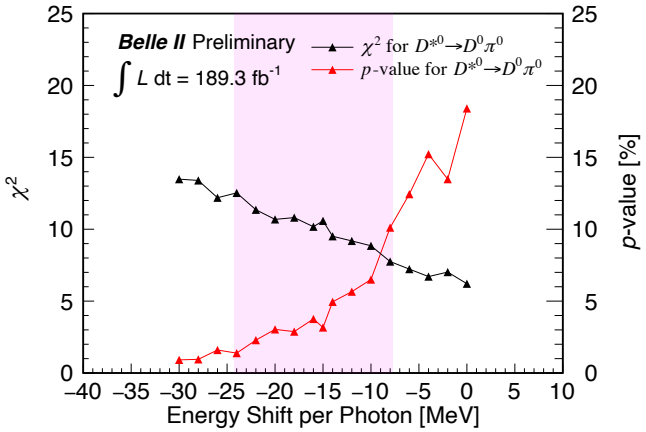


Figure 6.6: Values of χ^2 from E_{ECL} distributions in $D^{*0} \to D^0\pi^0$ at different energy shifts for fake photon corrections. The shaded region represents the uncertainty range considered for the correction.

ment between the data and the simulation.

6.4 Validation in ΔM_{D^*} side-band region

The ΔM_{D^*} sideband region is utilized to examine the fitting observables of the fake D^* candidates. The sideband ranges are specified in Table 5.7. In this region, fake D^* candidates are predominantly enhanced with a purity of more than 95%. Figure 6.9 compares simulation with data for E_{ECL} and M^2_{miss} in the M_{bc} sideband region. Their χ^2 values and corresponding p-values are evaluated in the ranges $0.0 \leqq E_{\mathrm{ECL}} \leqq 0.8$ GeV and $-1.0 \leqq M^2_{\mathrm{miss}} \leqq 8.0$ $(\mathrm{GeV}/c^2)^2$, and listed in Table 6.5. All the p-values exceed 5%, demonstrating good agreement between the data and the simulation, except for the M^2_{miss} distribution in $D^{*+} \to D^0\pi^+$. The $D^{*+} \to D^0\pi^+$ mode has the best signal-to-background

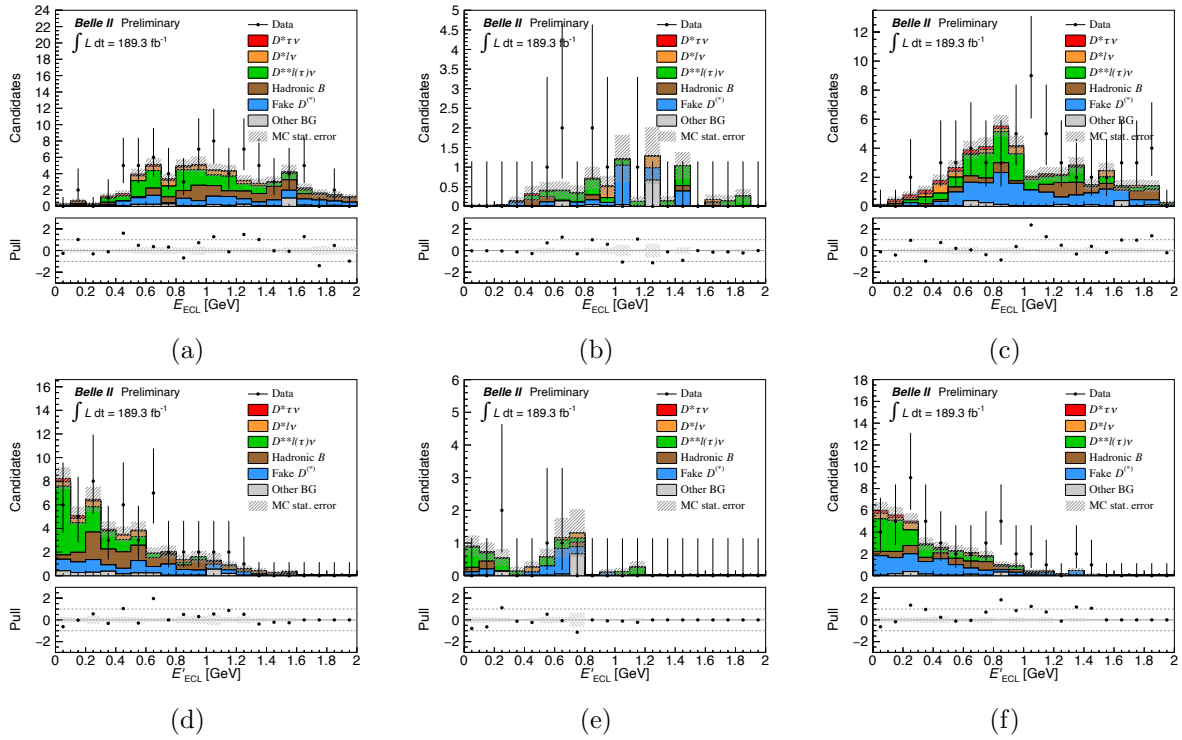
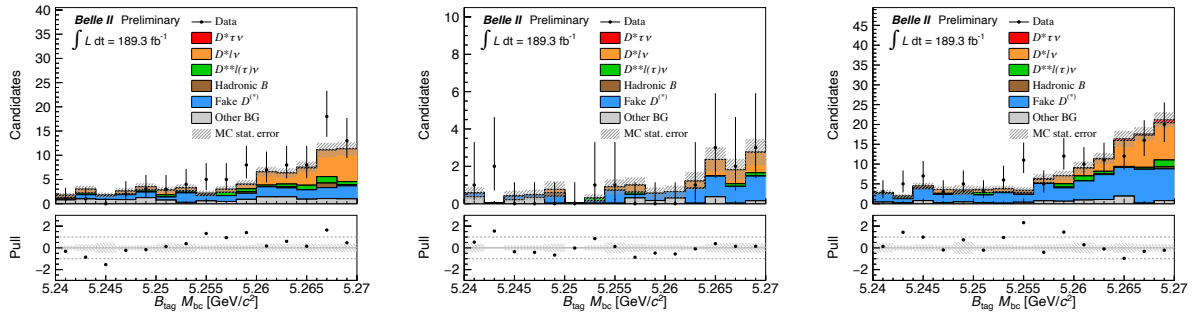


Figure 6.7: Comparison of (a)–(c) $E_{\rm ECL}$ and (d)–(f) $E'_{\rm ECL}$ distributions between data and simulation in the $N_{\pi^0_{\rm ROE}}$ sideband region with the B→D**lnu enhanced selection of $1.0 < M^2_{\rm miss} < 5.0$ $({\rm GeV}/c^2)^2$. $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ modes (right) are shown.

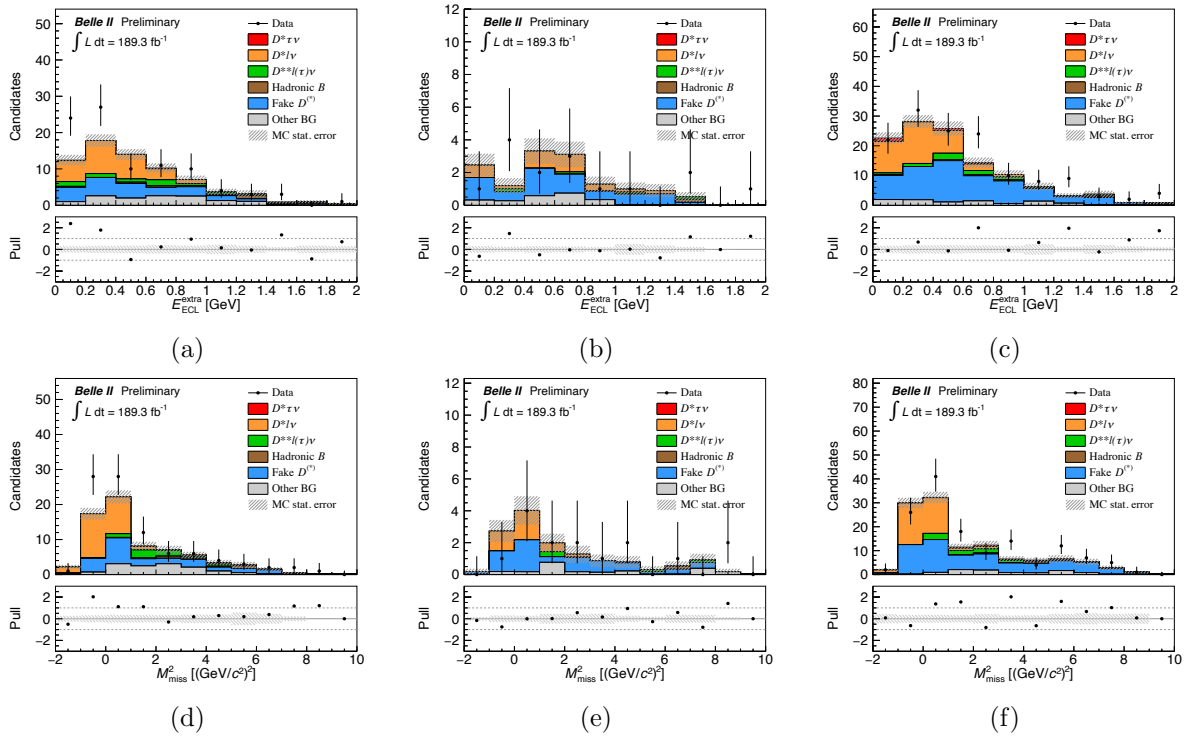
Table 6.4: Results of statistical tests for $E_{\rm ECL}$ and $E'_{\rm ECL}$ distributions with the selection for enhanced $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ background events by $1.0 < M^2_{\rm miss} < 5.0$ $({\rm GeV}/c^2)^2$ in the $N_{\pi^0_{\rm ROE}}$ sideband region. The simulation is compared with the sideband data in the ranges $[0.0, 2.0]$ GeV and $[0.0, 0.8]$ GeV, respectively. Only statistical uncertainties of the data and the simulation are considered in the evaluation.

Observable	NDF	χ^2 (p-value [%])		
		$D^{*+} \to D^0\pi^+$	$D^{*0} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
$E_{\rm ECL}$	19	16.0 (65.9)	16.7 (61.2)	23.5 (21.6)
$E'_{\rm ECL}$	7	10.8 (14.8)	8.07 (32.6)	4.15 (76.2)



(a) (b) (c)

Figure 6.8: Comparison of M_{bc} distributions between data and simulation in the M_{bc} sideband region. (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ modes are shown.



(a) (b) (c)

(d) (e) (f)

Figure 6.9: Comparison of (a)–(c) E_{ECL} and (d)–(f) M^2_{miss} distributions between data and simulation in the M_{bc} sideband region. $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ modes (right) are shown.

Table 6.5: Results of statistical tests for $M_{\rm bc}$ of $B_{\rm tag}$ candidates, $E_{\rm ECL}$, and $M^2_{\rm miss}$ distributions in $M_{\rm bc}$ sideband region. The simulation is compared with the sideband data in ranges of $[5.24, 5.27]$ GeV/c^2, $[0.0, 0.8]$ GeV, and $[-1.0, 8.0]$ $(\text{GeV}/c^2)^2$, respectively. Only statistical uncertainties of the data and the simulation are considered in the evaluation.

Observable	NDF	χ^2 (p-value [%])		
		$D^{*+} \to D^0\pi^+$	$D^{*0} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
$B_{\rm tag}$ $M_{\rm bc}$	14	10.2 (75.0)	11.0 (68.3)	20.0 (13.2)
$E_{\rm ECL}$	3	7.00 (7.20)	5.79 (12.2)	3.79 (28.5)
$M^2_{\rm miss}$	8	4.27 (83.2)	4.35 (82.4)	10.2 (14.0)

Table 6.6: Results of statistical tests for $E_{\rm ECL}$ and $M^2_{\rm miss}$ distributions in ΔM_{D^*} sideband region. The simulation is compared with the sideband data in ranges of $[0.0, 0.8]$ GeV and $[-1.0, 8.0]$ $(\text{GeV}/c^2)^2$, respectively. Only statistical uncertainties of the data and the simulation are considered in the evaluation.

Observable	NDF	χ^2 (p-value [%])		
		$D^{*+} \to D^0\pi^+$	$D^{*0} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
$E_{\rm ECL}$	7	11.5 (11.8)	3.79 (80.4)	14.0 (5.14)
$M_{\rm miss^2}$	17	30.6 (2.20)	6.07 (99.2)	8.11 (96.4)

ratio and holds the smallest fake D^* rate, leading to a fewer entries even in the ΔM_{D^*} sideband region. Additional χ^2 tests using histograms with a broader bin width are performed in Figure 6.11 for the $D^{*+} \to D^0\pi^+$ mode. The χ^2 value reaches 6.30 for five bins in the range of $-2.0 < M^2_{\rm miss} < 8.0$ $(\text{GeV}/c^2)^2$, and the p-value increases to 17.8%. Therefore, The low p-value could be due to the small number of entries in the bins.

6.5 Validation in $M^2_{\rm miss} < 1.0$ $(\text{GeV}/c^2)^2$

In the region $M^2_{\rm miss} < 1.0$ $(\text{GeV}/c^2)^2$, we examine the six kinematic variables: momentum transfer q^2, D^* momentum in the c.m. frame $|\vec{p}^{\,*}_{D^*}|$, lepton momentum in the c.m. frame $|\vec{p}_\ell{}^*|$, momentum of $\pi^+_{\rm slow}$ candidates $|\vec{p}_{\pi^+_{\rm slow}}|$, momentum of low-momentum $\pi^0_{\rm slow}$ candidates $|\vec{p}_{\pi^0_{\rm slow}}|$, momentum of K^0_S candidates from D decays $|\vec{p}_{K^0_S}|$. These variables are compared in Figures 6.12, 6.13, and 6.14. Their χ^2 values and p-values are listed in Table 6.7. The test observables are evaluated in ranges $[4.0, 8.0]$ $(\text{GeV}/c)^2$, $[0.50, 2.0]$ GeV/c, $[0.70, 2.2]$ GeV/c,

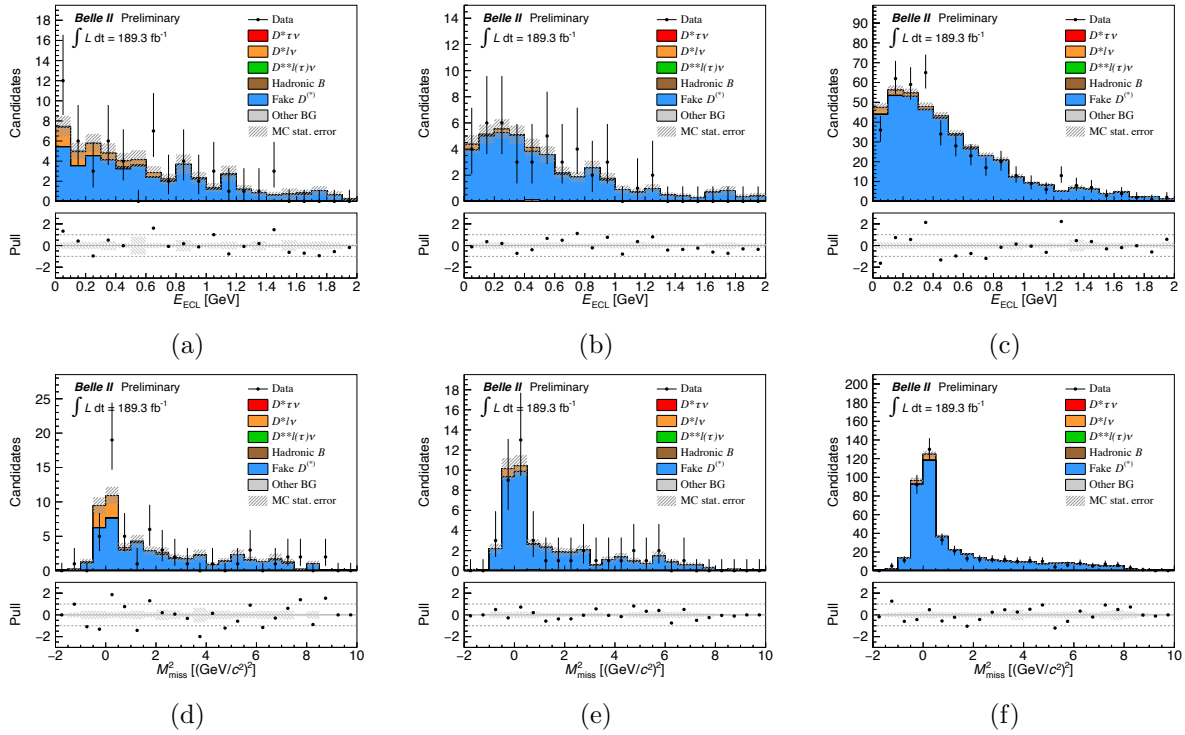


Figure 6.10: Comparison of (a)–(c) E_{ECL} and (d)–(f) M^2_{miss} distributions between data and simulation in the ΔM_{D^*} sideband region. $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ modes (right) are shown.

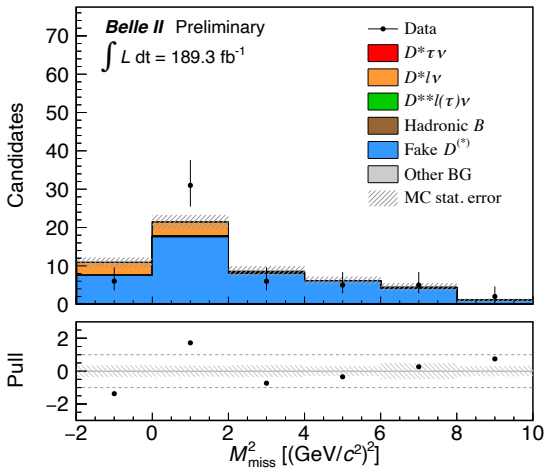


Figure 6.11: Comparison of M^2_{miss} distributions between the data and the simulation for $D^{*+} \to D^0\pi^+$ mode with a broader bin width in the ΔM_{D^*} sideband region.

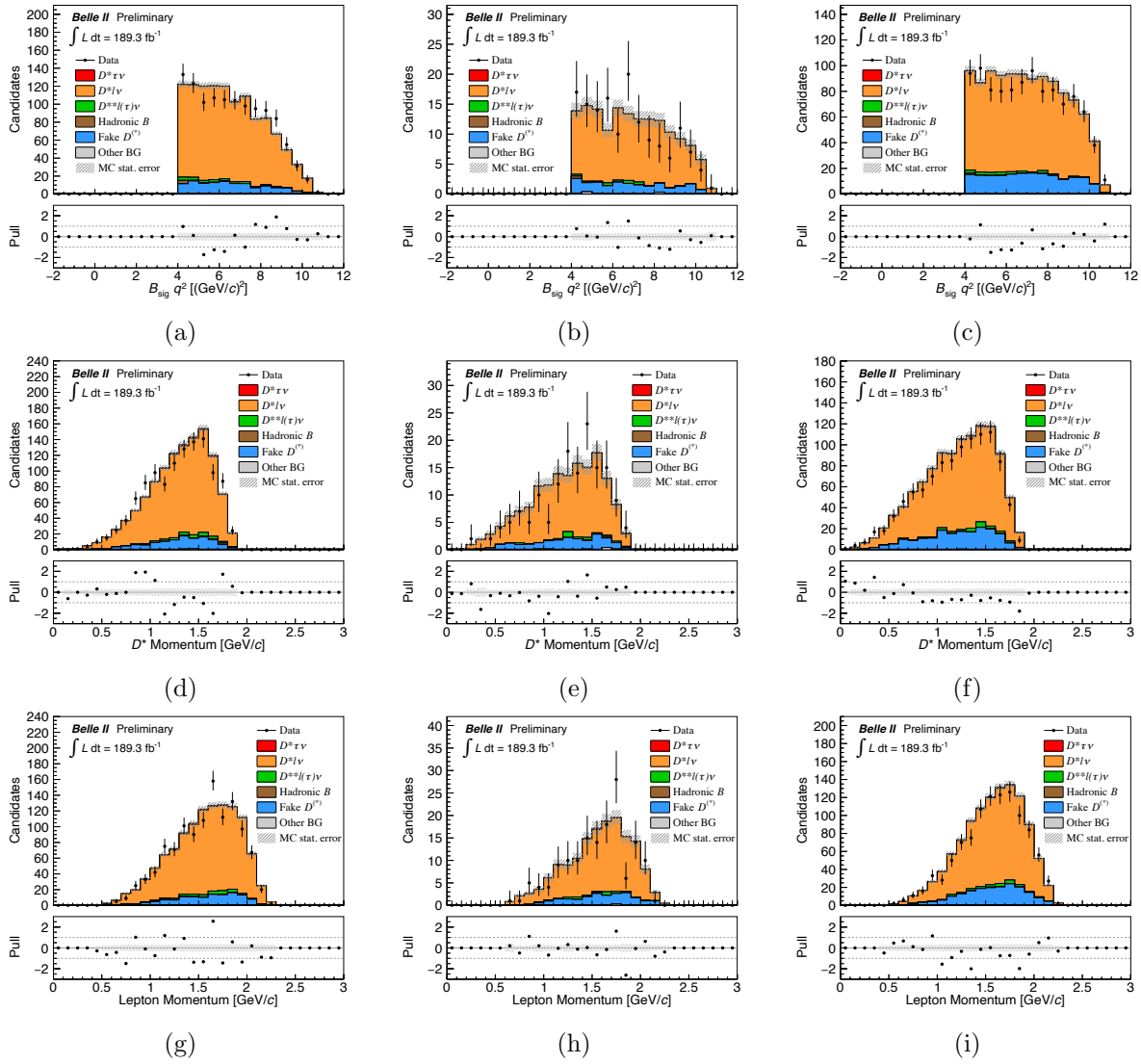
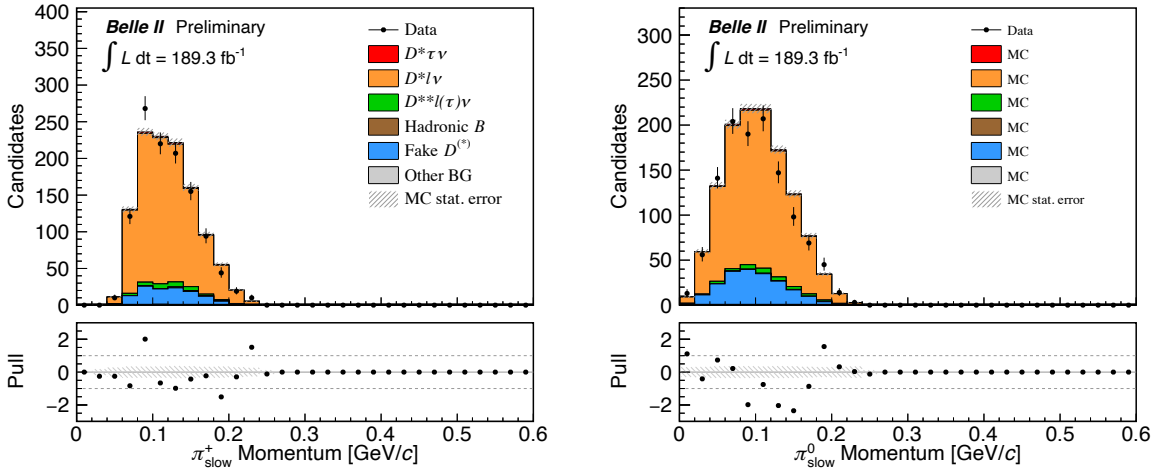


Figure 6.12: Comparison of (a)–(c) the q^2, (d)–(f) $|\vec{p}^{\,*}_{D^*}|$, and (g)–(i) $|\vec{p}^{\,*}_{\ell}|$ distributions between the data and the simulation in $M^2_{\rm miss} < 1.0$ (GeV/c^2)^2. $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ modes (right) are shown.

$[0.0, 0.20]$ GeV/c, $[0.0, 0.20]$ GeV/c, and $[0.10, 1.50]$ GeV/c, respectively. All the p-values exceed 5%, indicating good agreement between the data and the simulation.



(a) (b)

Figure 6.13: Comparison of (a) the $|\vec{p}_{\pi^+_{\rm slow}}|$ and (b) $|\vec{p}_{\pi^0_{\rm slow}}|$ distributions between the data and the simulation in $M^2_{\rm miss} < 1.0$ (GeV/c^2)^2.

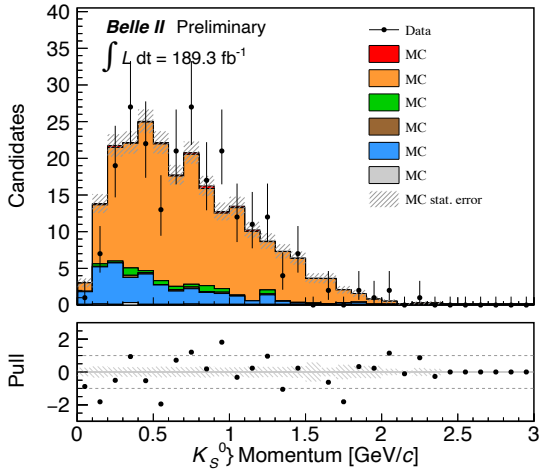


Figure 6.14: Comparison of the $|\vec{p}_{K^0_S}|$ distributions between the data and the simulation in $M^2_{\rm miss} < 1.0$ (GeV/c^2)^2.

Table 6.7: Results of statistical tests for distributions of kinematic variables in the region of $M^2_{\rm miss} < 1.0$ (GeV/c^2)^2. The simulation is compared with the data. Only statistical uncertainties of the data and the simulation are considered in the evaluation.

Observable	NDF	χ^2 (p-value [%])		
		$D^{*+} \to D^0\pi^+$	$D^{*0} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
q^2	11	14.2 (22.4)	11.3 (42.1)	8.00 (71.3)
$\lvert\vec{p}^{\,*}_{D^*}\rvert$	14	23.2 (5.67)	11.2 (59.2)	5.00 (98.6)
$\lvert\vec{p}^{\,*}_{\ell}\rvert$	14	20.8 (10.8)	14.0 (45.2)	111.1 (68.0)
$\lvert\vec{p}_{\pi^+_{\rm slow}}\rvert$	8	8.19 (41.5)	—	—
$\lvert\vec{p}_{\pi^0_{\rm slow}}\rvert$	9	—	8.11 (11.3)	
$\lvert\vec{p}_{K^0_S}\rvert$	13		17.61 (17.2)	

7 Signal extraction

7.1 Fit strategy

7.1.1 Fitting method

$R(D^*)$ is extracted through a two-dimensional extended binned maximum likelihood fit. The two-dimensional probability density functions (PDFs) are constructed independently in each of three fit categories of the D^* modes: $D^{*+} \to D^0\pi^+$, $D^{*+} \to D^+\pi^0$, and $D^{*0} \to D^0\pi^0$. A simultaneous fit is performed across the three D^* modes. The variables used in the fit are $E_{\rm ECL}$ and $M^2_{\rm miss}$ in the ranges of 0.0 to 2.0 GeV and -2.0 to 10.0 $(\mathrm{GeV}/c^2)^2$, respectively.

