Studies of $B \to D^{**}\ell\nu_\ell$ decays with Belle II

Nada Gharbi

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

June 2024

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, 24.07.2025 .....
Date

.....
Signature

1. Gutachter: Dr. Markus Prim
2. Gutachter: Prof. Dr. Florian Bernlochner

Acknowledgements

Special thanks go to my supervisors (and first reviewers) Dr Markus Tobias Prim and Dr Valerio Bertacchi for their constant support and availability throughout the Master's thesis. Discussions and different perspectives influenced the design of the approaches in this thesis. They provided me with an understanding of complex theoretical backgrounds and their implementation, and inspired me to approach problems with optimism and analytical thinking.
Besides that, I want to explicitly thank Dr Valerio Bertacchi for his contribution to the reconstruction script the semi-leptonic tagged FEI analysis.
I would also like to thank Professor Dr Florian Bernlochner, who also supervised my Master's thesis and acted as the second reviewer. I am extremely grateful to him for his helpful suggestions, in-depth discussions and constructive criticism during the preparation of this thesis.
Special thanks also go to Ilias Tsaklidis and Agrim Aggarwal for developing and providing the `SysVar` package, which was used in this thesis to apply correction weights and thereby improve the agreement between simulation and data in the comparison plots.
I would also like to thank Florian Herren and Raynette van Tonder, who provided me with the fitting code of the model-independent description of $B \to D\pi\ell\nu_\ell$ decays in Reference [1].
I am also grateful to Simon Calo for developing the reconstruction script for the kinematic variables, which is based on the Diamond Frame technique.
Finally, I would like to thank my family, friends, and the working group for their consistent support throughout my work.

Contents

CHAPTER 1

Introduction

Semileptonic decays of B mesons provide a crucial testing ground for the interplay between weak and strong interactions, offering valuable input for determining fundamental parameters such as CKM matrix elements. In particular, decays of the type $B \to D^{**}\ell\nu$, where D^{**} denotes an orbitally excited charm meson, are essential for understanding the dynamics of heavy-quark transitions and the structure of the hadronic final state.

A key challenge in studying these transitions is modelling the hadronic matrix elements that govern the decay amplitudes. The BLR (Bernlochner, Ligeti and Robinson) model [2] has been developed to describe the form factors associated with semileptonic $B \to D^{**}\ell\nu$ decays. These form factors are usually based on the assumption of narrow-width D^{**} states, which simplifies the theoretical treatment but does not reflect the complex nature of the broad excited states.

In simulation tools such as EvtGen [3], the D^{**} states are commonly treated as Breit-Wigner resonances, with their nominal masses and widths taken from the Particle Data Group (PDG) [4]. While this approach is relatively robust for narrow states, it becomes increasingly problematic for broad resonances such as the $D_0^*(2300)$ and $D_1'(2430)$. Using a Breit-Wigner form to describe these broad states can lead to biased signal yields when fitting invariant mass distributions, potentially misidentifying or misquantifying the contributions from individual resonances.

Recent theoretical efforts have highlighted the limitations of the narrow-width approximations and Breit-Wigner parameterisation. For example, extensions of Heavy Quark Effective Theory (HQET) aim to generalize the form factor description beyond the narrow-width limit [5], and alternative approaches such as the extended Boyd-Grinstein-Lebed (BGL) parameterisation offer more flexibility in modeling the dynamics of broad states [1]. Moreover, studies of the $D\pi$ phase motion suggest that the broad D^{**} states are not well modeled by simple resonances. The structure identified as the $D_0^*(2300)$, for example, appears to be dynamically generated by two poles far from the real axis, as shown in unitarized chiral perturbation theory (UChPT) [6, 7]. These findings challenge the conventional picture of such states as pure $q\bar{q}$ mesons.

In light of these challenges, a more model-independent strategy is desirable from the experimental perspective. The most robust approach involves extracting background-subtracted invariant mass spectra of the $D^{(*)}\pi$ system without imposing strong assumptions about the signal composition. This allows for flexible post-hoc modeling of the underlying amplitudes, enabling fits to the $M(D^{(*)}\pi)$ distributions using arbitrary models. From such fits, it is then possible to derive event by event weights and systematically disentangle the individual contributions from different D^{**} states. This data-driven strategy provides a

path to more accurate determinations of signal yields, branching fractions, and a deeper understanding of the nature of excited charm mesons.

This thesis presents a first feasibility study of the semileptonic decays $B \to D^{**}\ell\nu$ at Belle II using two different reconstruction strategies: an untagged approach and a semileptonic (SL) tagged approach. The goal is to evaluate the sensitivity and practical challenges associated with these methods, as well as to establish whether a full measurement effort based on these channels is promising. This study is performed using Monte Carlo (MC) simulations only.
After a theoretical overview of semileptonic B decays and the relevant form factor parameterizations, we describe the Belle II detector setup and the MC samples used in this study.
The core analysis begins with the untagged reconstruction approach, where only the signal side decay is reconstructed. This method is technically simple and has high statistics, but suffers from significant background contamination and tag-side cross-feed effects. A dedicated boosted decision tree (BDT) was trained for background suppression, though the separation power between signal and background remained limited.
Subsequently, we focus on the SL tagged approach, where the signal-side B decay is reconstructed in conjunction with a semi-leptonic tag-side B candidate. This method benefits from better kinematic constraints and background suppression. Following the event selection, the signal extraction is performed using the sPlot technique. Since this method requires a discriminating variable that is statistically independent of the observables of interest, we train a second BDT with variables uncorrelated to $M(D^{(*)}\pi)$, q^2, and $\cos\theta_V$. The output BDT score is then used as the discriminating variable in the fit.
As a final step, we explore corrections for detector and reconstruction effects. Migration matrices are constructed to unfold detector smearing effects, including inter-mode migrations between $D\pi$ and $D^*\pi$ final states. This study, demonstrates the potential of SL tagged analyses at Belle II and identifies the main areas where further work is needed. The results provide a foundation for continued development of a full measurement strategy for $B \to D^{**}\ell\nu$ decays.

CHAPTER 2

Theoretical overview

2.1 Orbitally Excited D Mesons

Heavy quark symmetry [8]-[9], introduced in the early 1990s, led to a substantial leap forward in our understanding of the spectroscopy and semileptonic decays of mesons containing a single heavy quark. In the infinite heavy quark mass limit ($m_Q \to \infty$), the spin and flavor of the heavy quark decouple from the dynamics of the light degrees of freedom, which consist of light quarks and gluons. As a result, these light degrees of freedom, characterized by quantum numbers $s_\ell^{\pi_\ell}$, determine the structure of the meson, forming spin doublets with total angular momentum

$$J = s_\ell \pm 1/2$$

and parity $P = \pi_\ell$.
This leads to the emergence of heavy quark spin and flavor symmetry, simplifying the treatment of heavy-light systems. This behavior can be understood from the fact that in a meson with a heavy quark, the spin interactions between the heavy quark and the light degrees of freedom are suppressed by powers of the heavy quark mass. Consequently, the heavy quark effectively appears to the light quark as a static color source. In this limit, the dynamics of the light quark field depend only on the velocity of the heavy quark and not on its spin or mass (or flavour), giving rise to heavy quark symmetry.
A meson is considered a heavy meson if the mass of the heavy quark satisfies $M_Q \gg \Lambda_{\mathrm{QCD}}$, where Λ_{QCD} is the QCD scale parameter. Corrections to this idealized scenario arise at subleading order in $\Lambda_{\mathrm{QCD}}/m_Q$, where the finite mass of the heavy quark breaks these symmetries. These corrections are proportional to inverse mass differences, such as $1/m_{Q_i} - 1/m_{Q_j}$, and can be systematically incorporated within the framework of Heavy Quark Effective Theory (HQET) [10].

For charmed mesons, the lowest-lying states are characterized by a light degrees-of-freedom spin-parity of $s_\ell^{\pi_\ell} = \frac{1}{2}^-$, corresponding to an orbital angular momentum $L = 0$. These form a doublet with total spin-parity $J^P = 0^-$ and $J^P = 1^-$, which are identified with the pseudoscalar D and vector D^* mesons, respectively [9]. For $L = 1$, the orbitally excited states, collectively denoted as D^{**}, include the mesons D_0^*, D_1', D_1, and D_2^*. These group into two doublets: the $s_\ell^{\pi_\ell} = \frac{1}{2}^+$ doublet, consisting of D_0^* and D_1', and the $s_\ell^{\pi_\ell} = \frac{3}{2}^+$ doublet, consisting of D_1 and D_2^* [11]. The mass spectrum of the charmed mesons is illustrated in Fig. 2.1 and again quantitative summarized with its quantum numbers in Table 2.1. HQET not only classifies the D^{**} states by their light-quark angular momentum, but also constrains their decay

modes, as conservation of quantum numbers determines whether a given state decays to $D\pi$ or $D^*\pi$.

Table 2.1: Masses and widths of D meson states [11].

Notation	$s_l^{\pi^l}$	J^P	m (GeV)	Γ (GeV)
D	$\frac{1}{2}^-$	0^-	1.87	
D^*	$\frac{1}{2}^-$	1^-	2.01	
D_0^*	$\frac{1}{2}^+$	0^+	2.40	0.28
D_1'	$\frac{1}{2}^+$	1^+	2.44	0.38
D_1	$\frac{3}{2}^+$	1^+	2.42	0.03
D_2^*	$\frac{3}{2}^+$	2^+	2.46	0.04
D'	$\frac{1}{2}^-$	0^-	2.54	0.13
D'^*	$\frac{1}{2}^-$	1^-	2.61	0.09

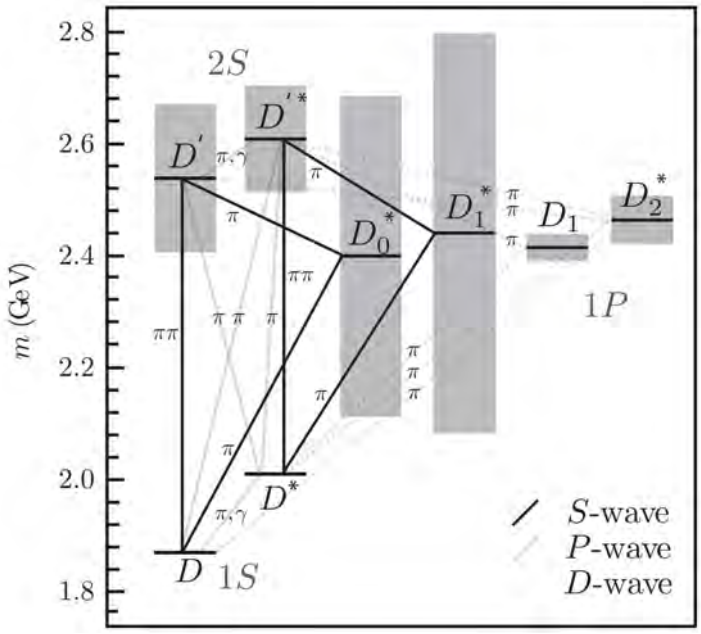


Figure 2.1: Masses of D^{**} (or 1P states) with the corresponding widths, shown with grey bands [11]. Note that D_1^* is described with D_1' in the text.

2.2 Two-pole structure of $D_0^*(2300)$

Resonances play a central role in the study of hadronic interactions and spectroscopy. They correspond to poles of the scattering amplitude on unphysical Riemann sheets of the complex energy plane. For two-body scattering, a resonance is typically identified with a pole on the second Riemann sheet,

$$E_{\text{pole}} = M_R - \frac{i}{2}\Gamma_R, \tag{2.1}$$

where M_R and Γ_R denote the mass and width of the resonance, respectively. In terms of the center-of-mass momentum k, the position of the pole is often expressed as

$$k = \sqrt{E_{\text{pole}} - E_{\text{threshold}}}, \tag{2.2}$$

A true resonance manifests as a rapid phase motion of the scattering amplitude and appears as a pole close to the real axis. The extraction of such poles requires an analytic continuation into the complex plane, which in turn demands a theoretical framework consistent with unitarity and analyticity [12, 13].

Traditionally, resonances such as the $D_0^*(2300)$ have been modeled using the BW parametrization [14], which has the (non-relativistic) form

$$\text{BW}(E) \propto \frac{1}{(E - M_R)^2 + \Gamma_R^2/4}. \tag{2.3}$$

While effective for narrow, isolated resonances, this approach breaks down in several critical ways in the case of the broad $D_0^*(2300)$:

- **Narrow width assumption:** The BW form assumes $\Gamma \ll M$, which is violated by the broad width of the $D_0^*(2300)$, with $\Gamma \sim 200$–300 MeV.
- **Single isolated resonance:** The BW model neglects interference with other resonant and non-resonant contributions.
- **No nearby thresholds:** Coupled-channel effects from $D\eta$, D_sK are ignored in the BW description.
- **Pole near the real axis:** The BW assumes a pole close to the real axis.
- **Negligible background coupling:** BW assumes only weak coupling (or no strong coupling) to continuum states or background amplitudes.

This highlights the need to go beyond traditional parametrizations in order to achieve a more accurate understanding of scalar charm mesons and broad states.

The scalar charmed meson $D_0^*(2300)$, listed in the Review of Particle Physics (RPP) [4] as the lightest charm-non-strange scalar state, has long served as the starting point for many theoretical analyses. However, recent high-precision data from the LHCb experiment, specifically for the decay $B^- \to D^+\pi^-\pi^-$, reveal serious inconsistencies between this state and experimental observations. A detailed reanalysis using unitarized chiral perturbation theory (UChPT) [15, 16] suggests the presence of a lower-mass scalar resonance. The UChPT approach incorporates chiral symmetry, unitarity, and coupled-channel dynamics. It predicts a scalar resonance around 2.1 GeV that emerges dynamically from the $D\pi$ S-wave interaction and aligns with LHCb data [17].
Figure 2.2 shows the $D\pi$ S-wave amplitude extracted from LHCb data, along with the corresponding fit from UChPT.
UChPT further predicts a second pole in the complex energy plane at $(2451^{+35}_{-26} - i\,134^{+7}_{-8})$ MeV [15], which is located deeper on the unphysical Riemann sheet. This second pole is located close to the $D_s\bar{K}$ threshold and enhances the cusp-structure in the S-wave amplitude. The interference between the shallow pole near 2.1 GeV and the deeper pole near the $D_s\bar{K}$ threshold leads to the distorted line shape seen in experiment, which cannot be reproduced by a single Breit-Wigner (BW) resonance.

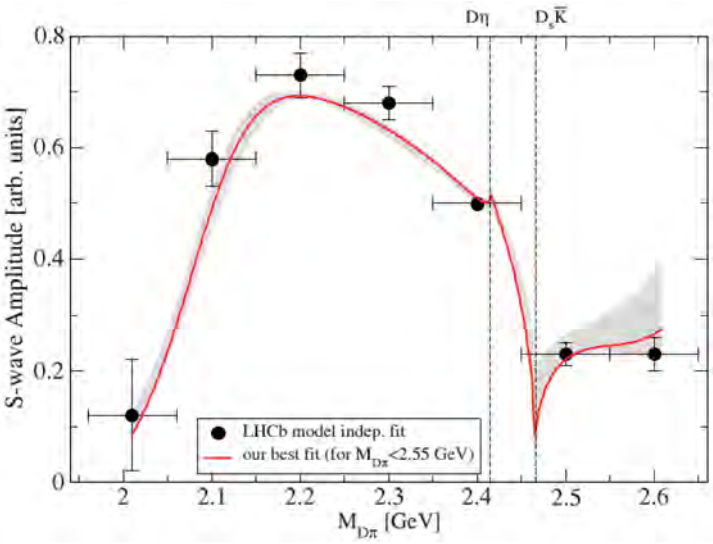


Figure 2.2: Comparison of the S-wave amplitude based on UChPT to the S-wave anchor points of the $B \to D\pi\pi$ decay with LHCb [17]. Figure taken from [15].

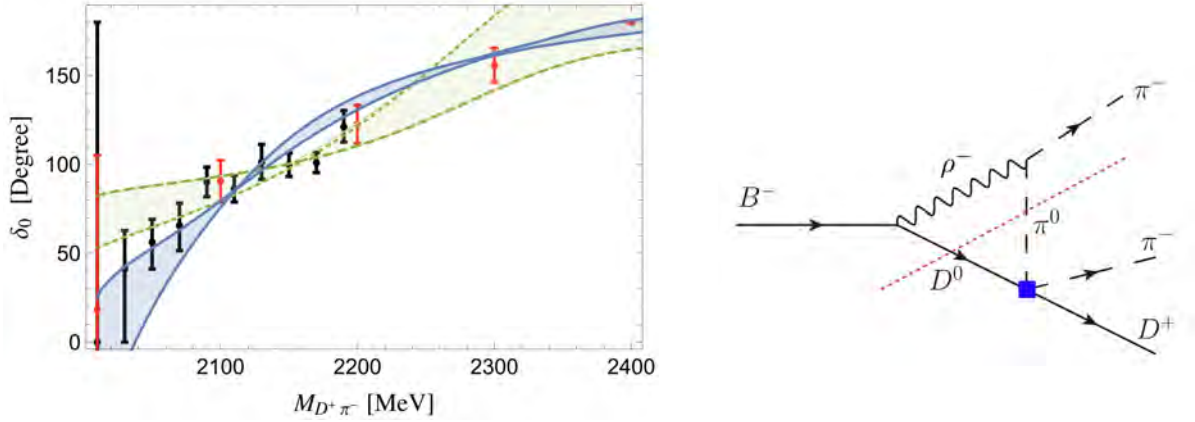


Figure 2.3: (Left) Phase motion of the $D^{**} \to D\pi$ transition. The red points represent LHCb data, while the black points are extracted via an angular-moment fit under the assumption of negligible D-wave contributions. The green curve corresponds to a Breit–Wigner parametrization, and the blue curve shows the UChPT prediction compared to the data. (Right) Triangle diagram dominating the decay $B^- \to D^+\pi^-\pi^-$ via intermediate ρD states. Figures taken from [16].

What makes the agreement between UChPT and data even more apparent is the phase motion shown in Fig. 2.3, where the UChPT prediction aligns significantly better with the data than a conventional BW parametrization[1]. This is quantitatively supported by the fit quality, where the UChPT fit yields a p-value of 0.1, compared to 3×10^{-5} for the BW description [16]. It is important to note, however,

[1] Note that the upper and lower bounds of these curves in the plot are not independent but correlated, forming a theoretical envelope (shaded region) that captures the range of possible phase motions. The short-dashed green curve defines the lower edge of this envelope and best describes the low-energy region, whereas the long-dashed green curve describes the high-energy regime more accurately. All other curves interpolate between these two extremes.

that the extracted phase δ_0 does not directly correspond to the $D\pi$ scattering phase shift. Rather, it represents the phase of the amplitude relative to the $D\pi$ system in this particular decay process. The observed amplitude is dominated by the triangle diagram shown in Fig. 2.3, which introduces its own phase motion. As such, this implies theoretical uncertainties in the extraction of the $D\pi$ phase from such hadronic decays.
Moreover this underlines the advantage of studying the semi-leptonic $B \to D^{**}\ell\nu$ decays, which provides more direct and model-independent access to the $D\pi$ phase through Watson's theorem [18].

2.3 Semi-leptonic $B \to D^{**}\ell\nu_\ell$ decays

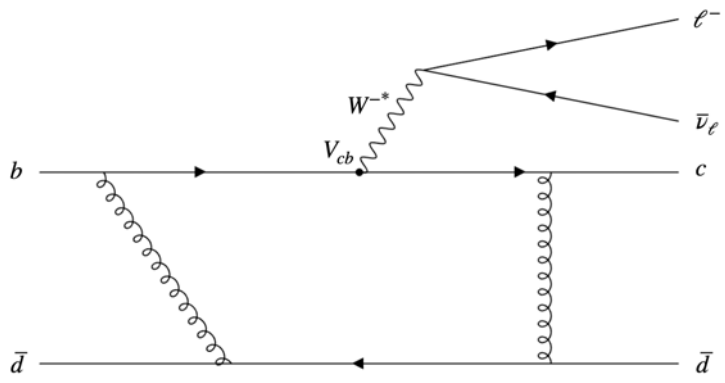


Figure 2.4: Feynman diagram of the $\overline{B}^0 \to D^{**+}\ell^-\bar{\nu}_\ell$ decay.

The effective Lagrangian of $b \to c\ell\nu$ transitions is given in the standard model of particle physics (SM) by [19, 20]

$$\mathcal{L}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{cb}(\bar{c}\gamma_\mu P_L b)(\bar{\nu}_\ell \gamma^\mu P_L \ell) + \text{h.c.} \tag{2.4}$$

with Fermis constant $G_F^{-1} = 8m_W^2/(\sqrt{2}g_2^2) = \sqrt{2}v^2$, where $v \approx 246.22\,\text{Gev}$ [4] denotes the Higgs field vacuum expectation value, g_2 the $SU(2)$ weak coupling constant and m_W the mass of the W boson. $P_{L,R} = (1 \mp \gamma^5)$ are the projection operators and and $\overline{c}(= c^\dagger\gamma^0)$, b, ℓ and $\overline{\nu}_\ell(= \nu_\ell^\dagger\gamma^0)$ the corresponding quark and lepton fields. V_{cb} stands for the quark-mixing Cabibbo-Kobayashi-Masakawa (CKM) matrix element.
The matrix element for the $\overline{B}^0 \to D^{**+}\ell\overline{\nu}_\ell$ decay can be written as

$$\mathcal{M}(\overline{B}^0 \to D^{**+}\ell^-\bar{\nu}_\ell) = -i\frac{G_F}{\sqrt{2}} V_{cb} L^\mu H_\mu \tag{2.5}$$

where higher-order electro-weak corrections are neglected. Therefore the matrix element can be factorised in leptonic L_μ and hadronic current H^μ, describing the hadronic and leptonic part of the decay. While the leptonic part of this decay is fully understood and takes the form

$$L^\mu = \bar{\ell}\gamma^\mu(1-\gamma_5)\nu_\ell, \tag{2.6}$$

the hadronic matrix element is less well understood and the predominant theory uncertainty arises from

its description [20]. The hadronic current is given by

$$H_\mu = \langle D^{**+}(p')|\bar{c}\gamma_\mu(1-\gamma_5)b|\overline{B}^0(p)\rangle. \tag{2.7}$$

This current can be interpreted as a combination of a vector and an axial-vector current, often referred to as a $V-A$ structure, reflecting the chiral nature of the weak interaction (with $V^\mu = \bar{c}\gamma^\mu b$ and $A^\mu = \bar{c}\gamma^\mu\gamma_5 b$).
The hadronic matrix element involves non-perturbative quantum chromodynamics (QCD) effects and, therefore, cannot generally be computed using perturbation theory. In other words, it is not possible to start from the quark-level diagram and systematically include higher-order corrections in powers of the strong coupling constant, α_s, since $a_s > 1$ in our regime [20].
The non-perturbative nature of the hadronic matrix element implies that it must be parameterised in terms of a finite set of scalar functions known as form factors. These form factors encapsulate the QCD dynamics and depend only on kinematic invariants such as q^2, the squared momentum transfer between the initial and final hadronic states. While the Lorentz structure of the matrix element is determined by the symmetry, the form factors themselves must be determined either from theory or from experiment. There are different theoretical frameworks to do so [20]:

- **Heavy Quark Effective Theory (HQET)** utilises the heavy-quark symmetry arising in the limit $m_{b,c} \to \infty$ and provides systematic expansions in $1/m_Q$ and α_s. The most widely used description of $B \to D^{**}\ell\nu$ decays within HQET is the Leibovich–Ligeti–Stewart–Wise (LLSW) parameterization [10]. In the heavy-quark limit, certain form factors vanish at zero recoil ($w = 1$), where the final-state meson is at rest in the B frame. As the allowed kinematic range is narrow ($1 < w \lesssim 1.32$), $O(\Lambda_{\mathrm{QCD}}/m_{b,c})$ corrections become particularly important. The LLSW framework expands the decay amplitudes around $w = 1$, expressing the Isgur–Wise function $\tau(w)$ as
$$\tau(w) = \tau(1)\left[1 + \tau'(w-1) + O((w-1)^2)\right],$$
with $\tau(1)$ and its slope τ' as free parameters to be determined from experiment. Additionally, subleading HQET corrections introduce symmetry-breaking contributions through new form factors such as $\eta_{1,2}$ and $\tau_{1,2,3}$. The original LLSW study provides differential decay rates as functions of w and the angular variables. The Bernlochner, Ligeti, Robinson (BLR) approach [2] builds upon this framework by including higher-order QCD corrections (such as α_s terms) and extending the formalism to accommodate potential non-SM contributions. In both the LLSW and BLR approaches, one can analytically integrate over the angular variables to obtain the single differential decay rate $d\Gamma/dw$.

- **Lattice QCD (LQCD)** enables non-perturbative first-principles calculations, which are particularly reliable at high q^2, although the coverage and accessible decay channels are limited.

- **QCD summation rules** and **Light Cone Summation Rules (LCSR)** provide semi-analytical estimates at low to medium q^2 based on operator-product expansion and dispersion relations.

- **Dispersive methods** use analyticity and unitarity to constrain form factors in a model-independent way, leading to convergent series expansions such as the Boyd, Grinstein and Lebed (BGL) [21] parameterisation.

While these approaches are well established for exclusive decays with single-hadron final states such as $B \to D^{(*)}\ell\nu$, more recent developments have extended these techniques to multi-hadron final states. In particular, a model-independent description of $B \to D\pi\ell\nu$ decays has been proposed that generalises the BGL formalism to semileptonic decays with broad hadronic resonances and coupled channels. This framework enables data-driven form factor extractions with robust, systematically improvable uncertainties and marks a major step forward in the precision study of semi-leptonic transitions involving states such as the D^{**}.

2.3.1 A model-independent description of $B \to D\pi\ell\nu_\ell$ decays

The description of the model-independent formalism of $B \to D\pi\ell\nu_\ell$ decays is taken from Reference [1]. Motivated by $B \to \pi\pi\ell\nu$ decays [22, 23] one involves a partial-wave decomposition in the hadronic helicity angle $\cos\theta_V$ and finds for the hadronic matrix elements

$$\langle D(p_D)\pi(p_\pi)|V^\mu|B(p_B)\rangle = i\epsilon^{\mu\nu\rho\sigma} p^\rho_{D\pi} p^\sigma_B \sum_{l>0} L^{(l),\nu}\, g_l(q^2, M^2_{D\pi}), \tag{2.8}$$

$$\begin{aligned}\langle D(p_D)\pi(p_\pi)|A^\mu|B(p_B)\rangle = {}& \frac{1}{2}\sum_{l>0}\left(L^{(l),\mu} + \frac{4}{\lambda_B}\left[(p_B\cdot p_{D\pi})q^\mu - (p_{D\pi}\cdot q)p^\mu_B\right] L^{(l),\nu} q_\nu\right) f_l(q^2, M^2_{D\pi}) \\ & + \frac{M_{D\pi}(M^2_B - M^2_{D\pi})}{\lambda_B}\left[(p_B + p_{D\pi})^\mu - \frac{M^2_B - M^2_{D\pi}}{q^2}q^\mu\right]\sum_{l>0} L^{(l),\nu} q_\nu \mathcal{F}_{1,l}(q^2, M^2_{D\pi}) \\ & + M_{D\pi}\frac{q^\mu}{q^2}\sum_{l>0} L^{(l)} q_\nu \mathcal{F}_{2,l}(q^2, M^2_{D\pi}) \\ & + \left[(p_B + p_{D\pi})^\mu - \frac{M^2_B - M^2_{D\pi}}{q^2}q^\mu\right] f_+(q^2, M^2_{D\pi}) \\ & + \frac{M^2_B - M^2_{D\pi}}{q^2} q^\mu f_0(q^2, M^2_{D\pi}).\end{aligned} \tag{2.9}$$

where the groups $\{g_l, f_0, \mathcal{F}_{2,l}, f_+, f_l, \mathcal{F}_{1,l}\}$ denotes the form factors and λ_B the threshold factor which can be calculated by

$$\lambda_B = M^4_B + M^4_{D\pi} + q^4 - 2(M^2_B M^2_{D\pi} + M^2_{D\pi} q^2 + q^2 M^2_B).$$

The vector $L^{(l)}$ is associated with the angular momentum of the hadron system in its final state in the B-meson rest frame and is given by

$$\begin{aligned} L^{(l)}_\mu\, q^\mu &= M_B W^l P_l(\cos\theta_V) \\ L^{(l)}_\mu\, p^\mu_{D\pi} &= 0 \end{aligned} \tag{2.10}$$

with $W = |\vec{q}||\vec{p}_D/(M_B M_{D\pi})|$ and the Legendre polynomials P_l.
After angular integration and omitting helicity-suppressed terms one finds for the double differential

decay rate for massless leptons

$$\frac{\mathrm{d}^2\Gamma}{\mathrm{d}M_{D\pi}^2\,\mathrm{d}q^2} = \frac{G_F^2|V_{cb}|^2}{(4\pi)^5} M_B \left(W \frac{\lambda_B}{M_B^2} \frac{4|f_+|^2}{3} + M_{D\pi}^2 \sum_{l>0} \frac{4W^{2l+1}}{3(2l+1)} \left[(M_B^2 - M_{D\pi}^2)^2 \frac{|\mathcal{F}_{1,l}|^2}{\lambda_B} + \frac{(l+1)}{l} q^2 \left(|g_l|^2 + \frac{|f_l|^2}{\lambda_B} \right) \right] \right). \tag{2.11}$$

The fivefold differential decay rate for $B \to D\pi\ell\nu$, including (next to q^2 and $M_{D\pi}$) the helicity angle of the charged lepton $\cos\theta_l$, the hadronic helicity angle $\cos\theta_V$ and the azimuthal angle between the $\ell\nu$ and $D\pi$ planes χ, can be found in the supplemental material of Reference [1].

To constrain the shape of the form factors in a model-independent way, one uses analyticity and unitarity. These allow a rigorous application of dispersion relations, leading to bounds on the form factor coefficients in their z-expansion, with

$$z(q^2, q_0) = \frac{q_0^2 - q^2}{\left(\sqrt{q_+^2 - q^2} + \sqrt{q_+^2 - q_0^2} \right)^2}$$

where $q_+^2 = (M_B + M_D + m_\pi)^2$.

The expansion ensures convergence in the physical region of q^2 and is structured such that only a few coefficients need to be fitted. The form factors are factorized into a leptonic and hadronic part,

$$f_l(q^2, M_{D\pi}^2) = \hat{f}_l(q^2) \cdot h_l(M_{D\pi}^2),$$

where $\hat{f}_l(q^2)$ encodes the leptonic dependence and is approximately independent of $M_{D\pi}$. One then finds

$$\hat{f}_l(q^2, M_{D\pi}^2) = \frac{1}{\phi_l(q^2) B_f(q^2)} \sum_k a_k^{(l)} z(q^2)^k,$$

where B_f and ϕ_l are known Blaschke and outer functions, and the unitarity constraint takes the form

$$\sum_k |a_k^{(l)}|^2 < 1.$$

Importantly, the unitarity bounds are diagonal and apply separately to three form factor groups: $\{f_0, \mathcal{F}_{2,l}\}$, $\{f_+, f_l, \mathcal{F}_{1,l}\}$, and $\{g_l\}$. They cannot be mixed arbitrarily in the bound, which simplifies the fitting procedure while maintaining theoretical control.

For narrow resonances such as the D^* and D_2^*, a relativistic Breit-Wigner lineshape with Blatt-Weisskopf damping factors is used to describe the contribution of the corresponding partial waves. These resonances dominate the vector and tensor components of the decay and are well separated from the broad nonresonant background.

In contrast, the S-wave ($l = 0$) is significantly affected by nonperturbative strong interactions and receives contributions from multiple coupled hadronic channels. This is handled using a coupled-channel treatment of $D\pi$, $D\eta$, and D_sK scattering, encoded in the T-matrix. The scattering amplitudes are modeled with the previous discussed unitarized chiral perturbation theory (UChPT).

The form factor for the S-wave is then expressed via a Muskhelishvili–Omnès solution involving the T-matrix, which resums final-state interactions in a model-independent way and ensures correct analytic

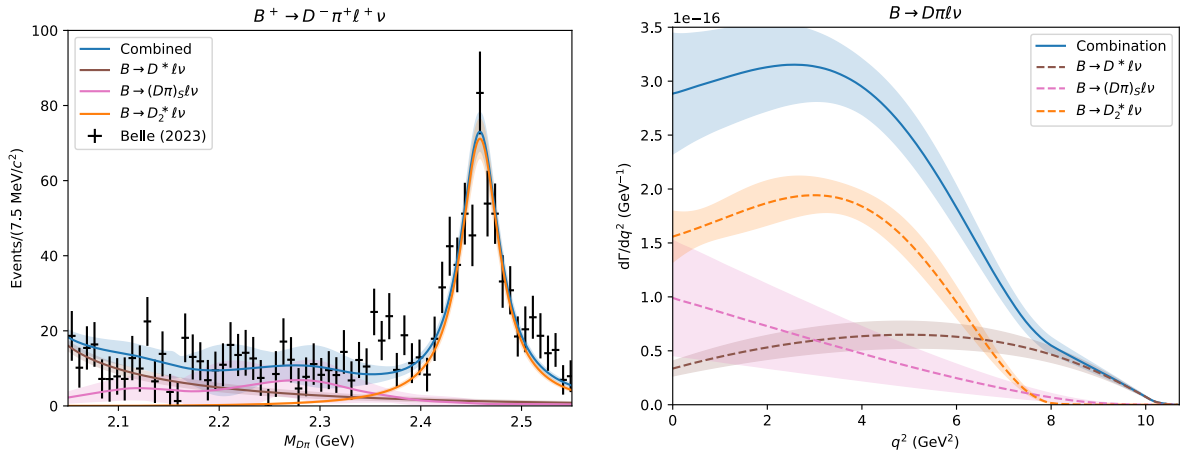


Figure 2.5: (Left) Fit of measured $M_{D\pi}$-spectrum with hadronic tagged Belle data from [24]. (Right) Predicted partial-wave decomposition of the $B \to D\pi\ell\nu$ q^2-spectrum from the $M_{D\pi}$ fit. Figures taken from [1]

behavior:

$$f^{(S)}(q^2, M_{D\pi}^2) = \Omega(M_{D\pi}^2) \cdot P(q^2, M_{D\pi}^2),$$

where Ω is the Omnès matrix determined from the T-matrix and P is a polynomial in q^2 and $M_{D\pi}^2$ with coefficients constrained by unitarity.

This model was fitted to hadronic tagged Belle data [24] for the $M_{D\pi}$ invariant mass spectrum, as shown in Fig. 2.5. The fitted form factors were then used to compute the partial-wave decomposition of the $B \to D\pi\ell\nu$ differential decay rate in q^2.
The extracted branching fractions are

$$\mathrm{Br}(B \to D_2^*(\to D\pi^\pm)\ell\nu) = (1.90 \pm 0.11) \times 10^{-3}, \tag{2.12}$$

$$\mathrm{Br}(B \to (D\pi)_S \ell\nu) = (1.03 \pm 0.27) \times 10^{-3}. \tag{2.13}$$

for the D- and S-wave channel correspondingly.
The D-wave branching fraction exceeds Belle's result in Reference [24], as their smooth nonresonant background model overlaps with the D_2^* peak. In the present approach, the S-wave and D^* components are negligible near the resonance. The S-wave result is consistent with Ref. [25] but smaller than values typically assumed in experimental analyses.

With this analysis, we aim to contribute to and improve the statistics of the invariant $D\pi$ mass spectrum, and, potentially with a measurement of q^2 and the helicity angles, provide additional experimental constraints on the fit, thereby reducing the theoretical uncertainties.

2.4 Overview of Kinematic Variables

As mentioned above the dynamics of semi-leptonic B decays are fully described by a set of Lorentz-invariant kinematic variables q^2, w, $\cos\theta_V$, $\cos\theta_\ell$ and χ, which encode both the momentum transfer and the angular distributions of the decay products and serve as inputs for form factor extractions and

partial wave analyses. In this section, we quantify these kinematic variables.

The squared momentum transfer to the lepton-neutrino system is given by

$$q^2 = (p_B - p_{D^{*(*)}})^2 = (p_\ell + p_\nu)^2, \tag{2.14}$$

where p_B, $p_{D^{*(*)}}$, p_ℓ, and p_ν are the four-momenta of the B meson, the charmed meson system (either D^*, or an excited D^{**}), the charged lepton, and the neutrino, respectively. This variable determines the momentum exchanged by the virtual W^* boson.

The recoil parameter w corresponds to the product of the four-velocities of the initial and final hadrons. The recoil parameter w is given by

$$w = v_B \cdot v_{D^{*(*)}} = \frac{m_B^2 + m_{D^{*(*)}}^2 - q^2}{2 m_B m_{D^{*(*)}}}, \tag{2.15}$$

The value $w = 1$ corresponds to zero recoil of the final-state hadron, and larger values correspond to higher momentum transfer. Depending on the theoretical framework, form factor parameterizations are typically expressed either as functions of q^2 or w, with w being particularly convenient in HQET-based approaches.

In addition to q^2 and w, a complete description of the decay also requires three angular variables. These are defined as illustrated in Figure 2.6 and are particularly relevant for angular analyses and partial-wave decompositions. They are defined as follows [26]:

- θ_ℓ: The angle between the flight direction of the lepton in the virtual W^* rest frame and the direction of the virtual W^* in the B rest frame.
- θ_V: The angle between the flight direction of the D meson in the D^* rest frame and the direction of the D^* in the B rest frame.
- χ: The tilting or dihedral angle between the decay plane defined by the $D^* \to D\pi$ decay and the plane defined by the $W^* \to \ell\nu$ decay.

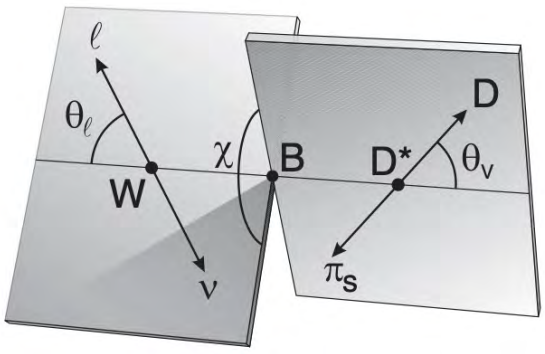


Figure 2.6: Illustration of the three kinematic angles θ_V, θ_ℓ, and χ, exemplarily shown for a $B \to D^* \ell\nu$ decay. Figure adapted from [26].

CHAPTER 3

Experimental Setup

This chapter provides a general summary of the experimental setup that was used to perform the analysis. It begins with a description of the SuperKEKB accelerator, the high-luminosity electron-positron accelerator that serves as the basis for data collection at Belle II. It then describes the main features of the Belle II detector, which is designed to take precise measurements of B meson and other flavour sector decays, for example D meson and τ lepton decays.
The use of Monte Carlo (MC) simulations plays a crucial role in the analysis, both for the estimation of efficiencies and for the modelling of signal and background processes. At this stage, the analysis is only performed on the centrally produced MC samples. In this chapter, we thus describe the MC samples used for our study. Furthermore, the chapter outlines the experimental data set and presents the tagging techniques employed at Belle II.

Both, experimental and simulated data are reconstructed and analyzed using the Belle II analysis software framework (basf2) [27], primarily written in `C++` and `Python`.
A more detailed summary of the experimental setup can be found in references [28], [29], and [30]. This thesis only discusses the basic foundations out of it.

3.1 SuperKEKB accelerator

The SuperKEKB accelerator, located in Tsukuba, Japan, is an asymmetric electron-positron collider and the successor to the KEKB accelerator. A schematic view of SuperKEKB is provided in Fig. 3.1. It operates at a center-of-mass energy of $\sqrt{s} = 10.58\,\text{GeV}$, corresponding to the mass of the $\Upsilon(4S)$ resonance, which decays almost exclusively into pairs of B^+B^- and $B^0\bar{B}^0$ mesons. This makes SuperKEKB a *B factory*, enabling high-precision studies of B-meson decays and CP violation. The collider operates with asymmetric beam energies of 7 GeV for electrons and 4 GeV for positrons, resulting in a Lorentz boost of $\beta\gamma = 0.28$ of the center-of-mass frame in respect to the laboratory frame. Since the mass of the $\Upsilon(4S)$ is close to twice the B meson mass, the $B\bar{B}$ pair is produced nearly at rest in the center-of-mass frame, but appears boosted in the lab frame. This boost extends the average B-meson flight length to around $130\,\mu\text{m}$, allowing spatial separation between the production and decay vertices of the two B mesons, crucial for time-dependent CP studies. Electrons are generated using a photo-cathode RF gun, while positrons are produced by striking a tungsten target. Both beams are accelerated in a linear accelerator before injection into separate storage rings and are then collided at the interaction point, surrounded by

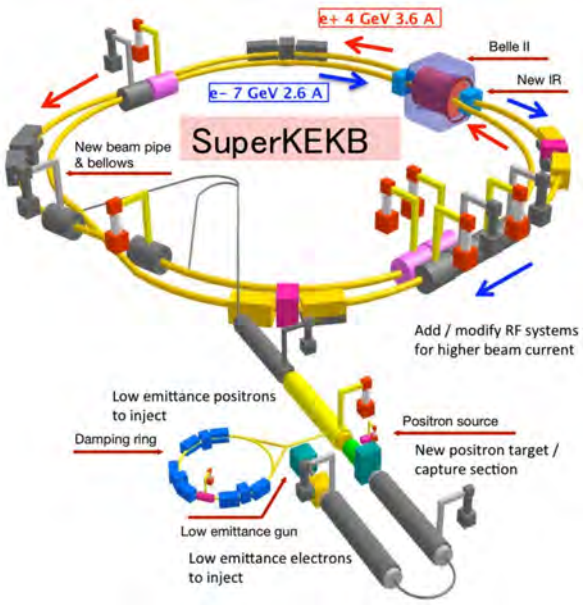


Figure 3.1: Schematic view of the SuperKEKB accelerator complex. Figure from Reference [31]

the Belle II detector. Operating at an e^+e^- collider offers advantages such as low background, absence of pile-up, and a well-defined initial state. These features facilitate analyses involving missing energy and full-event reconstruction. In addition to running at the $\Upsilon(4S)$, SuperKEKB also collects off-resonance data to study backgrounds from light and charm quark production. The luminosity, $\mathcal{L}$, quantifies the rate at which beam particles interact per unit of time and area. It depends on the number of particles per bunch, the number of bunches, and their collision frequency. SuperKEKB aims for a luminosity of $6 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, which is 30 times higher than that of KEKB. The last published luminosity record at SuperKEKB, $4.71 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, was reached on 22 June 2022 [32].

3.2 Belle II detector

The Belle II detector, located at the interaction point (IP) of SuperKEKB, is a cylindrical, general-purpose detector designed to study the physics of B mesons. It covers almost the entire solid angle, comprising several nested sub-detectors arranged in a barrel structure and sealed by front and rear end caps. As an upgraded version of its predecessor, Belle [33], Belle II was built to operate at up to 40 times higher luminosity, with a background level increased by a factor of 10–20 [30].

Most of its subsystems are embedded in a uniform 1.5 T magnetic field, which is generated by a superconducting solenoid. These subsystems are optimised to identify particles produced in positron-electron collisions, and to measure their energies and momenta across a range from tens of MeV to several GeV.

The detector's right-handed coordinate system has its origin at the IP. The z-axis aligns with the direction of the electron beam (forward), the y-axis points vertically upwards and the x-axis extends radially outwards. Due to the asymmetric beam energies at SuperKEKB, Belle II exhibits a forward–backward asymmetry. Cylindrical coordinates (r, Φ, z) are typically used, where the polar angle θ is measured

from the z-axis and defines detector acceptance. The subdetectors are described in the following section and are illustrated in Fig. 3.2.

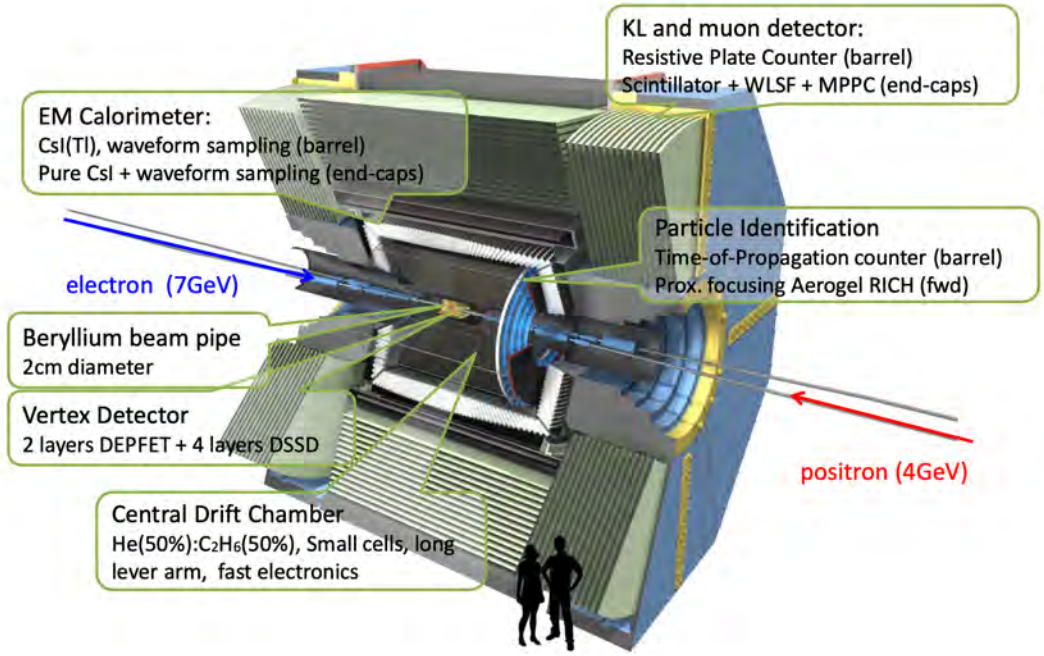


Figure 3.2: Schematic overview of the Belle II detector with a closeup indicating all the different sub detectors on the right. Figures from Reference [34].

Subdetectors of Belle II

All quantified values in the following section were explicitly taken from [28].

Vertex Detector (VXD) The innermost subsystem is the Vertex Detector (VXD), which provides precise measurements of particle trajectories near the IP and is important for tracking low-momentum particles. This is a crucial feature for soft pion reconstruction. It consists of two components: the PiXel Detector (PXD) and the Silicon Vertex Detector (SVD).
The PXD, constructed from DEPFET silicon sensors with pixel sizes as small as $50 \times 55\ \mu\text{m}^2$, forms the inner part of the VXD. It consists of two layers located at radii of 14 mm and 22 mm from the IP and provides extremely high spatial resolution, essential for reconstructing decay vertices.
Surrounding the PXD is the SVD, made of four layers of double-sided silicon microstrip sensors, extending the vertex detector radius up to 140 mm.

Central Drift Chamber (CDC) Externally to the VXD lies the Central Drift Chamber (CDC), a He-C_2H_6 gas-filled detector also used for tracking charged particles and measuring their momenta. Charged particles ionize the gas as they pass through. The free electrons drift to sense wires, producing signals. The curvature of these tracks in the magnetic field provides information about the momentum and charge of the particles. Together with the VXD, the CDC forms the tracking system of the Belle II detector.

Particle Identification: TOP and ARICH Particle identification in Belle II is handled by two Cherenkov-based detectors: the Time-of-Propagation (TOP) counter in the barrel region and the Aerogel Ring-Imaging Cherenkov (ARICH) detector in the forward region.
The TOP detector consists of quartz bars that detect Cherenkov photons produced by charged particles. The high refractive index of quartz causes Cherenkov light to undergo total internal reflection, allowing the Cherenkov angle, which depends on the particle velocity, to be determined from the photon's travel time between its emission point and the photodetectors.
The ARICH uses aerogel as the radiator medium and images Cherenkov rings directly. Charged particles passing through the aerogel emit Cherenkov light forming a ring, whose radius depends on the Cherenkov angle and thus also the particle's velocity. The TOP and ARICH information combined with momentum measurements from tracking allows the determination of the particle mass and thus species the particle (identification). Together, these detectors provides efficient separation of pions, kaons, protons, electron and muons.

Electromagnetic Calorimeter (ECL) The electromagnetic calorimeter (ECL) is designed to measure the energy of electrons and photons in electromagnetic showers. Consisting of over 8,000 thallium-doped caesium iodide (CsI(Tl)) scintillation crystals, it is arranged in the barrel and endcap regions. When particles interact with the crystals, scintillation light is emitted whose intensity is proportional to the deposited energy. The ECL complements the tracking detectors in particle identification; for instance, it can distinguish electrons showering in the calorimeter from muons, which typically pass through with minimal energy loss.

KLong and Muon Detector (KLM) The outermost subsystem is the KLong and Muon (KLM) detector. It consists of alternating layers of iron and active detector elements, resistive plate chambers and scintillators. surrounding the ECL. The KLM identifies muons based on their penetration depth and reconstructs neutral long-lived kaons (K_L^0) through hadronic interactions in the iron. Muons, being minimally ionizing, typically traverse all inner subdetectors and are distinguished from hadrons and electrons by their unique signatures in the KLM.

More details about the detector and its single components can be found in the Technical Design Report (TDR) [29].

3.3 Monte Carlo simulation

Monte Carlo simulations are essential for the interpretation of data from high-energy physics experiments such as Belle II. They reproduce both the underlying physical processes and the detector response, allowing comparisons with real data and guiding signal extraction strategies.
Simulation begins with event generation, where particle interactions are modeled based on theoretical inputs. At Belle II, `EvtGen` [3] is primarily used to simulate physics processes, including B and D decays. The event generator operates based on input information about the produced particles, including their masses, spins, decay widths, lifetimes, and decay channels, the corresponding branching fractions and theoretical decay models [30]. Decays which are not included in `EvtGen` are handled by `PYTHIA` [35], which also simulates quark hadronization and continuum $e^+e^- \to q\bar{q}$ processes, in conjunction with `KKMC` [36]. `PHOTOS` [37] adds final-state photon radiation.

Untagged	
MC kind	luminosity
MCrd	$14.96\,\text{fb}^{-1}$
SL tagged	
MC kind	luminosity
MCrd	$1.443\,\text{ab}^{-1}$
MCri	$2.8\,\text{ab}^{-1}$

Table 3.1: Summary of the used MC samples and their corresponding luminosity's provided by the Belle II collaboration. Values are taken from [40] and [43].

The second stage, detector simulation, uses `Geant4` [38], which models the interactions of generated particles with the detector material, including effects like ionization, Bremsstrahlung, and Cherenkov radiation. These interactions are converted into corresponding signals in each subdetector.
MC simulations provide access to both generator-level truth and reconstructed-level observables, enabling efficiency calculations, background estimation, and validation of selection strategies for instance in the signal extraction.