7.1.2 Yield parameterization

$R(D^*)$ is described using the number of events and reconstruction efficiencies,

$$R(D^*) = \frac{\mathcal{B}(\overline{B} \to D^*\tau^-\nu)}{\mathcal{B}(\overline{B} \to D^*\ell^-\nu)}, \ (\ell = e, \ \mu) \tag{7.1}$$

$$= \frac{N_{D^*\tau\nu}}{(N_{D^*\ell\nu}/2)} \cdot \frac{\varepsilon_{D^*\ell\nu}}{\varepsilon_{D^*\tau\nu}}, \tag{7.2}$$

where $N_{D^*\tau(\ell)\nu}$ is the observed number of $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ ($D^*\ell^-\overline{\nu}_\ell$) candidates in the data and $\varepsilon_{D^*\tau(\ell)\nu}$ is the reconstruction efficiency of correctly reconstructed $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ ($D^*\ell^-\overline{\nu}_\ell$) candidates. The factor 2 in the denominator averages the summed yield from the two light leptons, $\ell \in \{e, \mu\}$. Here, the reconstruction efficiencies are defined as follows,

$$\varepsilon_{D^*\tau(\ell)\nu} = \frac{N^{\rm rec}_{D^*\tau(\ell)\nu}}{N^{\rm gen}_{D^*\tau(\ell)\nu}}, \tag{7.3}$$

where $N^{\rm rec}_{D^*\tau(\ell)\nu}$ and $N^{\rm gen}_{D^*\tau(\ell)\nu}$ are the number of reconstructed and generated $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ ($D^*\ell^-\overline{\nu}_\ell$) decays, respectively. $N^{\rm gen}_{D^*\tau(\ell)\nu}$ is derived by Eq. (7.4) and Eq. (7.5),

$$N^{\rm gen}_{D^*\tau\nu} = N^{\rm gen}_{B\overline{B}} \cdot \mathcal{B}(\overline{B} \to D^*\tau^-\overline{\nu}_\tau)^{\rm gen} + N^{\rm gen}_{B\overline{B}} \cdot \frac{\mathcal{B}(\overline{B} \to D^*\tau^-\overline{\nu}_\tau)^{\rm gen}}{\mathcal{B}(\overline{B} \to D\tau^-\overline{\nu}_\tau)^{\rm gen} + \mathcal{B}(\overline{B} \to D^*\tau^-\overline{\nu}_\tau)^{\rm gen}}, \tag{7.4}$$

$$N^{\rm gen}_{D^*\ell\nu} = N^{\rm gen}_{B\overline{B}} \cdot 2\left[\mathcal{B}(\overline{B} \to D^*e^-\overline{\nu}_e)^{\rm gen} + \mathcal{B}(\overline{B} \to D^*\mu^-\overline{\nu}_\mu)^{\rm gen}\right]. \tag{7.5}$$

Here $N^{\rm gen}_{B\overline{B}}$ is the number of $B\overline{B}$ pairs generated in the simulation, and $\mathcal{B}(B \to X)^{\rm gen}$ refers to a simulated branching fraction for $B \to X$ decay. In the signal MC samples, one B meson decays generically, including the decays to $D^{(*)}\tau^-\overline{\nu}_\tau$, while its counterpart B meson decays exclusively to $D^{(*)}\tau^-\overline{\nu}_\tau$. Thus, in Eq. (7.4), the first and second terms correspond to the number of generated $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ decays from generic B decays and that from the exclusive signal B decays, respectively. On the other hand, both B mesons decay through generic processes in the generic MC samples. The number of generated $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$ decays is therefore twice the individual counts from one of the $B\overline{B}$ pair, which is represented by the factor of 2 in Eq. (7.5).

For each D^* mode, the yields of event categories or sub-categories are parameterized as follows.

1. Signal events
   The yield $N_{D^*\tau\nu}$ is parameterized by Eq. (7.6).

$$N_{D^*\tau\nu} = R(D^*) \cdot \frac{N_{D^*\ell\nu}}{2} \cdot \frac{\varepsilon_{D^*\tau\nu}}{\varepsilon_{D^*\ell\nu}}. \tag{7.6}$$

   The $R(D^*)$ is shared among the three D^* mode. The yield is unconstrained in the fit, while the reconstruction efficiencies of the signal and the normalization events are taken as nuisance parameters. The nuisance parameter, hereafter, is constrained by a single Gaussian function of which mean is the estimated value and σ is its uncertainty.

   When $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ candidates accompany a fake lepton candidate from tau decay which passes the requirement of the lepton identification, they can also be considered as signal events. The yield $N^{\ell\text{-misID}}_{D^*\tau\nu}$ is scaled to the yield $N_{D^*\tau\nu}$ with a fixed ratio in the signal $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ MC samples.

2. Normalization events
   The yield $N_{D^*\ell\nu}$ is parameterized for neutral and charged B modes by Eq. (7.7) and Eq. (7.8), respectively.

$$N_{D^*\ell\nu} = \begin{cases} 2\mathcal{B}(\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell) \times \left[2N_{B\overline{B}} f_{00}\right] \times \varepsilon_{D^*\ell\nu}, & (7.7) \\ 2\mathcal{B}(B^- \to D^{*0}\ell^-\overline{\nu}_\ell) \times \left[2N_{B\overline{B}}\,(1 - f_{00})\right] \times \varepsilon_{D^*\ell\nu}, & (7.8) \end{cases}$$

   where f_{00} is the branching fraction of $\Upsilon(4S) \to B^0\overline{B^0}$ and $N_{B\overline{B}}$ is the number of $B\overline{B}$ events in the data. The yield is unconstrained in the fit, while f_{00}, $N_{B\overline{B}}$, and $\varepsilon_{D^*\ell\nu}$ are taken as nuisance parameters. $\mathcal{B}(\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell)$ are required to be the same between the fit categories of $D^{*+} \to D^0\pi^+$ and $D^{*+} \to D^+\pi^0$.

3. Background events with a correctly reconstructed D^* candidate

   (a) $\overline{B} \to D^{**}\ell^-\overline{\nu}$
   The yield $N_{D^{**}\ell\nu}$ is unconstrained in the fit to be determined by data because the branching fractions for some of the $\overline{B} \to D^{**}\ell^-\overline{\nu}$ decays have not been measured.

   (b) Hadronic B decays
   The yield N_{HadB} is fixed with simulation.

   (c) "Other" background events
   The yield N_{other} is fixed with simulation.

4. Background events with a fake D^* candidate
   The yield $N_{\text{Fake}D^*}$ is taken as nuisance parameters. The constraint follows a single Gaussian function with σ corresponding to an uncertainty of the data-simulation ratio in the ΔM_{D^*} side-band regions.

Table 7.1: Summary of the fit parameters. The index i takes three fit categories of the D^* modes. The number of parameters are in parentheses. The number of $\varepsilon^i_{D^*\ell\nu}$ is counted only for the signal events because they are shared in parametrization with the normalization events. "Total" row shows the total number of parameters of each type in parentheses.

PDF component	Parameter					
	Unconstrained		Nuisance		Fixed	
1. Signal events	$R(D^*)$	(1)	$\varepsilon^i_{D^*\tau\nu}$, $\varepsilon^i_{D^*\ell\nu}$	(6)	$\frac{N^{\ell\text{-misID}\,i}_{D^*\tau\nu}}{N^i_{D^*\tau\nu}}$	(3)
2. Normalization events	$\mathcal{B}(\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell)$, $\mathcal{B}(B^- \to D^{*0}\ell^-\overline{\nu}_\ell)$	(2)	f_{00} $N_{B\overline{B}}$ $(\varepsilon^i_{D^*\ell\nu})$	(2)	—	
3. Background events with a correct D^*						
(a) $\overline{B} \to D^{**}\ell^-\overline{\nu}$	$N^i_{D^{**}\ell\nu}$	(3)	—		—	
(b) Hadronic B decay	—		—		N^i_{HadB}	(3)
(c) Other background events	—		—		N^i_{other}	(3)
4. Background events with a fake D^*	—		$N^i_{\text{Fake}D^*}$	(3)	—	
Total		(6)		(11)		(15)

Table 7.2: Expected values of the nuisance parameters.

Parameter		Constraint		
f_{00}		0.484 ± 0.012		
$N_{B\overline{B}}$	$[10^6]$	198.0 ± 3.0		
		$D^{*+} \to D^0\pi^+$	$D^{*+} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
$\varepsilon_{D^*\tau\nu}$	$[10^{-5}]$	1.805 ± 0.018	0.277 ± 0.007	1.565 ± 0.017
$\varepsilon_{D^*\ell\nu}$	$[10^{-5}]$	5.363 ± 0.075	0.686 ± 0.027	4.192 ± 0.063
$N_{\text{Fake}D^*}$		$160.7^{+23.7}_{-25.5}$	$27.7^{+4.4}_{-4.8}$	$258.9^{+13.1}_{-13.6}$

Treatment of the fit parameters is summarized in Table 7.1. We assume $R(D^*)$ holds a common value for B^0 and B^+ decays under the isospin symmetry, setting $R(D^*) = R(D^{*+}) = R(D^{*0})$. The parameters indexed with i are unconstrained in the fit independently in each D^* mode. The fit involves a total of 17 floated parameters, out of which 11 parameters are subject to external constraints as nuisance parameters. The constraints on the nuisance parameters are summarized in Table 7.2. The nuisance parameters are treated as mutually independent. The values of the fixed parameters are provided in Table 7.3.

7.1.3 Probability density function

We determine the two-dimensional PDF from the simulation. There are empty bins in the created PDFs due to the limitations of the MC statistics. An adaptive kernel density estimation (KDE) method is utilized to infer the probability in those empty bins reasonably.

Table 7.3: Values of the parameters fixed in the fit for $R(D^*)$.

Parameter	Constrained value		
	$D^{*+} \to D^0\pi^+$	$D^{*+} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
$N^{\ell\text{-misID}}_{D^*\tau\nu}/N_{D^*\tau\nu}$ [%]	6.6	6.6	6.9
N_{HadB}	38.8	4.5	21.2
N_{other}	12.4	1.9	6.7

Table 7.4: Configurations of the adaptive kernel density estimation.

PDF component	`RooFit` class	ρ_{local}	
		E_{ECL}	M^2_{miss}
1. Signal events	`Roo2DKeyPDF`	2.0	2.0
2. Signal events with a misidentified ℓ	`Roo2DKeyPDF`	2.0	2.0
3. Normalization events	—	—	—
4. Background events with a correct D^*			
(a) $\overline{B} \to D^{**}\ell^-\overline{\nu}$	`RooNDKeyPDF`	1.0	0.4
(b) Hadronic B decay	`RooNDKeyPDF`	1.0	0.3
(c) "Other" background events			
$B^0 \leftrightarrow B^+$ cross feed	`RooNDKeyPDF`	2.0	0.1
"Other" $B\overline{B}$ background events	`RooNDKeyPDF`	1.0	0.4
Continuum events	`RooNDKeyPDF`	1.0	1.0
5. Background events with a fake D^*	`RooNDKeyPDF`	2.0	0.9

This estimation is facilitated by the `Roo2DKeyPDF` and `RooNDKeyPDF` classes in the `RooFit` package. The adaptive KDE is applied on every PDF for all event categories, except for the normalization events, in each of the D^* modes. In the adaptive KDE, there is a provision to customize user-specific width scale factors of local densities, denoted as ρ_{local}. Different width scale factors are adopted among event categories as listed in Table 7.4. Optimization of these factors is achieved by comparing the distributions with and without the adaptive KDE. For the "other" background events, different width scale factors are employed in the three components: $B^0 \leftrightarrow B^+$ cross feed, other $B\overline{B}$ background, and continuum events. The PDFs with the adaptive KDE are converted to histogram PDFs, discretized 20 and 24 uniform bins along the axes of E_{ECL} and M^2_{miss}, respectively. Figure 7.1 illustrate the constructed PDFs for (sub)categories in the $D^{*+} \to D^0\pi^+$ mode.

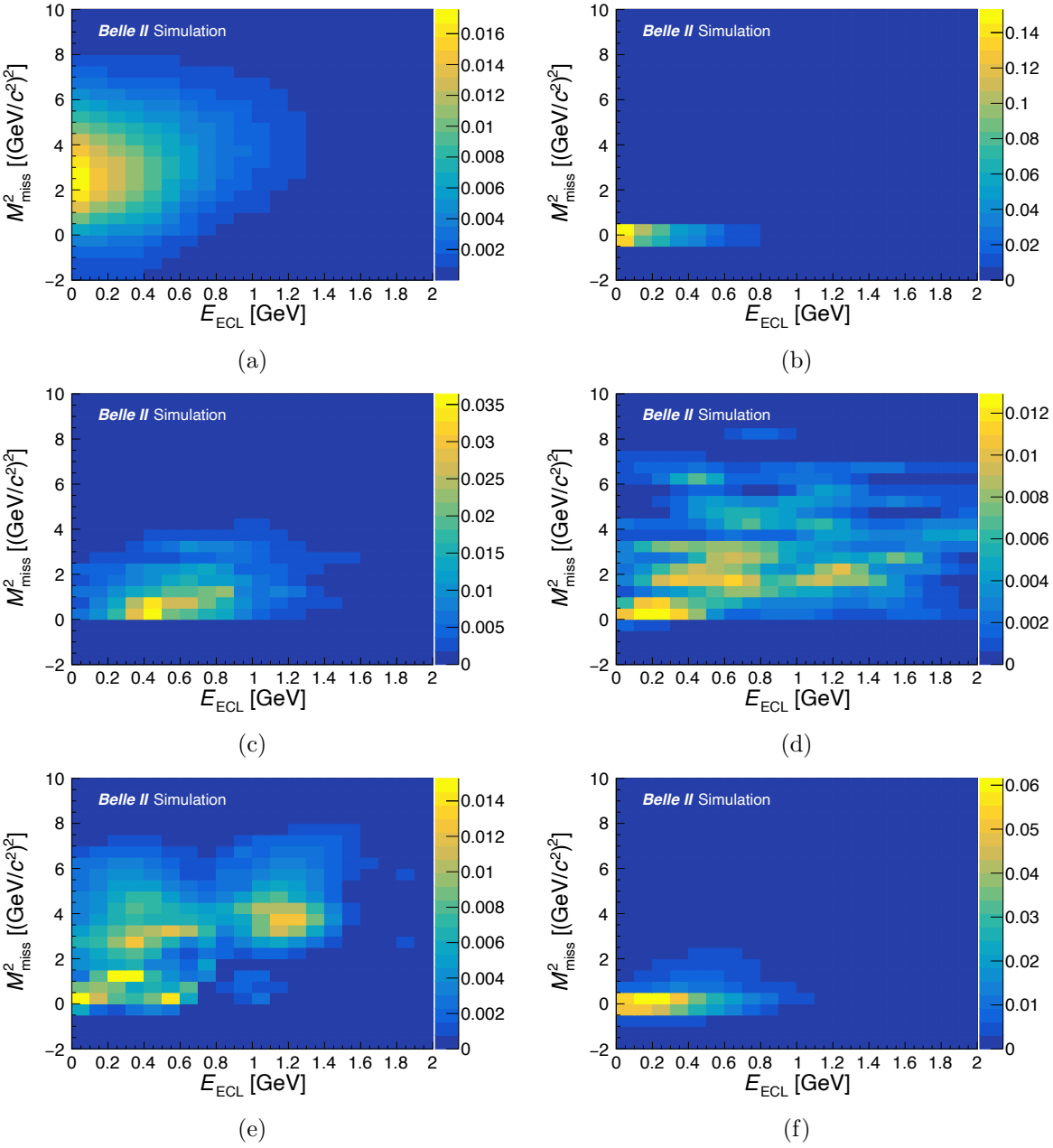


Figure 7.1: Two-dimensional histogram PDFs of E_{ECL} and M^2_{miss} from (a) $\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau$, (b) $\overline{B} \to D^{*+}\ell^-\overline{\nu}_\ell$, (c) $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$, (d) hadronic B decays, (e) "other" background events with a correctly reconstructed D^* candidate, and (f) background events with a fake D^* candidate in the $D^{*+} \to D^0\pi^+$ mode . The z-axis is in an arbitrary unit.

7.1.4 Likelihood function

An extended binned maximum likelihood L is defined as

$$L\left(\vec{n}\middle|\vec{\lambda}\right) \equiv \prod_{i=1}^{N_{\text{mode}}} \prod_{j=1}^{N_{\text{bin}}} \left[\frac{(\mathcal{P}_{ij})^{n_{ij}}}{n_{ij}!} \cdot e^{-\mathcal{P}_{ij}}\right] \cdot \mathcal{C}(\vec{\lambda}_{\text{nuisance}}), \tag{7.9}$$

where n_{ij} represents the observed number of events in the j-th bin for the i-th D^* mode, and $\vec{n}$ represents the collection of these observed counts. The indices i and j run up to $N_{\text{mode}} = 3$ and $N_{\text{bin}} = 480$, respectively. $\vec{\lambda}$ consists of 17 fit parameters floated in the fit, which includes $R(D^*)$ and a set of 11 nuisance parameters under Gaussian constraints, $\vec{\lambda}_{\text{nuisance}} \in \vec{\lambda}$. The expected number of events in the j-th bin for the i-th D^* mode $\mathcal{P}_{ij}$ is calculated with a sum of the PDF components from the six (sub)categories:

$$\mathcal{P}_{ij}\left(\vec{\lambda}\right) = \sum_{k=1}^{N_{\text{cat}}} \left[\nu_i^k\left(\vec{\lambda}\right) f_{ij}^k\right]. \tag{7.10}$$

Here, ν_i^k is the expected yield of the k-th event (sub)category in the i-th D^* mode, and f_{ij}^k corresponds to the probability of j-th bin for the k-th event (sub)category. The total number of event (sub)categories, N_{cat}, is six. The constraints for the nuisance parameters are given as

$$\mathcal{C}(\vec{\lambda}_{\text{nuisance}}) = \prod_{\lambda_l \in \vec{\lambda}_{\text{nuisance}}} \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_{\lambda_l,L} + \sigma_{\lambda_l,R}} \exp\left(-\frac{(\lambda_l - \mu_{\lambda_l})^2}{2\sigma_{\lambda_l}^2}\right), \tag{7.11}$$

where μ_{λ_l} and σ_{λ_l} indicate the mean and standard deviation of a double-sided Gaussian PDF for a nuisance parameter λ_l. The asymmetric standard deviation is defined as

$$\sigma_{\lambda_l} = \begin{cases} \sigma_{\lambda_l,L} & (\lambda_l < \mu_{\lambda_l}), \quad (7.12) \\ \sigma_{\lambda_l,R} & (\lambda_l \geqq \mu_{\lambda_l}). \quad (7.13) \end{cases}$$

7.2 Performance tests

The performance of our fitting procedure is validated through two distinct methods: the Asimov fit, as detailed in Section 7.2.1, and fitting with pseudo-experimental data sets, described in Section 7.2.2. Both approaches consistently yield fitted $R(D^*)$ values that align with the true $R(D^*)$ values. When the true $R(D^*)$ value is set to the SM expected value of 0.254, our fitting method demonstrates a statistical uncertainty in the range of approximately 15 to 16% in these tests.

7.2.1 Asimov fit

An Asimov data set, corresponding to an integrated luminosity of 189.26 fb^{-1} is generated using histogram PDFs with the SM $R(D^*)$ value of 0.254 assumed. A fit was then performed on this Asimov data set, yielding a result of $R(D^*) = 0.254^{+0.042}_{-0.039}$. The central value of this fit result is equal to the true value. The asymmetric errors represent statistical uncertainties of 15.3% and 16.3% on the positive and negative sides, respectively. Additionally, the pull values in all bins, as illustrated in Figures 7.2 to 7.4, are adequately close to zero, in line with the expectations.

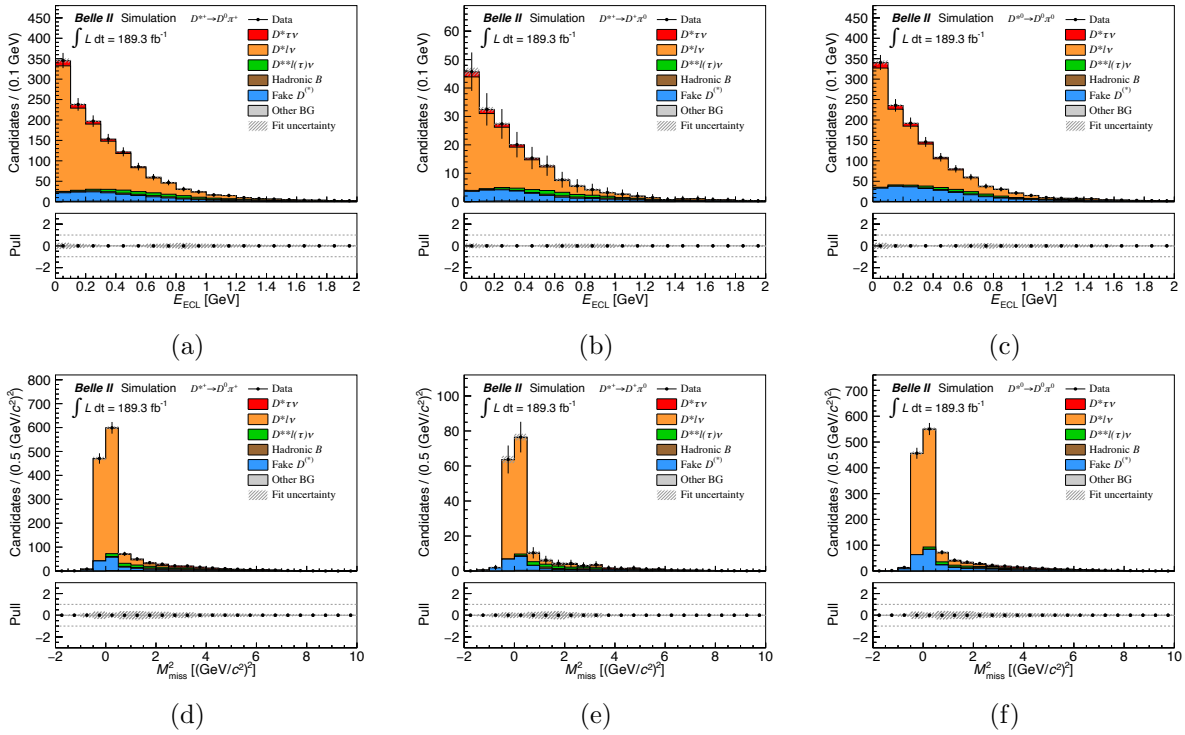


Figure 7.2: Fit results to (a)–(c) the E_{ECL} and (d)–(f) the M^2_{miss} distributions in the $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ (right) modes of the Asimov data set, where $R(D^*)$ is set at 0.254.

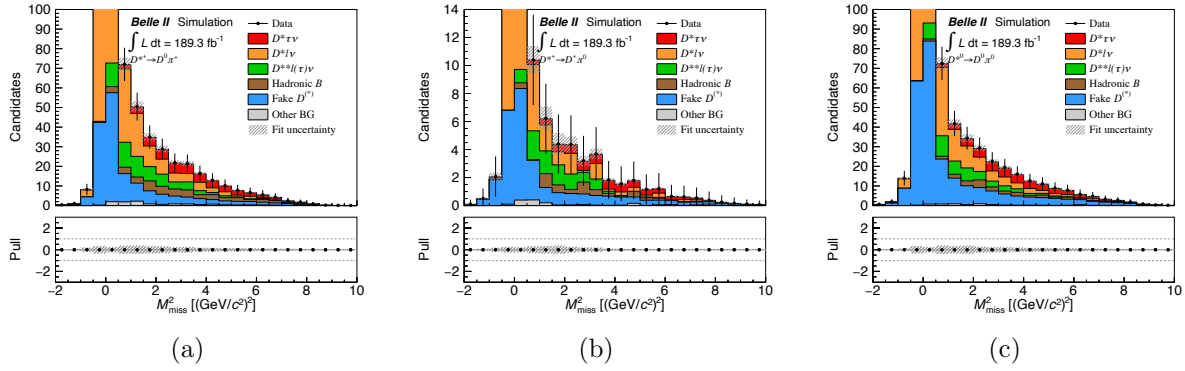
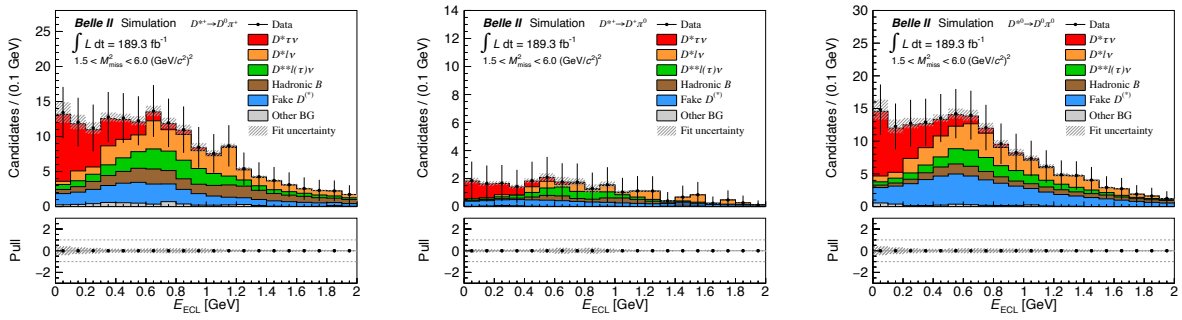


Figure 7.3: Fit results to the M^2_{miss} distributions with zoomed vertical axes in the (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ modes of the Asimov data set, where $R(D^*)$ is set at 0.254.



(a) (b) (c)

Figure 7.4: Fit results to the $E_{\rm ECL}$ distributions for a signal-enhanced selection of $1.5 < M^2_{\rm miss} < 6.0$ $({\rm GeV}/c^2)^2$ in the (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ modes of the Asimov data set, where $R(D^*)$ is set at 0.254.

7.2.2 Fits with pseudo experiment data sets

We generate 5,000 sets of pseudo-experimental data, each containing an integrated luminosity of 189.26 fb^{-1}, using the `generate` function in the `RooFit` package. These sets, functioning as toy MC samples, are based on histogram PDFs with the SM expectation value of $R(D^*)$. Figure 7.5 presents the $R(D^*)$ values and their asymmetric errors derived from 5,000 fits of these pseudo-experimental data sets. The asymmetric errors expect the statistical uncertainties with respect to the SM expectation of $R(D^*)$ as Eq. 7.14.

$$\sigma_{R(D^*)_{\rm SM}} = {}^{+0.041}_{-0.039} \left({}^{+16.1\%}_{-15.3\%}\right) \tag{7.14}$$

They agree with those from the Asimov fit discussed in Section 7.2.1. In addition, the pull distribution is obtained using the fitted $R(D^*)$ values and uncertainties of these 5,000 pseudo experiments as illustrated in Figure 7.6. A Gaussian fit to this pull distribution yields mean and standard deviation values consistent with 0.0 and 1.0, respectively, within a 2σ range of their statistical uncertainties.

To evaluate the linearity between the true and fitted $R(D^*)$ values, we conduct fits on pseudo-experimental data sets with varying the true $R(D)$ value from 0.100 to 0.600 in steps of 0.010. For each true value, 5,000 data sets are fitted, and the mean and statistical uncertainty of the fitted $R(D^*)$ values are determined. This test demonstrates linearity as illustrated in Figure 7.7. The bias between the true and fitted values is quantified by fitting a linear function to the test results, represented as:

$$R(D^*)_{\rm true} = 1.008 \cdot R(D^*) - 0.003. \tag{7.15}$$

Here, $R(D^*)_{\rm true}$ denotes the true $R(D^*)$ value. The uncertainty associated with this linear function is found to be at most 0.1% of $R(D^*)_{\rm true}$ within the tested range. While we do not adjust the $R(D^*)$ values obtained from the fit, the observed bias is accounted for as a systematic uncertainty arising from the fitting procedure.

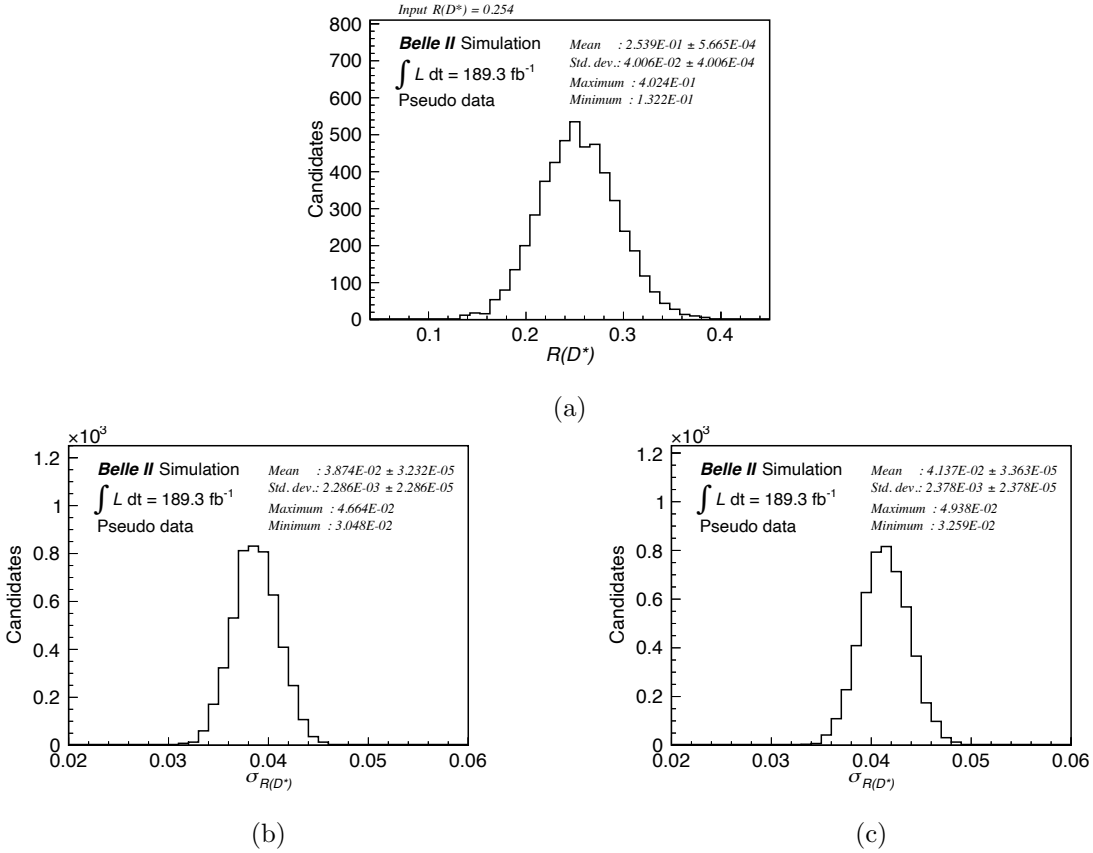


Figure 7.5: Results of the 5,000 pseudo experiments at the SM expectation: $R(D^*) = 0.254$. The distribution of (a) the fitted $R(D^*)$ values and (b) their negative and (c) positive $R(D^*)$ errors are shown.