Belle II provides two types of MC simulation samples: run-independent (ri) and run-dependent (rd) [39]. Run-independent MC is generated using averaged detector conditions and generic beam-background overlays, independent of specific data-taking runs. These samples are primarily used for fast validation and systematic studies, as they are less resource-intensive to produce.
In contrast, run-dependent MC samples are generated with run-by-run accuracy. They incorporate realistic beam backgrounds, trigger configurations, and detector conditions that match each specific data-taking period. This leads to a significantly improved agreement between data and simulation, making run-dependent MC the preferred choice for precision measurements. However, this accuracy comes at the cost of increased computational complexity and resource usage during production.

For the untagged analysis we use MC run-dependent samples with an integrated luminosity of $\mathcal{L} = 14.96\,\text{fb}^{-1}$ [40].
In the SL tagged analysis, we use the *SL FEI Skim* MC run-dependent and run-independent samples with an integrated luminosity of approximately $1.443\,\text{ab}^{-1}$ and $2.8\,\text{ab}^{-1}$ respectively [41]. Skims are reduced versions of data and MC samples that are created by applying analysis specific selection criteria. In our case, the selection criteria associated with the SL FEI, i.e. selections for the tagged B, are applied. More details about the selection criteria and the FEI algorithm will be explained later in Section 5.2 and Section 3.5 respectively. The advantage of skimmed samples are that they only contain analysis relevant events and are therefore about 90 % [42] smaller than the original datasets. For Belle II, where an integrated luminosity up $50\,\text{ab}^{-1}$ of data is expected [42], skimming is essential for practical data processing. Both samples were scaled to the integrated luminosity available during the Long Shutdown 1 (LS1), so that the expected number of events matches the yield anticipated from experimental data.
All samples (including the skimmed samples) are provided by the collaboration.
A summary of the used MC samples and corresponding luminosity's can be found in Table 3.1.

Furthermore, dedicated run-independent signal MC samples were generated for the specific purpose of this analysis. These samples are fully integrated into the `basf2` framework and were produced using the `EvtGen` event generator. For the untagged analysis, we generated one million events for each B decay

mode, evenly split between the electron and muon channels, as well as among the relevant resonant and non-resonant contributions. The D^* and D modes were generated exclusively, including only the specific signal $D^{(*)}$ decay modes discussed in Section 5.3. For visualisation purposes, the branching fraction ratios between the resonances were manually adjusted to reflect the expected physical proportions in the corresponding distributions.
In the SL tagged analysis, one million events were generated per resonance in the signal B decay modes, 500 000 for electron and 500 000 for muon channels. As before, the relative weights of the resonances were manually adjusted in the plots to match the expected branching fraction ratios.

3.3.1 Modelling of semileptonic B meson decays

Semileptonic $B \to X\ell\nu$ decays are described using form factor parameterisations that model their kinematic distributions. In particular, $B \to D^{(*)}\ell\nu$ decays are modeled using the BGL parameterisation [21], with parameters fitted to experimental data (in Reference [44] and [45]).
Decays involving orbitally excited charmed mesons (D^{**}) are modeled using the BLR form factor model [2], already roughly discussed in Section 2.3. The BLR parameterisation relies on the narrow-width approximation, as any other form factor model for semi-leptonic B decays so far. Altough it does not apply to broad resonances such as the D_0^* and D_1', the BLR model is the most accurate model available. This analysis is anticipated to contribute to a more accurate and robust parameterisation for the broad D^{**} states.

3.3.2 Topological Analysis with `TopoAna`

`TopoAna` [46] is a tool for analysing physical processes that can be used to extract and classify generator-level information from MC samples of high-energy physics experiments. The software was developed in `C++` with `ROOT` and LaTeXsupport and offers customised algorithms for component analysis and signal identification. It accurately characterises all the single decay modes in individual events based on generator information and outputs the results in ROOT files, plain text, LaTeXor PDF format. `TopoAna` is frame-independent and has already been adopted by experiments such as BESIII, Belle and Belle II.
TopoAna was used in our analysis to better understand and analyse the source of the peaking background, as well as to classify self cross feed events and tag cross feeed events. An explanation of what we refer as self- or tag cross feed events will be provided subsequently in the description of the single event categories in Section 4.3.

3.4 Experimental data sample

Between 2019 and 2022, prior to the LS1, the Belle II detector collected a total integrated luminosity of 427.87 fb^{-1} [43]. This period marked the first phase of data collection at Belle II before the experiment paused in summer 2022 for major upgrades to both the detector and the SuperKEKB accelerator.
Out of the total recorded luminosity, 365.37 fb^{-1} were collected at the $\Upsilon(4S)$ resonance, which is of primary interest for B physics. An additional 42.74 fb^{-1} were recorded at about 60 MeV below the resonance energy, forming the so-called *off-resonance* sample. These off-resonance data are crucial for studying non-resonant backgrounds such as $e^+e^- \to q\bar{q}$ (with $q = u, d, s, c$), which do not involve $B\overline{B}$ production. Moreover, 19.76 fb^{-1} were collected at energies slightly above the $\Upsilon(4S)$, enabling studies in quarkonium and dark sector physics.

In both, the untagged and SL tagged analysis, the continuum fraction after our applied selection cuts, is quite small and therefore we do not need to calibrate it on off-resonance data. For the SL tagged analysis we look in particular into real Belle II data in Section 6.2.

3.5 Tagging at Belle II and the Full Event Interpretation (FEI)

The decay of a $\Upsilon(4S)$ resonance produces a pair of B-mesons. While one B-meson decays into the mode of interest (denoted as B_{sig}), the second B-meson, referred to as B_{tag}, can be treated using different tagging strategies. In exclusive tagging, the B_{tag} is first fully reconstructed in hadronic or semileptonic decay modes. Using the known initial state and the reconstructed tag-side kinematics, one can then infer essential properties of the signal-side B meson, such as its charge, flavour, and momentum, thereby enabling a more constrained and precise analysis of the signal decay.

- **Untagged:** The B_{tag} is not explicitly reconstructed. Instead, the signal-side B (B_{sig}) is reconstructed first, and the rest of the event (ROE) is interpreted as the B_{tag}. This approach retains the highest efficiency, as it implicitly includes all possible decay modes of the tag-side B without applying stringent constraints on its reconstruction.
- **Semileptonic tag (SL tag):** The B_{tag} is reconstructed in a semileptonic decay mode. While this provides partial information about the event, the presence of neutrinos leads to missing energy and incomplete reconstruction.
- **Hadronic tag:** The B_{tag} is reconstructed in fully hadronic decay modes. Although this approach has lower efficiency due to the small branching fractions of hadronic decays, it allows for complete reconstruction of the tag-side B meson. This is particularly useful when the signal B decays to final states with invisible particles, such as neutrinos, as it enables indirect access to the full event information.

An overview of the $\Upsilon(4S)$ decay with the example of a semi-leptonic tagged B-meson is illustrated in Fig. 3.3. The tagged B-meson is reconstructed using an algorithm called the *Full Event Interpretation* (FEI). The FEI reconstructs B_{tag} candidates across thousands of exclusive decay channels through a stepwise hierarchical method [47]. Starting from well-identified stable particles, like pions and kaons, it incrementally builds up more complex particles, such as D mesons, ultimately forming complete B meson candidates. At each stage of this reconstruction, pre-trained multivariate classifiers (MVAs) evaluate how likely a given candidate reflects the true physical particle. These probabilities, combined with kinematic and vertex information, are passed forward through the hierarchy. The outcome is a single *signal probability* for each B meson candidate, which serves as a quantitative criterion for selecting valid candidates during analysis [47].

In the initial stage of this work, an untagged analysis approach was pursued. However, it proved challenging to sufficiently suppress background contributions using this method, and the background rejection procedure became increasingly impractical. To maintain a reasonably high signal efficiency while achieving better background control, the strategy was shifted to a semileptonic (SL) tagged approach.

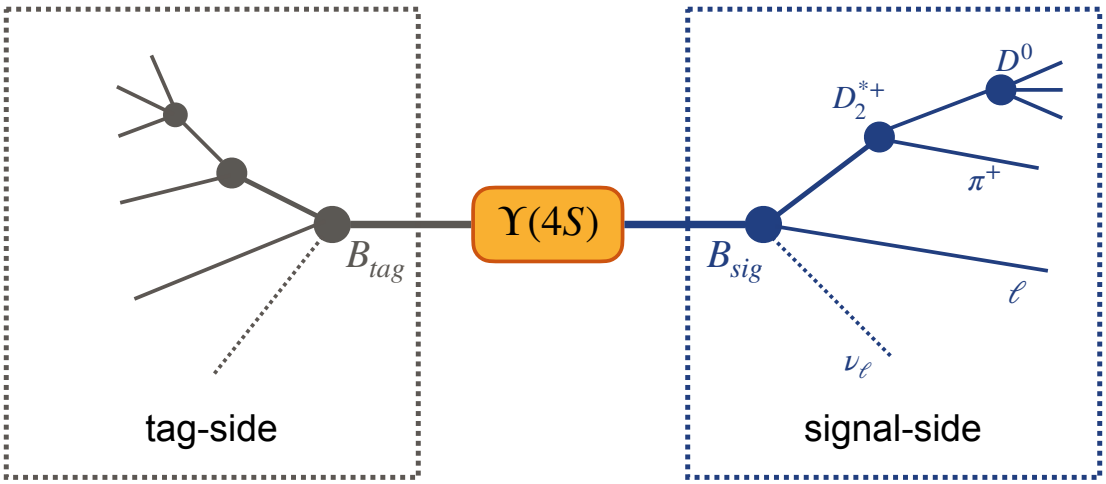


Figure 3.3: Schematic overview of a $\Upsilon(4S)$ decay. On the left a semi-leptonic tag side decay and on the right one of our semi-leptonic signal side decay. Since the two sides overlap spatially within the detector, the assignment of a measured track to either side cannot be determined a priori. Figure motivated by [47].

CHAPTER 4

Untagged event reconstruction and selections

The untagged $\overline{B}^0 \to D^{**+}\ell\nu_\ell$ analysis with $\ell = e$[1] is motivated by a previous untagged $\overline{B} \to D^{*+}(\to D^0\pi^+)\ell\nu_\ell$ analysis with Belle II [48], presented by Chaoyi Lyu. Since Lyu's analysis focused exclusively on the D^* region, he applied a simple selection criterion on $\Delta M \in [0.141, 0.156]$ GeV/c^2 to exclude the D^{**} region, which is of interest in the present study. As a result, the feasibility of using an untagged approach for the D^{**} region remained untested. This analysis represents the first attempt to explore the D^{**} region using an untagged strategy. Following this, we reconstruct the same D^* and D^0 decay modes as in Lyu's analysis and apply the same selection criteria, with minor modifications to certain cuts. However, these selections still result in significant background contamination. To improve the signal-to-background ratio, we attempted to train a BDT and apply a cut on the signal probability output. This step, however, proved to be non-trivial due to the complexity of the background and the limited discriminative power of the BDT given the set of input variables used in our study. As a consequence, the untagged approach was ultimately not pursued further and we shifted our focus to a semi-leptonic tagged strategy as a more viable approach, which will be discussed in Chapter 5.

4.1 Reconstructed decay modes

In this analysis we focus on three different B decay modes (including their charge conjugated modes)

1. $\overline{B}^0 \to \underbrace{D^0\pi^+}_{D^{*+},\, D_0^{*+},\, D_2^{*+}} \quad \ell^-\overline{\nu}_\ell$

   This decay mode is motivated by the untagged Belle II analysis of the channel $\overline{B} \to D^{*+}(\to D^0\pi^+)\ell^-\nu_\ell$ conducted by Chaoyi Lyu. The branching fraction of the semi-leptonic decay $B \to D^*$ is approximately an order of magnitude larger than that of the corresponding $B \to D^{**}$ decays. As a result, the distributions of kinematic variables are generally dominated by contributions from the D^* resonance.

   To better isolate the different resonant contributions, the analysis is sometimes, though not always, performed separately in two regions of the $D^0\pi^+$ invariant mass:

[1] This thesis focuses exclusively on the electron channel; however, the analysis strategy is equally applicable to the muon channel.

- $M(D^0\pi^+) < 2.03\ \mathrm{GeV}/c^2$ (dominated by D^{*+}),
- $M(D^0\pi^+) > 2.03\ \mathrm{GeV}/c^2$ (more sensitive to D^{**+} states, i.e. here D_0^{*+} and D_2^{*+}).

This approach can help to manage the differing kinematic features of the resonances when appropriate.

Furthermore, in the decay $D^{*+} \to D^0\pi^+$, the pion is commonly referred to as a "slow pion" due to the small mass difference between the D^* and D mesons. This limited phase space results in the pion having a low momentum.

2. $B^- \to \underbrace{D^+\pi^-}_{D_0^{*0}, D_2^{*0}}\ \ell^-\overline{\nu}\ell$

   This decay channel has the advantage that it does not include contributions from the D^* resonance. As a result, background from D^* decays, typically arising from a soft pion, is expected to be reduced. Moreover, the kinematic variables and their distributions are no longer dominated by the D^* component, which may allow for a clearer investigation of the excited charm states D_0^{*0} and D_2^{*0}. In addition, the D^+ meson has a longer lifetime, and consequently a longer flight distance, than the D^0. It is expected that this feature may enable better separation of signal and background through the use of vertex information.

3. $B^- \to \underbrace{D^{*+}\pi^-}_{D_1^0,\, D_1'^0,\, D_2^{*0}}\ \ell^-\overline{\nu}_\ell$

   This decay channel includes an additional intermediate resonance, namely $D^{*+}(\to D^0\pi^+)\pi^-$, resulting in a higher number of final-state particles. This increased multiplicity may help suppress combinatorial background, making the overall background potentially easier to manage.

In the further course of this thesis, any reference to the $D\pi$-system includes the combinations $D^0\pi^+$ and $D^+\pi^-$ from the first two decay modes as well as the $D^{*+}\pi^-$-system from the third decay mode.
For D^* we chose the following decay mode

- $D^{*+} \to D^0\pi^+_{\mathrm{slow}}$

and for the corresponding D modes, we chose

- $D^0 \to K^-\pi^+$
- $D^+ \to K^-\pi^+\pi^+$

These exclusive decay chains are chosen due to the relatively high branching ratios, the clear signature and the restricted kinematics of the D^* decay, which allows its selection with high purity [49].
In an untagged analysis, the second B meson is not explicitly reconstructed, as previously discussed in Section 3.5. Consequently, the rest of the event (ROE) comprises all particles that are not associated with the reconstructed signal-side decay. This includes both the second B meson and potential misreconstructed particles originating from the signal B decay.
As an example, an illustration of an event for the first B decay mode is shown in Fig. 4.1.

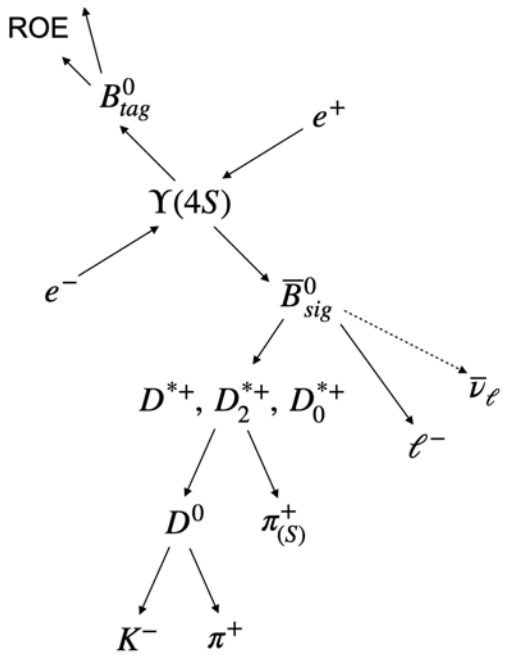


Figure 4.1: Schematic overview of an untagged $\overline{B}^0 \to D^{*(*)+}\ell^-\overline{\nu}_\ell$ event. This figure is adapted from Reference [50].

4.2 Reconstruction and selections on B_{sig}

First the events must pass the High Level Trigger (HLT) [51] [2].
To ensure that the final-state particles (FSP) are well-reconstructed within the detector, loose selection criteria are applied to their associated tracks. These include particle identification requirements as well as geometrical constraints to ensure the tracks originate near the interaction point (IP). The quality of the track origin is assessed using the variables `dr` and `dz`, which describe the point of closest approach (POCA) to the IP. Specifically, `dr` corresponds to the transverse distance, and `dz` to the longitudinal (z-axis) component of the POCA. Tracks are required to satisfy $|\text{dr}| < 0.5\,\text{cm}$ and $|\text{dz}| < 2\,\text{cm}$. Furthermore, the polar angle θ of the track must lie within the acceptance region of the CDC, defined as $17° < \theta < 150°$. For the lepton reconstruction we use the `stdCharged.stdLep()` function, provided by the `basf2` software with arguments recommended by the lepton IDentification (ID) group in Reference [53]. `stdCharged.stdLep()` takes the reconstructed charged tracks in an event, keeps only those that match the chosen lepton-ID working point, collects them in a list, and tags each candidate with the corresponding identification efficiency and fake-rate scale factors (including their statistical and systematic uncertainties) fetched from the calibration database according to the chosen input global tag. Beforehands we already apply particle identification[3] cuts larger than 99% for electrons and 90% for muons. In addition, electron

[2] The HLT at Belle II is a computing cluster that performs full event reconstruction (excluding PXD hits) in real time using the `basf2` software, filtering and categorizing events to reduce data storage by over 60% [52]. It operates in filtering or monitoring mode and produces fast skims such as the hadronic HLT skim (hlt_hadron), which selects hadronic events.

[3] Electron candidates are selected using particle identification likelihoods that incorporate information from the CDC, ARICH, ECL, and KLM detectors. For muon identification, the TOP detector is additionally used alongside the CDC, ARICH, ECL, and KLM subsystems.

candidates are corrected for Bremsstrahlung energy losses using the `correctBrems()` function provided in the `basf2` framework. This correction requires a list of photon candidates, which must satisfy region-specific energy thresholds to suppress background contributions. The function then adds the four-momenta of the photons found close to the electron's trajectory to the electron candidate. This process recovers the radiated energy and stores links to the used photons. In this way, it recovers the energy radiated through bremsstrahlung and produces an updated electron candidate with a momentum closer to the true value. Specifically, photons must have an energy greater than 75 MeV if located in the forward endcap of the ECL, above 50 MeV if in the barrel region, and above 100 MeV if in the backward endcap [54]. Furthermore we add a momentum cut in the Center of MaSs frame (CMS) for both leptons, specifically to pass $1.2\,\mathrm{GeV/c^2} <= p_\ell^{\mathrm{CMS}} <= 2.4\,\mathrm{GeV/c^2}$.
For the charged final state hadrons, $\pi^\pm$ and $K^\pm$, we require, next to the already mentioned geometrical detector track cosntrain, the number of CDC hits associated to the track to be larger than 20 to ensure high performance on the hadron ID [55]. We also require the corresponding hadron IDs to be larger than 50%.
The pion arising from the $D^{(*)*}$ resonance, is required to satisfy a hadron ID cut of smaller than 0.1. No additional track quality cuts, such as the good track requirement or minimum number of CDC hits, are applied, as these are often too strict for low-momentum pions. Note that the pion is always slow in the case where the $D^{(*)}\pi$ system originates from a D^* resonance (i.e. for our first decay mode $\overline{B}^0 \to D^{*+}(\to D^0\pi^+)\ell^-\overline{\nu}_\ell$), and can potentially (but less likely) be slow if it arises from a higher excited charm states such as D^{**}.
In the specific case of the semi-leptonic decay channel $B^- \to D^{*+}(\to D^0\pi^+_{\mathrm{slow}})\pi^-\ell^-\overline{\nu}_\ell$, additional requirements are applied to enhance the reconstruction of the D^* resonance. The momentum of the slow pion from the D^* decay is required to be less than $0.4\,\mathrm{GeV/c}$, and the mass difference $\Delta M = M(D^*) - M(D^0)$ must lie within the range $[0.141, 0.156]\ \mathrm{GeV/c^2}$.
Futhermore we apply a mass constrain to the D meson in all three channels and restrict its absolut mass difference to the nominal PDG value to be smaller than 15 MeV.
Moreover we require the cosine of the angle of the B meson and the $Y = D^{*(*)}\ell$ system, noted as $\cos\theta_{BY}$, to lie within -4 and 1. A signal candidate should yield a physical value, i.e. a value within the range $[-1, 1]$. However, we retain the wider range to potentially use this variable for signal extraction at a later stage.
In this analysis, we apply a vertex fit to improve the reconstruction quality of the decay chain. Two primary strategies for vertex fitting are commonly used: the *kinematic fit* (kfit), which is applied individually to each intermediate resonance (such as D or $D^{*(*)}$), and the *tree fit*, which fits the entire decay chain simultaneously starting from the B meson. For our purposes, we adopt the kfit approach, applying a mass constraint to the D meson vertex, but not to the $D^{*(*)}$ vertex. The motivation behind this choice is that the kfit allows us to obtain individual χ^2 probabilities for each fitted vertex, which provides more detailed information than the single χ^2 value obtained from a global tree fit. Applying a mass constraint to the D meson forces its invariant mass to match the nominal value from the PDG, thereby improving the mass resolution of the reconstructed $D\pi$ system. As a consequence, this can reduce uncertainties that affect subsequent steps such as unfolding. It is important to note that any selection criteria based on the reconstructed D mass must be applied prior to the vertex fit, as the constraint overrides the measured invariant mass. We also explored various vertex fit configurations, including applying or omitting the mass constraint on the D vertex, and employing a tree fit at the B meson level with or without updating all daughter particles. However, these alternatives did not lead to significant improvements in reconstruction

performance, particularly in terms of signal-to-background ratio. Therefore, we do not discuss them in detail in this thesis.
As already mentioned in Section 4.1, the ROE consists of including all reconstructed particles in the event that are not associated with the signal candidate. As a result, a typical ROE contains not only the accompanying B meson (e.g. the tag side B when the signal side is reconstructed), but also a variety of other particles. These can include hadronic splittings, scattered photons from radiation processes, radiation-induced background particles and decay products of long-lived particles such as kaons or pions [56]. To improve event reconstruction, the ROE is often "cleaned up" using *ROE masks*, which filter out irrelevant or misreconstructed particles based on defined criteria. In this analysis, we evaluated the impact of applying a ROE mask by comparing results with and without its use. The differences observed were negligible and did not significantly affect the reconstruction performance. Therefore, in alignment with the Belle II untagged $\overline{B}^0 \to D^{*+}\ell\nu_\ell$ analysis, which also did not employ a ROE mask, we proceed without applying one.
To reduce the multiplicity of reconstructed candidates per event, we limit the number of candidates to a maximum of six. In events with more than six candidates, we retain only those with the highest χ^2 probability from the B vertex fit.
A sum up of all applied selection criteria for the untagged $B \to D\pi\ell\nu_\ell$ analysis is given in Table 4.1.

4.3 Event categorisation and first look

In Fig. 4.2, the invariant mass spectrum of the reconstructed $D\pi$ system is shown for all three B decay modes, after applying the full selection criteria described in Section 4.2. This distribution is shown for simulated data only, or more precise, using the `MC15rd bucket26` sample only.
Figure 4.3 presents the distribution of the number of signal candidates per event. Multiple reconstructions can occur within a single event, for example, due to the interchange of two pions, one originating from the tag side and the other from the signal side. The plot is limited to a maximum of six candidates per event, as only the first six candidates with the highest vertex fit χ^2 probability were retained in the reconstruction. In simulation we can categorize the events using their generator information. For that we find the following categories:

- **Signal events**: An event is categorized as a signal event if it corresponds to one of the signal decay modes described in Section 4.1, and if all reconstructed particles match the true particles in the decay chain according to the Monte Carlo truth information. Each corresponding intermediate resonance (as well as the non-resonant case) is treated and shown separately in the plot and labelled respectively.

- **Self cross feed**: It is possible for an event to originate from a true signal decay but be misreconstructed due to incorrect assignments of final-state particles. If the misreconstruction occurs solely within the signal-side decay products, we refer to the event as a signal event with Self Cross Feed (SCF). An example of this would be the interchange of a pion originating from a D^{**} resonance with one from the D meson. Such cases can only occur if the exchanged particles have identical charges, since differing charges naturally prevent such misassignments. For instance, in the decay mode $B^- \to D^+\pi^-\ell^-\overline{\nu}_\ell$, SCF is excluded by construction, since the pions involved carry opposite charges. If events are labeled as SCF in this decay mode, they typically arise from kaon–pion misidentification. However, the overall SCF fraction is found to be very small.