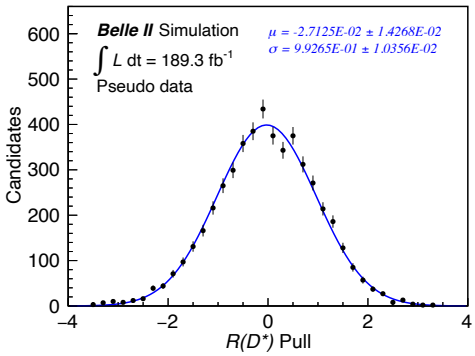


Figure 7.6: Pull distribution of the $R(D^*)$ fit using the pseudo experiment data sets at the SM expectation: $R(D^*) = 0.254$.

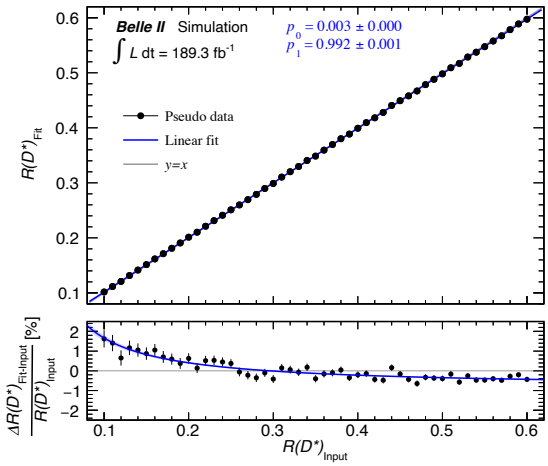


Figure 7.7: Linearity of $R(D^*)$ obtained from the signal extraction fit against the true $R(D^*)$ ranging from 0.100 to 0.600 in 0.010 steps. The blue line is the fitted linear function. The bottom panel shows the normalized differences of the fitted $R(D^*)$ values from the true $R(D^*)$ values.

8 Results

8.1 Fit results

We observe the data as shown in Figure 8.1. From the fit to the observed events, we find

$$R(D^*) = 0.262\ ^{+0.041}_{-0.039}\ (\text{stat.})\ ^{+0.035}_{-0.032}\ (\text{syst.}). \tag{8.1}$$

Figures 8.2 to 8.4 illustrate the projected distributions of the fit results. The p-value of the goodness of fit stands at 4.4%, determined by the χ^2 distribution from the fits with pseudo data sets at the fitted $R(D^*)$ as shown in Figure 8.5. Tables 8.1 and 8.2 list the resulting values of the floated and nuisance parameters. Figure 8.6 shows the observed correlation between those parameters. Table 8.3 presents the observed signal and normalization events for each D^* mode, calculated based on the fit result. Table 8.4 summarizes the systematic uncertainties observed in the real data. The details of these systematic uncertainties are discussed in Section 9.

The likelihood values relative to the maximum value in the nominal fit, L_{max}, are computed as a function of the $R(D^*)$ values. Figure 8.7 displays the negative log-likelihood ratios. The significance of the signal events is calculated using $\sqrt{-2\log\left(L_0/L_{\text{max}}\right)}$, where L_0 represents the likelihood for a background-only hypothesis, corresponding to $R(D^*) = 0$. The signal decays are observed with a significance of 9.7 σ, which accounts for statistical uncertainty only. When incorporating the systematic uncertainty, the significance is assumed to be proportional to $\Delta R(D^*)$, with a coefficient of $1/\sigma$. The total uncertainty σ is determined as a quadrature sum of statistical and systematic uncertainties in Eq. (8.1), calculated by

$$\sigma = \sqrt{\sigma^2_{\text{stat}} + \sigma^2_{\text{syst}}}. \tag{8.2}$$

The $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$ decays are observed with a significance of 5.2 σ, including the systematic uncertainty.

Table 8.1: Observed (expected) values of the floated parameters. Only statistical uncertainties are given.

Parameter		Observed (expected) value		
$R(D^*)$		$0.262\ ^{+0.041}_{-0.039}$		
$\mathcal{B}(\overline{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell)$	[%]	$5.27\ ^{+0.25}_{-0.24}\ (5.03 \pm 0.11)$		
$\mathcal{B}(B^- \to D^{*0}\ell^-\bar{\nu}_\ell)$	[%]	$5.50\ ^{+0.28}_{-0.27}\ (5.41 \pm 0.11)$		
		$D^{*+} \to D^0\pi^+$	$D^{*+} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
$N_{D^{**}\ell\nu}$		$34.7^{+19.2}_{-18.2}$ (61.6 ± 2.2)	$5.8^{+5.6}_{-4.7}$ (9.0 ± 0.9)	$64.5^{+19.3}_{-18.3}$ (46.0 ± 2.0)

Table 8.2: Observed (expected) values of the nuisance parameters.

Parameter		Observed (expected) value		
f_{00}		0.484 ± 0.012 (0.484 ± 0.012)		
$N_{B\bar{B}}$ $[10^6]$		198.0 ± 3.0 (198.0 ± 3.0)		
		$D^{*+} \to D^0\pi^+$	$D^{*+} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
$\varepsilon_{D^*\tau\nu}$	$[10^{-5}]$	1.805 ± 0.018 (1.805 ± 0.018)	0.277 ± 0.007 (0.277 ± 0.007)	1.565 ± 0.017 (1.565 ± 0.017)
$\varepsilon_{D^*\ell\nu}$	$[10^{-5}]$	5.368 ± 0.075 (5.363 ± 0.075)	0.683 ± 0.025 (0.686 ± 0.027)	$4.190^{+0.063}_{-0.062}$ (4.192 ± 0.063)
$N_{\mathrm{Fake}D^*}$		$164.7^{+21.9}_{-22.6}$ $(160.7^{+23.7}_{-25.5})$	$28.8^{+4.1}_{-4.2}$ $(27.6^{+4.4}_{-4.8})$	$258.9^{+12.7}_{-12.7}$ $(251.4^{+13.1}_{-13.6})$

Table 8.3: Observed (expected) yields of the signal and normalization events. Only statistical uncertainties are given. The signal events include those with a misidentified lepton candidate.

Parameter	Observed (expected) yield		
	$D^{*+} \to D^0\pi^+$	$D^{*+} \to D^+\pi^0$	$D^{*0} \to D^0\pi^0$
$N_{D^*\tau\nu}$	50.9 ± 7.8	7.8 ± 1.2	49.2 ± 7.5
$N_{D^*\ell\nu}$	1084.6 ± 36.7 (1041.0 ± 11.2)	137.9 ± 6.6 (133.2 ± 4.3)	940.9 ± 36.0 (927.2 ± 10.7)

8.2 Consistency checks with split samples

The robustness of the fit results is tested using split data sets. Three tests are conducted on data sets divided into two by reconstructed modes or run periods as follows:

- Reconstructed modes of lepton candidates in the B_{sig} side, either $\overline{B} \to D^* e^- \overline{\nu}_e$ or $\overline{B} \to D^* \mu^- \overline{\nu}_\mu$.
- Charge of B_{sig} candidates, either $\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell$ or $B^- \to D^{*0}\ell^-\overline{\nu}_\ell$.
- Run periods, with runs taken either before or after noon of Apr. 16th, 2021 (JST).

In the first test, where the data set is separated by reconstructed lepton modes on the B_{sig} side, the different PDFs are prepared for each split data set by also separating the reconstructed MC candidates based on the lepton modes. For the other tests, the same PDFs are used in the fit. The $R(D^*)$ fits are performed on the split data sets under the same configuration as the nominal fit. Figure 8.8 compares the fit results only with statistical uncertainties. All tests reveal consistency between the fit results of the two split

Table 8.4: Summary of observed systematic uncertainties in $R(D^*)$.

Source	Uncertainty
PDF shapes	$^{+9.1\%}_{-8.3\%}$
MC statistics	$^{+7.5\%}_{-7.5\%}$
$\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ branching fractions	$^{+4.8\%}_{-3.5\%}$
Fixed backgrounds	$^{+2.7\%}_{-2.3\%}$
Hadronic B decay branching fractions	$^{+2.1\%}_{-2.1\%}$
Reconstruction efficiency	$^{+2.0\%}_{-2.0\%}$
Kernel density estimation	$^{+2.0\%}_{-0.8\%}$
Form factors	$^{+0.5\%}_{-0.1\%}$
Peaking background on ΔM_{D^*}	$^{+0.4\%}_{-0.4\%}$
$\tau^- \to \ell^-\nu_\tau\bar{\nu}_\ell$ branching fractions	$^{+0.2\%}_{-0.2\%}$
$R(D^*)$ fit method	$^{+0.1\%}_{-0.1\%}$
Total systematic uncertainty	$^{+13.5\%}_{-12.3\%}$

data sets within 1σ. Moreover, they all agree with the nominal fit result in Eq. (8.1) within their statistical uncertainties. Consequently, this confirms the robustness of the nominal fit result.

8.3 Branching fractions of $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$

In the signal extraction fit, the branching fractions of $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$ decays are also measured. We obtain

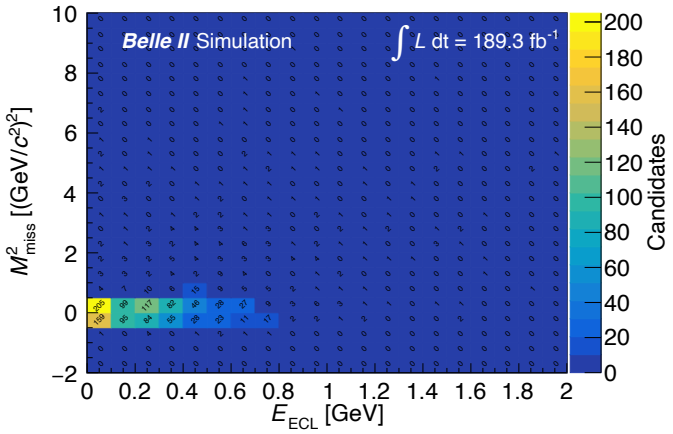
$$\mathcal{B}(\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell) = 5.27\ ^{+0.25}_{-0.24}\ (\text{stat.})\ ^{+0.14}_{-0.12}\ (\text{syst.})\ \%, \tag{8.3}$$

$$\mathcal{B}(B^- \to D^{*0}\ell^-\overline{\nu}_\ell) = 5.50\ ^{+0.28}_{-0.27}\ (\text{stat.})\ ^{+0.14}_{-0.14}\ (\text{syst.})\ \%. \tag{8.4}$$

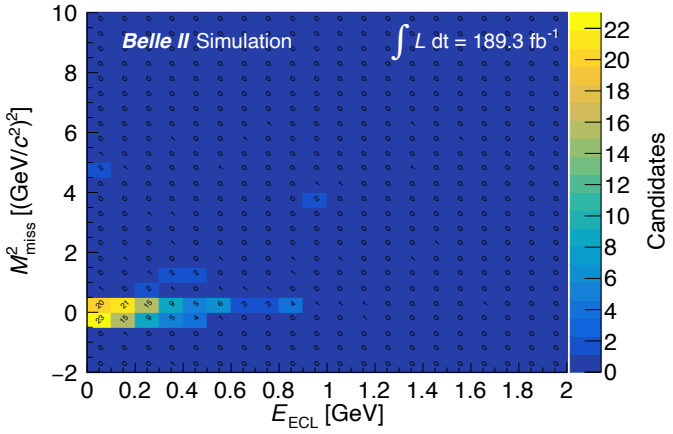
The systematic uncertainties are evaluated as listed in 8.5, following the same procedures as those for the $R(D^*)$ described in Section 9. This is the first measurement of the branching fraction of $B^- \to D^{*0}\ell^-\overline{\nu}_\ell$ at the Belle II experiment. Both branching fractions are in agreement with the world averages of the isospin-averaged branching fractions, which are 5.03 ± 0.11 for the neutral and 5.41 ± 0.11 for the charged B decays, respectively.

Table 8.5: Summary of observed systematic uncertainties in $\mathcal{B}(\overline{B} \to D^* \ell^- \overline{\nu}_\ell)$ from the signal extraction fit.

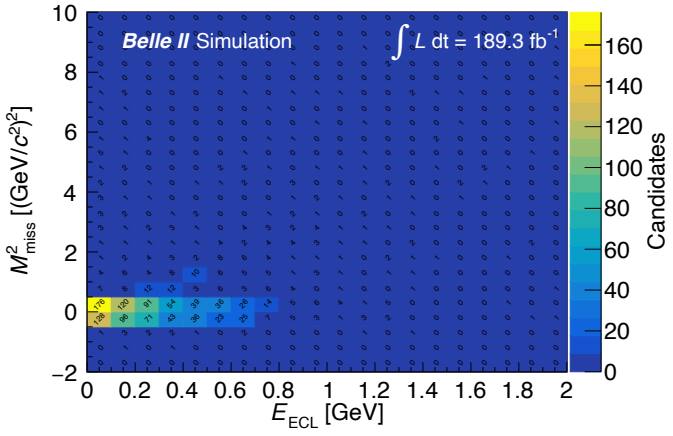
Source	Uncertainty	
	$\mathcal{B}(B^0)$	$\mathcal{B}(B^+)$
PDF shapes	$^{+1.1\%}_{-0.7\%}$	$^{+1.2\%}_{-0.9\%}$
MC statistics	$^{+2.2\%}_{-2.2\%}$	$^{+2.0\%}_{-2.0\%}$
$\overline{B} \to D^{**} \ell^- \overline{\nu}_\ell$ branching fractions	$^{+0.1\%}_{-0.2\%}$	$^{+0.9\%}_{-1.2\%}$
Fixed backgrounds	$^{+0.1\%}_{-0.1\%}$	$^{+0.1\%}_{-0.1\%}$
Hadronic B decay branching fractions	$^{+0.1\%}_{-0.2\%}$	$^{+0.1\%}_{-0.1\%}$
Kernel density estimation	$^{+0.7\%}_{-0.0\%}$	$^{+0.5\%}_{-0.4\%}$
$R(D^*)$ fit method	$^{+0.2\%}_{-0.0\%}$	$^{+0.1\%}_{-0.0\%}$
Total systematic uncertainty	$^{+2.6\%}_{-2.3\%}$	$^{+2.6\%}_{-2.5\%}$



(a)



(b)



(c)

Figure 8.1: Observed two-dimensional distributions of $E_{\rm ECL}$ and $M^2_{\rm miss}$ from the (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, (c) $D^{*0} \to D^0\pi^0$ modes in the data. The figures in the plots are the number of entries in each bin.

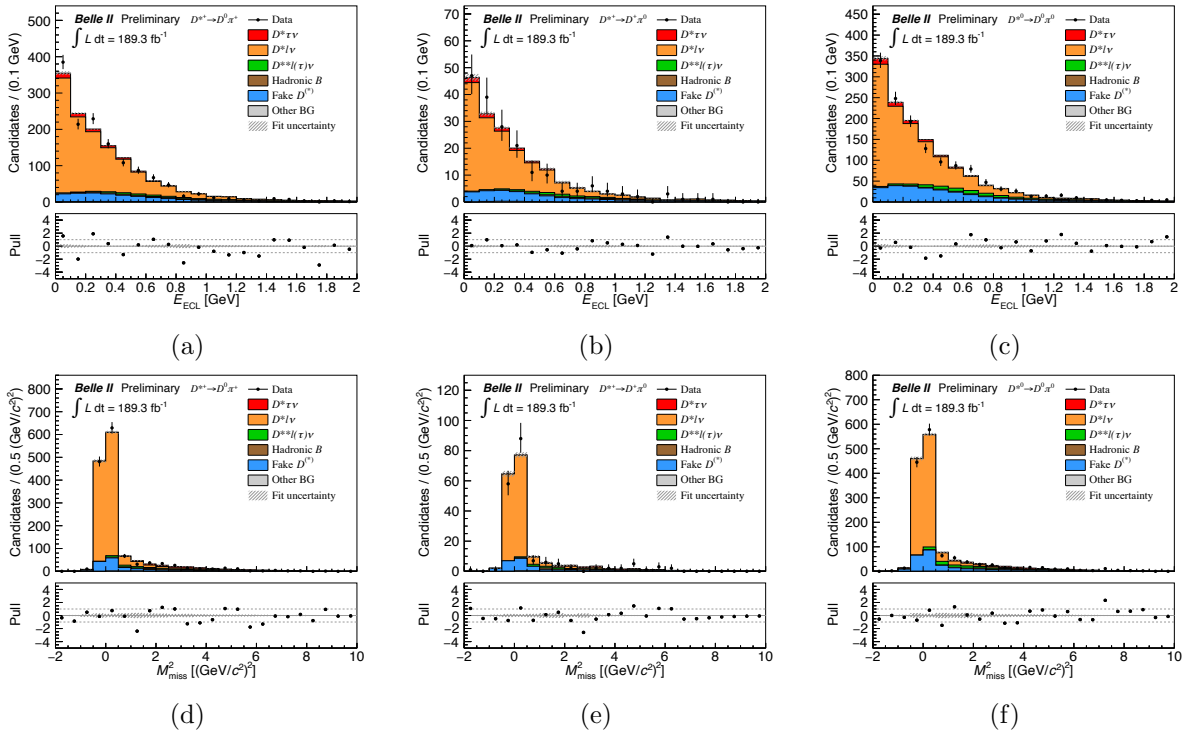


Figure 8.2: Results of the signal extraction fit with the real data projected on the E_{ECL} ((a)–(c)) and M^2_{miss} axes ((d)–(f)) in the $D^{*+} \to D^0\pi^+$ (left), $D^{*+} \to D^+\pi^0$ (middle), and $D^{*0} \to D^0\pi^0$ (right) modes. The black points are the data and the stack histograms are the fitted PDFs. The bottom panels show the pull values of the data yield to the total fitted yield in each bin.

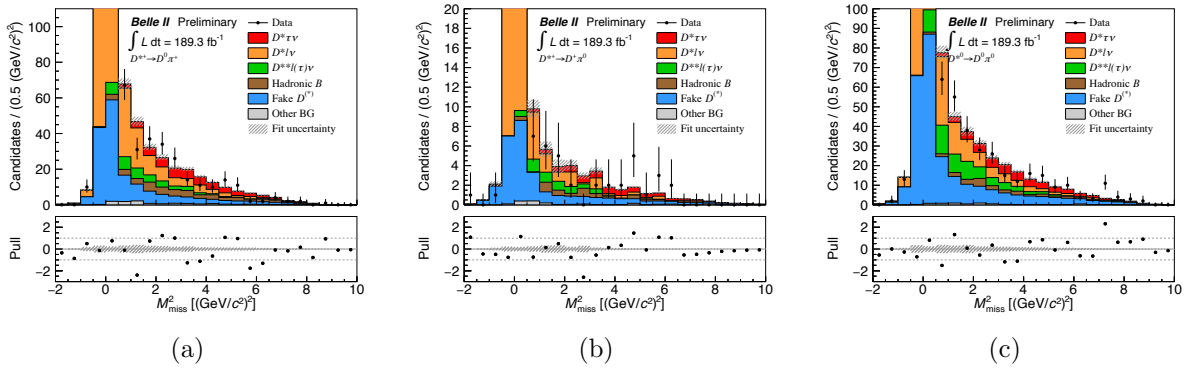
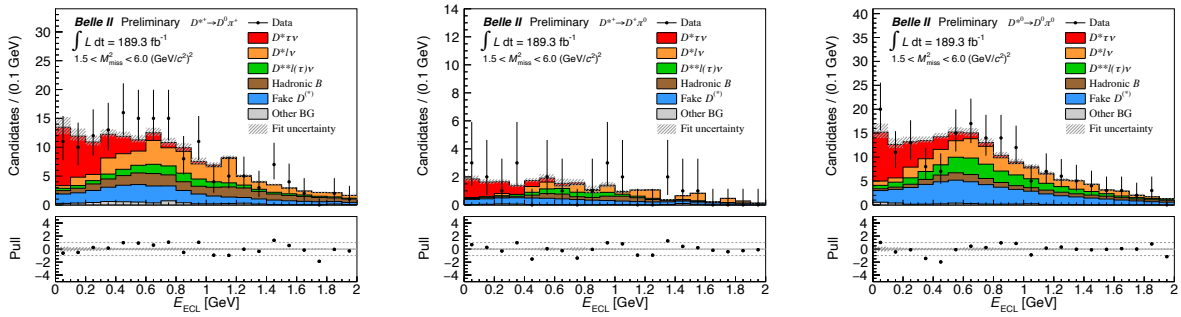


Figure 8.3: Results of the signal extraction fit with the real data projected on the M^2_{miss} distributions with zoomed vertical axes in the (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ modes. The black points are the data and the stack histograms are fitted PDFs. The bottom panels show the pull values of the data yield to the total fitted yield in each bin.



(a) (b) (c)

Figure 8.4: Results of the signal extraction fit with the real data projected on the E_{ECL} distributions for a signal-enhanced selection of $1.5 < M^2_{\mathrm{miss}} < 6.0$ $(\mathrm{GeV}/c^2)^2$ in the (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$ modes.

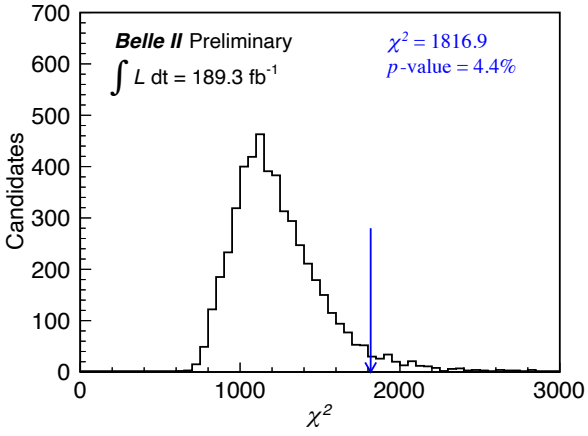


Figure 8.5: Distribution of χ^2 values from the signal extraction fits with pseudo data sets at $R(D^*) = 0.262$. The blue arrow indicates the χ^2 value of 1816.9 obtained in the data.

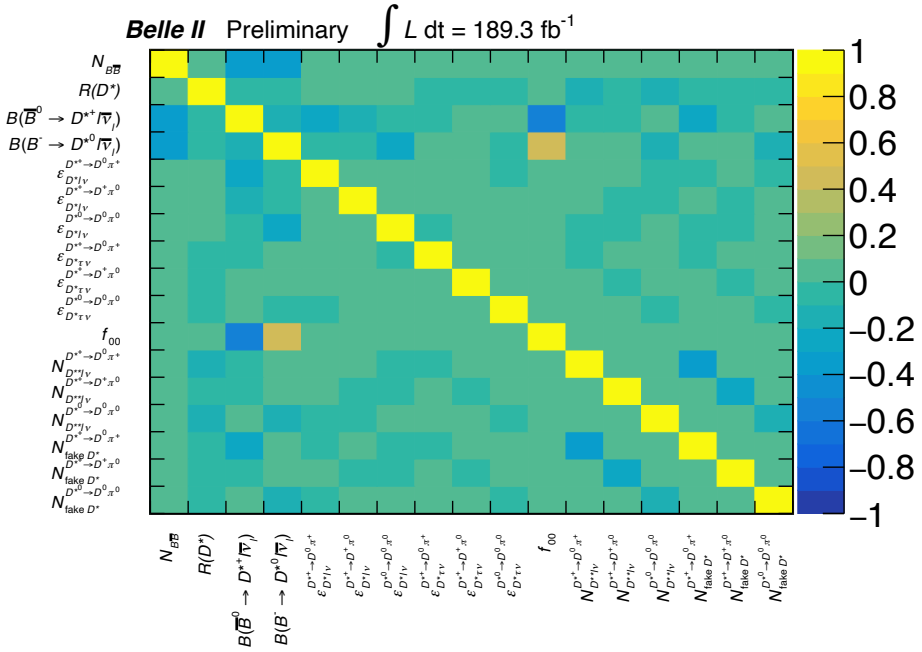
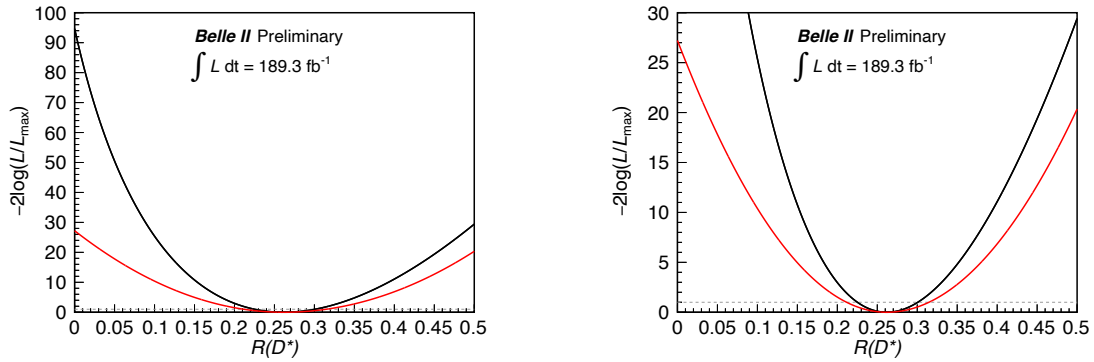


Figure 8.6: Correlation matrix of the floated and nuisance parameters observed in the $R(D^*)$ fit to the data.



(a) (b)

Figure 8.7: Distribution of the negative log-likelihood ratio as a function of the $R(D^*)$ values. The right plot enlarges the small region of the negative log-likelihood ratio in the left plot. The black and red curves show the log-likelihood ratios with the statistical uncertainty and the total uncertainty, respectively. The horizontal dashed grey lines indicate $-2\log(L/L_{\max}) = 1$.

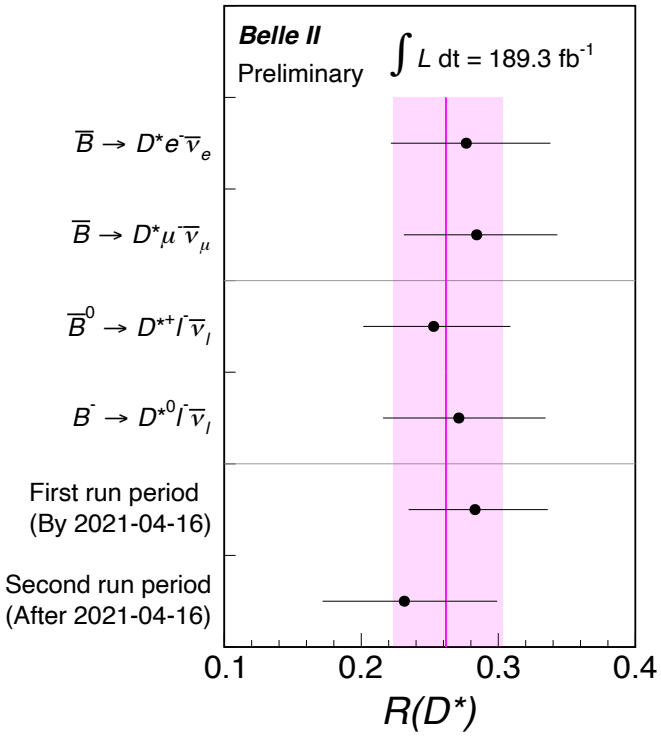


Figure 8.8: Comparison of fit results with split data sets. Fit results from two split data sets based on reconstructed modes of lepton candidates in the B_{sig} side (top), charge of B_{sig} candidates (middle), run periods (bottom) are compared in pairs. All the uncertainties are statistical only. The magenta band indicates the result of the nominal fit and its statistical uncertainty.