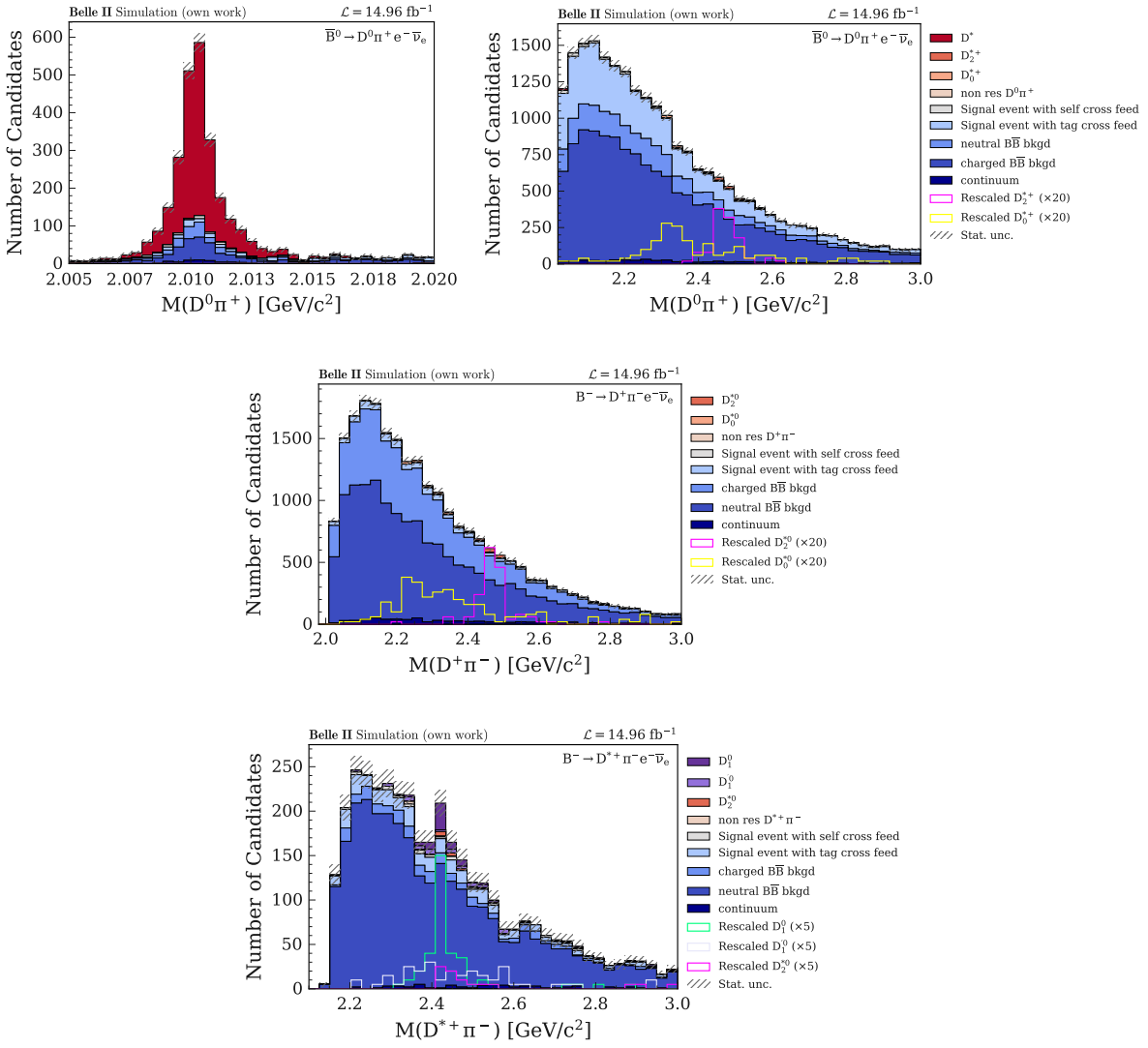


Figure 4.2: Invariant mass spectrum of the $D\pi$ system right after applying all the selection cuts in 4.1. Note that the two upper plots correspond to the same B decay mode, but are split in such a way that the D^* resonance is separated to prevent it from dominating the plot.

Table 4.1: Selection cuts employed to all decay channels

	Variable	Selection criterion
Good tracks	d_r	< 2 cm
	$\lvert d_z \rvert$	< 4 cm
	polar angle θ	$\in [17°, 150°]$
Leptons	e-ID (or μ-ID)	> 0.99 (0.9)
	p_ℓ^{CMS}	$\in [1.2, 2.4]$ GeV/c^2
	standard list	
	brems correction for e	
	good tracks	
Charged hadrons (K^+ and π^+)	kaonID (or pionID)	> 0.5
	nCDCHits	> 20
	good tracks	
π_{slow}	pionID	> 0.1
	$p_{\pi_{\mathrm{slow}}}^{\mathrm{CMS}}$	< 0.4 GeV/c
Intermediate particles	$\lvert M(D^{0/+}) - M_{D^{0/+}}^{\mathrm{PDG}} \rvert$	< 15 MeV/c^2
	$M(D^{*+}) - M(D^0)$	$\in [0.141, 0.156]$ GeV/c
Continuum suppression	$p_{D^{*(*)+}}^{\mathrm{CMS}}$	<2.5 GeV/c
	$\cos\theta_{TBT0}$	< 0.7
Other selections	$\cos\theta_{BY}$	$\in [-4, 2]$
	High level trigger	on
	vertex kfit	
	(with mass constrain on D)	

Importantly, SCF events do not significantly distort the invariant mass distribution of the $D\pi$ system, nor the momentum transfer variable q^2. Therefore, SCF events are included in the signal category and are treated as signal unless explicitly shown as a separate component in the plots. One exception arises in observables sensitive to particle kinematics and angular correlations. In particular, the helicity angle may be affected by the misassignment of pions in SCF events, and caution is required when interpreting results involving this variable.

- **Tag cross feed**: However, if, for example, a pion is misreconstructed from the tag side and is included in our reconstructed signal decay, it changes all variables, invariant $D\pi$ mass, q^2 and also the helicity angels. these events are labelled by us as Tag Cross Feed (TCF) and continue to count as background components.

- **Neutral B$\overline{\mathrm{B}}$ background**: Events that do not contain a signal decay, but originate from $\Upsilon(4S) \to B^0\overline{B}^0$ decays and are misreconstructed as signal, are categorized as neutral $B\overline{B}$ background.

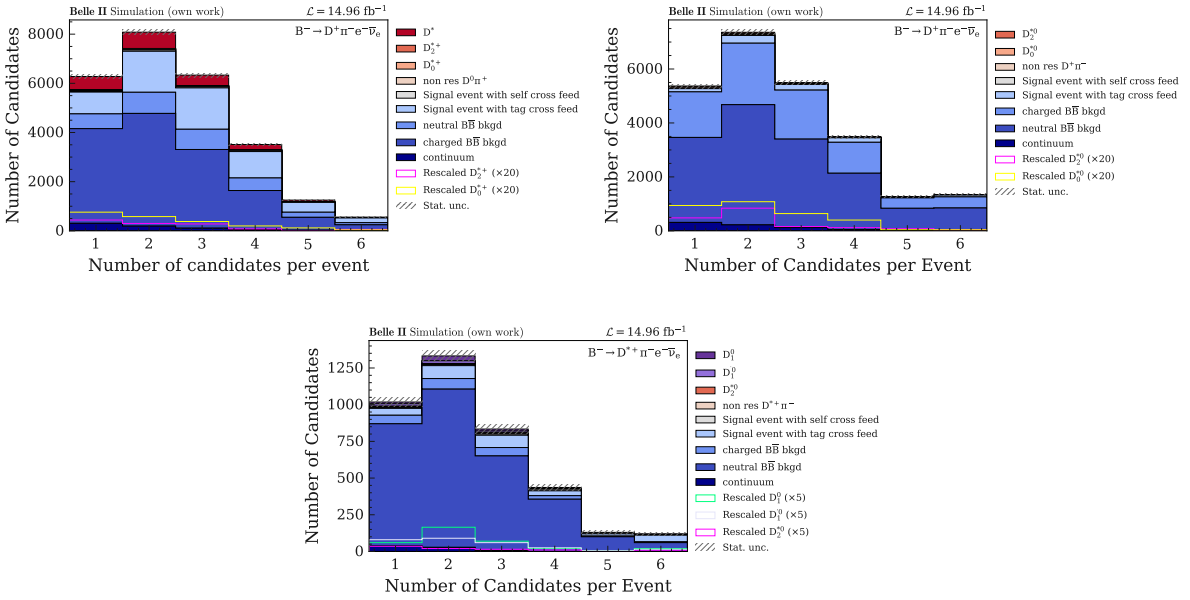


Figure 4.3: Number of candidates per event right after applying all the selection cuts in 4.1. We find an average of 2.50, 2.58 and 2.40 number of candidate per event for the $D^0\pi^+$, $D^+\pi^-$ and $D^{*+}\pi^-$ modes respectively.

- **Charged B$\overline{\text{B}}$ background**: The same applies to misreconstructed events originating from $\Upsilon(4S) \to B^+B^-$ decays, which do not follow a signal decay mode but are falsely reconstructed as signal. These events are classified as charged $B\overline{B}$ background. In some cases, the neutral and charged $B\overline{B}$ background components are combined and shown as a single category labeled as $B\overline{B}$ background in the plots.

- **Continuum event**: Continuum events originating from $e^+e^- \to q\bar{q}$ processes, with $q = u, d, s, c$, are also labeled accordingly. However, they do not play a significant role in this analysis.

All event categories are defined based on Monte Carlo (MC) truth-matching variables. The signal category is identified manually by requiring that the `mcPDG` codes of all particles, including intermediate resonances, match the expected values, and that the full decay tree structure corresponds to the intended signal decay (here we make use of the variables `genMotherID` and `genParticleID`). Events classified as SCF and TCF are identified using a combination of MC truth information and the `TopoAna` package. The remaining background components correspond to predefined event classification variables, also available within the `basf2` framework.

A first look of the invariant mass spectrum of the $D\pi$ system reveals that the distribution is heavily dominated by background, with the signal components beeing barely visible. As the figure shows, the purity for the $B^- \to D^{*+}\pi^-\ell^-\bar{\nu}_\ell$ mode is the highest, with fewer counts generally observed. This meets our expectations because, compared to the Dπ modes, there is one more intermediate resonance and therefore less combinatorical background. Additionally, we were able to exclude background candidates from the D^* using the mass constraint (specifically the cut on ΔM).
In the following section, to suppress therefore the background and enhance the signal sensitivity, we try to train BDT.

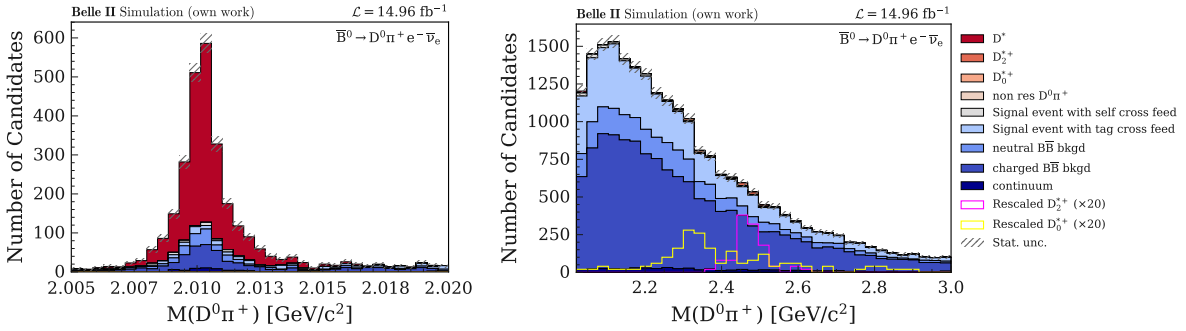
4.4 Training a BDT for background suppression

For the BDT training, we make use of the `Python` library `XGBoost` [57] for implementing gradient boosted decision trees, and `Optuna` [58] for hyperparameter optimization. For the latter, we set the maximum depth of the tree to 3 to prevent overfitting and set the learning objective to `binary:logistic`. While the model itself is a BDT, this objective specifies that the training is guided by the logistic loss function commonly used in logistic regression. As a result, the model outputs probabilities between 0 and 1, which we refer as the BDT signal probability later on. The two hyperparameters that we optimise with `Optuna` are the regularisation terms α and λ, which correspond to L1 and L2 regularisation respectively and are also known as Lasso [59] and Ridge [60] regression. L1 regularisation promotes sparsity by encouraging some weights to become zero, while L2 regularisation penalises large weights to improve generalisation. This tuning was motivated by initial signs of overfitting during training with default parameter values.
To select the best hyperparameters for the BDT training, we use stratified k-fold cross-validation within the `OptunaSearchCV` framework. We choose $k = 10$, which is commonly used to balance bias and variance. In this approach, the training data is randomly split into ten folds, each preserving the original class distribution. For each trial, the model is trained on nine folds and validated on the remaining one, cycling through all folds. The average validation score, based on the area under the ROC (Receiver Operating Characteristic) curve (AUC - Area Under the Curve), is used to guide the hyperparameter search. The receiver ROC curve illustrates the trade-off between the true positive rate and the false positive rate at various classification thresholds, here it scans through different thresholds of the signal probability of the BDT output. It provides a visual representation of a model's ability to distinguish between classes, independent of any specific threshold.
With this, we try to make the BDT training as robust and general as possible and not sensitive to any training set.

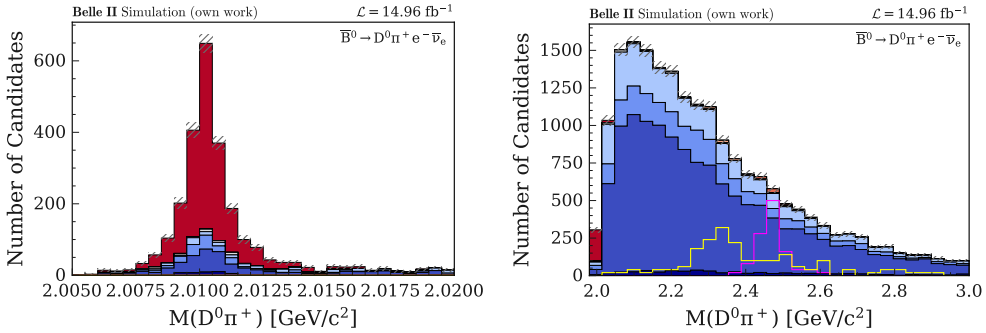
In preparing the training and test data for the BDT, we include, in addition to the run-dependent MC `bucket26` sample with a luminosity of $14.96\,\text{fb}^{-1}$, our self-generated run-independent signal MC to increase the statistics of the signal channel. Since the two datasets are highly imbalanced, with a total of roughly 25000 (2300 signal) reconstructed candidates in the generic MC and 130000 (47000 signal candidates) in the signal MC, we apply event weights during training to compensate for this imbalance and prevent the model from becoming biased toward the signal class. To do so, we assign a weight ratio of 25:1 between signal and generic MC events, motivated by the corresponding difference in their effective luminosities. We then split the combined dataframes in 50% for the training and 50% for the test set. This ensures that we have enough statistic left to evaluate later signal probability cuts to the unseen generic part of the testset.

The two classes in the binary BDT classification are defined as signal and background. The signal class includes events in which the B meson decays semi-leptonically to a D^{**} state, including both resonant and non-resonant contributions. Thus, all D^{**} resonances are grouped into a single signal category.
For the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ decay mode, where a contribution from the D^* resonance is also present, we apply a veto to exclude D^*-like events from the training sample. This is done by first reconstructing a D^* candidate using selection criteria adapted from Lyus analysis, specifically requiring $0.141 < \Delta M < 0.156$ and $E_{\text{vis}}^{\text{CMS}} > 4$ GeV. Events passing this selection are removed from the dataset to prevent the BDT from being dominated by the abundant and well-separated D^* contribution, which is not the target of the classification. A plot of the invariant mass $M(D\pi)$ distribution before and after applying the veto is

shown in Fig. 4.4.



(a) Invariant mass $M(D\pi)$ distribution before applying the D^* veto. (Left) the D^* region and (right) the D^{**} region (cut at $M(D\pi) > 2.02$).



(b) Invariant mass $M(D\pi)$ distribution after applying the D^* veto. (Left) the sub-sample of the D^* veto and (right) the left over candidates with the D^{**} resonances used for the BDT training.

Figure 4.4: Invariant mass distributions of $D\pi$ candidates in the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ sample, before and after removing events in the D^* region, motivated by Chaoyis selection cuts.

As part of the BDT training process, we need to find input variables for the BDT training that are suitable and have some separation power between signal and background candidates. Throughout this thesis, we explored different sets of input variables for the BDT training. Commonly used variables in Belle II analyses include the energy difference between the B candidate (or ROE) and half of the center-of-mass energy, as well as the beam-constrained mass M_{bc}. We also considered the beam-constrained mass of the ROE, defined analogously with respect to $E_{\text{CMS}}/2$. In addition, we included the χ^2 probabilities of various vertex fits, hoping they could help distinguish correctly reconstructed candidates. Further input variables are included, such as the momenta of the pion and the lepton, as well as the invariant mass of the D meson prior to the vertex fit. The variables discussed above are those ultimately used for the BDT results presented in this thesis. Adding further inputs, such as the angle between the D meson and the pion or ROE variables with an applied mask, did not improve the model performance and are excluded to keep the input set compact for the thesis. Nevertheless, should a more detailed exploration of these variables be of interest, the corresponding distributions can be found in the appendix in Fig. A.4, Fig. A.5 and Fig. A.6. The same applies to additional distributions of vertex-related variables, specifically the transverse distance of the vertices in respect to the IP and the significance of the vertex-fit

distance. These distributions are shown in Fig. A.1, Fig. A.2 and Fig. A.3.
The distributions of the selected input variables used for the BDT training are shown in Fig. 4.5 for the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ mode, in Fig. 4.6 for $B^- \to D^+\pi^-e^-\bar{\nu}_e$, and in Fig. 4.7 for the $B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$ channel. The input variable distributions of the signal MC sample are provided in the appendix in Fig. B.1, Fig. B.2, and Fig. B.3 to allow for better comparison to avoid overloading the main part of the thesis with a large number of distributions. From a naive first look at the one-dimensional input variable distributions one does not reveal any clear separation between signal and background candidates. However, by combining information from the full set of input features, we expect the BDT to learn a multidimensional decision boundary that exploits differences in the underlying probability density functions (PDFs) of signal and background events.
Having these defined input variables and trained the BDT, we now focus on evaluating its classification performance.

4.4.1 BDT performance

There are two key aspects we pay particular attention to when evaluating our BDT model. Firstly, it should perform equally well on the unseen test set as it does on the training set. Secondly, the signal should ideally be clearly distinguishable from the background in the signal probability distribution: background candidates should peak near 0, while signal candidates should peak near 1.
As previously mentioned, a visual means of evaluating model performance is the ROC curve, shown in Fig. 4.8. One can immediately observe that the performance on the training and test sets is nearly, but not exact, identical, with similar, but not exact, AUC values for both. The lowest performance is seen for the $B^- \to D^{*+}\pi^-\ell^-\bar{\nu}_\ell$ channel, which may be attributed to limited statistics. However, the ROC curve itself is not our primary focus: more informative are the signal probability distributions shown in Fig. 4.9.
Unfortunately, these distributions reveal that our BDT model does not consistently classify signal candidates with high confidence. This is particularly pronounced in the $B^- \to D^+\pi^-\ell^-\bar{\nu}_\ell$ channel, where the signal probability distribution also peaks near zero. Regarding overfitting, the model's performance on the test and training sets appears generally reliable. Nonetheless, it can be seen that the model slightly overfits specific structures in the training data, especially for signal candidates in the low-probability region, despite the use of regularisation techniques in the training.
It is also informative to examine the feature importance of the input variables. For this evaluation, we use the *gain* metric, which is implemented in the `XGBoost` framework. This metric indicates the average improvement in the loss function across all splits where the respective feature is used [61]. The corresponding gain values and the resulting ranking are shown in Fig. 4.10. It can be observed that the ranking of the features differs across the three B decay modes. For the $D\pi$ channels, the χ^2 probability of the vertex fit appears to provide the most information gain. In contrast, the χ^2 probabilities of the $D^{*(*)}$ and D vertices are generally less informative. However, in the semileptonic $D^+\pi^-$ mode, the χ^2 probability of the D^+ vertex shows a relatively high information gain, which is consistent with expectations: due to its longer lifetime compared to the D^0, the D^+ has a greater flight distance before decaying, resulting in a more distinct vertex. Additionally, the variable ΔE, which denotes the difference between the reconstructed energy E and half the centre-of-mass energy, proves to be particularly informative in the $D^*\pi$ channel.

Even though the signal probability distribution does not exhibit a clear separation between signal and background, we can still attempt to apply a cut on the signal probability and subsequently examine

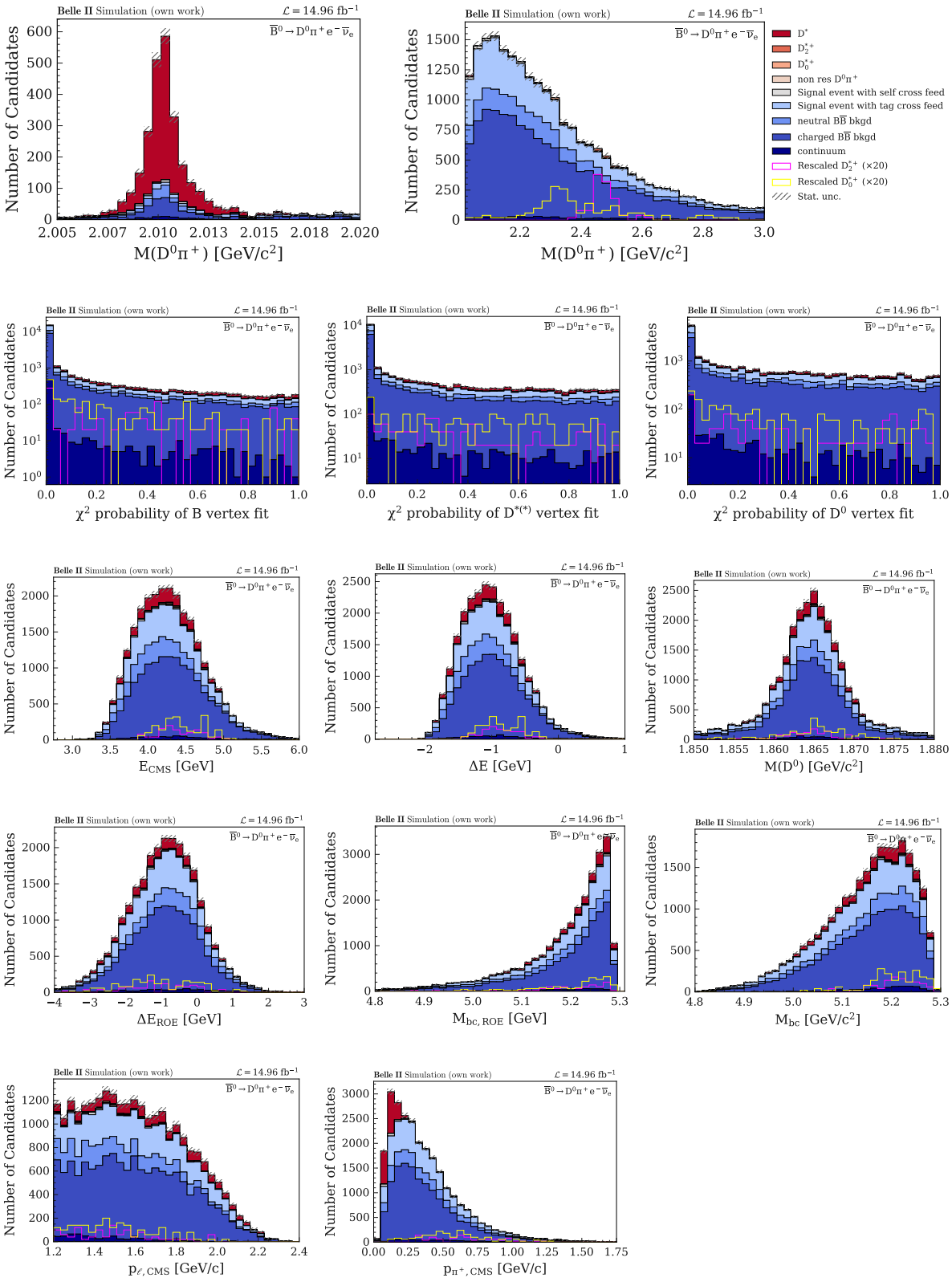


Figure 4.5: BDT input variables for the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