9 Systematic uncertainty

Systematic uncertainties in $R(D^*)$ are summarized in Table 8.4. We account for uncertainties in detector responses, underlying physics processes, and discrepancies between the real data and the simulation. Systematic uncertainties, except those from the kernel density estimation, are assessed through refitting, testing procedures, or a combination of them that involve variations in the sources of these uncertainties. The procedures are outlined as follows:

- Refitting
  1. Sources of systematic uncertainty are varied randomly, which leads to the construction of a new PDF. This step includes updating the constraints on nuisance parameters and the values of fixed parameters.
  2. With this new PDF, a refit of $R(D^*)$ to the data is then performed.
  3. We repeat this refitting process (steps 1 and 2) 1,000 times. The variation in the fitted $R(D^*)$ values from the nominal result is recorded as $\Delta R(D^*)$.
  4. The systematic uncertainty is quantified by fitting a double-sided single Gaussian function to the distribution of $\Delta R(D^*)$ from these 1,000 fits or taking the minimum and maximum shifts within the 1,000 fits. The Gaussian function is characterized by a mean μ and asymmetric errors σ_L and σ_H. The function provides the uncertainty range determined by $\mu+\sigma_H$ and $\mu-\sigma_L$ for the positive and negative errors in $R(D^*)$, respectively.
- Testing
  1. We vary the sources of systematic uncertainty by a predetermined amount and construct a new PDF in this test configuration. This step updates the constraints on nuisance parameters and fixed parameter values as well.
  2. The $R(D^*)$ fit to the data is performed with this new PDF.
  3. The deviation from the nominal $R(D^*)$ fit result is treated as a systematic uncertainty in $R(D^*)$.

9.1 Shapes of probability density functions

In a region of $M^2_{\text{miss}} < 1.0\ (\text{GeV}/c^2)^2$, we have observed discrepancies between data and simulation in the shapes of the fit variables, as discussed in Section 5. To address these discrepancies, corrections are introduced, including an energy shift for fake γ candidates in the ROE and smearing factors for the resolution of M^2_{miss}. The uncertainties associated with these correction factors, listed in Table 9.1, contribute to changes in the PDF shapes.

To estimate the systematic uncertainty in the PDF shapes due to these corrections, the energy shifts for fake γ candidates and the smearing factors for the M^2_{miss} resolution are varied according to their respective uncertainties. The resolutions of M^2_{miss} are varied randomly using single Gaussian functions, consistently fluctuated between the neutral and

Table 9.1: Correction factors for the shapes of $E_{\rm ECL}$ and $M^2_{\rm miss}$.

Fit variable	Correction factor
$E_{\rm ECL}$	$\delta E_\gamma = -15^{+7}_{-9}$ MeV
$M^2_{\rm miss}$	$\Delta\sigma_{M^2_{\rm miss}} = \begin{cases} 0.061^{+0.015}_{-0.017}\ ({\rm GeV/c^2})^2 \text{ for the neutral } B \text{ mode} \\ 0.060^{+0.028}_{-0.054}\ ({\rm GeV/c^2})^2 \text{ for the charged } B \text{ mode} \end{cases}$

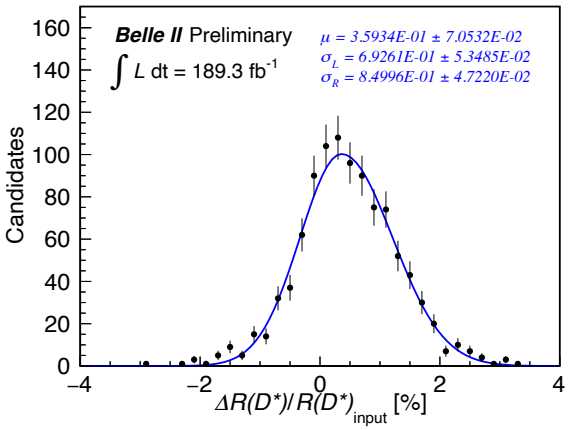


Figure 9.1: Distribution of $\Delta R(D^*)$ from 1000 fits with the energy shift of -15 MeV for the $E_{\rm ECL}$ correction in the evaluation of systematic uncertainty due to the PDF shapes. The fitted double-sided single Gaussian function is shown by the blue curve.

charged B modes based on the uncertainties. This variation is performed 1,000 times for each energy shift of fake γ candidates. For each energy shift, $\pm1\sigma$ ranges in $\Delta R(D^*)$ distributions are obtained from 1,000 refits. An example is shown in Figure 9.1 for the energy shift of -15 MeV. These 1,000 refits are executed at energy shifts ranging from -24 MeV to -8.0 MeV in 2.0 MeV steps as well as at the nominal energy shift of -15 MeV. Among all tested energy shifts, the largest positive and negative deviations at the $\pm1\sigma$ points are then identified as the uncertainty. Figure 9.2 illustrates the dependence of $\Delta R(D^*)$ on these energy shifts. The uncertainty is evaluated to be $^{+0.024}_{-0.022}$ $\binom{+9.1\%}{-8.3\%}$.

9.2 Simulation sample size

The PDFs are constructed using MC samples of limited size. The shape of these PDFs may vary due to statistical fluctuations in the generation process. To estimate uncertainties arising from the limited MC statistics, MC distributions are produced using the bootstrapping method [51]. This method involves randomly resampling the simulated events with replacement from the original MC sample in each candidate category of every D^* mode. Simultaneously, the number of simulated events in each candidate category is fluctuated using a Poisson distribution, with the mean set to the number of entries in the original MC samples. The reconstruction efficiencies, $\varepsilon_{D^*\tau\nu}$ and $\varepsilon_{D^*\ell\nu}$, are updated accordingly.

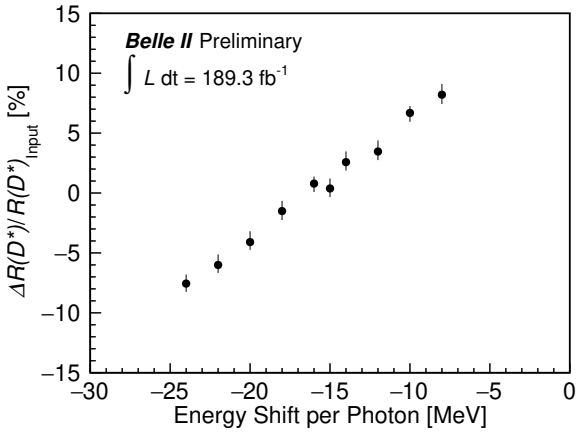


Figure 9.2: Results of the evaluation of systematic uncertainties from the PDF shapes as a function of energy shifts, ranging from -24 to -8 MeV. The uncertainty of points corresponds to $\pm 1\sigma$ of $\Delta R(D^*)$ in the evaluation at each of the energy shifts.

Adaptive KDE is applied to the resampled MC distributions to construct new PDFs based on the resampled events. The new PDF is then fitted to the data, and these $R(D^*)$ fits are repeated 1,000 times.

Figure 9.3 illustrates the distribution of $R(D^*)$ differences from the nominal fit result. A shift towards the positive direction in the mean value of the $\Delta R(D^*)$ distribution is observed, which is attributable to low statistics of background events in the simulation. To demonstrate the bias caused by the limited size of the MC sample, resampling is tested while scaling the number of simulated events taken by bootstrapping from one-tenth to three times the original MC sample size in increments of 0.1. This approach simulates evaluations with an increased or decreased MC sample size for the PDF construction. The $\Delta R(D^*)$ distributions at half and double the sample size in resampling are presented in Figure 9.4. Figure 9.5 summarizes the mean and standard deviations obtained by single Gaussian fits to the $\Delta R(D^*)$ distributions as a function of the resampling size factor, $r_{N_{\text{resample}}}$, relative to the original MC size. The results indicate that both the bias in the mean values and the standard deviations are dependent on the resampling size. Functions of $p_0 + (p_1/r_{N_{\text{resample}}})$ and $p_0/\sqrt{r_{N_{\text{resample}}}}$ are fitted to the mean and standard deviations for values of $r_{N_{\text{resample}}}$ between 0.1 and 1.0, respectively. The evaluated mean and standard deviations in the higher $r_{N_{\text{resample}}}$ region follow the extrapolated fitted functions. As the resampled size increases, the bias diminishes to a negligible level. In this analysis, a symmetric systematic uncertainty is conservatively assigned by taking sum of the mean and right-hand sigma of the Gaussian fit in the $\Delta R(D^*)$ distribution at $r_{N_{\text{resample}}}$ equal to 1.0. The uncertainty is estimated to be $^{+0.020}_{-0.020}$ $\binom{+7.5\%}{-7.5\%}$.

9.3 Branching fractions of $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$

The composition uncertainties of $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ decays in the PDFs may introduce the systematic uncertainty in $R(D^*)$. These branching fractions are fluctu-

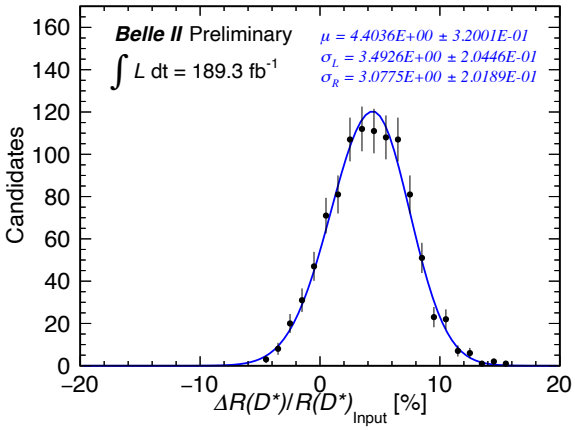
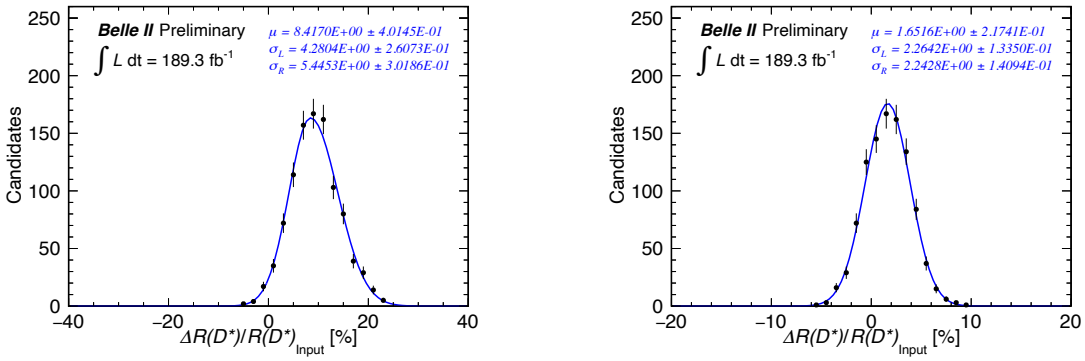


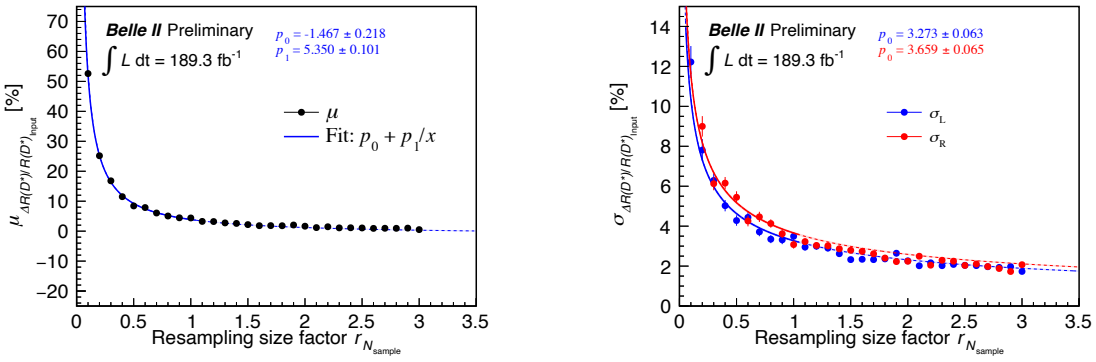
Figure 9.3: Distribution of $R(D^*)$ differences from the nominal fit result after the bootstrap application. The fitted double-sided single Gaussian is shown by the blue curve.



(a) (b)

Figure 9.4: Distributions of $R(D^*)$ differences from the nominal fit result after the bootstrap application scaling to a half (doubled) sample size. The fitted double-sided single Gaussian function is shown by the blue curve.

ated based on the uncertainties listed in Table 4.3. The fluctuation is applied to the four categories of decay modes as detailed in Table 9.2. For measured resonant $\overline{B} \to D^{**}\ell\overline{\nu}_\ell$, namely $\overline{B} \to \{D_1, D_0^*, D_1', D_2^*\}\ell^-\overline{\nu}_\ell$, $\overline{B} \to D_s^{(*)}K\ell^-\overline{\nu}_\ell$, and non-resonant $\overline{B} \to D^{(*)}\pi\ell^-\overline{\nu}_\ell$ decays, single Gaussian functions are used to vary their branching fractions. For modes not yet observed, branching fractions are varied randomly within a uniform distribution from 0% to 200% of the estimated values for each decay mode category, representing a conservative evaluation. The branching fractions of the unmeasured modes are fully correlated within their respective categories during these fluctuations. This variation process and the $R(D^*)$ fitting are repeated 1,000 times for all decay mode categories. The resulting $R(D^*)$ differences for each category are depicted in Figure 9.6. For the category of measured resonant $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D_s^{(*)}\ell^-\overline{\nu}_\ell$ decays, uncertainties are determined using the standard deviation range of a single double-sided Gaussian function fitted to the $\Delta R(D^*)$



(a) (b)

Figure 9.5: Dependence of the mean and standard deviation against the factor of the resampling size, $r_{N_{\text{resample}}}$. The points are evaluated mean and standard deviations at each $r_{N_{\text{resample}}}$. The fitted function of $p_0 + p_1/r_{N_{\text{resample}}}$ are shown by the blue line in Figure (a). The fitted functions of $p_0/\sqrt{r_{N_{\text{resample}}}}$ are shown by the blue and red lines for left- and right-hand sigmas in Figure (b), respectively.

Table 9.2: Systematic uncertainties from the branching fractions of $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$.

Decay mode	Fluctuation function	Uncertainty	
Measured resonant $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ $\overline{B} \to D_s^{(*)}K\ell^-\overline{\nu}_\ell$	Single Gaussian function	$^{+0.000}_{-0.000}$	$\binom{+0.1\%}{-0.1\%}$
Non-resonant $\overline{B} \to D^{(*)}\pi\ell^-\overline{\nu}_\ell$	Single Gaussian function	$^{+0.000}_{-0.005}$	$\binom{+0.1\%}{-1.9\%}$
The $\overline{B} \to X_c\ell^-\overline{\nu}_\ell$ gap	Uniform (±100%)	$^{+0.012}_{-0.004}$	$\binom{+4.7\%}{-1.5\%}$
$\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$	Uniform (±100%)	$^{+0.003}_{-0.007}$	$\binom{+1.1\%}{-2.5\%}$
Total (quadratic sum)		$^{+0.013}_{-0.009}$	$\binom{+4.8\%}{-3.5\%}$

distributions. In the other categories, , where distribution shapes are not Gaussian-like, the minimum and maximum shifts from the 1,000 $R(D^*)$ fits are conservatively assigned as negative and positive systematic uncertainties. These estimated uncertainties are summarized in Table 9.2. The total uncertainty resulting from the $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell/D^{**}\tau^-\overline{\nu}_\tau$ compositions is calculated as a quadratic sum of the uncertainties from all the decay mode categories, estimated to be $^{+0.013}_{-0.009}$ $\binom{+4.8\%}{-3.5\%}$.

9.4 Background categories with fixed yields

Candidates in the "other " background category are accompanied by an incorrectly reconstructed B_{tag} candidate. To account for possible discrepancies between data and simulation

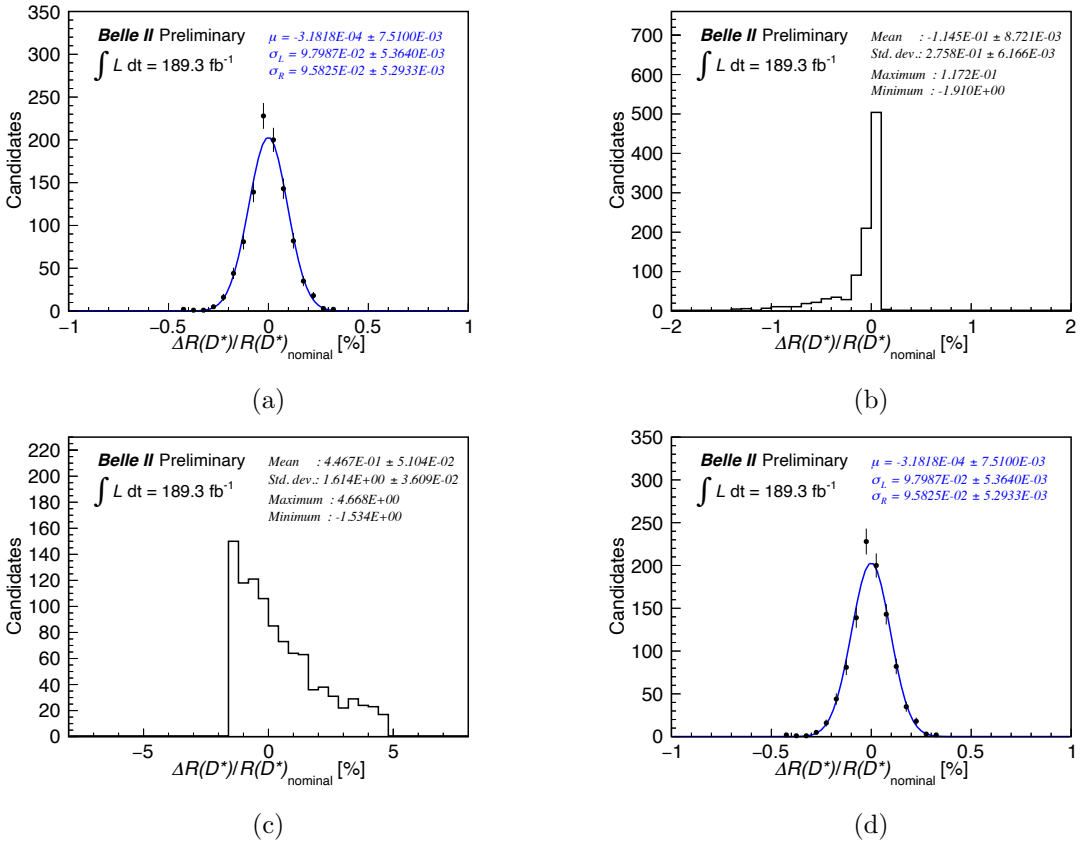


Figure 9.6: Distributions of $\Delta R(D^*)$ in 1000 fits involving the fluctuations within uncertainties on the branching fractions of $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ decays. (a) Measured resonant $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D_s^{(*)}K\ell^-\overline{\nu}_\ell$ decays. The fitted double-sided single Gaussian function is shown by the blue curve. (b) Non-resonant $\overline{B} \to D^{(*)}\pi\ell^-\overline{\nu}_\ell$ decays. (c) Gap component of $\overline{B} \to X_c\ell^-\overline{\nu}_\ell$ decays filled by $\overline{B} \to D^{**}(\to D^{(*)}\pi\pi)\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D^{**}(\to D^{(*)}\eta)\ell^-\overline{\nu}_\ell$ decays. (d) $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ decays.

in the fractions of the wrong B_{tag} candidates, the fixed yields of this category are varied across all D^* modes. This variation employs uniform distributions ranging from 0% to 200% of the expected yields in the simulation. The variation process is independently conducted for three divided event types: $B^0 \leftrightarrow B^+$ cross-feed, "other" $B\overline{B}$ background, and continuum events. Each type undergoes 1,000 iterations of this variation process.

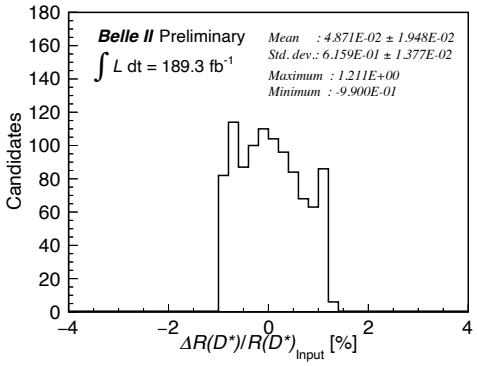
The distributions of $\Delta R(D^*)$ for each event type are illustrated in Figure 9.7. The maximum and minimum shifts observed in the $\Delta R(D^*)$ are considered as the systematic uncertainty, as summarized in Table 9.3. These uncertainties are then combined in a quadratic sum for all three categories, resulting in a total systematic uncertainty of $^{+2.7\%}_{-2.3\%}$.

9.5 Branching fractions of hadronic B decays

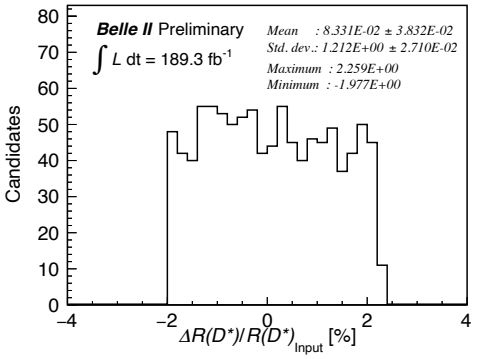
Similarly to the uncertainties from the branching fractions of $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$, the composition uncertainties of hadronic B decays in the PDFs may also contribute to the systematic

Table 9.3: Systematic uncertainties from the "other" background categories.

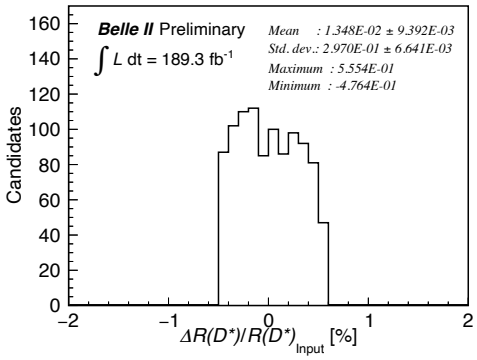
Type	Fluctuation function	Uncertainty
$B^0 \leftrightarrow B^+$ cross-feed events	Uniform ($\pm 100\%$)	$^{+0.003}_{-0.003}$ $\left(^{+1.2\%}_{-1.0\%}\right)$
"Other" $B\overline{B}$ background events	Uniform ($\pm 100\%$)	$^{+0.006}_{-0.005}$ $\left(^{+2.3\%}_{-2.0\%}\right)$
Continuum events	Uniform ($\pm 100\%$)	$^{+0.002}_{-0.001}$ $\left(^{+0.6\%}_{-0.5\%}\right)$
Total (quadratic sum)		$^{+0.007}_{-0.006}$ $\left(^{+2.7\%}_{-2.3\%}\right)$



(a)



(b)



(c)

Figure 9.7: Distributions of $\Delta R(D^*)$ from the evaluation of systematic uncertainty due to the "other" background categories. (a) $B^0 \leftrightarrow B^+$ cross-feed events. (b) "Other" $B\overline{B}$ background events. (c) Continuum events.

Table 9.4: Systematic uncertainties from branching fractions of hadronic B decays.

Decay mode category	Fluctuation function	Uncertainty	
$\overline{B} \to D_s^{(*)} D^*$	Single Gaussian	± 0.000	($\pm 0.2\%$)
$\overline{B} \to D^* n\pi(\pi^0)$	Single Gaussian	± 0.000	($\pm 0.0\%$)
$\overline{B} \to D^* D^{(*)} K$	—	Negligible	
Other listed B decays	—	Negligible	
Unlisted B decays	Uniform ($\pm 100\%$)	± 0.006	($\pm 2.1\%$)
Total		$\pm 0.006\%$	($\pm 2.1\%$)

uncertainty in $R(D^*)$. Branching fractions of the hadronic B decay are fluctuated based on the uncertainties listed in Table 4.4. In each decay mode category, the branching fractions of measured hadronic B decays are simultaneously randomized using single Gaussian functions. Those of unlisted B decays are varied randomly with a uniform distribution from 0% to 200% of the estimated branching fractions. Uncertainties among $B \to D_s^{(*)} D^*$ decays are considered to be fully correlated, reflecting the systematic variation due to cross-feed events in their branching fraction measurement [52]. In contrast, uncertainties among $B \to Dn\pi(\pi^0)$ decays are treated as uncorrelated. Within unlisted B decays with the same number of first B daughters, the branching fractions are correlated for a conservative assessment. No correlation is considered between any other decay modes during this fluctuation. This process of variation and the subsequent fitting of $R(D^*)$ are repeated 1,000 times for all decay mode categories. Systematic uncertainties in the categories of $\overline{B} \to D^* D^{(*)} K$ and the other listed B decays are ignored, as these candidates either represent a small fraction of reconstructed candidates or have their branching fractions measured with sufficiently small uncertainties. The resulting distributions of $R(D^*)$ differences are displayed in Figure 9.8. Table 9.4 summarizes the uncertainties evaluated by single Gaussian fits to the $\Delta R(D^*)$ distributions. In total, an uncertainty of $^{+0.006}_{-0.006}$ $\left(^{+2.1\%}_{-2.1\%}\right)$ is estimated a quadratic sum of all categories.

9.6 Reconstruction efficiencies

Efficiencies of the particle reconstruction possess inherent uncertainties which can lead to variations in the PDF shapes, efficiencies of signal and normalization events, and expected background yields. These uncertainties are largely mitigated by taking the ratio between the signal and normalization modes. Nevertheless, differences in kinematic distributions between these two modes still introduce systematic uncertainties in $R(D^*)$. The following efficiencies and their uncertainties are considered:

- π^+_{slow} efficiency,
- Lepton identification efficiency,
- Hadron identification efficiency,

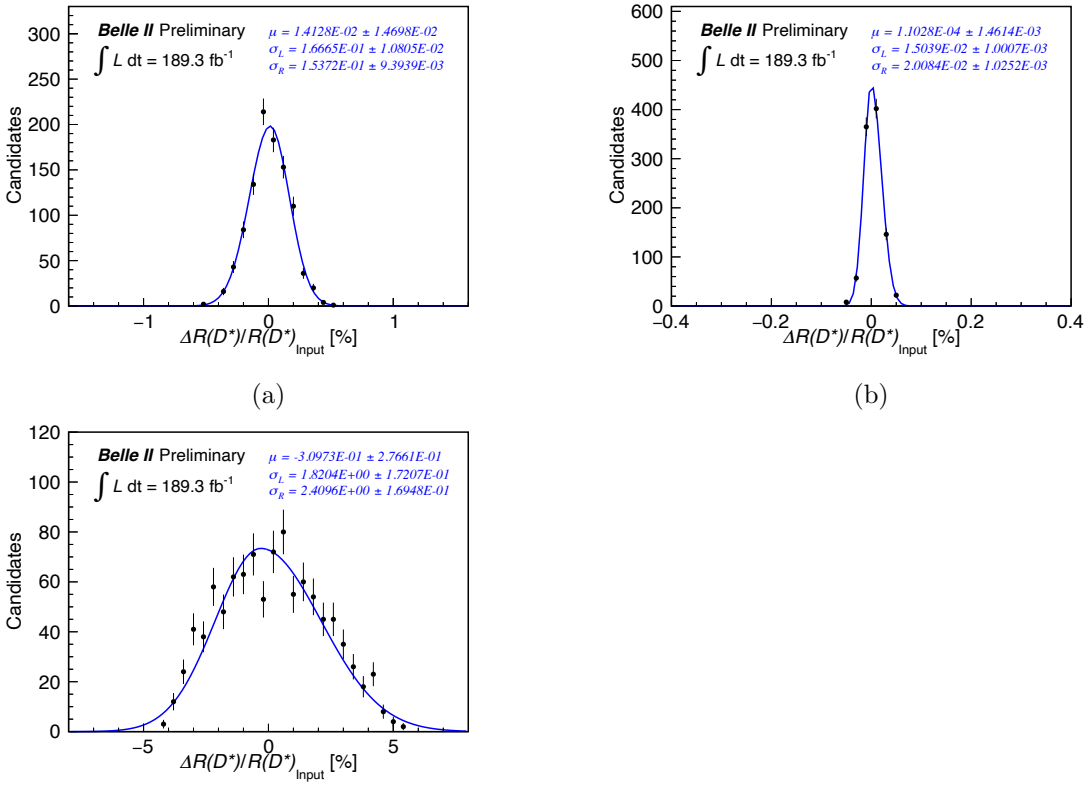


Figure 9.8: Distributions of $\Delta R(D^*)$ in 1000 fits involving the fluctuations within uncertainties on the branching fractions of hadronic B decays. (a) $\overline{B} \to D_s^{(*)} D^*$, (b) $\overline{B} \to D^* n\pi(\pi^0)$, (c) unlisted B decays. The fitted double-sided single Gaussian function is shown by the blue curve.

- π^0 reconstruction efficiency,
- K_S^0 reconstruction efficiency,
- B_{tag} reconstruction efficiency in FEI, and
- Tracking efficiency.

Each of these efficiencies is varied by $\pm 1\sigma$ of its respective uncertainties to evaluate differences in PDF shapes and yields. For the K_S^0 efficiency, an additional uncertainty of 10% is included to account for yield deviation observed between data and simulation in the q^2 side-band region. Fits are performed incorporating these configurations at the $\pm 1\sigma$ shifts. Differences in $R(D^*)$ relative to the fit result at the nominal efficiencies are regarded as uncertainties attributable to these efficiencies. These uncertainties are summarized in Table 9.5. The total uncertainty, calculated as a quadratic sum of all sources, is determined to be ± 0.005 ($\pm 2.0\%$). Notably, this total uncertainty is predominantly influenced by the lepton identification efficiency.

Table 9.5: Systematic uncertainties from reconstruction efficiencies.

Source	Uncertainty	
π^+_{slow} efficiency	± 0.002	($\pm 0.6\%$)
Lepton identification efficiency	± 0.005	($\pm 1.8\%$)
Hadron identification efficiency	± 0.000	($\pm 0.1\%$)
π^0 reconstruction efficiency	± 0.000	($\pm 0.1\%$)
K^0_S reconstruction efficiency	± 0.000	($\pm 0.1\%$)
B_{tag} reconstruction efficiency in FEI	± 0.002	($\pm 0.8\%$)
Tracking efficiency	± 0.000	($\pm 0.1\%$)
Total (quadratic sum)	± 0.005	($\pm 2.0\%$)

9.7 Kernel density estimation

The adaptive KDE smooths the PDF shape along user-specified width scale factors of local densities, ρ_{local}. The PDF shapes could change depending on the chosen values for the width scale factors in the KDE. To evaluate the potential biases from the selections of ρ_{local}, pseudo data samples are prepared without applying the adaptive KDE as the fitted distribution at 189.26 fb^{-1}. The nominal PDF, incorporating the adaptive KDE, is then fitted to these pseudo data samples. This fitting process is conducted 1000 times, with MC entries bootstrapped for each iteration to generate the pseudo data samples. In these samples, the $R(D^*)$ value is assumed to be 0.262, as obtained in the data $R(D^*)$ fit. Figure 9.9 displays the resulting $\Delta R(D^*)$ distribution from this evaluation. A single double-sided Gaussian function fitted to the $\Delta R(D^*)$ distribution yields a mean of $+0.6 \pm 1.4$ % in $\Delta R(D^*)$. This mean bias is regarded as the systematic uncertainty due to the adaptive KDE, resulting in $^{+0.004}_{-0.003}$ $\binom{+2.0\%}{-0.8\%}$.

9.8 Form factors

To update the simulation of the semi-leptonic B decays with the up-to-date form factor modeling and updated value of form factor parameters, event-by-event weights are calculated, which affect the PDF shapes. These modeling uncertainties induce changes in the PDF shapes. To vary form factor parameters according to their uncertainty, all parameters are rotated to an uncorrelated eigenbasis using the correlation matrix within the model. This rotation disentangles the correlation of the form factor parameters, allowing the uncertainties of the rotated parameters to be treated independently and summed up in quadrature. For each rotated parameter, the value is shifted by $\pm 1\sigma$ of its associated uncertainty, and a covariance matrix is constructed. The resulting covariance matrices of all rotated parameters are summed into a single covariance matrix for the form factor process. Corresponding event-by-event weights are fluctuated to modify the PDF shapes to fit with the incorporated uncertainty variation.

As described in Section 7, the reconstructed events are categorized into six templates

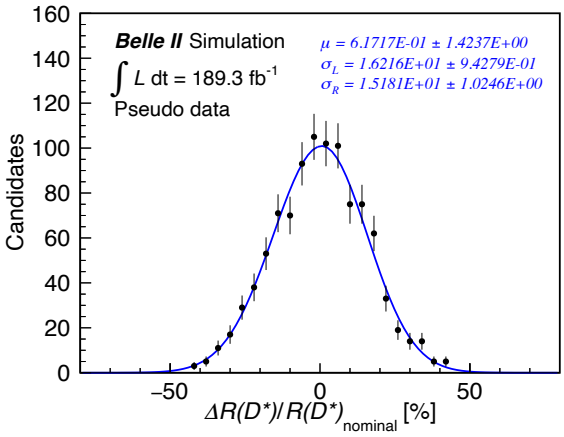


Figure 9.9: Distribution of $R(D^*)$ for the evaluation of systematic uncertainty due to the kernel density estimation using the pseudo data at fitted $R(D^*)$ of 0.262. The fitted double-sided single Gaussian function is shown by the blue curve.

for fitting. Excluding the continuum events template, five PDFs are influenced by the form factor uncertainties of semi-leptonic B decays. Different form factor models for semi-leptonic B decays are evaluated independently. For instance, when parameter values for the decay $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ and $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$, which is modeled by the BGL parameterization, are varied, other decays continue to use the nominal PDFs. Within each template, events are further divided into ten different semi-leptonic B decays identified by generator truth information. Thus, three fit categories of D^* modes, five templates per D^* mode, and ten semi-leptonic B decays per template result in a total of 200 covariance matrices for a single form factor model.

For the construction of the covariance matrix, the difference in entries of two-dimensional PDFs is determined for each bin between PDFs with the $\pm 1\sigma$ shifts of each rotated parameter and the nominal PDFs. Two absolute differences given by positive and negative 1σ shifts of the rotated parameters are averaged in each bin. A positive or negative sign corresponding to a larger absolute difference is assigned in that bin. These binned differences form a two-dimensional matrix of 20×24 bins in $E_{\rm ECL}$ and $M^2_{\rm miss}$ axes, respectively. The difference matrix is then flattened into one dimension array of 480 bins, denoted as $\vec{\delta}_{\rm FF}$. A covariance matrix $\Sigma_{\rm FF}$ is obtained by taking the outer product of $\vec{\delta}_{\rm FF}$, i.e., $\vec{\delta}_{\rm FF} \times \vec{\delta}_{\rm FF}$. Figure 9.10 shows covariance matrices for the template of the normalization events. Figures (**??**) to (**??**) correspond to each decay process categorized by the generator information.

As described in Section 5, while $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ and $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$ processes use the BLPRXP modeling, $\overline{B} \to D^{**} \ell^- \overline{\nu}_\ell$ and $\overline{B} \to D^{**} \tau^- \overline{\nu}_\tau$ processes utilize the BGL modeling. The systematic uncertainty is evaluated individually for each form factor model. For each model, a matrix of random percentage factors is generated following a multi-variate Gaussian distribution based on the covariance matrix. This matrix is applied to the nominal PDFs to vary the shapes with correlation, and the new PDF is then fitted to the data. This procedure is repeated for 1,000 fits, producing a distribution of $R(D^*)$ differences relative

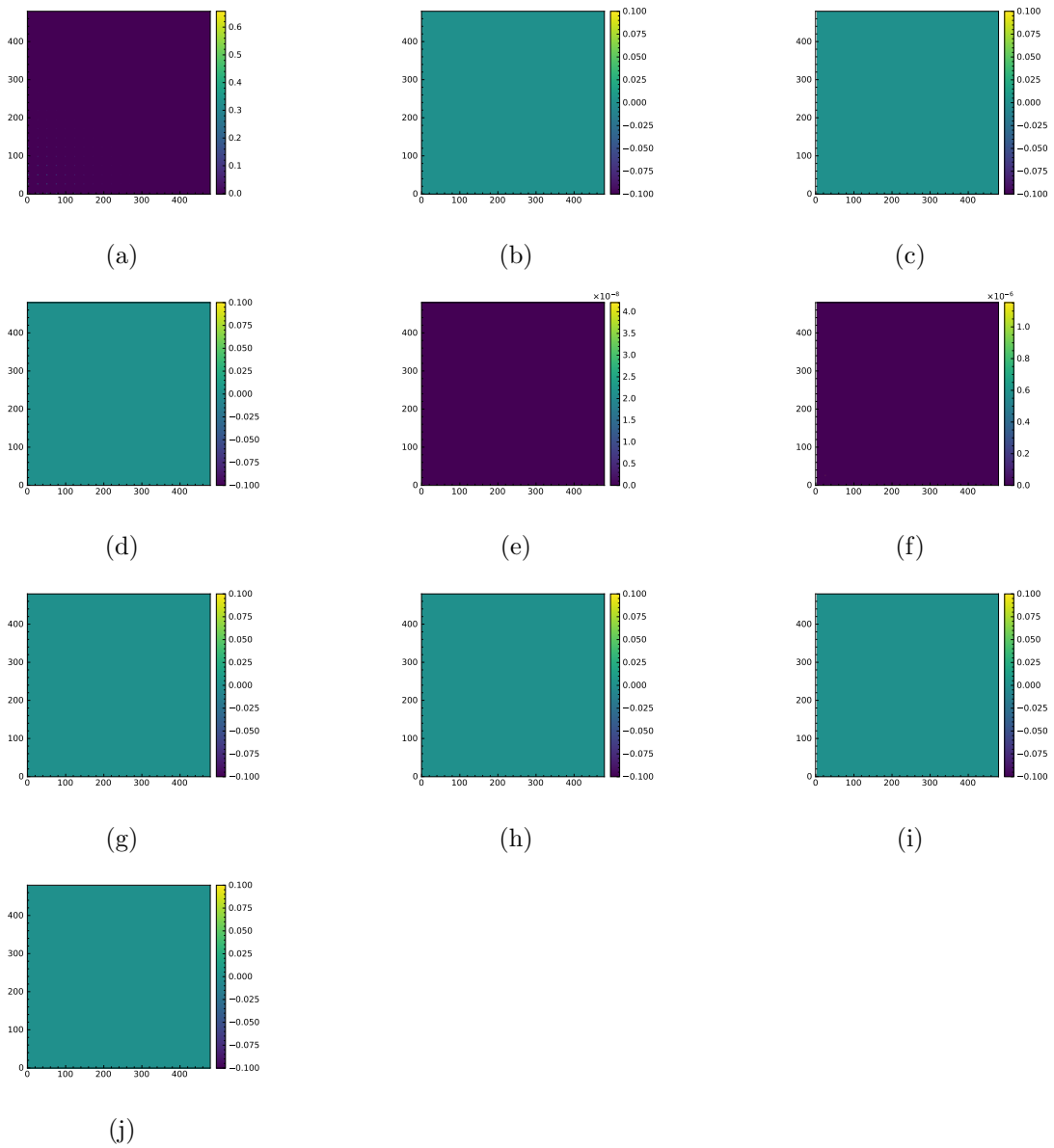


Figure 9.10: Covariance matrices for template of the normalization events in the $D^{*+} \to D^0\pi^+$ decay. The covariance matrices are created for each of the categorized semi-leptonic B decays identified by generator information. The categories of semi-leptonic B decays are (a) $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$, (b) $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$, (c) $\overline{B} \to D_1\ell^-\overline{\nu}_\ell$, (d) $\overline{B} \to D_2^*\ell^-\overline{\nu}_\ell$, (e) $\overline{B} \to D_0^*\ell^-\overline{\nu}_\ell$, (f) $\overline{B} \to D_1'\ell^-\overline{\nu}_\ell$, (g) $\overline{B} \to D_1\tau^-\overline{\nu}_\tau$, (h) $\overline{B} \to D_2^*\tau^-\overline{\nu}_\tau$, (i) $\overline{B} \to D_0^*\tau^-\overline{\nu}_\tau$, and (j) $\overline{B} \to D1'\tau^-\overline{\nu}_\tau$.

to the nominal $R(D^*)$. A single double-sided Gaussian is fitted to the $\Delta R(D^*)$ distribution and its standard deviations are considered as the systematic uncertainty induced by the form factor parameters. The uncertainties from four processes are added in quadrature as a total systematic uncertainty.

Table 9.6: Expected systematic uncertainties from form factor parameterization.

Decay	Uncertainty	
$\overline{B} \to D^* \ell^- \overline{\nu}_\ell$	$^{+0.001}_{-0.000}$	$\binom{+0.3\%}{-0.1\%}$
$\overline{B} \to D^* \tau^- \overline{\nu}_\ell$	$^{+0.001}_{-0.000}$	$\binom{+0.2\%}{-0.0\%}$
$\overline{B} \to \{D_1, D_2^*\} \ell^- \overline{\nu}_\ell / \{D_1, D_2^*\} \tau^- \overline{\nu}_\tau$	$^{+0.001}_{-0.000}$	$\binom{+0.3\%}{-0.0\%}$
$\overline{B} \to \{D_0^*, D_1'\} \ell^- \overline{\nu}_\ell / \{D_0^*, D_1'\} \tau^- \overline{\nu}_\tau$	$^{+0.001}_{-0.000}$	$\binom{+0.2\%}{-0.0\%}$
Total (quadratic sum)	$^{+0.001}_{-0.000}$	$\binom{+0.5\%}{-0.1\%}$

The systematic uncertainties of form factor variations of $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$, $\overline{B} \to D^* \tau^- \overline{\nu}_\tau$, $\overline{B} \to \{D_1, D_2^*\} \ell^- \overline{\nu}_\ell / \{D_1, D_2^*\} \tau^- \overline{\nu}_\tau$, and $\overline{B} \to \{D_0, D_1'\} \ell^- \overline{\nu}_\ell / \{D_0, D_1'\} \tau^- \overline{\nu}_\tau$ are listed in Table 9.6. The total systematic uncertainty is estimated to be $^{+0.001}_{-0.000}$ $\binom{+0.5\%}{-0.1\%}$.

9.9 Peaking backgrounds on ΔM_{D^*}

The number of peaking background events in fake D^* candidates is determined by fitting the ΔM_{D^*} distributions relative to the correctly reconstructed D^* candidates in the simulation. However, the ratio of peaking backgrounds to correct D^* candidates may differ in the data. In the simulation, over 90% of the peaking background events are found to originate from events with incorrectly reconstructed $\overline{B} \to D^* \ell^- \overline{\nu}_\ell$ decays. The mis-reconstruction primarily occurs in the following scenarios:

- The software fails to match π^+_{slow} with the generated π^+ particle in the simulation, even though the π^+_{slow} candidate is correctly reconstructed according to an event display.
- A π^0 candidate in a D decay is not correctly reconstructed, involving either a photon candidate of the beam-induced background or a fake γ produced by hadronic showers.

The impact of software failure in matching on $R(D^*)$ is expected to be minor, as it is mitigated by taking a ratio between the signal and normalization modes. However, the beam-induced background photons and fake γs may not be accurately modeled by the simulation, potentially leading to systematic uncertainty in $R(D^*)$. To estimate this impact on $R(D^*)$, the number of events with incorrectly reconstructed π^0 from a D meson is varied from 0% to 200% of the expected yields in the simulation. The fit of $R(D^*)$ is then conducted with PDFs that include this variation. An observed shift of 0.4% in the fitted $R(D^*)$ is estimated as the systematic uncertainty.

9.10 $\tau \to \ell^- \nu_\tau \overline{\nu}_\ell$ branching fractions

The branching fractions of leptonic τ decays have uncertainties, as listed in Table 4.5. These uncertainties can cause variations in the signal efficiency $\varepsilon_{D^* \tau \nu}$, resulting in changes

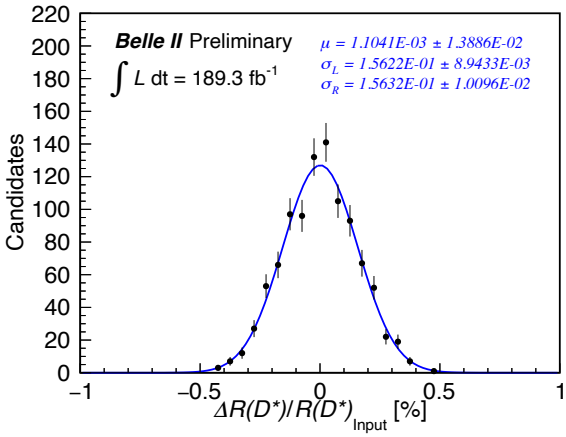


Figure 9.11: Distribution of $\Delta R(D^*)$ induced by the variation of the branching fractions of the leptonic τ decays. The fitted double-sided single Gaussian function is shown by the blue curve.

to $R(D^*)$. The uncertainties of the branching fractions for the $\tau^- \to e^- \nu_\tau \overline{\nu}_e$ and $\tau^- \to \mu^- \nu_\tau \overline{\nu}_\mu$ decays are varied consistently based on a single Gaussian function with a sigma corresponding to these uncertainties. This variation is conducted 1,000 times, and the resulting $\Delta R(D^*)$ shifts are evaluated with a fit of a single double-sided Gaussian function, as illustrated in Figure 9.11. The 1σ region of the $\Delta R(D^*)$ distribution is considered as the systematic uncertainty. The uncertainty is estimated to be $\pm 0.2\%$.

9.11 $R(D^*)$ fit method

A fitter bias is observed during the linearity test of the $R(D^{*)})$ fit method as described in Section 7.2.2. This fit bias is evaluated at $R(D^*) = 0.262$ using the linearity function of Eq. 7.15. The obtained bias of $^{+0.0\%}_{-0.1\%}$ is assigned as the systematic uncertainty induced by our fit method. Furthermore, there is a discrepancy observed between data and simulation in range of $1.8 < E_{\mathrm{ECL}} < 2.0$ GeV. When this range is excluded, the p-value for the goodness of fit for the $R(D^*)$ extraction increases from 4.4% to 14.4%, as shown in Figure 9.12. Reducing the fit range results in a $+0.1\%$ shift of the fitted $R(D^*)$. The systematic uncertainty is determined by a quadratic sum of these two contributions, yielding $\pm 0.1\%$.

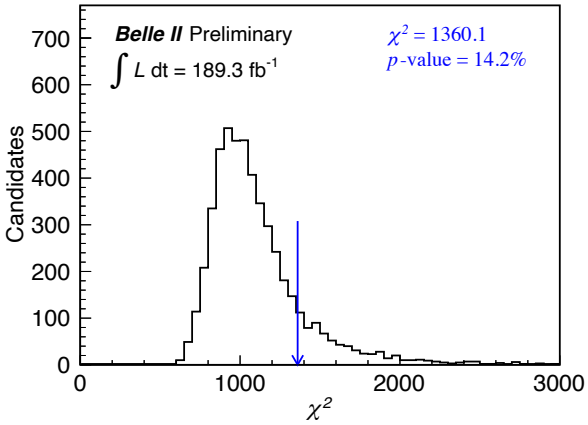


Figure 9.12: Distribution of χ^2 values from the signal extraction fits with pseudo data sets at $R(D^*) = 0.262$ where the E_{ECL} range is shortened up to 1.8 GeV. The blue arrow indicates the χ^2 value of 1360.1 obtained by the fit to the data within the reduced E_{ECL} range.

10 Discussion

10.1 Combination of result of $R(D^*)$ measurement

We combine our result of the $R(D^*)$ measurement with the previous measurements. Figures 10.1 and 10.2 illustrate summary of measurement results on $R(D^*)$ and combined $R(D^{(*)})$ averages. The Belle II $R(D^*)$ result agrees with the SM expectations at 0.254 ± 0.005. Thus, no significant LFU violation is found in $R(D^*)$. Our result is consistent with all the previous measurements within the uncertainty. The combined averages have an excess from the standard deviation by 3.3σ.

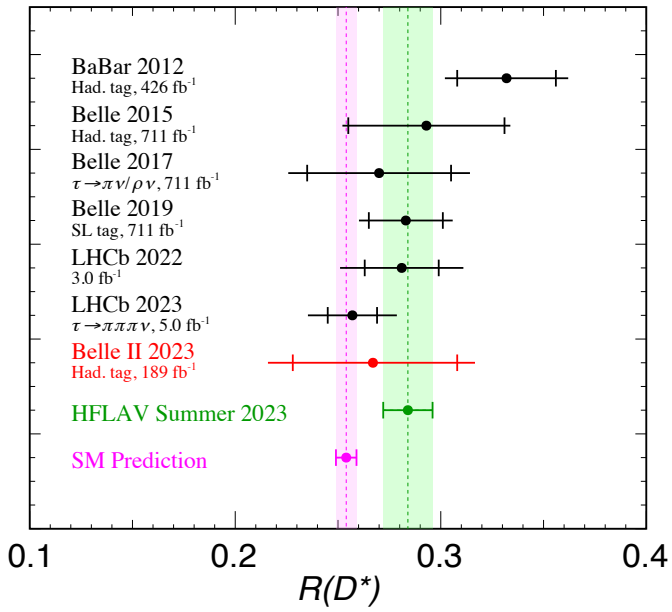


Figure 10.1: Comparison of the $R(D^*)$ measurements results. The red point represents our result, while the black points denote the results from previous measurements [19, 21, 23–26, 53]. The range between the small vertical lines signifies statistical uncertainty, and the extent of the horizontal line corresponds to total uncertainty. The magenta and green bands indicate the predicted $R(D^*)$ in the SM and the experimental world average, including our measurement result, respectively [27].

10.2 Constraints on New Physics parameters

We examine the allowed regions with the current world averages in Figure 10.2. Three scenarios are tested: Single-operator scenarios, a type-II two-Higgs doublet model scenario, and Leptoqruark scenarios. The significance is calculated using log-likelihood ratios with the total uncertainty of $R(D)$ and $R(D^*)$ with their correlation taken into account. The BSM contributions are estimated based on Eqs. (**??**). We assume that the efficiencies do not differ from the SM over the parameter space for simplicity.

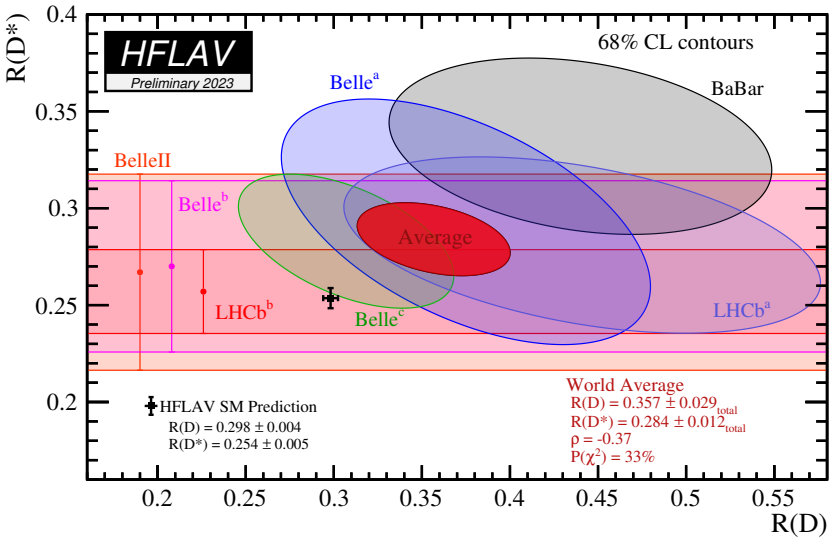


Figure 10.2: Summary of results of $R(D^{(*)})$ measurements performed by the BaBar, Belle, LHCb, and Belle II experiments [27]. The contour shows 68% confidence levels of measurements [19, 21, 23–26, 53]. The black point indicates the SM predictions.

10.2.1 Single-operator scenarios

As the simplest case, we test single-operator scenarios for five interaction types: C_{V_L}, C_{V_R}, C_{S_L}, C_{S_R}, C_T. Only one of the coefficients takes nonzero values while the others are zero. The allowed regions for each of them are shown in Figure 10.3.

10.2.2 Two-Higgs doublet model scenarios

The charged Higgs of two-Higgs doublet models can contribute to the semileptonic B decays as shown in Figure 10.4. The coefficients of scalar operators in type-II Higgs doublet model (2HDM) are expressed as [2]

$$C_{S_L} = -m_b m_\tau \frac{\tan^2\beta}{m_{H^+}^2}, \tag{10.1}$$

$$C_{S_R} = -m_b m_c \frac{1}{m_{H^+}^2}, \tag{10.2}$$

where m_b, m_c, and m_τ are masses of b, c quarks, and τ leptons, respectively. We adopt $m_b = 4.18$ GeV/c^2, $m_c = 0.901$ GeV/c^2, and $m_\tau = 1.76708$ GeV/c^2 at the scale $\mu_b = m_b$ [54]. As shown in Figure 10.5, we observed limited allowed regions for the type-II 2HDM and most of the regions are excluded by more than 3σ on the $\tan\beta$ and m_{H^+} plane.

10.2.3 Leptoqruark scenarios

The leptoquarks can contribute to the semileptonic B decays as shown in Figure 10.6. We consider three leptoquark types as possible candidates: a $SU(2)_L$-singlet vector U_1^μ, a

$SU(2)_L$-singlet scalar S_1, and $SU(2)_L$-doublet scalar R_2. The leptoquark mass is assumed to be 2 TeV. These leptoquarks can have nonzero coefficients as follows [29]:

$$U_1^\mu : C_{V_L}, C_{S_R}, \tag{10.3}$$
$$S_1 : C_{V_L}, C_{S_L} = -4C_T, \tag{10.4}$$
$$R_2 : C_{V_R}, C_{S_L} = +4C_T, \tag{10.5}$$
$$\tag{10.6}$$

at the LQ scale $\Lambda_{\rm LQ} = M_{\rm LQ}$. To estimate the contribution at the scale $\mu_b = m_b$, the coefficients are scaled down from the LQ scale to the scale $\mu_b = m_b$, and their relations are obtained as $C_{S_L}(\mu_b) = -8.7C_T(\mu_b)$ and $C_{S_L}(\mu_b) = +8.2C_T(\mu_b)$ for S_1-type and R_2-type leptoquarks, respectively [29]. Figure 10.7 shows constraints on the BSM parameters in the leptoquark scenarios. For each value of one parameter, the minimum significance that can be taken at any value of the other parameter is shown among each set of parameters. When we assume zero imaginary parts of the Willson coefficients, the correlation of allowed values of the parameters are shown in Figure 10.7.

10.3 Estimation of future sensitivity

We estimate the uncertainties of $R(D^*)$ in future analyses of the Belle II experiment based on two scenarios:

1. Conservative scenarios where the systematic uncertainties remain constant,

2. Optimistic scenarios where major systematic uncertainties decrease in relation to luminosity.

Initially, we estimate the expected statistical uncertainties using Asimov data sets. The $R(D^*)$ fit is performed with Asimov data sets at different luminosities. The size of the Asimov data set increases from 0.1 fb^{-1} to 5.0 fb^{-1} in 0.1 fb^{-1} steps. The obtained dependence of the statistical uncertainties on the luminosity $\mathcal{L}$ is fitted using a function $p_0 + p_1/\sqrt{\mathcal{L}}$. This fitted function serves to estimate statistical uncertainty at a given luminosity.

In the first scenario, the same level of systematic uncertainty is maintained without improvements in uncertainty sources, while statistical uncertainty decreases as luminosity increases. In the second scenario, reductions in discrepancies of the $E_{\rm ECL}$ and $M^2_{\rm miss}$ distributions between data and simulation are assumed, along with decreased systematic uncertainty of the PDF shape correction as luminosity increases. The uncertainty range of the energy shifts for fake photons narrows by the χ^2 tests using larger data sets. Additionally, the size of the simulation increases proportionally to the real data size. The uncertainty of the MC statistics at higher luminosity is estimated using fitted functions of bias and standard deviations in $\Delta R(D^*)$ from Figure 9.5.

Figure 10.9 displays extrapolated statistical and total uncertainties. Compared with Belle sensitivity, statistical uncertainty reduces by about 40% at 189 fb^{-1}. The grey line represents total uncertainties with constant systematic uncertainty size up to 5 ab^{-1}. The

statistical uncertainty is given by the black line. Improved statistical uncertainty enhances sensitivity on $R(D^*)$, even without systematic uncertainty improvements. The grey dashed line indicates total uncertainties with reduced systematic uncertainty in PDF shapes and MC statistics.