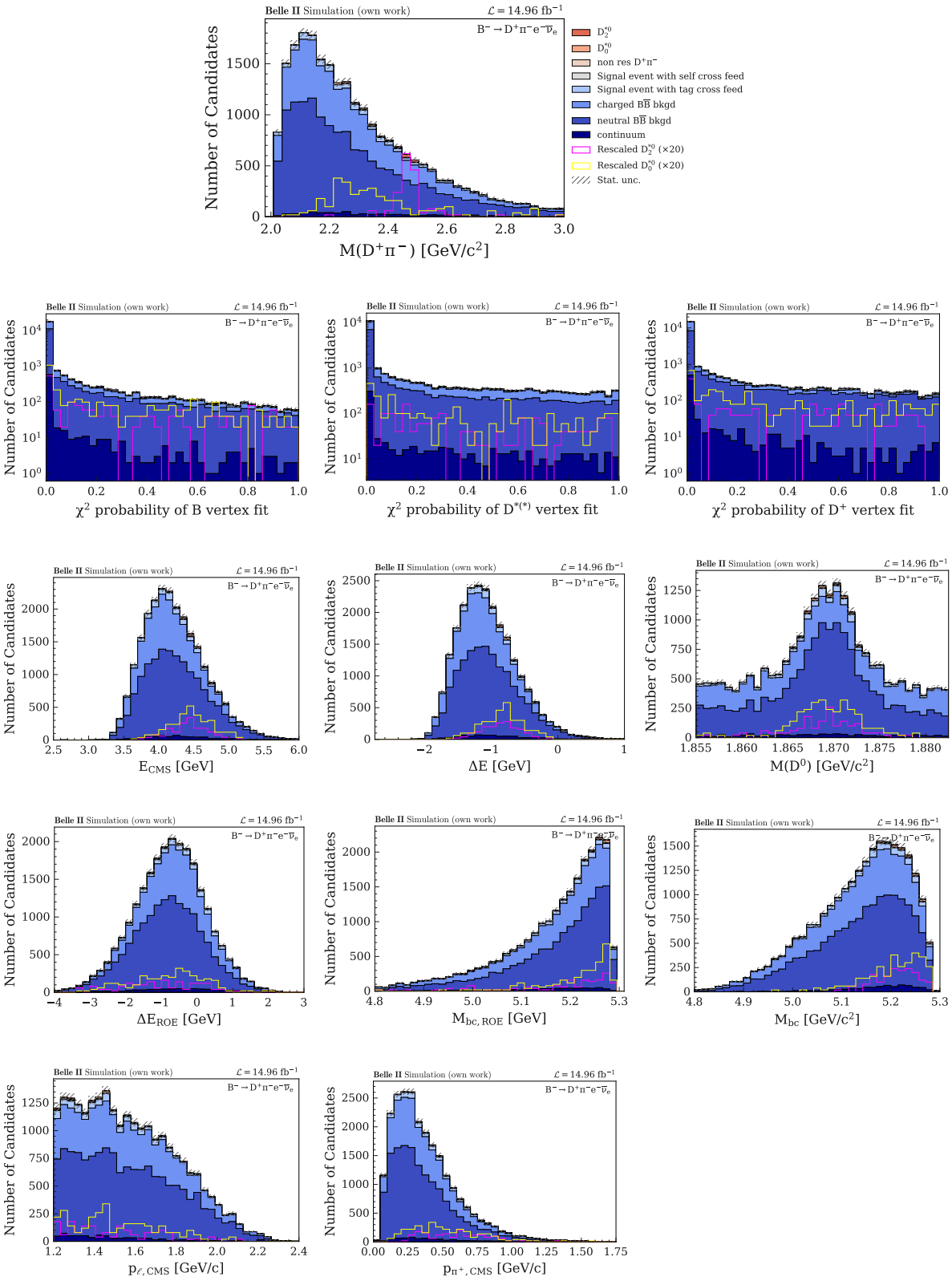


Figure 4.6: BDT input variables for the $B^- \to D^+\pi^-\ell^-\overline{\nu}_\ell$ decay mode. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

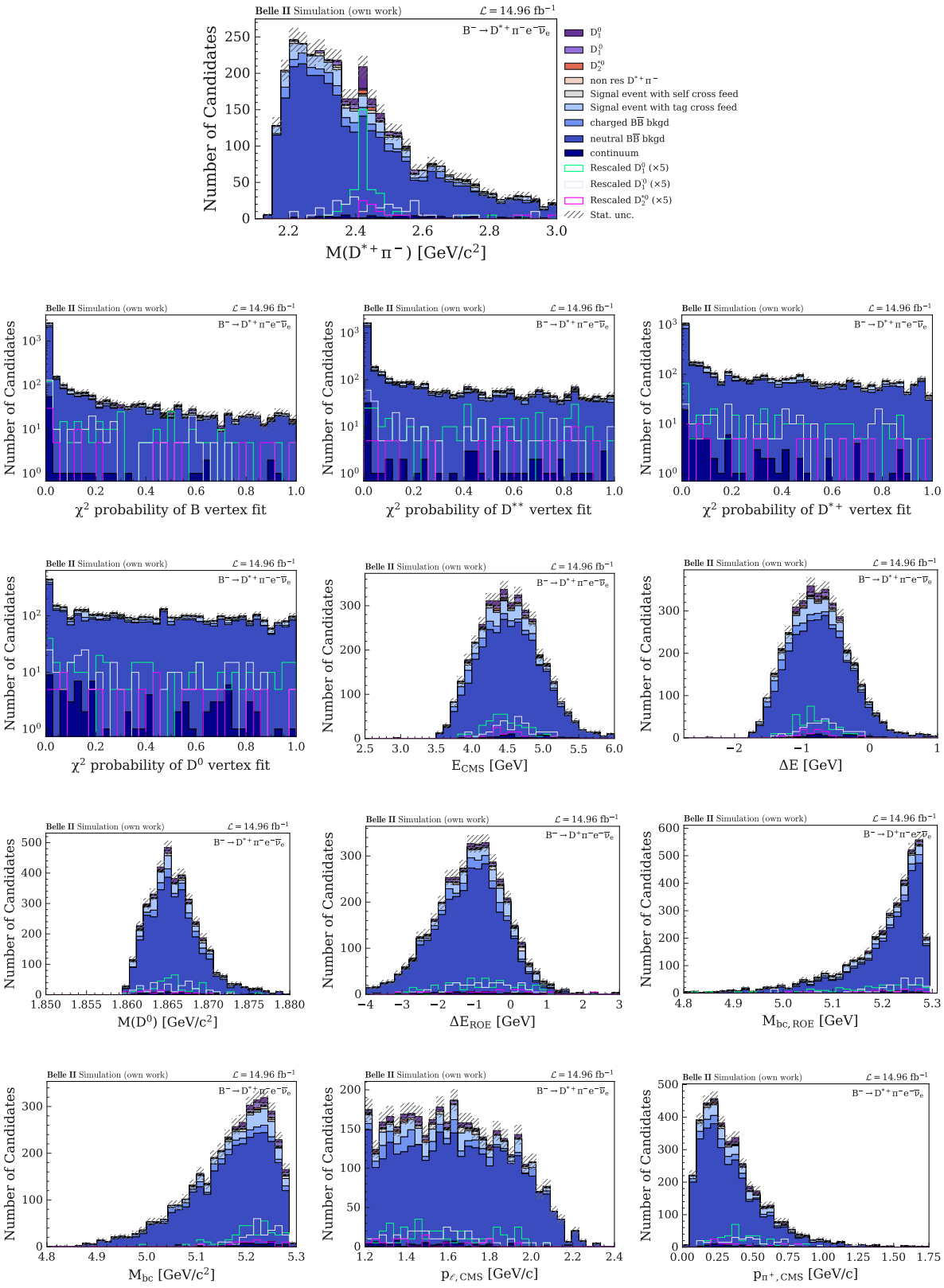


Figure 4.7: BDT input variables for the $B^- \to D^{*+}\pi^-\ell^-\overline{\nu}_\ell$ decay mode. Note that the invariant mass spectrum of the $D^*\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

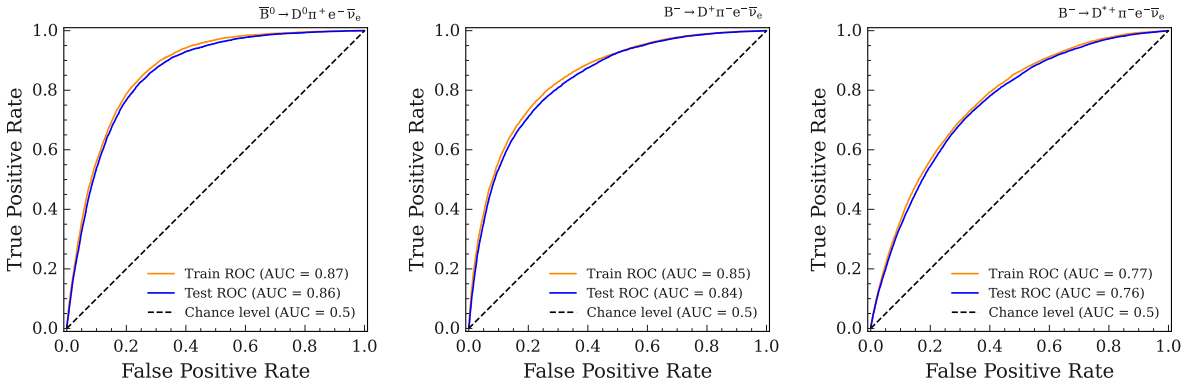


Figure 4.8: ROC curve for the train and test set for the untagged analysis for all three B decay modes. Note that in the case of the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode, decay modes from $D^{*+} \to D^0\pi^+$ decays are removed in advance. This is achieved by reconstructing the D^{*+} candidates using the same selection criteria as in the untagged Belle II $\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell$ analysis, and excluding these events from the dataset.

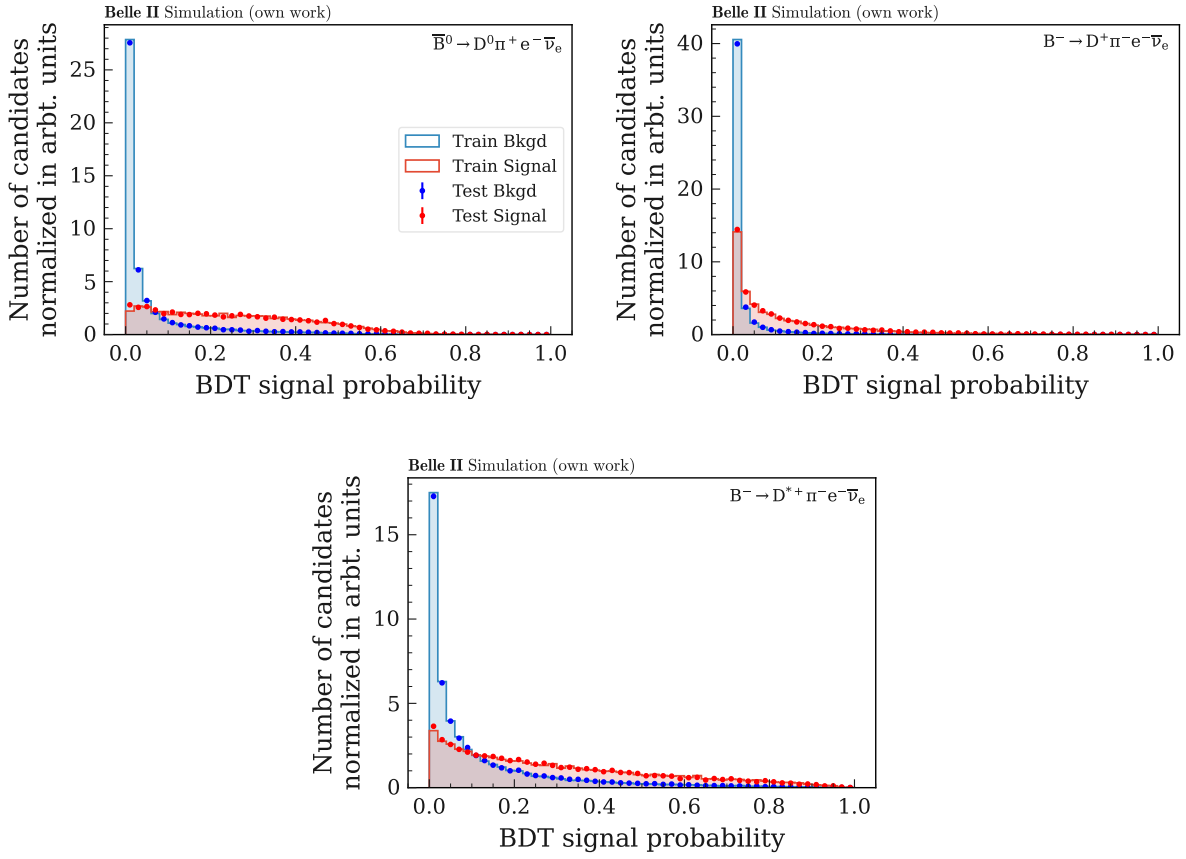
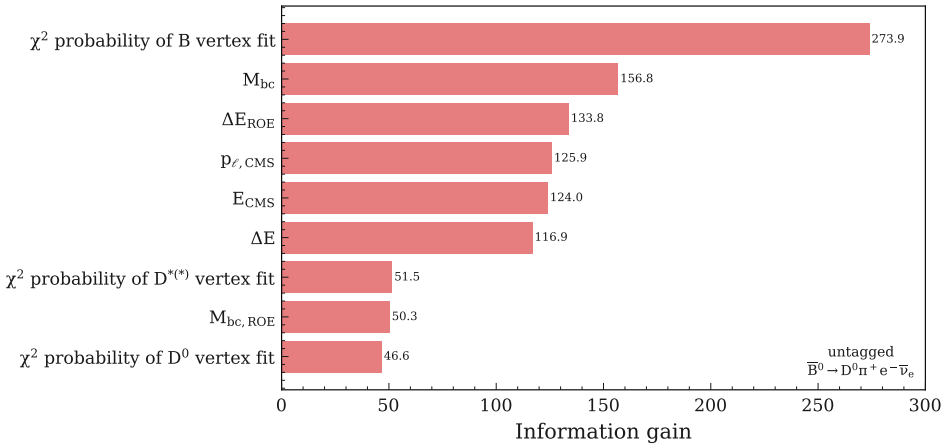
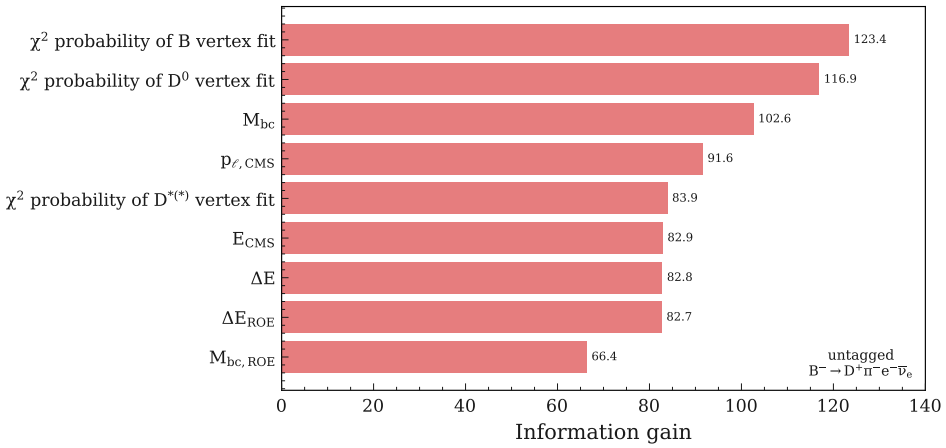


Figure 4.9: BDT signal probability for the train and test set for the untagged analysis for all three B decay modes. Note that in the case of the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode, decay modes from $D^{*+} \to D^0\pi^+$ decays are removed in advance. This is achieved by reconstructing the D^{*+} candidates using the same selection criteria as in the untagged Belle II $\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell$ analysis, and excluding these events from the dataset.





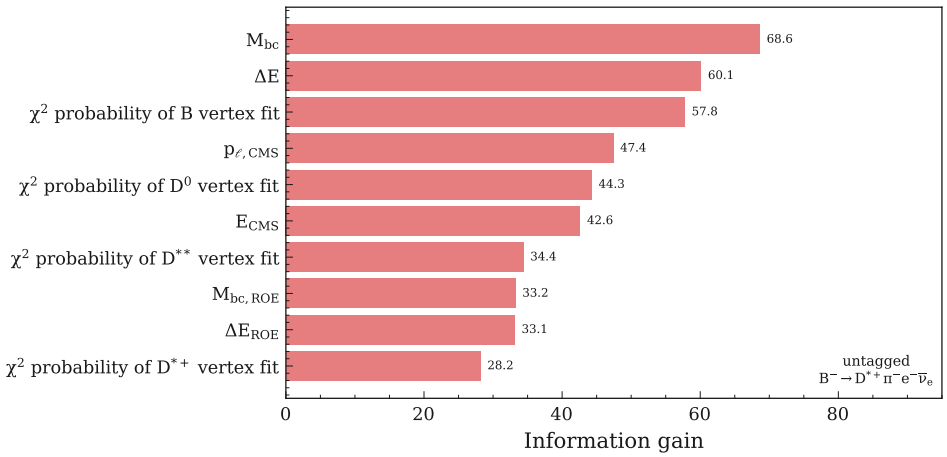


Figure 4.10: Feature importance ranking of the input variables of the BDT training for the untagged analysis for all three B decay modes. Note that in the case of the $\overline{B}^0 \to D^0 \pi^+ \ell^- \overline{\nu}_\ell$ decay mode, decay modes from $D^{*+} \to D^0 \pi^+$ decays are removed in advance. This is achieved by reconstructing the D^{*+} candidates using the same selection criteria as in the untagged Belle II $\overline{B}^0 \to D^{*+} \ell^- \overline{\nu}_\ell$ analysis, and excluding these events from the dataset.

the resulting invariant mass distribution. To determine an optimal cut, one that sufficiently suppresses background while retaining as much signal as possible, we employ a figure of merit (fom). For this purpose, we use the fom proposed by Punzi [62], defined as

$$\text{fom} = \frac{S}{\sqrt{S+B}}, \tag{4.1}$$

where S and B denote the number of signal and background events, respectively. The goal is to find the signal probability cut that maximises this fom. For this, we use the minimizer library `iMinuit` [63]. To increase statistical precision, the cuts are applied to the entire `MCrd15 bucket26` sample, including the candidates that were also used in the BDT training. It should therefore be noted that both the model and the cut may perform optimistically due to this overlap. The invariant mass spectrum of the $D^{(*)}\pi$ system, before and after applying the optimal BDT signal probability cut, is shown in Fig. 4.11. The signal and background yields before and after the cut, such as the quantitative value of the BDT signal probability cut are given in Table 4.2.

Table 4.2: Signal (S) and background (B) yields before and after the BDT signal probability cut.

Decay mode	Cut	S_{before}	B_{before}	S_{after}	B_{after}	$\frac{S}{S+B}_{\text{before}}$	$\frac{S}{S+B}_{\text{after}}$	$S_{\text{after}}/S_{\text{before}}$
$\bar{B}^0 \to D^0\pi^+\ell^-\bar{\nu}_\ell$	0.142	510 ± 23	$19\,486 \pm 140$	375 ± 19	$8\,586 \pm 93$	0.0255 ± 0.0011	0.0418 ± 0.0021	0.735 ± 0.050
$B^- \to D^+\pi^-\ell^-\bar{\nu}_\ell$	0.011	317 ± 18	$23\,975 \pm 155$	274 ± 17	$3\,379 \pm 58$	0.0130 ± 0.0007	0.0750 ± 0.0044	0.864 ± 0.071
$B^- \to D^{*+}\pi^-\ell^-\bar{\nu}_\ell$	0.038	193 ± 14	$3\,677 \pm 61$	172 ± 13	993 ± 32	0.0499 ± 0.0035	0.1476 ± 0.0104	0.891 ± 0.093

From this, one can see that the purity is improving and the background is being reduced. Nevertheless, the background contribution remains dominant. While one could further investigate the addition of more input variables or the tuning of more hyperparameters, it was expected that more discriminating power could be gained from the vertex fit variables. This we expected specifically for the $\bar{B}^0 \to D^0\pi^+ e\bar{\nu}_e$ channel, where we have a high tag cross feed contribution.
One variable that was excluded from the BDT training (to save it for the signal extraction fits) but nonetheless exhibits separation power is $\cos\theta_{BY}$. A comparison of the $\cos\theta_{BY}$ distributions before and after applying the signal probability cut shows that the BDT effectively suppresses background candidates outside the physical region, i.e. those with $\cos\theta_{BY} \notin [-1, 1]$. The remaining background is concentrated in the physical region and exhibits a peaking structure. This also implies that even a binned likelihood fit, based on the MC signal and background templates later on, may encounter difficulties, as the two templates provide limited separation capability.

Before exploring potential improvements to the BDT training, it may therefore be useful to first investigate the origin of the background contributions, what we will discuss in the next section.

4.5 Understanding of background components

We begin by examining the contribution from the charged $B\bar{B}$ component. Using `TopoAna`, which utilises generator-level information, the background can be efficiently categorised according to the true B decay modes. As the charged $B\bar{B}$ component in the $B^- \to D^{*+}\pi^- e\bar{\nu}_e$ channel is relatively small and not a dominant contributor, this decay mode is excluded from the charged $B\bar{B}$ background discussion. Figure 4.13 shows the invariant mass distributions for the $D\pi$ channels, where the charged $B\bar{B}$ background is further decomposed into its dominant decay contributions. In the $\bar{B}^0 \to D^0\pi^- e^-\bar{\nu}_e$ channel, the

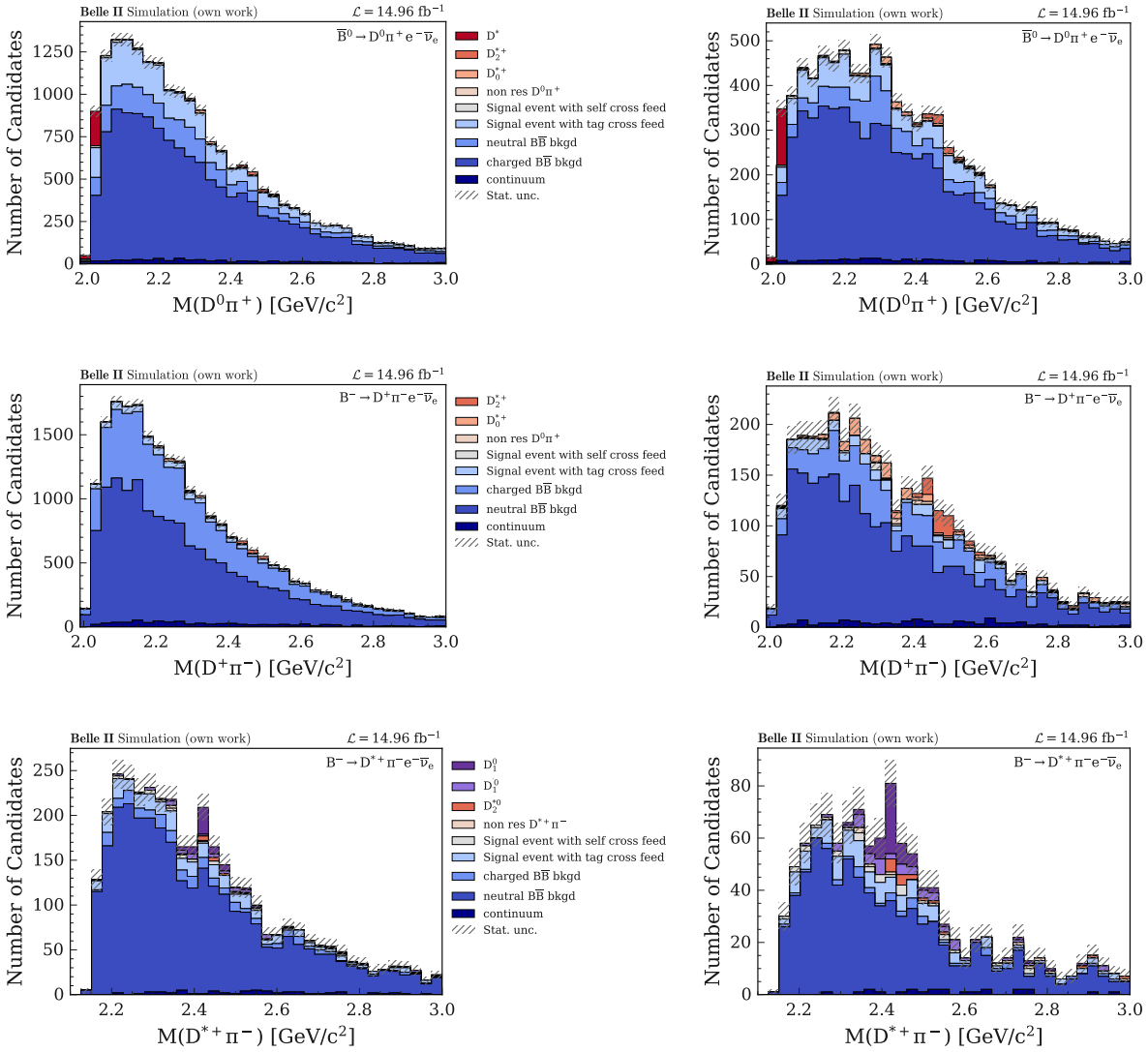


Figure 4.11: Invariant mass $M(D^{(*)}\pi)$ distribution (left) before and (right) after applying the BDT signal probability cut, optimized by maximizing the fom defined in Eq. (4.1) and quantitavly given in Table 4.2.