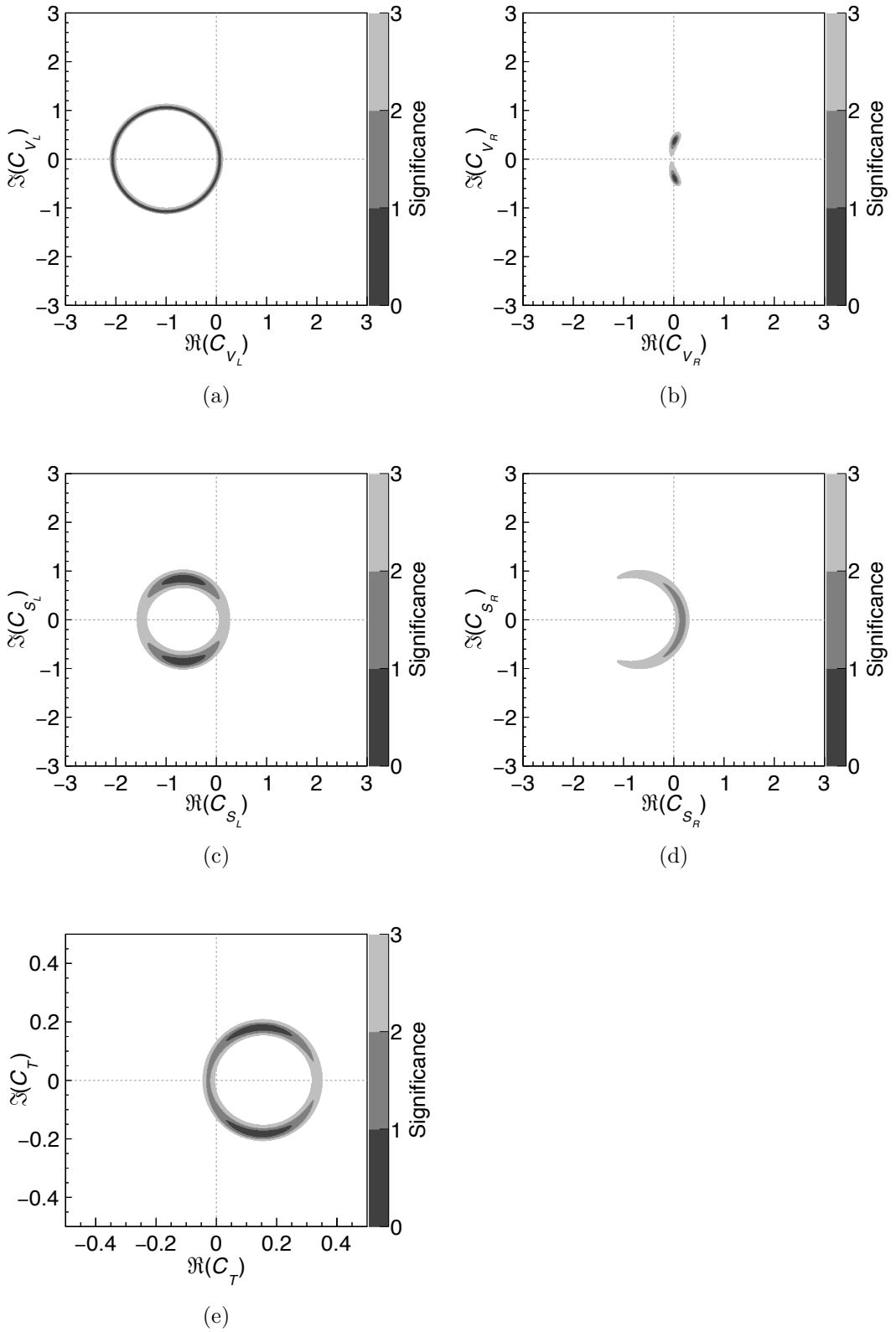


Figure 10.3: Favored regions according to the $R(D^{(*)})$ averages on the Wilson Coefficients for single BSM operators. The constraints are presented for (a) V_L-type, (b) V_R-type, (c) S_L-type, (d) S_R-type, and (e) T-type single operators, respectively. The vertical and horizontal axes represent the real and imaginary parts of the coefficients.

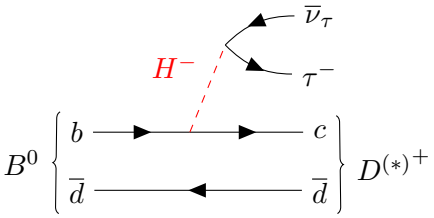


Figure 10.4: Feynman diagram of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays with a charged Higgs boson.

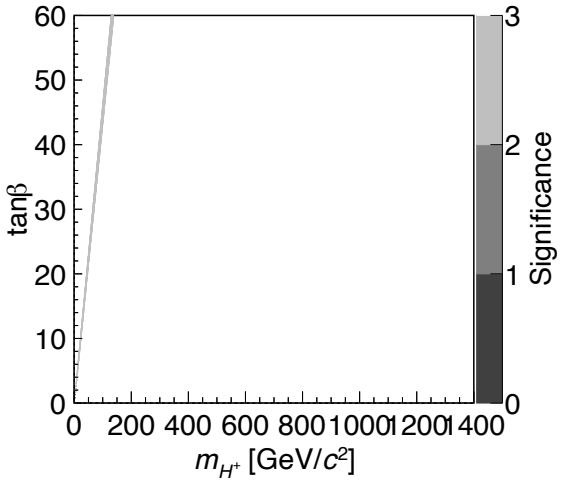


Figure 10.5: Favored regions according to the $R(D^{(*)})$ averages on masses and $\tan\beta$ of a charged Higgs boson in the two-Higgs type models (Type-II).

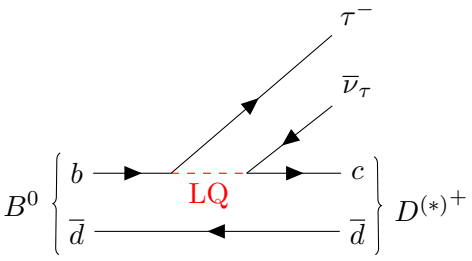


Figure 10.6: Feynman diagram of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays with a leptoquark.

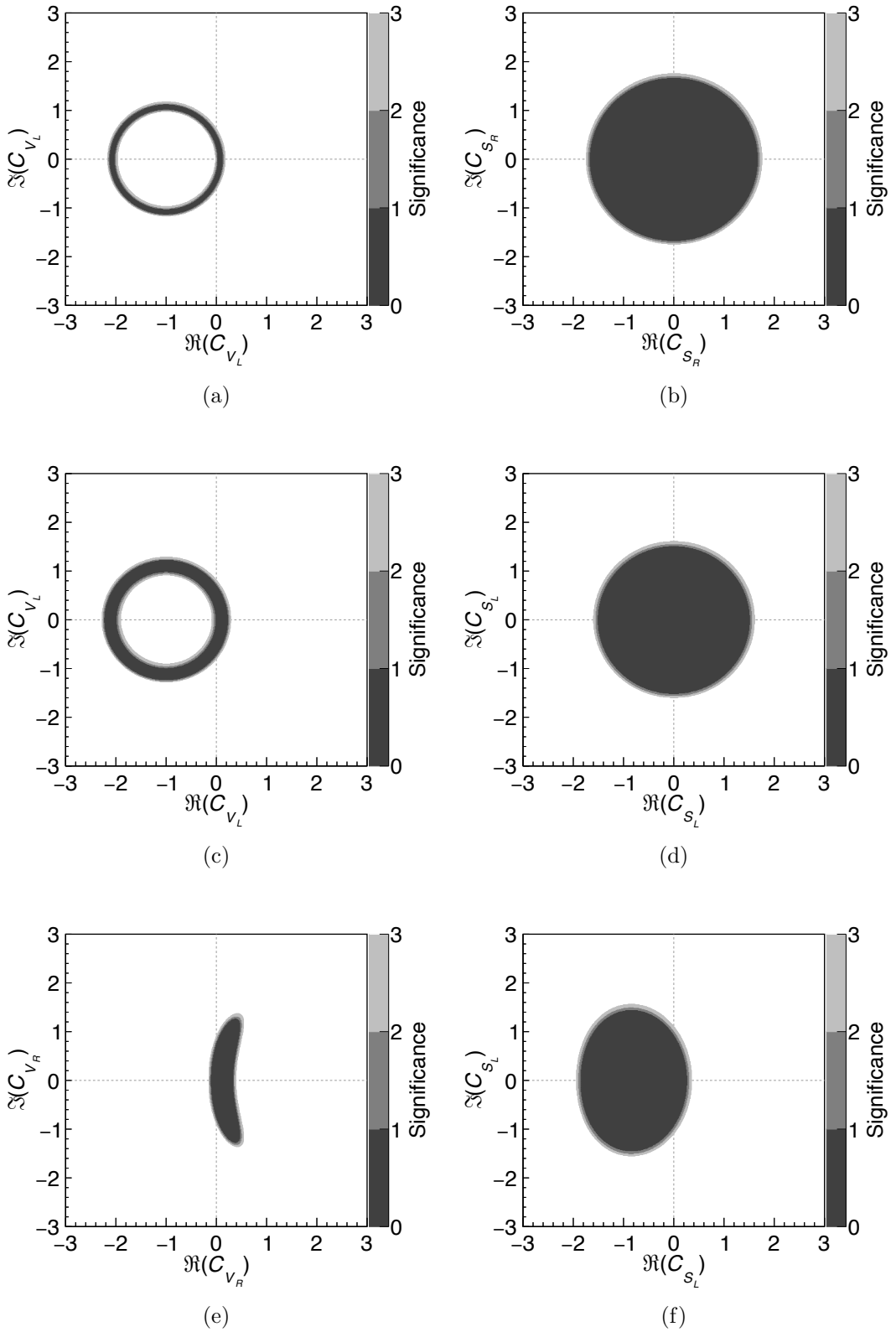


Figure 10.7: Favored regions according to the $R(D^{(*)})$ averages on the Wilson Coefficients for the leptoquark models. The constraints are presented for (a)–(b) U_1-type, (c)–(d) S_1-type, and (e)–(f) R_2-type leptoquark models, respectively. The vertical and horizontal axes represent the real and imaginary parts of the coefficients.

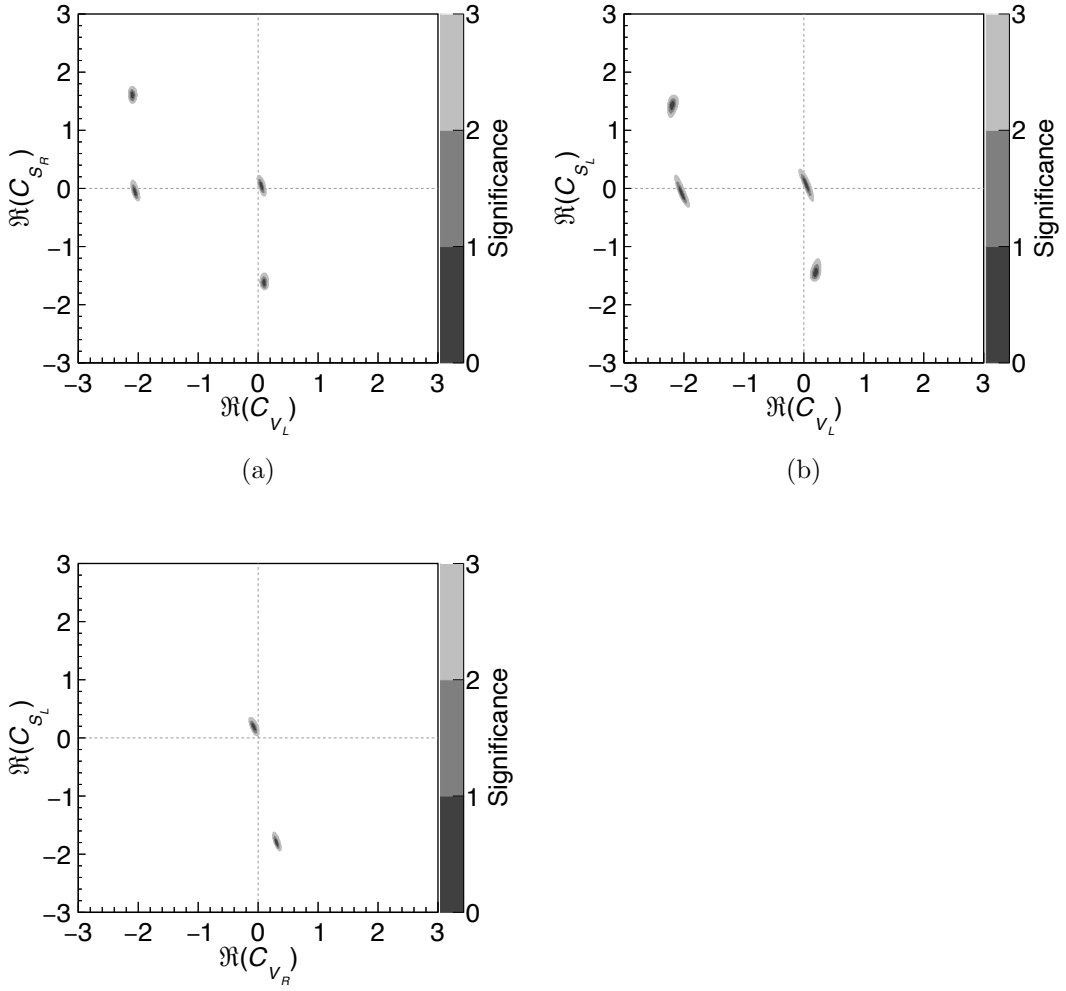


Figure 10.8: Favored regions according to the $R(D^{(*)})$ averages on the real Wilson Coefficients for the leptoquark models. The imaginary parts of the Wilson Coefficients are assumed to be zero. The constraints are presented for (a) U_1-type, (b) S_1-type, and (c) R_2-type leptoquark models, respectively. The vertical and horizontal axes represent the real parts of the two coefficients.

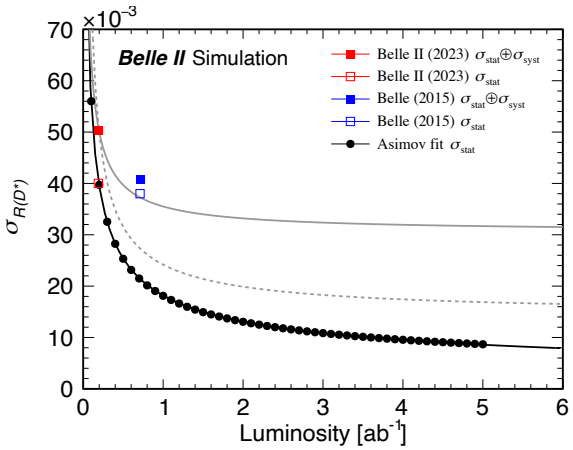


Figure 10.9: Dependence of uncertainties in the $R(D^*)$ measurement on the luminosity of the data set. Black points are estimated using Asimov data sets. The open (closed) red and blue rectangle points indicate the statistical (total) uncertainty in the $R(D^*)$ in our results at 189 fb^{-1} and previous measurement by the Belle experiment at 711 fb^{-1} [21].

11 Conclusion

The SM marks a lot of success in particle physics. The model holds lepton flavor universality as an axiom that states that the coupling constants in the interaction between leptons and gauge particles are the same regardless of the lepton flavor e, μ, or τ. This universality has been demonstrated by various experiments as well. On the other hand, in the semi-leptonic B decay, the ratio of their branching fractions $R(D^*)$ show an excess over the theoretical values at a significance of 3.2σ according to the average of results of measurements by the previous BaBar, Belle, and LHCb experiments. It has been suggested that this deviation could be a sign of BSM contributions.

To further investigate the observed deviation, we perform the first measurement of $R(D^*)$ and search for lepton flavor universality violation between $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ and $\overline{B} \to D^*\ell^-\overline{\nu}_\ell$ decays using a new electron-positron data samples, corresponding to 189 fb^{-1}, collected with the Belle II detector between 2019 and 2021 at the SuperKEKB accelerator. We introduce a new reconstruction technique, FEI, for B tagging and optimize selections to leverage the new reconstruction method. We establish calibration and calibration methods for major background components through these data-driven approaches. We demonstrate that the background sources are well-modeled in our simulation.

From the fit to the data, we determine

$$R(D^*) = 0.262\ ^{+0.041}_{-0.039}\ (\text{stat.})\ ^{+0.033}_{-0.032}\ (\text{syst.}). \tag{11.1}$$

This result is consistent with the SM predictions. Therefore, no significant violation of lepton flavor universality is observed. The results also agree with the previous measurements within the uncertainty. The world average of the $R(D^{(*)})$ measurements, including this measurement, showed a slight increase in the deviation from the SM predictions from 3.2σ to 3.3σ. Finally, we discuss possible BSM contributions in $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ based on the new world average of $R(D^{(*)})$ for single operators, two-Higgs doublet models, and leptoquark scenarios. We achieve a $R(D^*)$ measurement with improved sensitivity by 40% compared to the corresponding Belle measurement using the early data and establish a baseline for future analyses in the Belle II experiment.

Acknowledgments

As I reflect on the journey of my doctoral studies, I am filled with immense gratitude for the myriad of individuals whose support and guidance have been foundational to my success.

Foremost, I am immensely grateful to my supervisor, Prof. Toru Iijima, for his unwavering guidance, encouragement, and insight. His expertise and wisdom have been pivotal in both my academic and personal growth.

I express my sincere thanks to my collaborators in the $R(D^*)$ analysis, Dr. Qi-Dong Zhou, Dr. Taichiro Koga, and Dr. Racha Cheaib. Their collaboration, expertise, and shared passion for our research have significantly contributed to overcoming numerous challenges and achieving our analytical goals.

My heartfelt thanks are also directed towards my collaborators in our Japanese analysis group: Dr. Kodai Matsuoka, Dr. Koji Hara, Dr. Katsuro Nakamura, and Dr. Kenta Uno. They provided critical insights during complex phases of our analysis, which were instrumental to our success.

I am deeply grateful to the analysis reviewers: Prof. Robert Kowalewski, Prof. Pablo Goldenzweig, Prof. Anselm Vossen, Dr. Frank Meier, and Prof. Phillip Urquijo. Their constructive feedback and insightful critiques have substantially enhanced the quality of our work.

A special acknowledgment goes to the conveners of our analysis working group: Prof. Christoph Schwanda, Prof. Peter Mandeville Lewis, Prof. William Sutcliffe, Prof. Diego Tonelli, and Prof. Jim Libby. Their kind leadership and vision have been crucial in guiding our group towards meaningful and impactful research.

I extend warm thanks to my collaborators in the Bonn group: Prof. Florian Bernlochner, Mr. Henrik Junkerkalefeld, Ms. Alina Charlotte Manthei, and Mr. Ilias Tsaklidis. Their collaboration and shared expertise have been a source of inspiration and learning. Our collaborative work and discussion at Bonn University are one of my most cherished memories.

Additionally, I am sincerely thankful to the staff and students in our laboratory. Interacting with laboratory members has cultivated my motivation for the challenging analysis. In particular, I have greatly profited significantly from the members of the Belle II group: Dr. Kenji Inami, Mr. Yuki Sue, Mr. Ryogo Okubo, Ms. Akane Maeda, Mr. Michele Aversano, and Mr. Ryotaro Komori. Their support, analysis meetings, and dedication have been nurtured my research capabilities. Furthermore, I appreciate my esteemed colleagues, Ms. Haruka Asada and Dr. Moe Wakida, with whom I embarked on this doctoral journey. They were always willing to devote themselves to tough research with much effort, which kept me inspired even everywhere.

I cannot forget to express my gratitude to my colleagues, Dr. Yuma Uematsu, and Dr. Hanwook Bae. We shared the trials and challenges of research in the same period at the Belle II experiment. Let's again plan our next outing to the best ramen restaurants in Tsukuba as we did then!

This journey would have been incomplete without the collective effort, support, and encouragement of each one of you. My experiences at this phase of my academic career have been profoundly enriched and enlivened by your contributions. I am truly grateful for all of them. Thank you very much.

This work was supported by Japan Society for the Promotion of Science, Grant-in-Aid for JSPS Fellows Grant No. JP21J15570, and Grant-in-Aid for Scientific Research Grants No. 18H05226, No. 22K21347, and No. 23H05433.

We thank the SuperKEKB team for delivering high-luminosity collisions; the KEK cryogenics group for the efficient operation of the detector solenoid magnet; the KEK computer group and the NII for on-site computing support and SINET6 network support; and the raw-data centers at BNL, DESY, GridKa, IN2P3, INFN, and the University of Victoria for off-site computing support.

Appendix A Branching fractions of $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$

The heavier 1P states in the charm meson system, collectively known as D^{**}, significantly contribute to the $D^{(*)}(n\pi)$ states with $n > 0$. According to HQET, there are two narrow states, $D_1(2420)$ and $D_2^*(2460)$, which decay through D-wave transitions to $D^{(*)}$ with a width of approximately 20 MeV. In contrast, the broader states, $D_0(2400)$ and $D_1'(2430)$, decay through S-wave transitions to $D^{(*)}$ with widths of several hundred MeV. Measuring these heavier D^{**} mesons is complex due to the overlapping decay products, and typically only partial branching fractions have been measured. Therefore, total branching fractions for these modes are inferred from partial branching fractions, assuming isospin symmetry in the meson system. The composition of branching fractions used in our simulation is summarized in Table A.1.

For the non-resonant $B \to D^{(*)}\pi\ell^-\overline{\nu}_\ell$ decays, their contribution is consistent with zero after subtracting the resonant D^{**} contributions decaying to $D^{(*)}\pi$ from the inclusive branching fractions of $\overline{B} \to D^{(*)}\pi\ell^-\overline{\nu}_\ell$:

$$\mathcal{B}(\overline{B} \to D^*\pi\ell^-\overline{\nu}_\ell) - \mathcal{B}(\overline{B} \to D^{**}(\to D^*\pi)\ell^-\overline{\nu}_\ell) = (0.3 \pm 0.9) \times 10^{-3}, \tag{A.1}$$

$$\mathcal{B}(\overline{B} \to D\pi\ell^-\overline{\nu}_\ell) - \mathcal{B}(\overline{B} \to D^{**}(\to D\pi)\ell^-\overline{\nu}_\ell) = (-1.1 \pm 1.1) \times 10^{-3}. \tag{A.2}$$

Thus, the branching fractions for these non-resonant decays are set to zero in our simulation. These contributions are later considered in the evaluation of systematic uncertainties using the branching fractions from Eq. (A.1) and Eq. (A.2).

The total branching fractions of the inclusive $\overline{B} \to D^{(*)}\pi\pi\ell^-\overline{\nu}_\ell$ decays are derived using the ratios of branching fractions of $\overline{B} \to D^{(*)}\pi\pi\ell^-\overline{\nu}_\ell$ over $\overline{B} \to D^{(*)}\ell^-\overline{\nu}_\ell$ [55]. However, the $\overline{B} \to D\pi\pi\ell^-\overline{\nu}_\ell$ is saturated by the contributions from an observed mode of $\overline{B} \to D_1(\to D\pi\pi)\ell^-\overline{\nu}_\ell$,

$$\mathcal{B}(\overline{B} \to D\pi\pi\ell^-\overline{\nu}_\ell) - \mathcal{B}(\overline{B} \to D_1(\to D\pi\pi)\ell^-\overline{\nu}_\ell) = (0.6 \pm 0.9) \times 10^{-3}. \tag{A.3}$$

The residual contribution to $\overline{B} \to D\pi\pi\ell^-\overline{\nu}_\ell$ in Eq. (A.3) is attributed to decays through the broad through the broad D_0^* and D_1' resonances. It is assumed that the decays $\overline{B} \to D_0^*(\to D\pi\pi)\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D_1'(\to D\pi\pi)\ell^-\overline{\nu}_\ell$ contribute equally to these branching fractions. The expected branching fractions for $\overline{B} \to D^*\pi\pi\ell^-\overline{\nu}_\ell$ decays are

$$\mathcal{B}(\overline{B}^0 \to D\pi\pi\ell^-\overline{\nu}_\ell) = (2.0 \pm 1.1) \times 10^{-3}, \tag{A.4}$$

$$\mathcal{B}(B^- \to D\pi\pi\ell^-\overline{\nu}_\ell) = (2.2 \pm 1.1) \times 10^{-3}, \tag{A.5}$$

for neutral and charged B modes, respectively. These decays are also simulated in our simulation with equal contributions from the two resonant D^{**} decays via D_0^* or D_1'.

There exists a gap of branching fractions between the inclusive $\overline{B} \to X_c\ell^-\overline{\nu}_\ell$ and the sum of exclusive semi-leptonic B decays involving a charm meson. To fill this gap, we assign the decays $\overline{B} \to D_0^*(\to D\eta)\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D_1'(\to D^*\eta)\ell^-\overline{\nu}_\ell$ as contributing components, assigning a 100% uncertainty to their branching fractions.

Since $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ decays have not been observed yet, we estimate their branching fractions using those of $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$ and assuming $R(D^{**}) \equiv \mathcal{B}(\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{B} \to$

Table A.1: Compositions of branching fractions in each category of decay processes of the resonant D^{**} states in the simulation.

Decay process	Decay		$\mathcal{B}(D^{**})$ composition[%]
	D^{**0}	D^{**+}	
$D_1 \to D\pi\pi/D^*\pi$	$D_1^0 \to D^0\pi^+\pi^-$	$D_1^+ \to D^+\pi^+\pi^-$	17.2
	$D_1^0 \to D^0\pi^0\pi^0$	$D_1^+ \to D^+\pi^0\pi^0$	11.5
	$D_1^0 \to D^+\pi^-\pi^0$	$D_1^+ \to D^0\pi^+\pi^0$	11.5
	$D_1^0 \to D^{*+}\pi^-$	$D_1^+ \to D^{*0}\pi^+$	39.9
	$D_1^0 \to D^{*0}\pi^0$	$D_1^+ \to D^{*+}\pi^0$	20.0
$D_0^* \to D\pi$	$D_0^{*0} \to D^+\pi^-$	$D_0^{*+} \to D^0\pi^+$	66.7
	$D_0^{*0} \to D^0\pi^0$	$D_0^{*+} \to D^+\pi^0$	33.3
$D_0^* \to D\pi\pi$	$D_0^{*0} \to D^0\pi^+\pi^-$	$D_0^{*+} \to D^+\pi^+\pi^-$	42.9
	$D_0^{*0} \to D^+\pi^-\pi^0$	$D_0^{*+} \to D^0\pi^+\pi^0$	28.6
	$D_0^{*0} \to D^0\pi^0\pi^0$	$D_0^{*+} \to D^+\pi^0\pi^0$	28.6
$D_0^* \to D^*\pi\pi$	$D_0^{*0} \to D^{*0}\pi^+\pi^-$	$D_0^{*+} \to D^{*+}\pi^+\pi^-$	42.9
	$D_0^{*0} \to D^{*+}\pi^-\pi^0$	$D_0^{*+} \to D^{*0}\pi^+\pi^0$	28.6
	$D_0^{*0} \to D^{*0}\pi^0\pi^0$	$D_0^{*+} \to D^{*+}\pi^0\pi^0$	28.6
$D_0^* \to D\eta$	$D_0^{*0} \to D^0\eta^0$	$D_0^{*+} \to D^+\eta^0$	100.0
$D_1' \to D^*\pi$	${D_1'}^0 \to D^{*+}\pi^-$	${D_1'}^+ \to D^{*0}\pi^+$	66.7
	${D_1'}^0 \to D^{*0}\pi^0$	${D_1'}^+ \to D^{*+}\pi^0$	33.3
$D_1' \to D\pi\pi$	${D_1'}^0 \to D^+\pi^+\pi^-$	${D_1'}^+ \to D^+\pi^+\pi^-$	42.9
	${D_1'}^0 \to D^0\pi^+\pi^0$	${D_1'}^+ \to D^0\pi^+\pi^0$	28.6
	${D_1'}^0 \to D^+\pi^0\pi^0$	${D_1'}^+ \to D^+\pi^0\pi^0$	28.6
$D_1' \to D^*\pi\pi$	${D_1'}^0 \to D^{*0}\pi^+\pi^-$	${D_1'}^+ \to D^{*+}\pi^+\pi^-$	42.9
	${D_1'}^0 \to D^{*+}\pi^-\pi^0$	${D_1'}^+ \to D^{*0}\pi^+\pi^0$	28.6
	${D_1'}^0 \to D^{*0}\pi^0\pi^0$	${D_1'}^+ \to D^{*+}\pi^0\pi^0$	28.6
$D_1' \to D^*\eta$	${D_1'}^0 \to D^{*0}\eta^0$	${D_1'}^+ \to D^{*+}\eta^0$	100.0
$D_2^* \to D^{(*)}\pi$	$D_2^{*0} \to D^+\pi^-$	$D_2^{*+} \to D^0\pi^+\pi^0$	40.0
	$D_2^{*0} \to D^0\pi^0$	$D_2^{*+} \to D^+\pi^0$	20.0
	$D_2^{*0} \to D^{*+}\pi^-$	$D_2^{*+} \to D^{*0}\pi^+$	26.7
	$D_2^{*0} \to D^{*0}\pi^0$	$D_2^{*+} \to D^{*+}\pi^0$	13.3

$D^{**}\ell^-\overline{\nu}_\ell) = 0.085 \pm 0.012$ [43]. An uncertainty of 100% is assigned for all $\overline{B} \to D^{**}\tau^-\overline{\nu}_\tau$ decays.