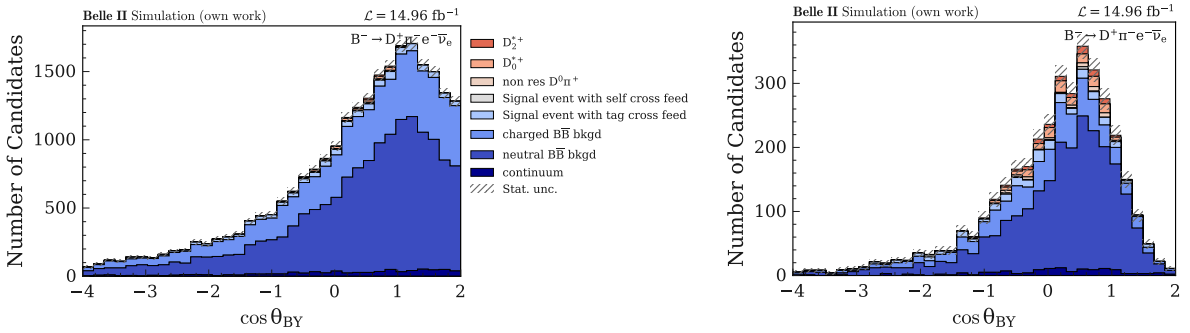


Figure 4.12: $\cos\theta_{BY}$ distribution (left) before and (right) after applying the BDT signal probability cut, optimized by maximizing the fom defined in Eq. (4.1) and quantitatively given in Table 4.2, exemplary shown for the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ channel.

decay mode $B^- \to D^0 e^- \bar{\nu}_e$ is found to be the most significant source of background. This suggests that, during reconstruction, a pion from the tag side is incorrectly combined with the D^0, resulting in a misreconstructed $D^{*(*)}$ state that mimics a $\bar{B}^0 \to D^0\pi^-e^+\bar{\nu}_e$ decay.
A similar situation arises for the $B^- \to D^{*0}e^-\bar{\nu}_e$ decay, in which the D^{*0} decays into a D^0 and a neutral pion π^0. If the π^0 is not reconstructed, the remaining D^0 may again be wrongly combined with a pion from the tag side, leading to a misidentification as a $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ candidate.
For the semi-leptonic $D^+\pi^-$ channel, the main components of the charged $B\bar{B}$ background are also the two decay channels of the $B^- \to D^0 e^-\bar{\nu}_e$ channel and $B^- \to D^{*0}e^-\bar{\nu}_e$. The issue already begins with the fact that in approximately 90% of cases the decay $D^0 \to K^-\pi^+$ is misreconstructed as a $D^+ \to K^-\pi^+\pi^+$ by incorrectly combining an additional pion from the tag side.
In summary, the dominant contribution to the charged $B\bar{B}$ background therefore originates from misreconstruction involving a pion from the tag side.

In the case of the neutral $B\bar{B}$ backgrounds the splitting in the true B decays can be seen in Fig. 4.14. The neutral $B\bar{B}$ background in the semileptonic $\bar{B}^0 \to D^0\pi^+$ channel is comparatively small. The dominant contribution arises from $\bar{B} \to D_1^{(\prime)+}e^-\bar{\nu}_e$ decays. These can be interpreted as downfeed processes, in which the $D_1^{(\prime)+}$ meson decays via $D_1^{(\prime)+} \to D^{*+}\pi^0$ or $D_1^{(\prime)+} \to D^0\pi^+\pi^0$, and the neutral pion is not reconstructed. As a result, such events are misreconstructed as signal candidates and feed into the $\bar{B}^0 \to D^0\pi^+$ channel.
In the $D^+\pi^-$ decay channel, the largest contribution to the background originates from $\bar{B}^0 \to D^+e^-\bar{\nu}_e$ and $\bar{B}^0 \to D^{*+}(\to D^+\pi^0)e^-\bar{\nu}_e$ decays. Similar to the charged $B\bar{B}$ background observed in the $D^0\pi^+$ channel, the D^+ is misreconstructed as part of the signal decay by wrong combining it with a pion from the tag side. In the case of the D^{*+} contribution, this misreconstruction is further facilitated by the failure to reconstruct the neutral pion from the $D^{*+} \to D^+\pi^0$ decay.
For the $D^{*+}\pi^-$ channel, most of the neutral $B\bar{B}$ background arises from $\bar{B}^0 \to D^{*+}e^-\bar{\nu}_e$ decays, which are misreconstructed as $B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$ due to the inclusion of an additional pion from the tag side.

Furthermore, especially with the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ channel, the tag cross feed component is comparatively high. As in the other $B\bar{B}$ background cases, it can be assumed that the misreconstruction is primarily caused by confusion with a pion from the tag side. In this particular case, however, we take a closer look at the origin of the misreconstruction, as shown in Fig. 4.15. The figure shows the distribution of the different error types during the reconstruction. It can be seen that in about 70% of the

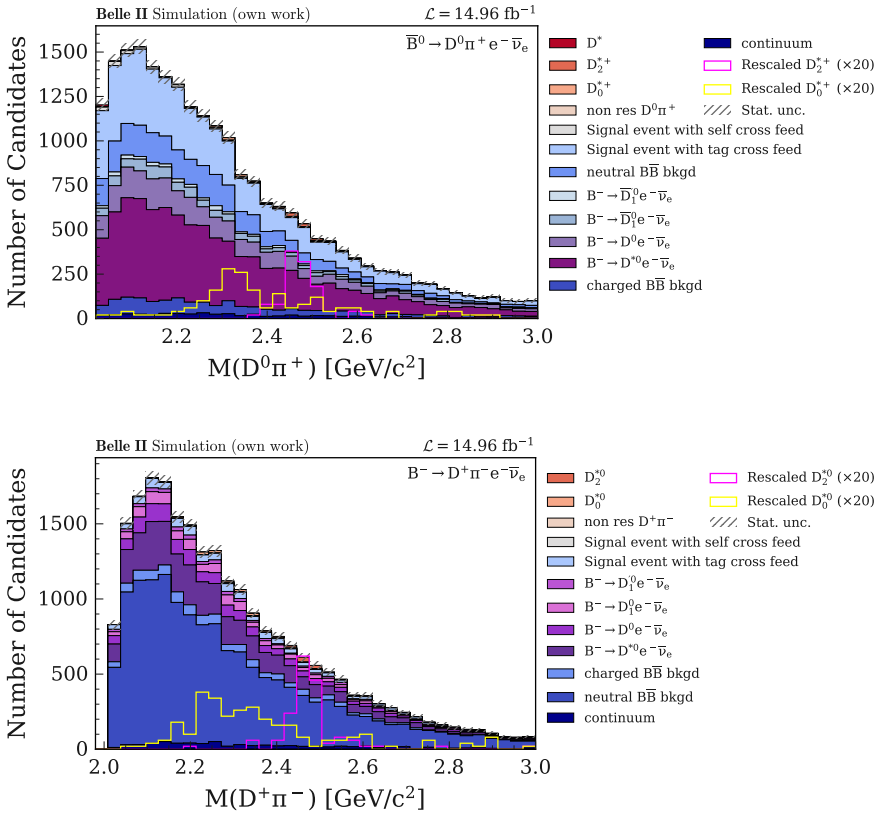


Figure 4.13: Invariant mass spectrum of the $D\pi$ system right after applying all the selection cuts in 4.1. In this case, the charged $B\overline{B}$ background is separated into distinct components, and the generator-level decay information is shown, as obtained by using TopoAna [46]. The semi-leptonic $\overline{B}^0 \to D^{*+}\pi^-$ decay is not shown, since the charged background component is quite small as one can see in Fig. 4.2.

cases the entire decay chain is correctly reconstructed, with the exception that an wrong (but real) pion is selected from the tag side. In contrast, it is relatively rare that the pion is correctly assigned while other components of the event are incorrectly reconstructed. The second most common scenario involves the true identification of the pion (in terms of particle type), but with an incorrect origin, accompanied by an incorrectly reconstructed D meson.

Based on the observation that the dominant source of the background comes from the misreconstruction with particles from the tag side, the analysis strategy was revised. Instead of focusing exclusively on the $B \to D^{**}\ell\bar{\nu}_\ell$ signal decay, the tag side will now also be reconstructed. This approach provides greater experimental control as it allows the assignment of almost[4] all final state particles to either the signal or tag side. To achieve higher selection efficiency, we use the semi-leptonic (SL) tag side $B \to D^{**}\ell\bar{\nu}_\ell$ strategy in the rest of the thesis (besides the hadronic tagged approach). Also the SL tagged $B \to D^{**}\ell\bar{\nu}_\ell$ analysis represents the first study of its kind within the Belle II experiment and has not yet been conducted before.

[4] It should be noted that these are only almost all final state particles, as certain particles, such as neutral pions from $\gamma\gamma$ decays, can still be missed in the reconstruction.

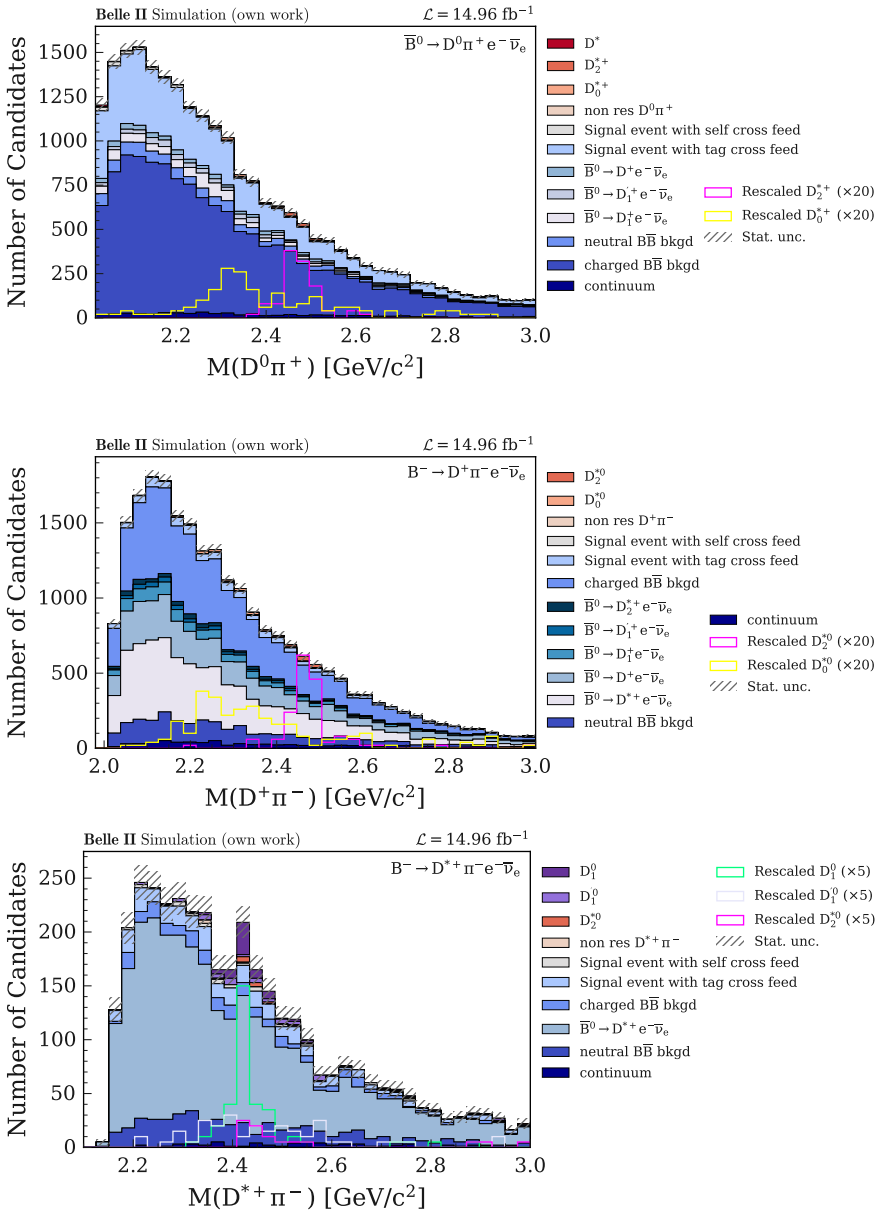


Figure 4.14: Invariant mass spectrum of the $D\pi$ system right after applying all the selection cuts in 4.1. In this case, the neutral $B\overline{B}$ background is separated into distinct components, and the generator-level decay information is shown, as obtained by using TopoAna [46].

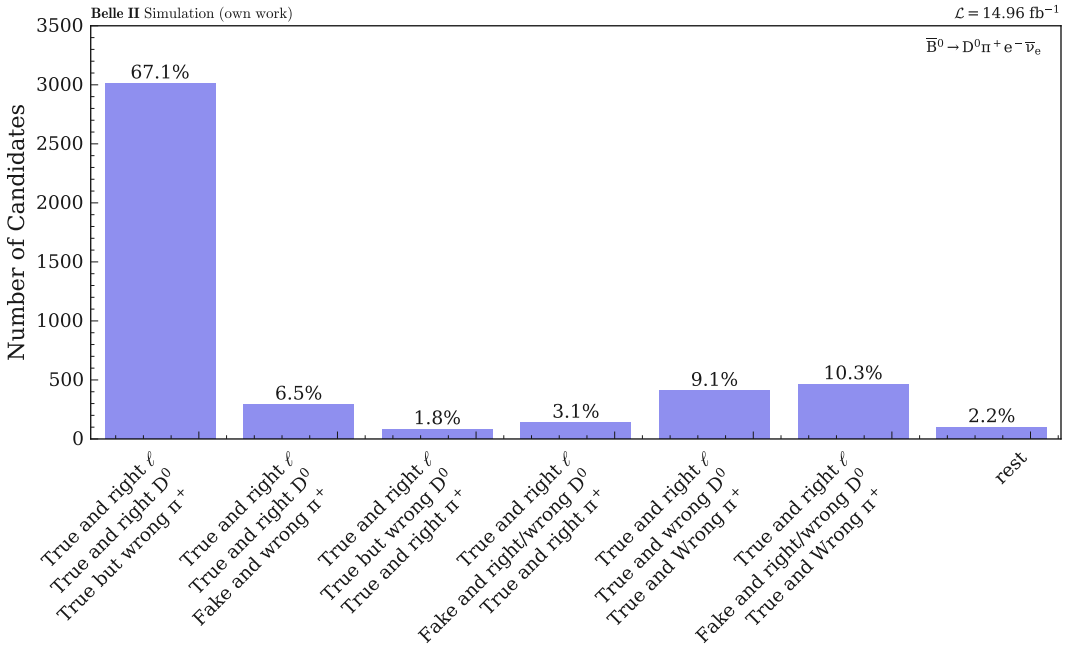


Figure 4.15: Distribution of the background source for neutral $B\overline{B}$ background. Here "true" or "fake" means that the particle refers to its right MC particle (i.e. in Belle II jargon `mcPDG` is correct) while "right" and "wrong" means that it does not match the decay chain (or in other words `genMotherID` is not consistent).

CHAPTER 5

Semi-leptonic tagged event reconstruction and selection

This chapter outlines the reconstruction and selection strategy for SL tagged $B \to D^{(*)}\pi\ell\nu_\ell$ decays with Belle II. The SL tagging approach benefits from higher signal efficiency but presents additional challenges due to the presence of two undetected neutrinos.
We describe the reconstructed decay channels and selection criteria applied to both tag and signal side, including additional hadronic D decay modes made possible by the improved background conditions. Special attention is given to specific background contributions such as misreconstructed events due to confusions with the tag side.
The treatment of multiple candidates per event is addressed using a Best Candidate Selection (BCS) strategy, and the performance of the reconstruction is evaluated by comparing signal yields with a hadronic tagged Belle measurements in 2020 [24]. Furthermore, the reconstruction of kinematic variables is discussed, including an approach to approximate the B rest frame, since with SL tagging one can not directly access the signal B rest frame.
As in the untagged analysis, this study focuses exclusively on the electron channel. For the decay mode $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$, the D^* mass region is excluded from the plots of the $D\pi$ invariant mass spectrum and focussed on the D^{**} regions to make this thesis more compact.

5.1 Reconstructed decay modes

In this analysis, we focus on four different B decay modes (including charge-conjugated modes):

1. $\overline{B}^0 \to \underbrace{D^0\pi^+}_{D^{*+},\, D_0^{*+},\, D_2^{*+}} \ell^-\overline{\nu}_\ell$
2. $B^- \to \underbrace{D^+\pi^-}_{D_0^{*0},\, D_2^{*0}} \ell^-\overline{\nu}_\ell$
3. $\overline{B}^0 \to \underbrace{D^{*0}\pi^+}_{D_1^+,\, D_1'^+,\, D_2^{*+}} \ell^-\overline{\nu}_\ell$

4. $B^- \to \underbrace{D^{*+}\pi^-}_{D_1^0, D_1'^0, D_2^{*0}} \ell^- \overline{\nu}_\ell$

Compared to the untagged analysis, we include an additional B decay mode involving a neutral pion. This is possible in the semileptonic tagged analysis due to the more manageable background conditions, which allow reconstruction of channels with one neutral particle while still maintaining a reasonable signal-to-background ratio.
The D^* mesons are reconstructed via the following decay modes:

- $D^{*0} \to D^0 \pi_s^0$
- $D^{*+} \to D^0 \pi_s^+$

For the same reasons discussed in the untagged analysis in Section 4.1, the chosen exclusive decay chains offer relatively high branching fractions, clean signatures, and, in the case of D^* decays, kinematic constraints that enable high-purity reconstruction.
For D^0 mesons, we include five hadronic decay modes:

- $K^-\pi^+$
- $K^-\pi^+\pi^0$
- $K^-\pi^+\pi^+\pi^-$
- K^+K^-
- $K_S^0\pi^+\pi^-$

and for the D^+ modes, we include the following decay channels

- $K^-\pi^+\pi^+$
- $K_S^0\pi^+$
- $K_S^0K^+$
- $K^+K^-\pi^+$

The choice of D meson decay modes is motivated by their relatively high branching ratio and a hadronic tagged Belle II analysis of $B \to D^{**}\ell\nu_\ell$ decays [64]. In that study, the signal-to-background ratio was evaluated for various D decay channels. We adopt these D decay modes, where only modes with favorable signal-to-background ratios are included. Additionally we avoid channels with more than one neutral particle (i.e. π^0 or K_S^0) in the final state.
Neutral pions are reconstructed exclusively via the dominant decay channel $\pi^0 \to \gamma\gamma$, which accounts for approximately 99% of all π^0 decays. For K_S^0 mesons, only the decay mode $K_S^0 \to \pi^+\pi^-$ is considered in the reconstruction.

5.2 Tag side selection

B_{tag} candidates are reconstructed using the SL tagged FEI algorithm, as outlined in Section 3.5. In this analysis, we restrict the selection to the following B_{tag} decay modes:

1. $B^0_{\text{tag}} \to D^- \ell^+ \nu_\ell$
2. $B^0_{\text{tag}} \to D^{*-} \ell^+ \nu_\ell$
3. $B^+_{\text{tag}} \to \bar{D}^0 \ell^+ \nu_\ell$
4. $B^+_{\text{tag}} \to \bar{D}^{*0} \ell^+ \nu_\ell$

Semileptonic $D\pi$ decay modes are excluded from the B_{tag} reconstruction, as they overlap with the signal decay modes in our study. Including these decay modes on the tag side would reduce the overall signal efficiency. This is because a tag-side best candidate selection is performed at a later stage of the analysis. True signal events may be preferentially reconstructed on the tag side in such cases, meaning they no longer contribute to the signal yield on the signal side. This effectively results in a loss of signal efficiency.
The reconstruction is based on semileptonic FEI (SL-FEI) skimmed samples, in which the B_{tag} candidates have already been reconstructed using a predefined set of selection criteria, as summarized in Table 5.1. In this analysis, we apply a further selection cuts on the B_{tag} candidates to further suppress background and improve purity. These modified selection criteria are adopted from a semi-leptonic tagged Belle II analysis [65] and are also listed in Table 5.1.

5.3 Signal side selection

The selection of the signal-side B meson in the SL tagged analysis largely follows the criteria established in the untagged analysis, as described in Section 4.2. However, due to the reduced combinatorial background resulting from the explicit reconstruction of the tag-side B meson, we apply a slightly looser selection in certain variables to increase signal efficiency.
One example is the lepton momentum, which was subject to a stringent lower threshold in the untagged analysis to suppress background from low-momentum leptons, particularly those originating from secondary decays or misidentification. In the SL-tagged analysis, this requirement is relaxed due to improved background conditions resulting from the reconstruction of the tag-side B meson. Specifically, we apply momentum requirements of $p_e \in [0.3, 2.4]$ GeV/c for electrons and $p_\mu \in [0.6, 2.4]$ GeV/c for muons. The tighter lower cut on the muon momentum is motivated by the generally higher misidentification rate in the muon list at low momenta. This is particularly relevant because muons and charged pions have similar masses, making it more difficult to distinguish between them based on kinematic and detector signatures alone. As a result, low-momentum pions are more likely to be misidentified as muons, leading to increased background contamination. A higher momentum threshold helps suppress such backgrounds while preserving acceptable signal efficiency.
Similarly, the selection on $\cos\theta_{BY}$ is loosened to a broader range of $[-15, 15]$. This allows for the

Table 5.1: Overview of the selection criteria applied to B_{tag} candidates. These cuts were either already included in the SL FEI skim procedure [66] or are adapted from the SL tagged Belle II analysis in [65]. z_0 and d_0 denote the z-coordinate and signed distance of the point-of-closest-approach respectively. E, p_T and θ stand for the energy, the transverse momentum and the polar angle of the corresponding particle. $n_{\text{cleaned tracks}}$ describes the number of charged tracks that satisfy the clean track criteria, while $n_{\text{cleaned ECL clusters}}$ refers to the number of ECL clusters that pass the corresponding quality selection.

	Variable	Selection criterion
Cleaned track	$\lvert z_0 \rvert$	$< 2.0\,\text{cm}$
	$\lvert d_0 \rvert$	$< 0.5\,\text{cm}$
	p_T	$> 0.1\,\text{GeV}$
Cleaned ECL cluster	E	$> 0.1\,\text{GeV}$
	θ	$\in [0.2967, 2.6180]$
B_{tag}	$\cos\theta_{BY}$	$\in [-1.75, 1.1]$
	$\log_{10}(\text{FEI signal probability})$	> -2.0
	p_ℓ^{CMS}	$> 1.0\,\text{GeV}$
	$n_{\text{cleaned tracks}}$	≥ 3
	$n_{\text{cleaned ECL clusters}}$	≥ 3
	$E_{\text{vis}}^{\text{CMS}}$	$> 4.0\,\text{GeV}$
π^+	Cleaned track	
γ	Cleaned ECL cluster	

possibility of optimizing the cut offline in the analysis[1].

In addition, the online selection in the untagged analysis limited the candidate multiplicity by retaining only the six best B candidates based on the vertex χ^2 probability. In the SL-tagged analysis, we do not apply such an online reduction. Instead, all candidates passing the selection are retained, and a best-candidate selection is performed offline.

The SL-tagged analysis also includes a broader set of D meson decay modes, including those with neutral particles such as π^0 and K_S^0. For π^0 reconstruction, a mass window of $M(\pi^0) \in [0.120, 0.145]\ \text{GeV}/c^2$ is applied. The neutral pion candidates are selected using the Belle II standard list provided via the `stdPi0s.stdPi0s()` function, configured to target a photon reconstruction efficiency of 40%. This function applies calibrated selection criteria optimized by the collaboration to balance signal efficiency and background suppression in $\pi^0 \to \gamma\gamma$ decays. Also, on this set of reconstructed π^0s, a mass-constrained kfit is applied to the π^0 candidates, which improves the resolution by constraining the invariant mass of the photon pair to the nominal π^0 mass.

For the reconstruction of K_S^0 mesons, we apply a set of selection criteria motivated by previous Belle II analyses. These include a requirement on the flight distance to be larger than 0, the significance of the displacement between the decay vertex and the interaction point ($\log_{10}(\text{"significance of distance"}) > 0$), and the alignment between the reconstructed momentum and the vertex vector ($\cos\theta(p_{K_S^0}, \text{vertex}) > 0.8$).

[1] Online cuts refer to selection criteria applied during the event reconstruction within the `basf2` framework and are embedded in the generated output (e.g., ntuples). In contrast, offline cuts are applied later during the analysis stage—typically in user-level code such as Jupyter notebooks—allowing for flexible tuning based on signal and background distributions.

Additionally, we apply a mass window cut on the reconstructed K_S^0 mass $|M(K_S^0) - M_{K_S^0}^{\text{PDG}}| < 10\,\text{MeV}/c^2$, following recommendations from Reference [67].
The selection is implemented using the Belle II standard function `stdV0s.stdKshorts()`, which reconstructs $K_S^0 \to \pi^+\pi^-$ candidates by combining oppositely charged pion tracks and applying pre-optimized vertex and kinematic selection criteria to ensure good purity and efficiency.
In the semi-leptonic $D^*\pi$ modes, a mass window is applied to the reconstructed D^{*0} candidates and the associated neutral slow pion. Additionally, a lower momentum threshold is imposed on the slow pion, analogous to the selection applied in $D^{*+} \to D^0\pi^+_{\text{slow}}$ decays.
For the vertex fitting procedure, we apply a tree fit to the signal-side B vertex, with all daughter particle four-momenta updated accordingly. In contrast to the untagged analysis, we do not apply a mass constraint on the $D^{(*)}$ vertices. This decision is based on the observation that the mass constraint had minimal impact on the fit quality in the untagged analysis and the the individual vertex-related variables did not strongly improve the separation power of signal and background candidates. Thus, this introduced an unnecessary constraint on a quantity that is correlated with the invariant mass $M(D^{(*)}\pi)$, which serves as a key observable in the partial-wave analysis and was therefore removed.
A summary of the selection criteria applied during ntuple production (referred to as online cuts) is provided in Table 5.2.

After applying the standard online selection criteria, designed to be generally applicable and not tailored to a specific analysis, the final ntuples are produced. In this thesis, additional offline selection criteria are developed and optimized to further suppress background and enhance signal sensitivity for the targeted decay modes. The key observables used in this offline optimization process are outlined in the following.