Appendix B M_D and ΔM_{D^*} fits for resolution corrections

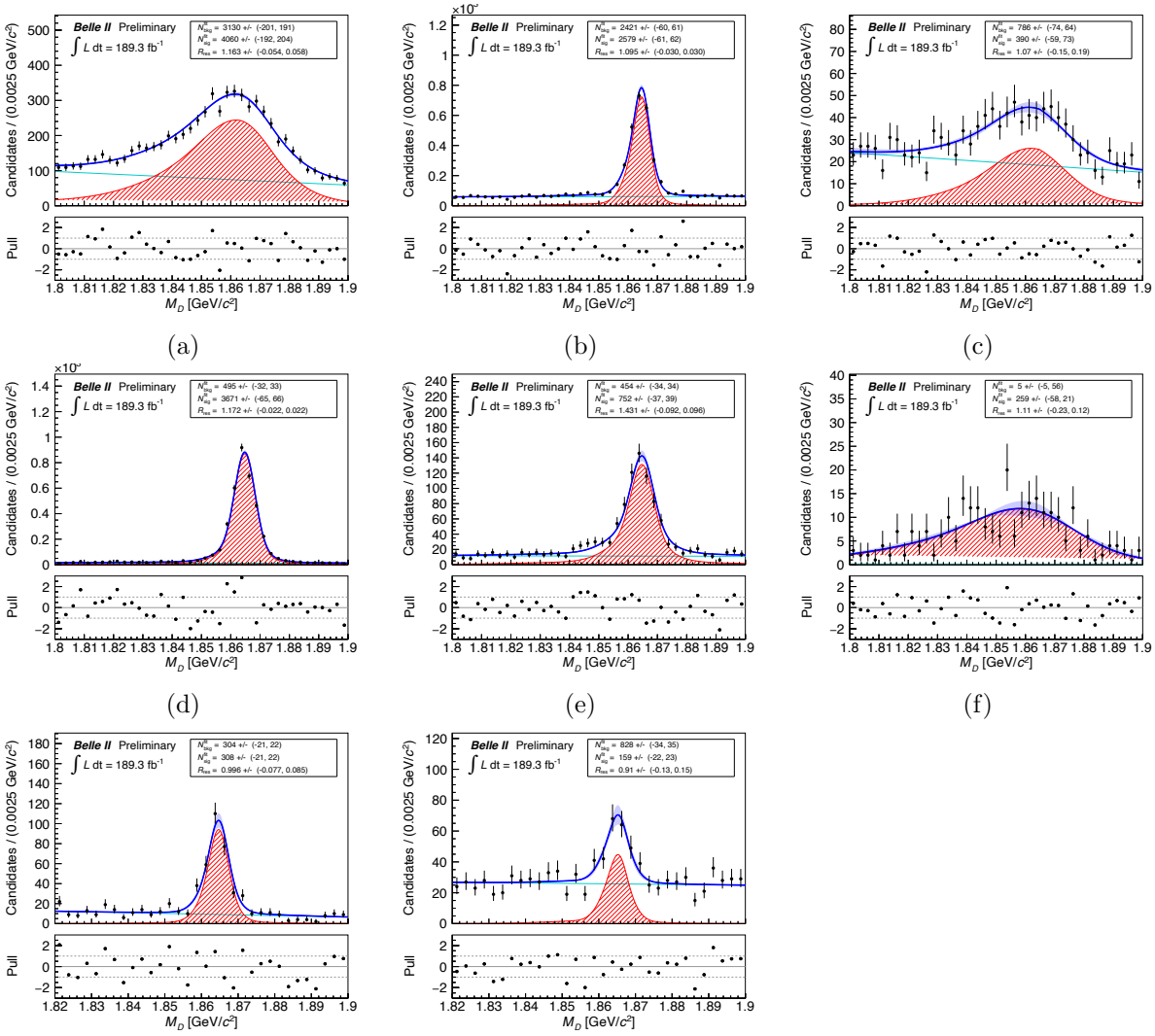
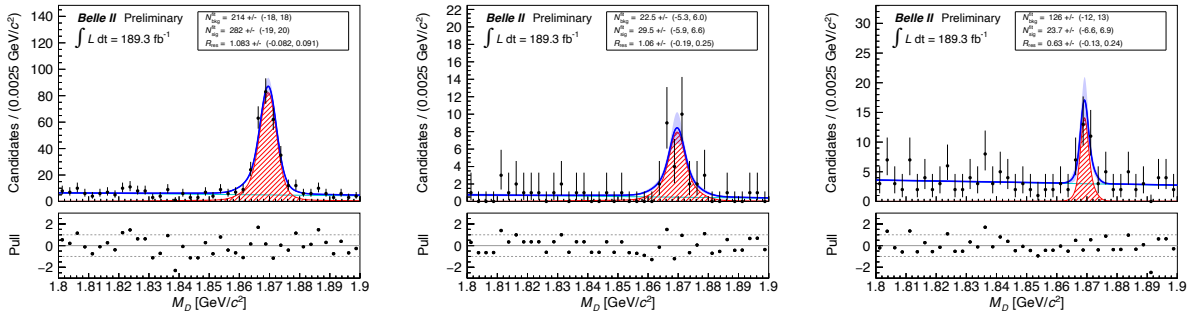


Figure B.1: Results of M_D fits for D^0 modes in the data. (a) $D^0 \to K^-\pi^+\pi^0$, (b) $D^0 \to K^-\pi^+\pi^-\pi^+$, (c) $D^0 \to K^0_S\pi^+\pi^-\pi^0$, (d) $D^0 \to K^-\pi^+$, (e) $D^0 \to K^0_S\pi^+\pi^-$, (f) $D^0 \to K^0_S\pi^0$, (g) $D^0 \to K^-K^+$, and (h) $D^0 \to \pi^+\pi^-$.



(a) (b) (c)

Figure B.2: Results of M_D fits for D^+ modes in the simulation. (a) $D^+ \to K^-\pi^+\pi^+$, (b) $D^+ \to K^0_S\pi^+$, and (c) $D^+ \to K^-K^+\pi^+$.

Table B.1: Correction factors r_{width} for the selection windows of the M_D and ΔM_{D^*} signal regions.

Decays	r_{width}
$D^0 \to K^-\pi^+\pi^0$	$1.10^{+0.05}_{-0.06}$
$D^0 \to K^-\pi^+\pi^-\pi^+$	$1.09^{+0.03}_{-0.03}$
$D^0 \to K^0_S\pi^+\pi^-\pi^0$	$1.02^{+0.19}_{-0.15}$
$D^0 \to K^-\pi^+$	$1.18^{+0.02}_{-0.02}$
$D^0 \to K^0_S\pi^+\pi^-$	$1.43^{+0.10}_{-0.10}$
$D^0 \to K^0_S\pi^0$	$1.08^{+0.23}_{-0.13}$
$D^0 \to K^-K^+$	$1.01^{+0.08}_{-0.09}$
$D^0 \to \pi^+\pi^-$	$0.95^{+0.16}_{-0.15}$
$D^+ \to K^-\pi^+\pi^+$	$1.09^{+0.09}_{-0.10}$
$D^+ \to K^0_S\pi^+$	$1.00^{+0.25}_{-0.19}$
$D^+ \to K^-K^+\pi^+$	$0.73^{+0.30}_{-0.17}$
$D^{*+} \to D^0\pi^+$	$1.14^{+0.07}_{-0.07}$
$D^{*+} \to D^+\pi^0$	$1.00^{+0.08}_{-0.07}$
$D^{*0} \to D^0\pi^0$	$1.10^{+0.08}_{-0.08}$

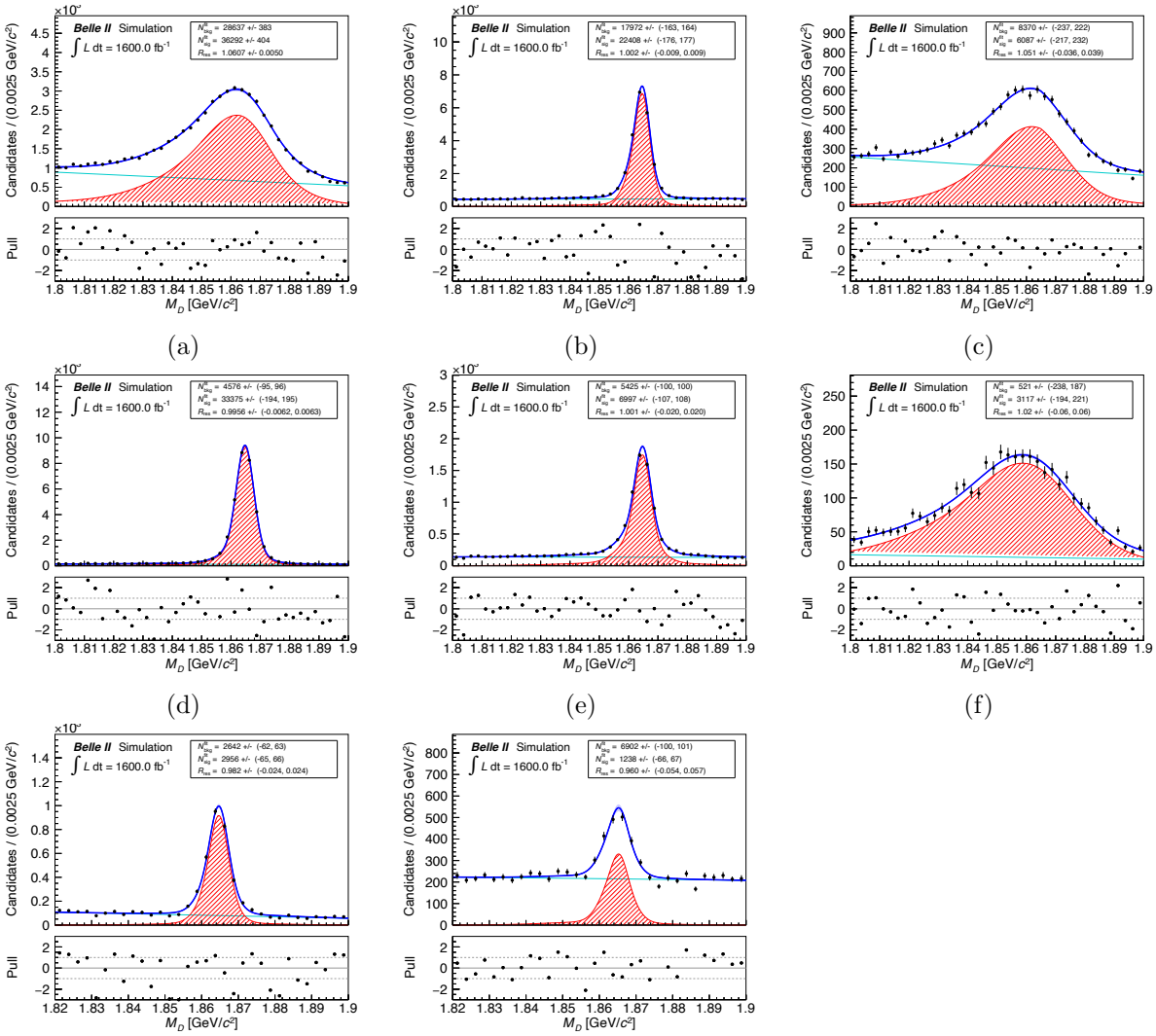
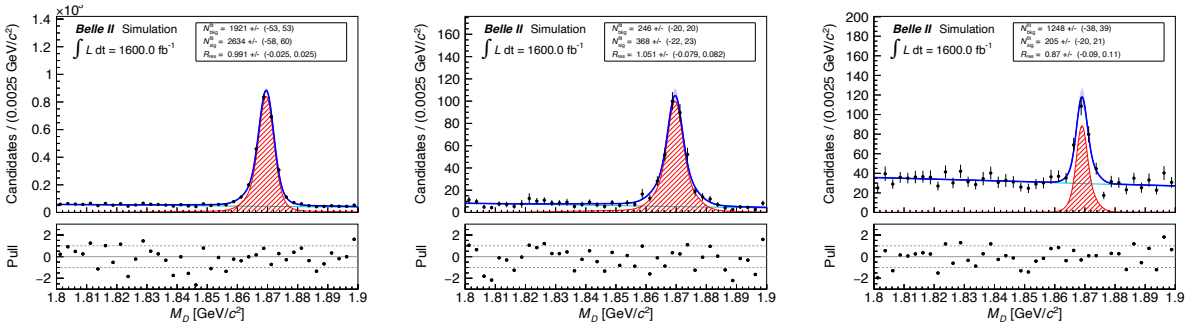


Figure B.3: Results of M_D fits for D^0 modes in the simulation. (a) $D^0 \to K^-\pi^+\pi^0$, (b) $D^0 \to K^-\pi^+\pi^-\pi^+$, (c) $D^0 \to K^0_S\pi^+\pi^-\pi^0$, (d) $D^0 \to K^-\pi^+$, (e) $D^0 \to K^0_S\pi^+\pi^-$, (f) $D^0 \to K^0_S\pi^0$, (g) $D^0 \to K^-K^+$, and (h) $D^0 \to \pi^+\pi^-$.



(a) (b) (c)

Figure B.4: Results of M_D fits for D^+ modes in the simulation. (a) $D^+ \to K^-\pi^+\pi^+$, (b) $D^+ \to K^0_S\pi^+$, and (c) $D^+ \to K^-K^+\pi^+$.

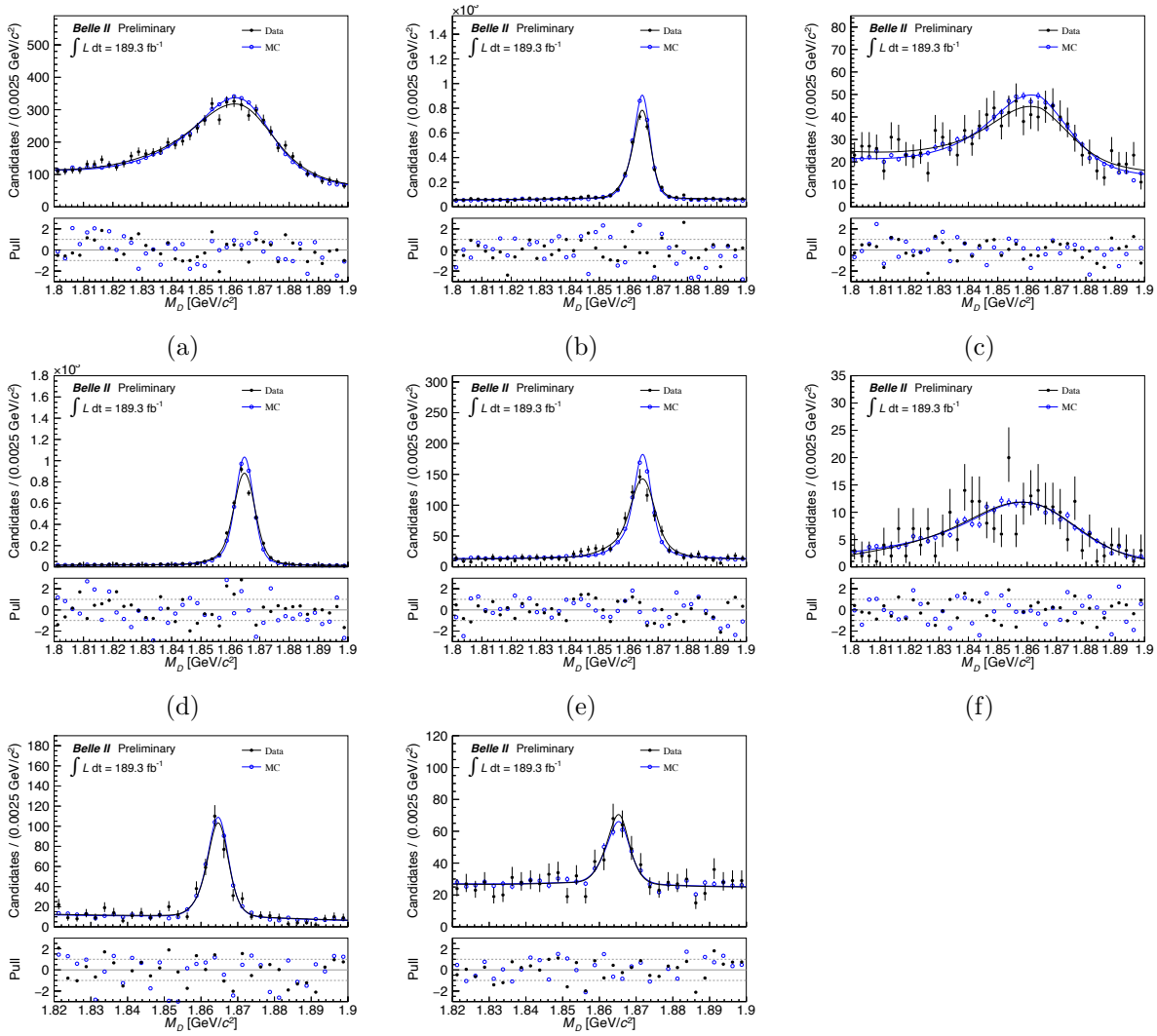
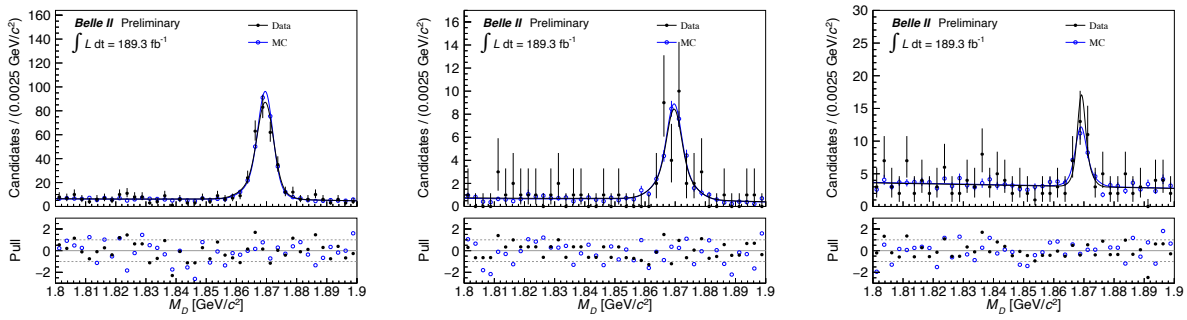


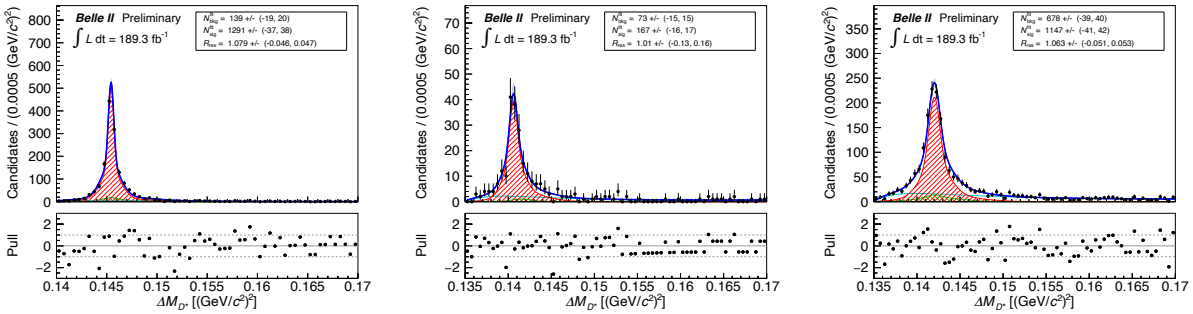
Figure B.5: Comparison of results of M_D fits for D^0 modes between the data and the simulation. (a) $D^0 \to K^-\pi^+\pi^0$, (b) $D^0 \to K^-\pi^+\pi^-\pi^+$, (c) $D^0 \to K^0_S\pi^+\pi^-\pi^0$, (d) $D^0 \to K^-\pi^+$, (e) $D^0 \to K^0_S\pi^+\pi^-$, (f) $D^0 \to K^0_S\pi^0$, (g) $D^0 \to K^-K^+$, and (h) $D^0 \to \pi^+\pi^-$.



(a) (b) (c)

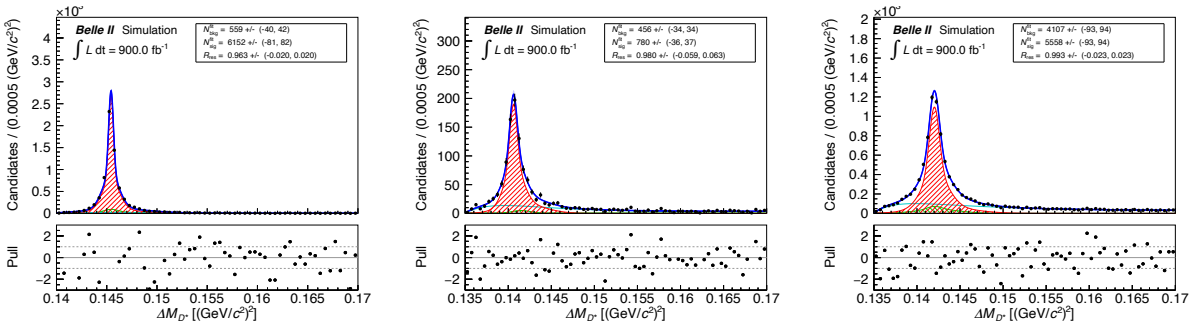
Figure B.6: Comparison of results of M_D fits for D^+ modes between the data and the simulation. (a) $D^+ \to K^-\pi^+\pi^+$, (b) $D^+ \to K^0_S\pi^+$, and (c) $D^+ \to K^-K^+\pi^+$.

Appendix C M_{D^*} fits for yield calibration of fake D^*



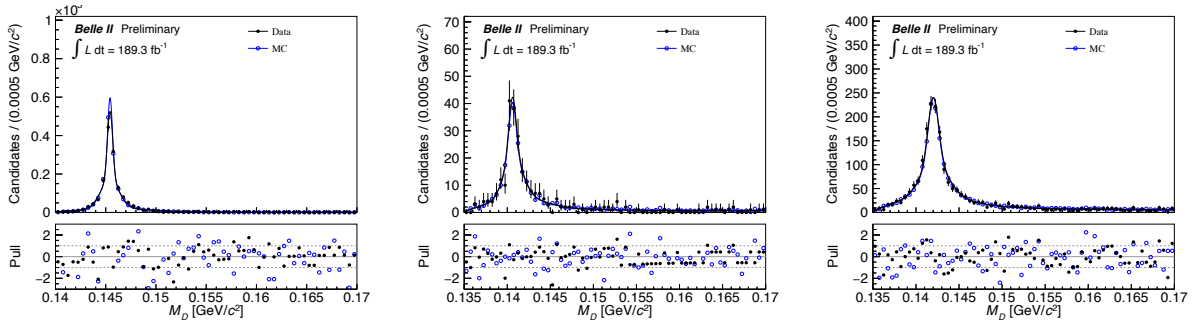
(a) (b) (c)

Figure C.1: Results of ΔM_D^* fits in the data for (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$.



(a) (b) (c)

Figure C.2: Results of ΔM_D^* fits in the simulation for (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$.



(a) (b) (c)

Figure C.3: Results of ΔM_D^* fits in the simulation for (a) $D^{*+} \to D^0\pi^+$, (b) $D^{*+} \to D^+\pi^0$, and (c) $D^{*0} \to D^0\pi^0$.

Appendix D E_{ECL} shapes

Other possible sources as the cause of the discrepancy in E_{ECL} are investigated to understand the tension. These include:

- π^0 from D decay, where the discrepancy in E_{ECL} would vary per D decay mode. This is no dependency observed.
- radiative photons, where the discrepancy in E_{ECL} would vary per lepton mode. This is no dependency observed.
- fake photons from hadronic split-off showers, where the tension in E_{ECL} would change if the amount of these fake photons are veried by validating the selection criteria of minC2TDist variable. This is also no change has been detected.
- fraction of misreconstructed (false `isSignal`) B_{tag} candidates. To explain the discrepancy in E_{ECL}, the miss-reconstruction fraction of data should be much smaller than MC by considering the E_{ECL} shape difference between correctly and wrongly reconstructed B_{tag} candidates. The fraction of data and MC is checked with M_{bc} and ΔE distributions of B_{tag}, and misreconstructed events in data is slightly larger than in MC. So, this hypothesis is not realistic.
- data/MC difference of FEI efficiency in each B_{tag} mode, which should be evident if the discrepancy in E_{ECL} shows dependence on the B_{tag} decay mode. This is also not observed.
- misreconstructed B_{sig} candidates, which can be checked if tight constraints are applied on the B_{sig} side. This is done and a dependence on whether or not B_{sig} is constrained is not observed.

Given the listed investigations and cross-checks, none of the hypotheses can explain the tension reasonably. The systematic uncertainty of this correction will be evaluated in Sec. **??**.

Appendix E Probability density function for signal extraction

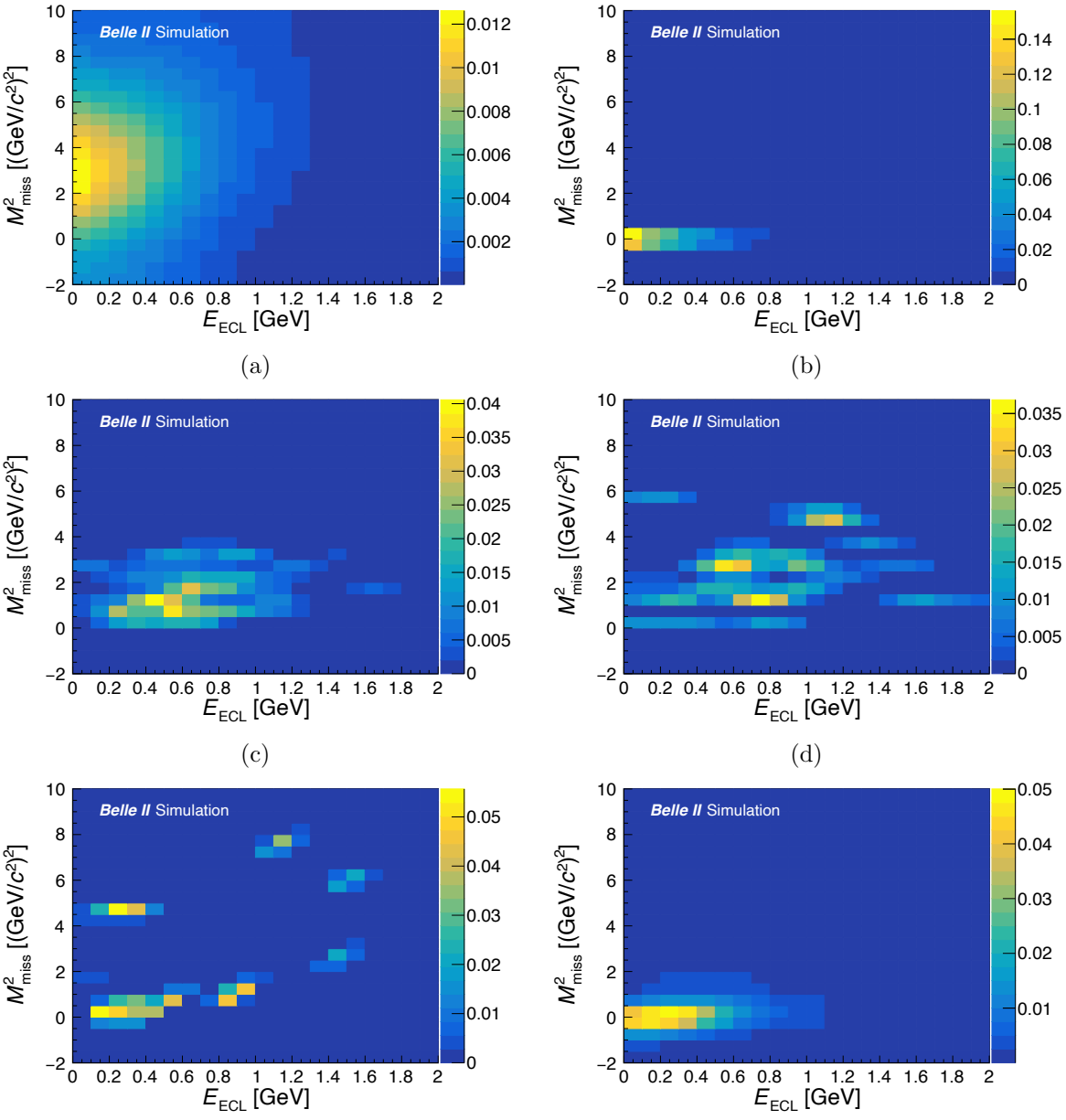


Figure E.1: Two-dimensional histogram PDFs of E_{ECL} and M^2_{miss} from (a) $\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau$, (b) $\overline{B} \to D^{*+}\ell^-\overline{\nu}_\ell$, (c) $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$, (d) hadronic B decays, (e) Sum of $B^+ \to B^0$ cross feed of semi-leptonic B decays, continuum events, and "other" background events with a correctly reconstructed D^* candidate, and (f) background events with a fake D^* candidate in the $D^{*+} \to D^+\pi^0$ mode . The z-axis is in an arbitrary unit.