$\cos^2\phi_B$: This variable encodes the geometric consistency of the reconstructed B mesons on the signal and tag sides. It is defined in terms of the angles $\cos\theta_{BY}^{\text{sig}}$ and $\cos\theta_{BY}^{\text{tag}}$ between the B meson and its visible decay products on both sides, here the $D^{(*)}\pi\ell$ system referred as Y, and the angle γ between the signal- and tag-side Y momenta and is given by

$$\cos^2\phi_B = \frac{\cos^2\theta_{BY}^{\text{sig}} + \cos^2\theta_{BY}^{\text{tag}} + 2\,\cos\theta_{BY}^{\text{sig}}\,\cos\theta_{BY}^{\text{tag}}\cos\gamma}{\sin^2\gamma}. \tag{5.1}$$

The individual $\cos\theta_{BY}$ terms are computed from energy and momentum conservation under the assumption of a massless neutrino:

$$\cos\theta_{BY} = \frac{2E_{\text{beam}}E_Y - m_B^2 - m_Y^2}{2|\vec{p}_B||\vec{p}_Y|}. \tag{5.2}$$

Properly reconstructed signal events are expected to yield physical values of $\cos^2\phi_B \leq 1$. However, due to detector resolution effects, a looser selection is applied. The optimal cut value is determined by maximizing Punzis figure of merit in Eq. (4.1). An illustration of the $\cos^2\phi_B$ geometry is provided in Fig. 5.1.

Number of charged tracks in the ROE: The ROE is expected to contain no additional charged particles in correctly reconstructed $\Upsilon(4S)$ signal events. To account for residual detector noise and

Table 5.2: Online cuts employed to all four B decay modes. Selection criteria adapted from [48].

	Variable	Selection criterion
Good tracks	d_r	< 2 cm
	$\lvert d_z \rvert$	< 4 cm
	polar angle θ	$\in [17, 150]$
Leptons	e-ID (or μ-ID)	> 0.9
	p_e^{CMS}	$\in [0.3, 2.4]$ GeV/c^2
	p_μ^{CMS}	$\in [0.6, 2.4]$ GeV/c^2
	standard lists	
	brems correction for e	
	good tracks	
Charged hadrons (K^+ and π^+)	kaonID (or pionID)	> 0.5
	nCDCHits	> 20
	good tracks	
π_{slow}	pionID	> 0.1
	$p_{\pi_{\mathrm{soft}}}^{\mathrm{CMS}}$	< 0.4 GeV/c
π^0	$M(\pi^0)$	$\in [0.120, 0.145]$ GeV/c^2
	standard list	
	(40% pion efficiency)	
K_S^0	flight distance	> 0
	$\log_{10}$("significance of distance")	> 0
	$\cos\theta(p_{K_S^0}, \mathrm{vertex})$	>0.8
	$\lvert M(K_S^0) - M_{K_S^0}^{\mathrm{PDG}} \rvert$	$<= 10$ MeV/c^2
	standard list	
Intermediate particles	$\lvert M(D^{0/+}) - M_{D^{0/+}}^{\mathrm{PDG}} \rvert$	< 15 MeV/c^2
	$M(D^{*+}) - M(D^0)$	$\in [0.141, 156]$ GeV/c
	$M(D^{*0}) - M(D^0)$	$\in [0.135, 156]$ GeV/c
Other selections	$\cos\theta_{BY}$	$\in [-15, 15]$
	High level trigger	on

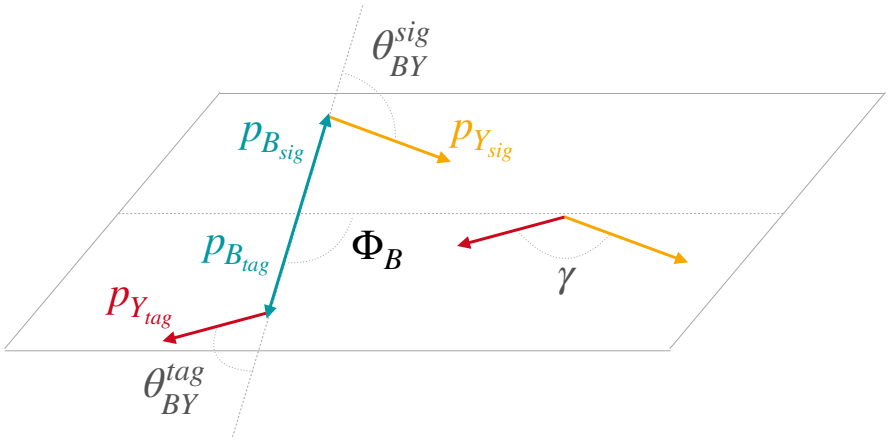


Figure 5.1: Illustration of the $\cos^2 \phi_B$ variable.

misreconstructed tracks, we allow up to one charged track in the ROE.

$E_{\text{extra}}^{\text{ECL}}$: This observable corresponds to the total energy deposited in the ECL by clusters not associated with the reconstructed decay products of the $\Upsilon(4S)$ system. In signal events, this quantity should be minimal and ideally reflect only calorimeter noise or undetected particles (e.g., neutrinos). Also here a cut is applied and optimized using Punzis figure of merit.

The selection cuts are applied in the order listed above. The order in which the selection cuts are applied is guided by the variable importance extracted from a BDT trained to discriminate signal from background. While the BDT itself is not used for event selection in this analysis, its variable ranking, specifically the frequency with which a variable appears at the first node of the decision tree, provides a measure of its discriminating power. Motivated by this ranking, a suitable order of variables is determined for the cut-based selection. For each variable, the cut value is then optimised using the Punzis figure of merit in Eq. (4.1), which is evaluated separately for each B decay mode. The final set of selection cuts, applied in the offline analysis, is summarised in Table 5.3. The distributions of the offline cut variables are shown in the appendix in Fig. C.1, Fig. C.2 and Fig. C.3.

It is found that the variable $\cos^2 \phi_B$ provides the highest separation power among all offline selection variables. Its impact on background suppression is substantial. This effect is illustrated in Fig. 5.2, which compares the $D^{(*)}$ mass spectrum before (i.e., directly after the reconstruction, applying only the online selection criteria) and after the application of the $\cos^2 \phi_B$ cut. A clear reduction in background is observed, leading to a significant improvement in the signal purity.

5.4 Event categorisation and first look

An initial view of the invariant $D^{(*)}\pi$ mass spectrum after applying all selection criteria discussed above is shown in Fig. 5.3. Similar to the untagged analysis (see Section 4.3), the events are categorized as follows:

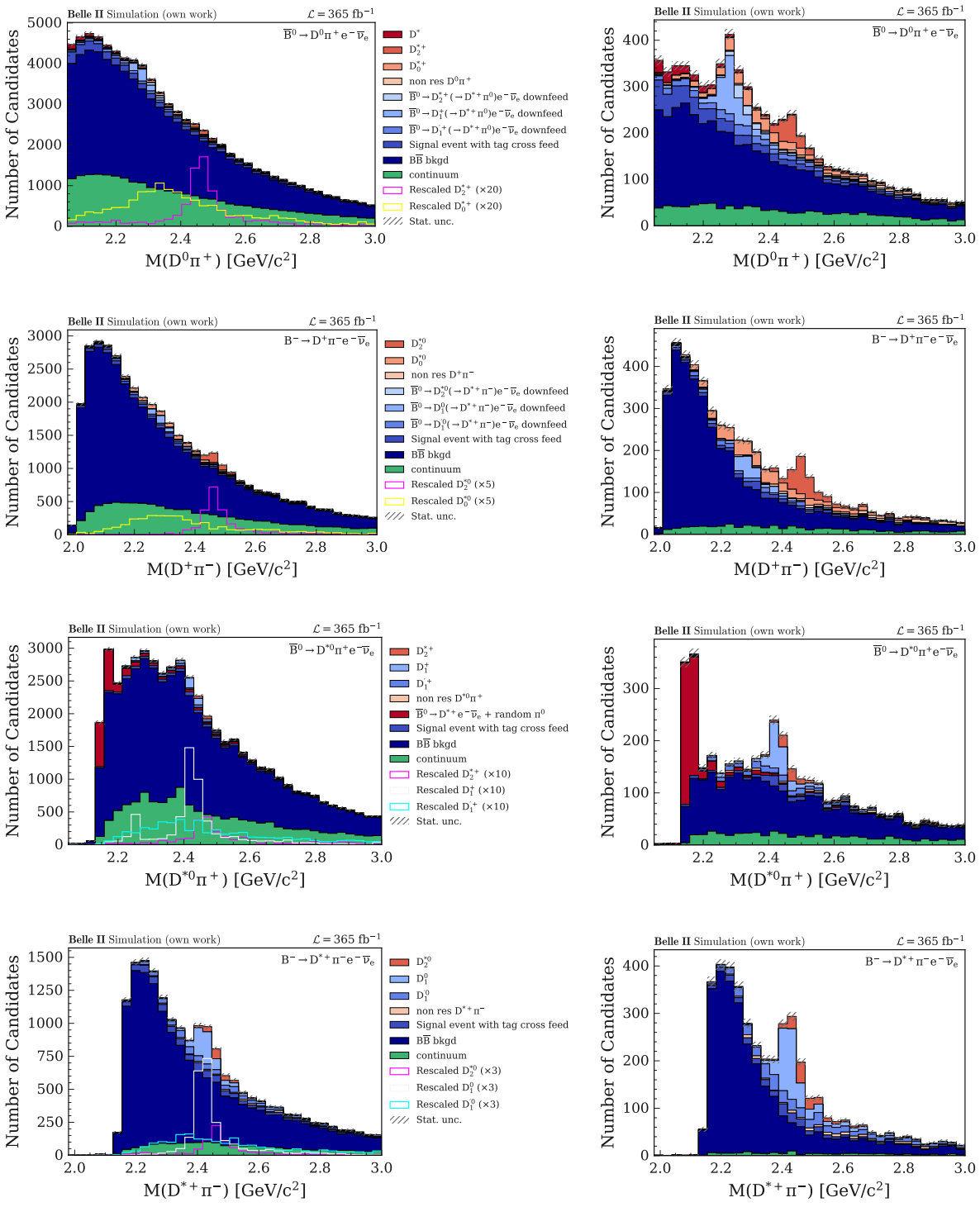


Figure 5.2: Impact on selection cut of $\cos^2 \phi_B$ variable for all four B decay modes. (Left) the $D^{(*)}\pi$ mass spectrum before the $\cos^2 \phi_B$ cut and without any other offline cut applied. (Right) the $D^{(*)}\pi$ mass distribution after the $\cos^2 \phi_B$ cut.

Table 5.3: Summary of optimized offline selection cuts used in the SL tagged analysis. The cut values are determined separately for each B decay mode by maximizing Punzis figure of merit in Eq. (4.1).

B decay mode	Variable	Selection cut
$\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$	$\cos^2\phi_B$	< 1.7
	n charged tracks in ROE	< 2
	$E^{\text{ROE}}_{\text{extra}}$	< 1.1
$B^- \to D^+\pi^-e^-\bar{\nu}_e$	$\cos^2\phi_B$	< 1.75
	n charged tracks in ROE	< 2
	$E^{\text{ROE}}_{\text{extra}}$	< 1.4
$\bar{B}^0 \to D^{*0}\pi^+e^-\bar{\nu}_e$	$\cos^2\phi_B$	< 1.65
	n charged tracks in ROE	< 2
	$E^{\text{ROE}}_{\text{extra}}$	< 1.5
$B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$	$\cos^2\phi_B$	< 2
	n charged tracks in ROE	< 2
	$E^{\text{ROE}}_{\text{extra}}$	< 1.52

- **Signal:** Events corresponding to the reconstructed signal decay modes, with all final-state particles correctly matched to the MC truth. Every resonances (such as D^*, D_0^*, D_2^*, D_1, D_1' and the corresponding non-resonant cases) are shown separately in the spectrum.

- **Self Cross Feed (SCF):** Events originating from a true signal decay but misreconstructed due to incorrect assignment of signal-side particles (e.g., pion–pion swaps within the decay chain) are refered as SCF events. SCF does not significantly affect the key observables such as $M(D\pi)$ or q^2, so it is typically included in the signal definition unless explicitly shown as its own category. Nevertheless, it has an effect on angle-based variables such as the hadronic helicity angle $\cos\theta_V$, which must be treated with caution in the later course of the analysis. In the decay mode $B^- \to D^+\pi^-e^-\bar{\nu}_e$, SCF is effectively excluded by design due to the opposite charges of the pions involved. In contrast, the channel $\bar{B}^0 \to D^{*0}\pi^0e^-\bar{\nu}_e$ has a significantly larger SCF contribution. This is primarily due to the presence of more than one neutral particle, which complicate the reconstruction and increase the likelihood of an incorrect assignment. Since suppressing the SCF in this channel proves difficult without significantly reducing the signal efficiency, these events are treated as part of the signal category. A more detailed discussion of the SCF fraction per B decay mode and their invariant $D^{(*)}\pi$-mass distribution can be found in Appendix C.3.

- **Tag Cross Feed (TCF):** Events in which particles from the tag side are incorrectly assigned to the reconstructed signal decay. These events distort kinematic variables and are treated as background.

- **Down-feed events:** For the semi-leptonic $B \to D\pi$ channels there are misreconstructed events originating from a $B \to D^{(*)}\pi\ell\nu$ event which we refer as down-feed events. This misreconstruction is due to the fact that it is actually a $\bar{B}^0 \to D_1^{(\prime)+}(\to D^{*0}\pi^+)\ell^-\bar{\nu}_\ell$ or $\bar{B}^0 \to D_2^{*+}(\to D^{*0}\pi^+)\ell^-\bar{\nu}_\ell$ event, where the D^{*0} decays into $D^0\pi^0$ and then the neutral pion is missed out in the reconstruction. The similar occurs in the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ channel, where some events originate from decays

such as $B^- \to D_2^{*0} e^- \bar{\nu}_e$ or $B^- \to D_1^{(\prime)0} e^- \bar{\nu}_e$, followed by the subsequent decay of $D_1^{(\prime)0}$, $D_2^{*0} \to D^{*+}(\to D^+\pi^0)\pi^-$. Again, if the neutral pion is not reconstructed, the event can be misidentified as a $D^+\pi^-$ final state, leading to a feed-down contribution.
These down-feed events lead to a peaking background in the $D^{(*)}\pi$ invariant mass spectrum. The peaks appear at masses lower than the nominal $D_1^{(\prime)}$ and D_2 resonance masses, due to the missing neutral pion in the reconstruction. In addition to that, the resulting peaks are broader.

- **Up-feed events:** In the $\bar{B}^0 \to D^{*0}\pi^+ e^- \bar{\nu}_e$ decay mode, it is possible that a $\bar{B}^0 \to D^0\pi^+ e^- \bar{\nu}_e$ event is wrongly reconstructed as a $B \to D^{*0}\pi$ decay because a randomly reconstructed π^0 is included. This type of misreconstruction is called an up-feed event. In most cases, the π^0 is not even a true π^0 and is incorrectly reconstructed from beam background photons.

- ***B*$\bar{B}$ background:** Events from generic $B\bar{B}$ decays that do not match any signal, cross-feed, down-feed or up-feed category.

- **Continuum background:** Events originating from $e^+e^- \to q\bar{q}$ with $q = u, d, s, c$. These contribute minimally to the background events.

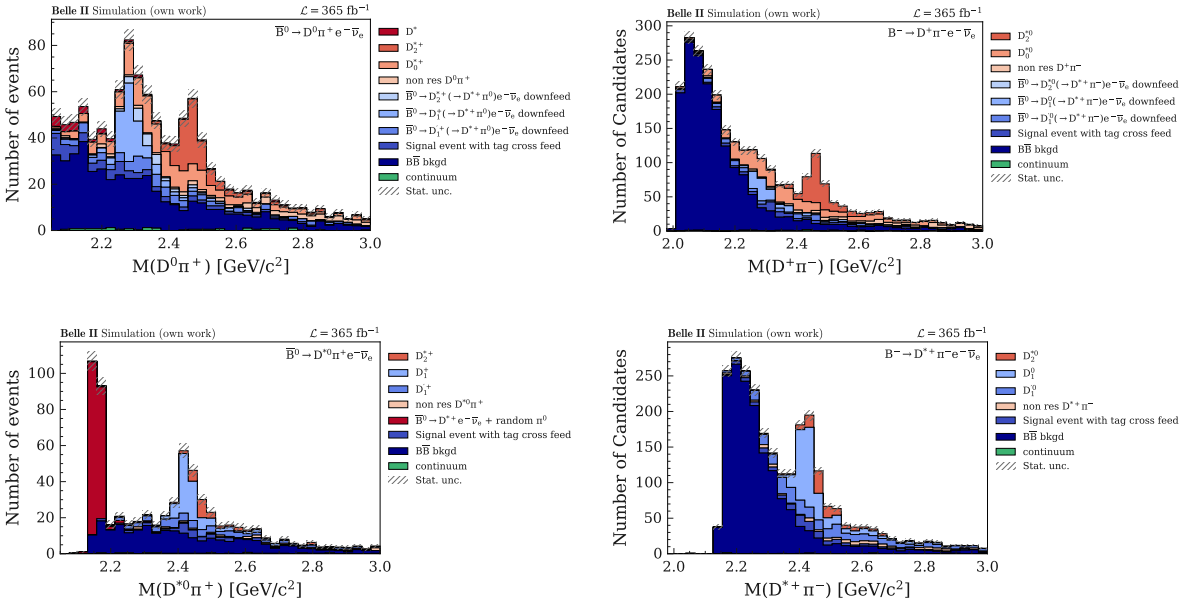


Figure 5.3: $D^{(*)}\pi$ mass spectrum after the event reconstruction and event selection cuts discussed in Section 5.3. Note that for the $\bar{B}^0 \to D^0\pi^+ e^- \bar{\nu}_e$ decay mode, the D^* resonance region is excluded from the plot.

What is apparent from a first look at the $D^{(*)}\pi$ invariant mass distribution is that for the charged B modes we have a neutral $B\bar{B}$ background tail. We will explore and try to suppress this in more detail in the next section.

5.5 Investigation of neutral $B\bar{B}$ background tail of charged B modes

To investigate the origin of the neutral background tail observed in the charged B modes, we examine the MC truth information. We first analyze the reconstructed decay on the tag side. There are two dominant B_{tag} decay modes that occur. In the reconstruction, we find a $B^+\text{tag} \to \bar{D}^0\ell^+\nu\ell$ decay, while the MC truth reveals that the true decay is actually a $B^0_{\text{tag}} \to D^{*-}(\to \bar{D}^0\pi^-)\ell^+\nu_\ell$ decay. In this case, the soft pion (π^-) from the D^{*-} decay is therefore missed out in the reconstruction.
On the signal side, we observe a reconstructed candidate for $B^-_{\text{sig}} \to D^+\pi^-\ell^-\bar{\nu}\ell$. In many of these cases, the π^- originates not from the signal-side decay but from the unreconstructed D^{*-} decay on the tag side. Thus one would suggest that, if the π^- from the tag side is combined with the reconstructed $\bar{D}^0$, the invariant mass of those two particles should peak around the D^- mass. This is indeed observed, as illustrated in Fig. 5.4. Based on this observation, we apply a D^* mass veto[2] to suppress these cross feed events with the tag side. This veto effectively removes background candidates with minimal loss of signal efficiency.
Secondly, we find cases where the tag-side reconstruction yields a $B^+\text{tag} \to \bar{D}^{*0}(\to \bar{D}^0\pi^0)\ell^+\nu\ell$ decay, while the true decay according to the MC truth is actually $B^0_{\text{tag}} \to D^+\ell^+\nu_\ell$. As in the previous scenario, the reconstructed $\bar{D}^0$ from the tag side is combined with a π^- from the signal side, leading to true D^- candidate. Therefore, to suppress this background, we again apply a D^{*-} mass veto on the $\bar{D}^0$ from the reconstructed tag side and the π^- from the signal side.
Both neutral $B\bar{B}$ background contributions arise from the same misreconstruction principle. However, the difference lies on a technical aspect. In the first case, the $\bar{D}^0$ originates directly from the B_{tag} decay, while in the second case, it is a daughter of the $\bar{D}^{*0}$, which itself is a daughter of B_{tag}. This results in two variables for which we can apply a D^* mass veto. The impact of applying the D^* mass is shown in Fig. 5.5. One can see qualitatively quite clear, that we lose a very large amount of the neutral $\bar{B}B$ background tail without any loss in signal efficiency. This selection cut was systematically studied and optimized within the scope of this master thesis, based on dedicated background studies and reconstruction performance.

5.6 Best Candidate Selection

After applying the analysis-specific selection criteria to suppress the dominant background contributions, multiple signal candidates remain within a single event. The distribution of the number of candidates per event is shown in Fig. 5.6. To proceed with the analysis, it is necessary to select a single candidate per event, ideally the one most consistent with originating from the true $\Upsilon(4S)$ decay. This procedure is referred to as the Best Candidate Selection (BCS).
In this analysis, the BCS is performed across all B decay channels. Specifically, we select the candidate with the highest FEI probability for the B_{tag}, under the assumption that a higher tag-side probability corresponds to a more accurately reconstructed event. This selection is done globally, allowing multiple candidates to remain on the signal side. To mitigate misreconstruction, particularly cases in which a $D^*\pi$ decay is mistakenly reconstructed as a $D\pi$ due to a missing pion, a veto is given to candidates reconstructed as $D^*\pi$ over $D\pi$, when both reconstructions are present. This strategy improves the reliability of the selected signal candidate and reduces down-feed contributions. After that we remain

[2] Quantitatively, this refers to the veto of the mass range $1.99 < M(D^*) < 2.03$ GeV.

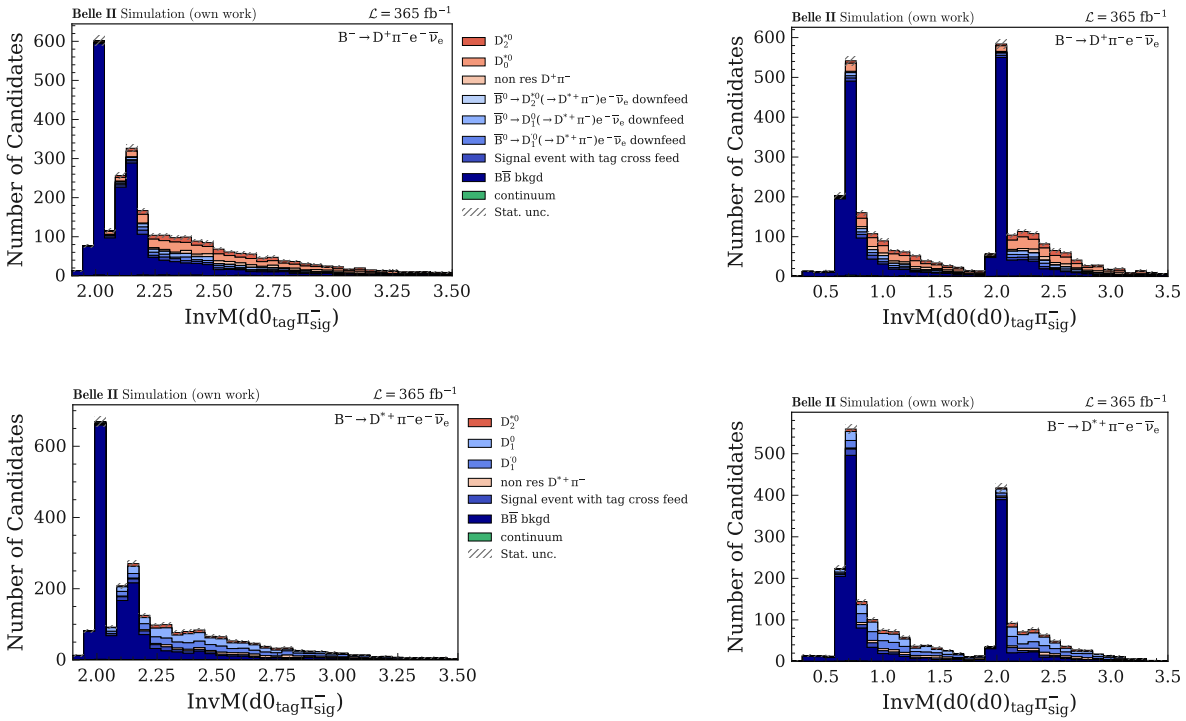


Figure 5.4: Distribution of combining the D^0 meson from the tag side with the pion from the signal side. One can see that one can apply a D^* mass veto without loosing many signal events.

with a few candidates per event, especially in the $D^0\pi^+$ channel and randomly choose the remaining candidate at that point. A random selection has the advantage of being statistically unbiased with respect to any kinematic or reconstruction variable.
The $D^{(*)}\pi$ invariant mass distribution before and after applying the BCS can be seen in Fig. 5.7. The FEI signal probability on which the BCS is performed on, is shown in Fig. C.8. Additionally, the distributions of q^2 and $\cos\theta_V$ after applying the BCS can be found in the appendix in Fig. C.9 and C.10 respectively. No significant improvement in purity is observed in the invariant mass distribution of the $D^{(*)}\pi$ system. This observation is further supported quantitatively by the comparison of signal yields before and after applying the Best Candidate Selection (BCS), summarized in Table 5.4. In the development of the BCS procedure, various selection criteria were explored to identify the most reliable candidate per event. These included, for example, selecting the candidate with the highest χ^2 vertex fit probability of the B meson, or performing a purely random selection to avoid kinematic bias. However, choosing the candidate based on the highest FEI signal probability for the B_{tag} consistently yielded the best overall performance in terms of signal purity.

5.7 Signal sensitivity

To evaluate the performance of our analysis, we compare the extracted signal yields with those obtained in the hadronic tagged Belle measurement from 2020 [24]. A summary of the corresponding signal and background yields is presented in Table 5.5.