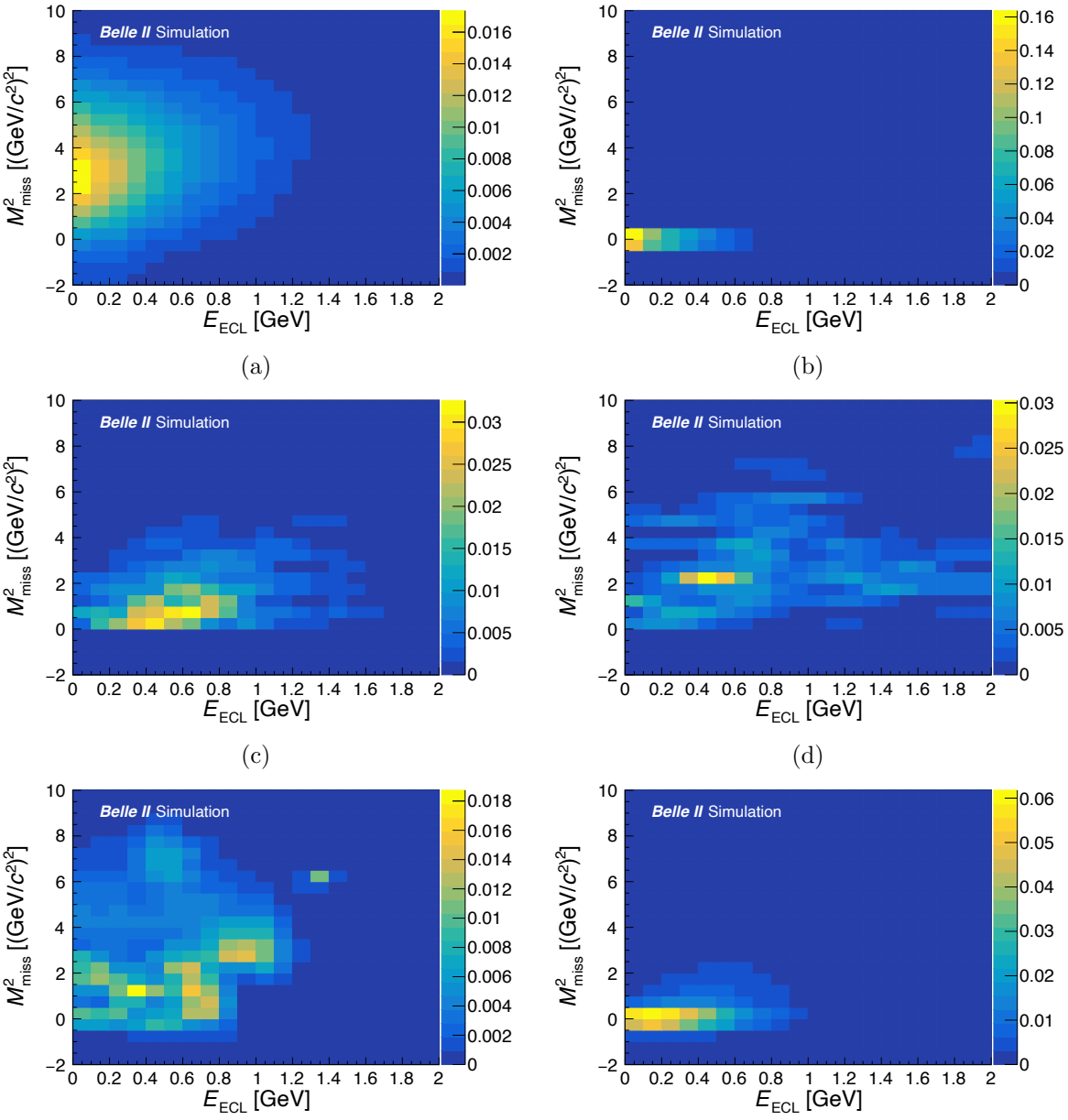


Figure E.2: Two-dimensional histogram PDFs of E_{ECL} and M^2_{miss} from (a) $\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau$, (b) $\overline{B} \to D^{*+}\ell^-\overline{\nu}_\ell$, (c) $\overline{B} \to D^{**}\ell^-\overline{\nu}_\ell$, (d) hadronic B decays, (e) Sum of $B^+ \to B^0$ cross feed of semi-leptonic B decays, continuum events, and "other" background events with a correctly reconstructed D^* candidate, and (f) background events with a fake D^* candidate in the $D^{*0} \to D^0\pi^0$ mode . The z-axis is in an arbitrary unit.

Appendix F Belle II constraints on New Physics parameters

F.1 Single-operator scenarios

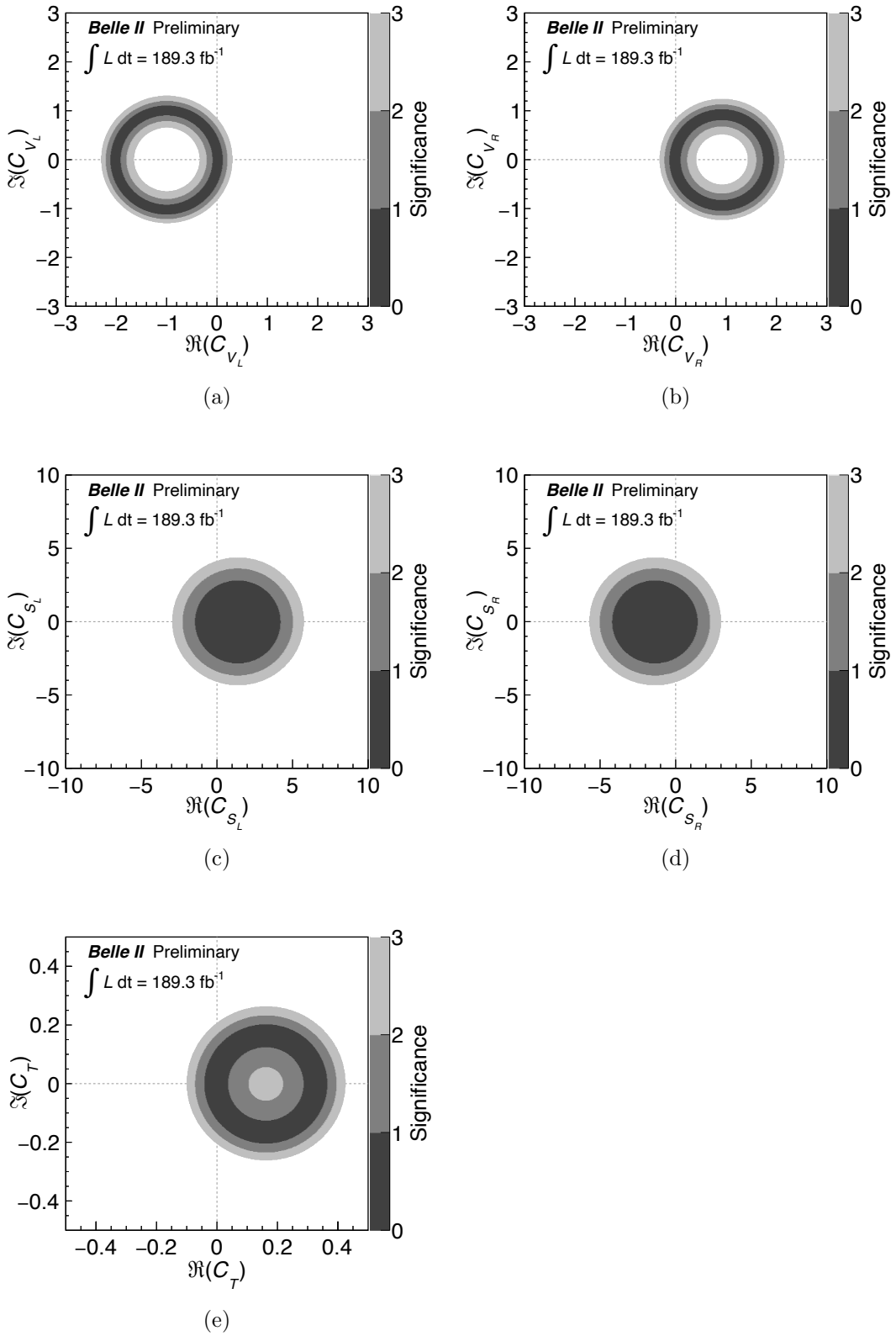


Figure F.1: Favored regions according to our $R(D^*)$ result on the Wilson Coefficients for single NP operators. The constraints are presented for (a) V_L-type, (b) V_R-type, (c) S_L-type, (d) S_R-type, and (e) T-type single operators, respectively. The vertical and horizontal axes represent the real and imaginary parts of the coefficients.

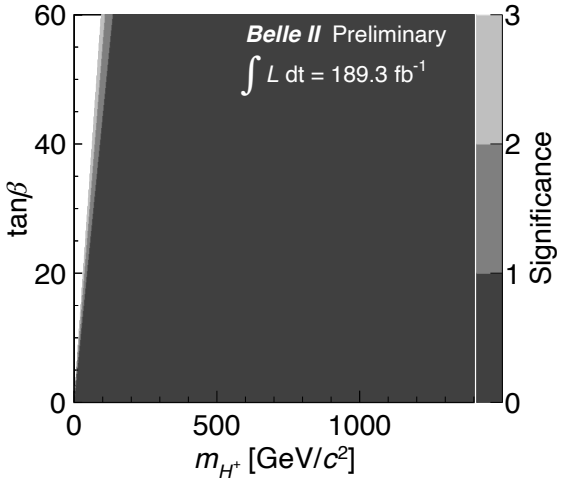


Figure F.2: Favored regions according to our $R(D^*)$ result on masses and $\tan\beta$ of a charged Higgs boson in the two-Higgs type models (Type-II).

F.2 Two-Higgs doublet model scenarios

F.3 Leptoqruark scenarios

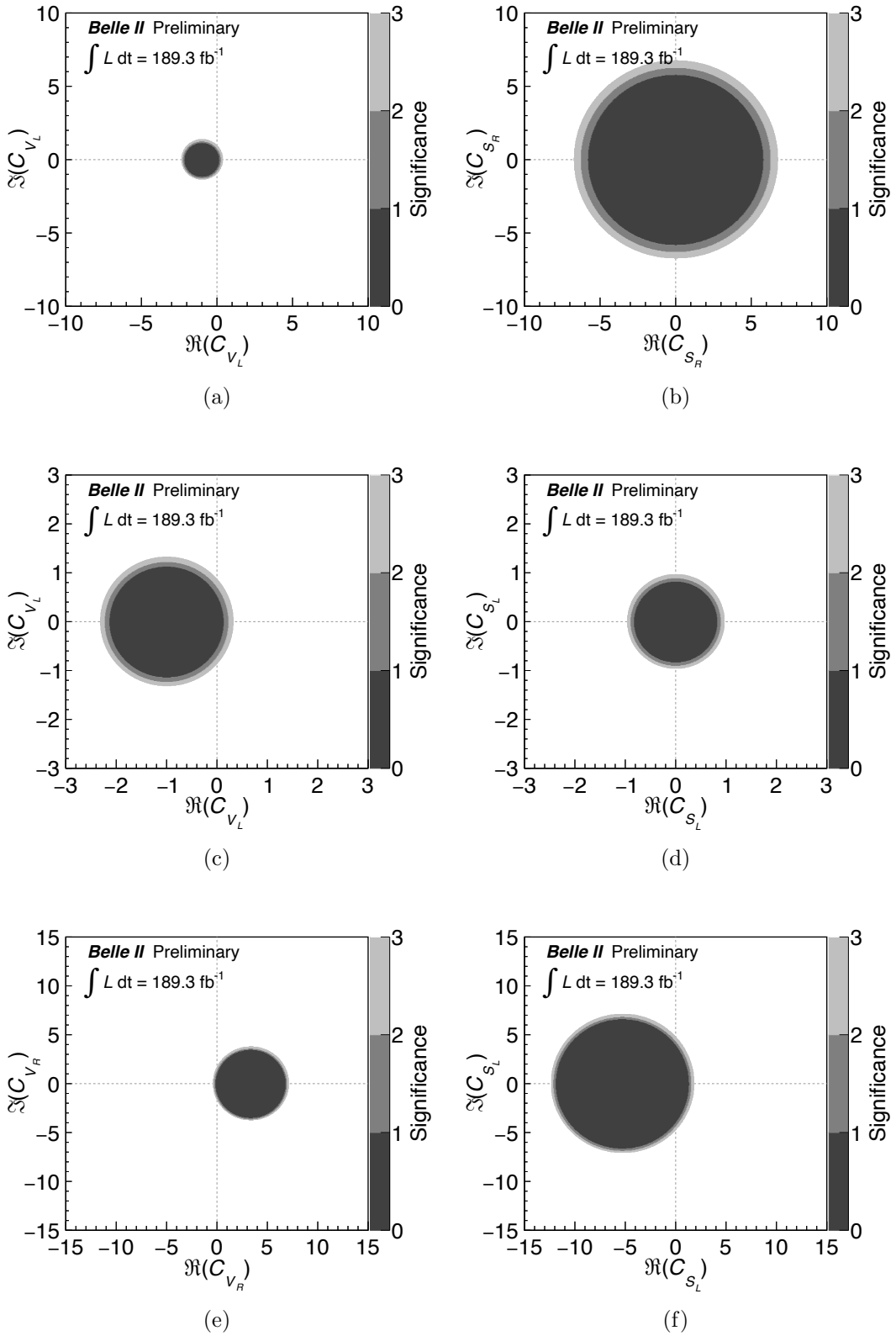


Figure F.3: Favored regions according to our $R(D^*)$ result on the Wilson Coefficients for the leptoquark models. The constraints are presented for (a)–(b) U_1-type, (c)–(d) S_1-type, and (e)–(f) R_2-type leptoquark models, respectively. The vertical and horizontal axes represent the real and imaginary parts of the coefficients.

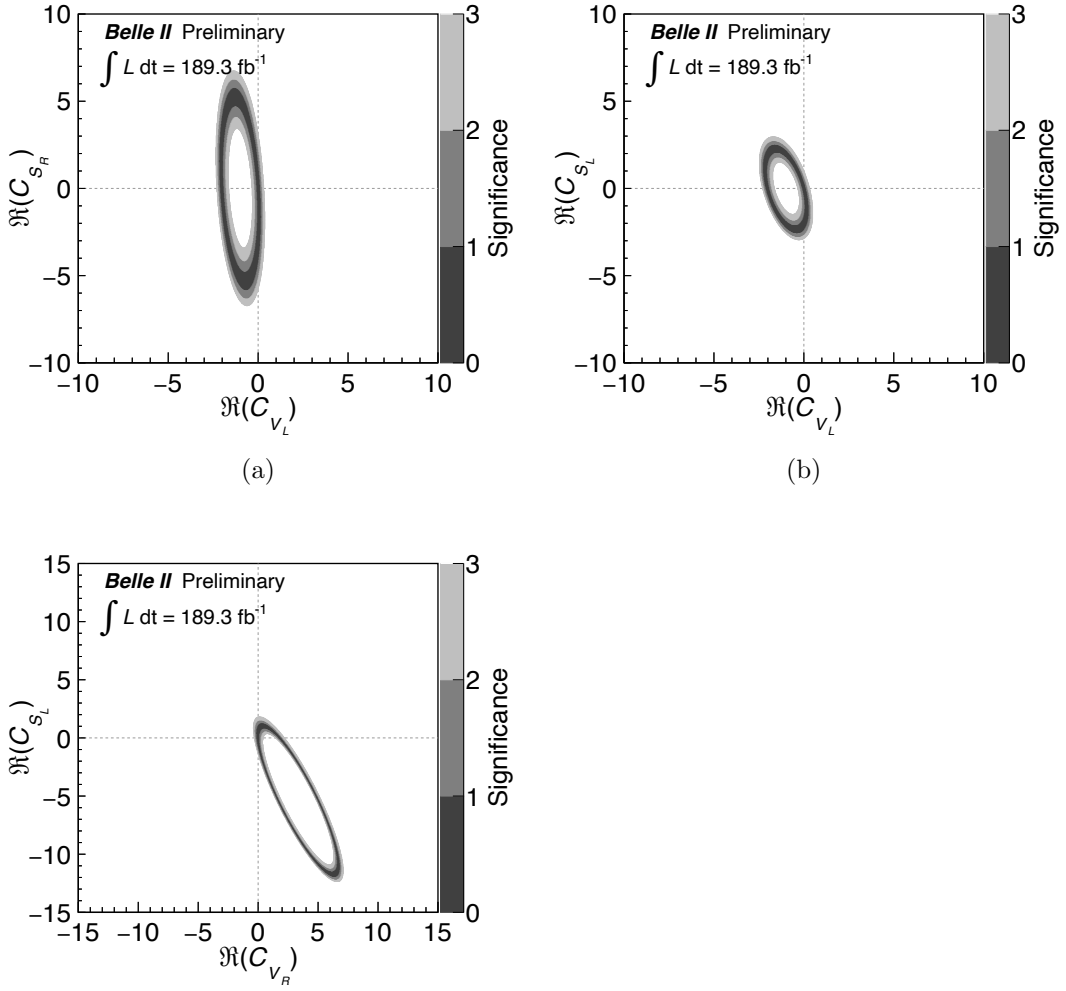


Figure F.4: Favored regions according to our $R(D^*)$ result on the real Wilson Coefficients for the leptoquark models. The imaginary parts of the Wilson Coefficients are assumed to be zero. The constraints are presented for (a)–(b) U_1-type, (c)–(d) S_1-type, and (e)–(f) R_2-type leptoquark models, respectively. The vertical and horizontal axes represent the real parts of the two coefficients.

References

[1] R.L. Workman et al. (The Particle Data Group Collaboration), Review of Particle Physics, *PTEP* **2022** (2022) 083C01.

[2] M. Tanaka and R. Watanabe, New physics in the weak interaction of $\bar{B} \to D^{(*)}\tau\bar{\nu}$, *Phys. Rev. D* **87** (2013) 034028 [arXiv:1212.1878].

[3] G. Aad et al. (The ATLAS Collaboration), Test of the universality of τ and μ lepton couplings in W-boson decays with the ATLAS detector, *Nature Phys.* **17** (2021) 813 [arXiv:2007.14040].

[4] G. Aad et al. (The ATLAS Collaboration), Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *Phys. Lett. B* **759** (2016) 601 [arXiv:1603.09222].

[5] (The CMS Collaboration), A precision measurement of the W boson decay branching fractions in pp collisions at $\sqrt{s} = 13$ TeV, .

[6] D. Bryman, V. Cirigliano, A. Crivellin and G. Inguglia, Testing Lepton Flavor Universality with Pion, Kaon, Tau, and Beta Decays, *Ann. Rev. Nucl. Part. Sci.* **72** (2022) 69 [arXiv:2111.05338].

[7] A. Pich, Precision Tau Physics, *Prog. Part. Nucl. Phys.* **75** (2014) 41 [arXiv:1310.7922].

[8] V.M. Aulchenko et al. (The KEDR Collaboration), Measurement of the ratio of the leptonic widths $\Gamma_{ee}/\Gamma_{\mu\mu}$ for the J/ψ meson, *Phys. Lett. B* **731** (2014) 227 [arXiv:1311.5005].

[9] E. Waheed et al. (The Belle Collaboration), Measurement of the CKM matrix element $|V_{cb}|$ from $B^0 \to D^{*-}\ell^+\nu_\ell$ at Belle, *Phys. Rev. D* **100** (2019) 052007 [arXiv:1809.03290].

[10] I. Adachi et al. (The Belle II Collaboration), Determination of —Vcb— using B¯0→D*+ℓ−ν¯ℓ decays with Belle II, *Phys. Rev. D* **108** (2023) 092013 [arXiv:2310.01170].

[11] L. Aggarwal et al. (The Belle II Collaboration), Test of Light-Lepton Universality in the Rates of Inclusive Semileptonic B-Meson Decays at Belle II, *Phys. Rev. Lett.* **131** (2023) 051804 [arXiv:2301.08266].

[12] S. Choudhury et al. (The Belle Collaboration), Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays, *JHEP* **03** (2021) 105 [arXiv:1908.01848].

[13] A. Abdesselam et al. (The Belle Collaboration), Test of Lepton-Flavor Universality in $B \to K^*\ell^+\ell^-$ Decays at Belle, *Phys. Rev. Lett.* **126** (2021) 161801 [arXiv:1904.02440].

[14] R. Aaij et al. (The LHCb Collaboration), Tests of lepton universality using $B^0 \to K^0_S\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays, *Phys. Rev. Lett.* **128** (2022) 191802 [arXiv:2110.09501].

[15] R. Aaij et al. (The LHCb Collaboration), Test of lepton universality in $b \to s\ell^+\ell^-$ decays, *Phys. Rev. Lett.* **131** (2023) 051803 [arXiv:2212.09152].

[16] R. Aaij et al. (The LHCb Collaboration), Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays, *Phys. Rev. D* **108** (2023) 032002 [arXiv:2212.09153].

[17] R. Aaij et al. (The LHCb Collaboration), Test of lepton universality with $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays, *JHEP* **05** (2020) 040 [arXiv:1912.08139].

[18] I. Adachi et al. (The Belle II Collaboration), Tests of Light-Lepton Universality in Angular Asymmetries of B0→D*-$\ell\nu$ Decays, *Phys. Rev. Lett.* **131** (2023) 181801 [arXiv:2308.02023].

[19] J.P. Lees et al. (The BaBar Collaboration), Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays, *Phys. Rev. Lett.* **109** (2012) 101802 [arXiv:1205.5442].

[20] J.P. Lees et al. (The 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** (2013) 072012 [arXiv:1303.0571].

[21] M. Huschle et al. (The 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** (2015) 072014 [arXiv:1507.03233].

[22] S. Hirose et al. (The Belle Collaboration), Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$, *Phys. Rev. Lett.* **118** (2017) 211801 [arXiv:1612.00529].

[23] S. Hirose et al. (The Belle Collaboration), Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ with one-prong hadronic τ decays at Belle, *Phys. Rev. D* **97** (2018) 012004 [arXiv:1709.00129].

[24] G. Caria et al. (The Belle Collaboration), Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method, *Phys. Rev. Lett.* **124** (2020) 161803 [arXiv:1910.05864].

[25] R. Aaij et al. (The LHCb Collaboration), Test of lepton flavor universality using B0→D*-τ+$\nu\tau$ decays with hadronic τ channels, *Phys. Rev. D* **108** (2023) 012018 [arXiv:2305.01463].

[26] (The LHCb Collaboration), Measurement of the ratios of branching fractions $\mathcal{R}(D^*)$ and $\mathcal{R}(D^0)$, *Phys. Rev. Lett.* **131** (2023) 111802 [arXiv:2302.02886].

[27] Y.S. Amhis et al. (The HFLAV Collaboration), Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021, *Phys. Rev. D* **107** (2023) 052008 [arXiv:2206.07501], https://hflav.web.cern.ch/.

[28] M. Bordone, M. Jung and D. van Dyk, Theory determination of $\bar{B} \to D^{(*)}\ell^-\bar{\nu}$ form factors at $\mathcal{O}(1/m_c^2)$, *Eur. Phys. J. C* **80** (2020) 74 [arXiv:1908.09398].

[29] S. Iguro, T. Kitahara and R. Watanabe, Global fit to $b \to c\tau\nu$ anomalies 2022 mid-autumn, arXiv:2210.10751.

[30] K. Akai, K. Furukawa and H. Koiso (The SuperKEKB Collaboration), SuperKEKB Collider, *Nucl. Instrum. Meth. A* **907** (2018) 188 [arXiv:1809.01958].

[31] T. Abe et al. (The Belle II Collaboration), Belle II Technical Design Report, arXiv:1011.0352.

[32] T. Keck et al., The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment, *Comput. Softw. Big Sci.* **3** (2019) 6 [arXiv:1807.08680].

[33] T. Keck, FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment, *Comput. Softw. Big Sci.* **1** (2017) 2.

[34] F. Abudinén et al. (The Belle II Collaboration), A calibration of the Belle II hadronic tag-side reconstruction algorithm with $B \to X\ell\nu$ decays, arXiv:2008.06096.

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

[36] S. Agostinelli et al. (The GEANT4 Collaboration), GEANT4–a simulation toolkit, *Nucl. Instrum. Meth. A* **506** (2003) 250.

[37] E. Barberio, B. van Eijk and Z. Was, PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays, *Comput. Phys. Commun.* **66** (1991) 115.

[38] N. Davidson, T. Przedzinski and Z. Was, PHOTOS interface in C++: Technical and Physics Documentation, *Comput. Phys. Commun.* **199** (2016) 86 [arXiv:1011.0937].

[39] T. Sjostrand, S. Mrenna and P.Z. Skands, A Brief Introduction to PYTHIA 8.1, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820].

[40] S. Jadach, B.F.L. Ward and Z. Was, The Precision Monte Carlo event generator K K for two fermion final states in e+ e- collisions, *Comput. Phys. Commun.* **130** (2000) 260 [arXiv:hep-ph/9912214].

[41] B.F.L. Ward, S. Jadach and Z. Was, Precision calculation for e+ e- —> 2f: The KK MC project, *Nucl. Phys. B Proc. Suppl.* **116** (2003) 73 [arXiv:hep-ph/0211132].

[42] F.U. Bernlochner, Z. Ligeti, M. Papucci, M.T. Prim, D.J. Robinson and C. Xiong, Constrained second-order power corrections in HQET: R(D(*)), —Vcb—, and new physics, *Phys. Rev. D* **106** (2022) 096015 [arXiv:2206.11281].

[43] F.U. Bernlochner and Z. Ligeti, Semileptonic $B_{(s)}$ decays to excited charmed mesons with e, μ, τ and searching for new physics with $R(D^{**})$, *Phys. Rev. D* **95** (2017) 014022 [arXiv:1606.09300].

[44] F.U. Bernlochner, Z. Ligeti and D.J. Robinson, Model independent analysis of semileptonic B decays to D^{**} for arbitrary new physics, *Phys. Rev. D* **97** (2018) 075011 [arXiv:1711.03110].

[45] F.U. Bernlochner, S. Duell, Z. Ligeti, M. Papucci and D.J. Robinson, Das ist der HAMMER: Consistent new physics interpretations of semileptonic decays, *Eur. Phys. J. C* **80** (2020) 883 [arXiv:2002.00020].

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

[47] S. Longo et al., CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders, *Nucl. Instrum. Meth. A* **982** (2020) 164562 [arXiv:2007.09642].

[48] F. Zernike, Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode, *Physica* **1** (1934) 689.

[49] V. Lakshminarayanan and A. Fleck, Zernike polynomials: a guide, *Journal of Modern Optics* **58** (2011) 545.

[50] N.D. Gagunashvili, Comparison of weighted and unweighted histograms, *PoS* **ACAT** (2007) 054 [arXiv:physics/0605123].

[51] K.G. Hayes, M.L. Perl and B. Efron, Application of the Bootstrap Statistical Method to the Tau Decay Mode Problem, *Phys. Rev. D* **39** (1989) 274.

[52] B. Aubert et al. (The BaBar Collaboration), Study of Resonances in Exclusive B Decays to anti-$D(*)D(*)K$, *Phys. Rev. D* **77** (2008) 011102 [arXiv:0708.1565].

[53] Y. Sato et al. (The Belle Collaboration), Measurement of the branching ratio of

$\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method, *Phys. Rev. D* **94** (2016) 072007 [arXiv:1607.07923].

[54] Z.-z. Xing, H. Zhang and S. Zhou, Updated Values of Running Quark and Lepton Masses, *Phys. Rev. D* **77** (2008) 113016 [arXiv:0712.1419].

[55] J.P. Lees et al. (The BaBar Collaboration), Observation of $\overline{B} \to D^{(*)}\pi^+\pi^-\ell^-\overline{\nu}$ decays in e^+e^- collisions at the $\Upsilon(4S)$ resonance, *Phys. Rev. Lett.* **116** (2016) 041801 [arXiv:1507.08303].