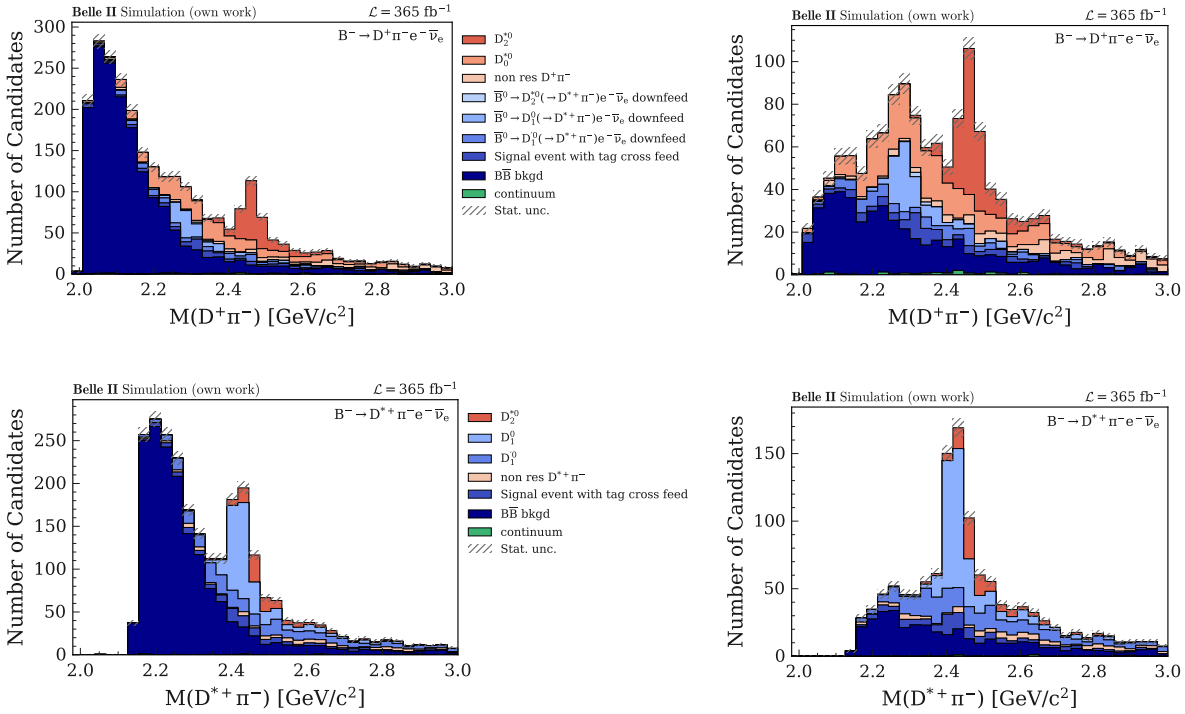


Figure 5.5: Impact of combining the D^0 meson from the tag side with the pion from the signal side and applying a D^* mass veto. (Left) $M(D^{(*)}\pi)$ distribution before a the cut and (right) after the cut is applied.

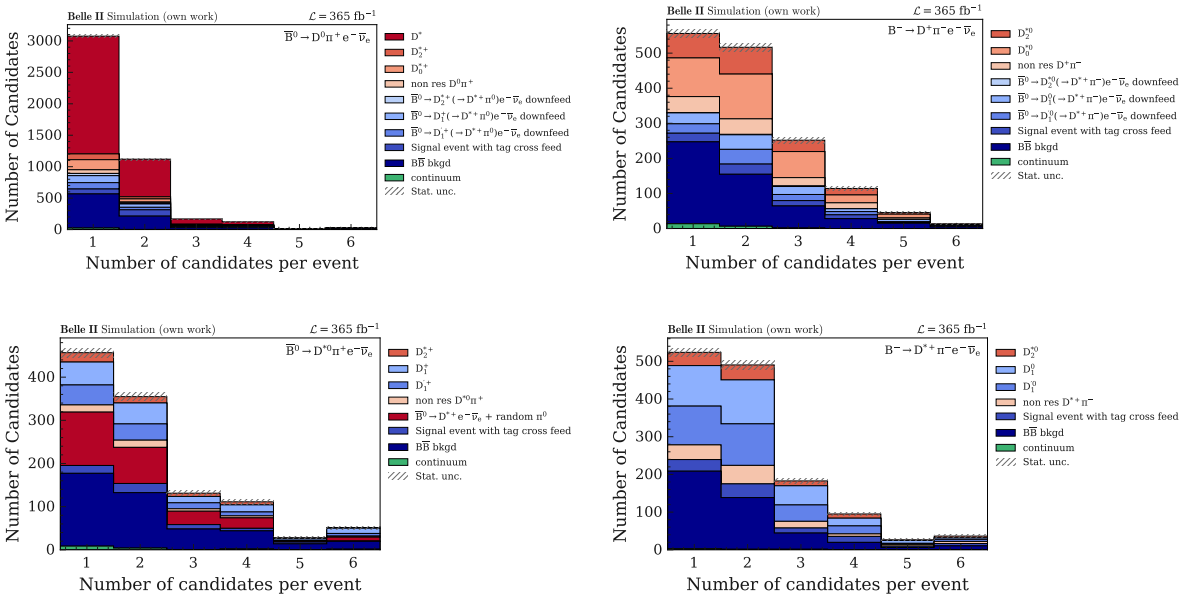


Figure 5.6: Number of candidates per event after all event selection cuts are applied. We find an average of 1.47, 2.18, 2.75, 2.34 number of candidate per event for the $D^0\pi^+$, $D^+\pi^-$, $D^{*0}\pi^+$ and $D^{*+}\pi^-$ modes respectively.

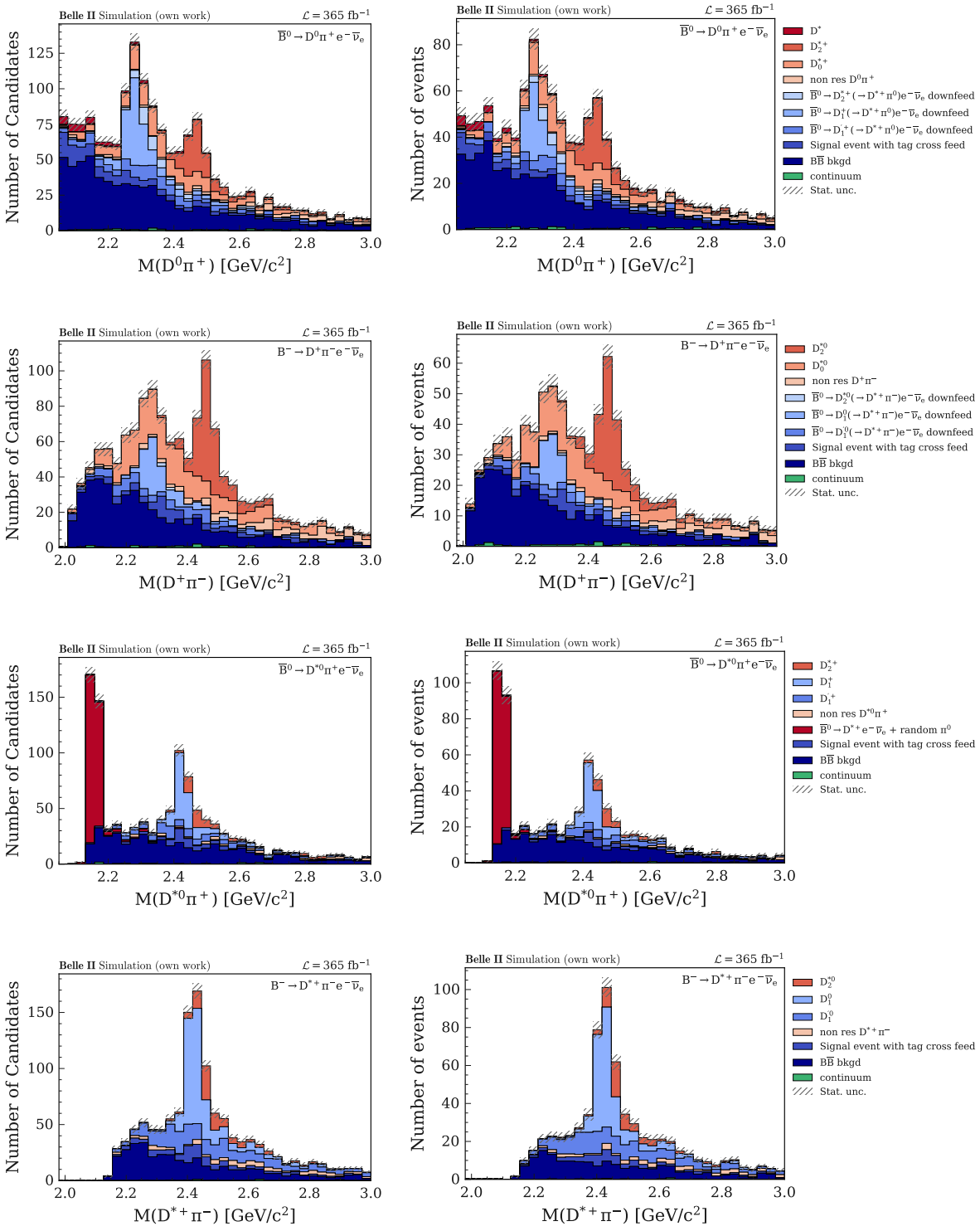


Figure 5.7: $M(D^{(*)}\pi)$ distribution before (left) and (right) after applying the BCS.

Table 5.4: Signal (S) and background (B) yields before and after applying the BCS. This yields corresponds to an integrated luminosity of $\mathcal{L} = 365\,\text{fb}^{-1}$.

Decay mode	S_{before}	B_{before}	S_{after}	B_{after}	$\frac{S}{S+B}_{\text{before}}$	$\frac{S}{S+B}_{\text{after}}$	$S_{\text{after}}/S_{\text{before}}$
$\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$	3065 ± 55	1477 ± 38	2409 ± 49	920 ± 30	0.675 ± 0.007	0.724 ± 0.008	0.786 ± 0.021
$B^- \to D^+\pi^-e^-\bar{\nu}_e$	693 ± 26	820 ± 29	409 ± 20	510 ± 23	0.458 ± 0.013	0.445 ± 0.016	0.590 ± 0.037
$\bar{B}^0 \to D^{*0}\pi^+e^-\bar{\nu}_e$	399 ± 20	821 ± 29	229 ± 15	489 ± 22	0.327 ± 0.013	0.319 ± 0.017	0.574 ± 0.048
$B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$	844 ± 29	546 ± 23	497 ± 22	276 ± 17	0.607 ± 0.013	0.643 ± 0.017	0.589 ± 0.033

For the $B \to D\pi\, e\, \nu_e$ decay modes, we observe higher signal yields in this work, achieved under more favorable background conditions. In the case of the $B^- \to D^{*+}\pi^- e\bar{\nu}_e$ mode, despite being the cleanest channel in our analysis, we also achieve higher signal yields compared to the Belle measurement, though this comes at the cost of roughly three times more background events.
The discrepancy becomes even more pronounced in the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ channel. Here, our signal yield is approximately half that of the Belle 2020 result, while the background yield is an order of magnitude larger. This behavior is not unexpected, as the presence of a neutral pion in the final state significantly deteriorates the background conditions in the SL tagged analysis, compared to hadronic tagging. This is primarily due to the poorer reconstruction efficiency and resolution of neutral particles, particularly $\pi^0 \to \gamma\gamma$ decays, in SL-tagged environments. The SL tagged analysis already involves two undetected neutrinos, one on each B meson side, resulting in limited kinematic constraints and further complicating the reconstruction and background suppression in final states involving neutral particles.

Table 5.5: Comparing the signal and background (B) yields (and their statistical uncertainties) of this work with the yields from the hadronic tagged Belle measurement in 2020 [24]. The signal and background yields from Belle correspond to the yields extracted from a fit in $U = E_{\text{miss}} - p_{\text{miss}}$. For better comparison the yields from this work are scaled to the Belle luminosity of $\mathcal{L} = 711\,\text{fb}^{-1}$.

	Signal		Background	
	Belle 2020	This work	Belle 2020	This work
$\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$	570 ± 35	680 ± 26	2641 ± 80	1792 ± 42
$B^- \to D^+\pi^-e^-\bar{\nu}_e$	721 ± 32	797 ± 28	1329 ± 53	1012 ± 32
$\bar{B}^0 \to D^{*0}\pi^+e^-\bar{\nu}_e$	798 ± 44	446 ± 21	77 ± 12	953 ± 31
$B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$	787 ± 35	968 ± 31	172 ± 19	538 ± 23

The impact of including neutral particles in the signal decay chain on background levels becomes further evident when examining the signal-to-background ratios of the individual D decay modes, as shown in Fig. 5.8. In particular, for the $D^{*0}\pi^+$ channel, it is clearly observed that the D decay mode involving additional neutral particles exhibit significantly worse signal-to-background performance. Here events labelled as background includes TCF, down- and up-feed, other $B\bar{B}$ background and continuum events.

5.8 Reconstruction of kinematic variables

A fundamental challenge of employing the SL tagged approach in this analysis lies in the inability to directly access the rest frame of the signal-side B meson. Unlike in hadronic tagged analyses, where the

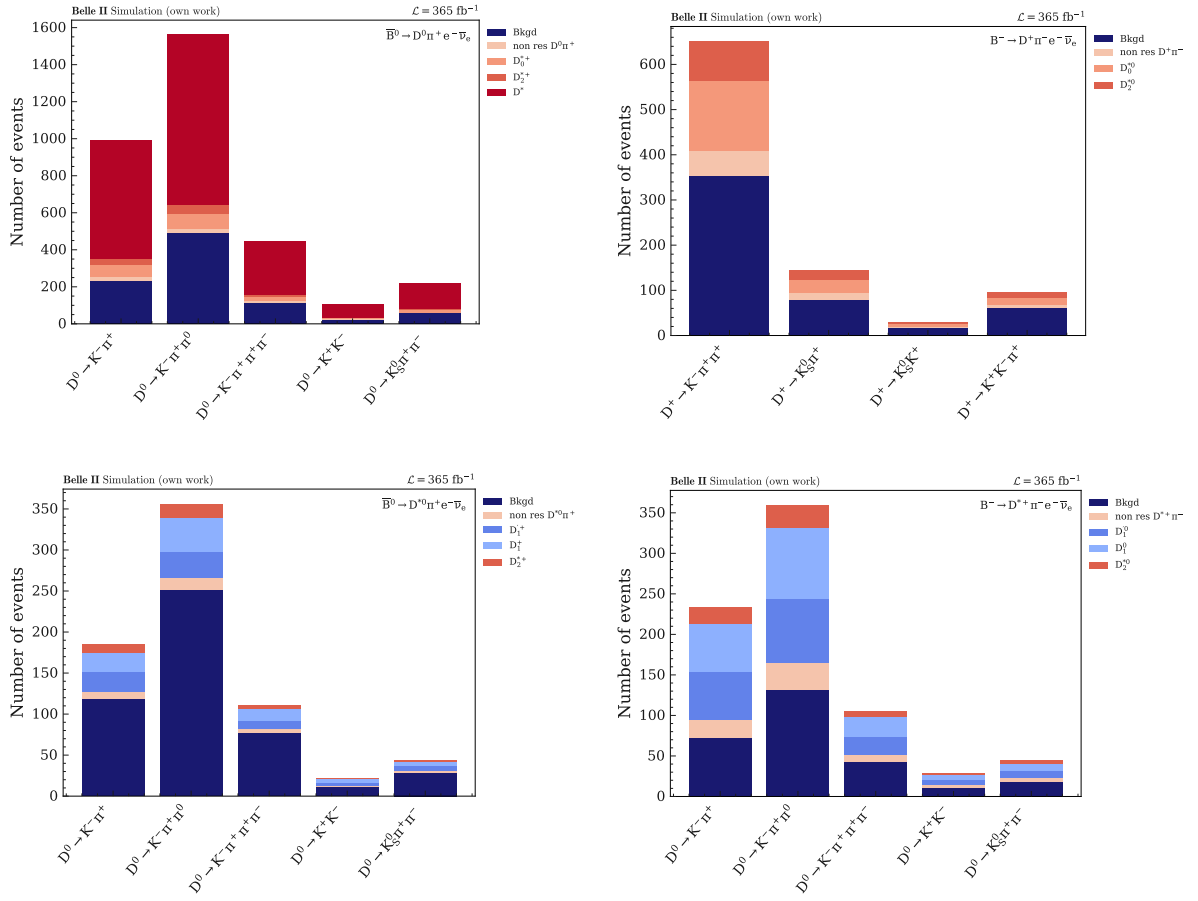


Figure 5.8: Signal and background yields for all four B decay modes separated in the different D modes.

second B meson is fully reconstructed and the missing neutrino on the signal side can be inferred from the beam-energy constraint, the SL tagged strategy suffers from an additional undetected neutrino in the tag-side decay. As a result, the full four-momentum of the signal-side B meson cannot be reconstructed, which limits the kinematic information available for the analysis.

While the magnitude of the B meson's momentum still can be estimated from the known beam energy, the direction of the B meson remains ambiguous. The variable $\cos\theta_{BY}$ encodes the angle between the B meson and the reconstructed Y system (here $D^{*(*)}\ell$), and thus constrains the B momentum vector to lie on the surface of a right circular cone. The axis of this cone is aligned with the momentum vector of the Y system, and its opening angle is given by $2\theta_{BY}$. However, since the azimuthal orientation around the Y axis is undetermined, the precise direction of the B meson on the cone cannot be uniquely reconstructed. This geometrical ambiguity poses a challenge for reconstructing kinematic observables that rely on knowledge of the full B meson momentum. Specifically if it comes to q^2 and the hadronic helicity angles $\cos\theta_V$, which are of interest in this analysis.

Several methods have been proposed to approximate the direction of the B meson momentum under these constraints. In this analysis, we adopt the so-called *diamond frame* approach. The details of this method are discussed in the following section.

5.8.1 Diamond frame approach

Since the only missing information of the B meson momentum is the azimuthal angle Φ_B, the diamond frame approach approximates the relevant kinematic variables by averaging over four representative values of Φ_B, 0, $\pm\pi/2$, and π. This discrete sampling effectively spans the cone of possible B directions described earlier. Each of these four configurations is then weighted by $\sin^2\theta_B$, where θ_B denotes the polar angle of the B meson. This weighting reflects the expected angular distribution of $B\bar{B}$ production in the $\Upsilon(4S)$ rest frame. An illustration of this method is shown in Fig. 5.9. The performance of the diamond

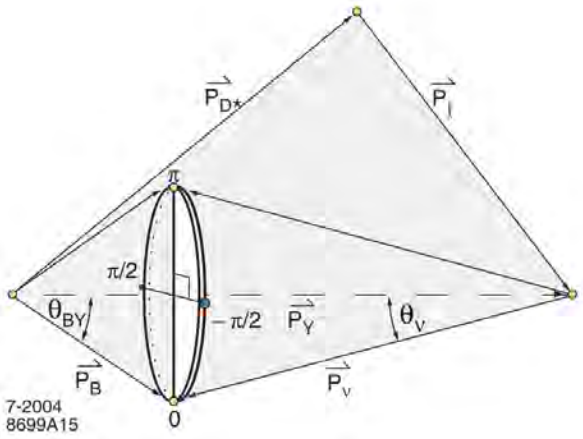


Figure 5.9: Illustration of the diamond frame approach. Figure taken from [26]

frame method can be evaluated by comparing the reconstructed kinematic variables obtained through this approach with the corresponding MC truth information, where the full B meson momentum is known. As shown exemplary for the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ decay mode in Fig. 5.10, this comparison demonstrates that the diamond frame method provides a reliable approximation of the B meson kinematics. In addition to the residuals of the q^2 and $\cos\theta_V$ variables obtained using the diamond frame approximation, we also show the resolution of the invariant $D\pi$ mass. While the $D\pi$ mass is not affected by the diamond frame construction, it is included here as it represents a key observable in the analysis.

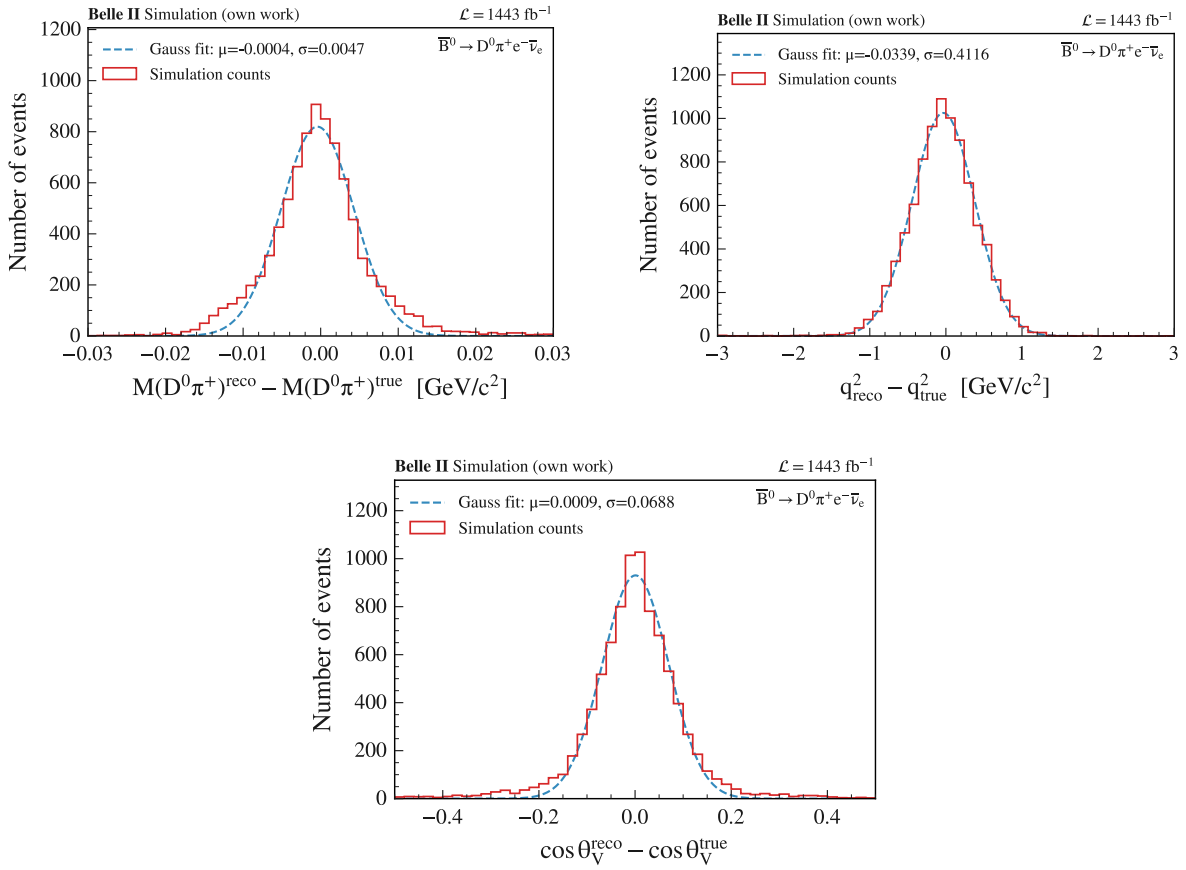


Figure 5.10: Distributions of residuals for kinematic variables $M(D^{(*)}\pi)$, q^2 and $\cos\theta_V$ exemplary shown for the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ mode. The two variables are q^2 and $\cos\theta_V$ reconstructed using the diamond frame approach. It is important to note that a Gaussian function is fitted to these distributions solely to obtain a rough estimate of the mean and standard deviation. This approach is used for convenience, even in cases where the distributions deviate significantly from a true Gaussian shape, and is not intended to model the underlying distribution accurately.

CHAPTER 6

Signal extraction fit setup

In order to extract the signal contribution from the selected events, we aim to perform a fit-based background subtraction. This method requires a variable that exhibits distinct distributions for signal and background, such that the signal yield can be extracted from the total sample. In the untagged study of $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$, the variable $\cos\theta_{BY}$ was used for this purpose due to its discriminating power. In the present analysis, however, $\cos\theta_{BY}$ is not suitable for signal extraction since this variable is highly correlated with $\cos^2\phi_B$, on which a tight selection cut has already been applied to suppress the background. As a consequence, the remaining $\cos\theta_{BY}$ distribution shows mostly contribution from peaking background components, reducing its effectiveness for clean signal extraction. The corresponding distributions are shown in Appendix Fig. D.1.
To overcome this limitation, we train a BDT using the variables that were employed for the event selection. With this, the goal is to exploit additional information contained in the probability density function (pdf) shape of these variables, which may still hold discriminative power even after already used for event selection. The output of the BDT, namely the signal probability score, is then used as the discriminating variable in a fit to extract the signal yield and thus subtract the background.

6.1 BDT training

Following the BDT training used for background suppression in the untagged analysis Section 4.4, we employ the `XGBoost` framework to train a BDT for signal extraction. Hyperparameter optimization is also again performed using the `Optuna` package, where we focus on the regularisation terms α and λ only. To prevent overfitting, we constrain the maximal depth of the decision trees to 3 in this analysis.
To enhance the statistical power of the training and increase the signal statistics, we augment the available MC15rd samples with additional, own produced run-independent signal Monte Carlo samples for this study. Similarly, to increase the background statistics, we supplement the training set with generic MC15ri samples. However, one must note that run-independent and run-dependent samples differ significantly in several variables, particularly in $E^{\mathrm{ROE}}_{\mathrm{extra}}$, as shown for the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ decay in Fig. 6.1. These discrepancies reflect differences in beam and detector conditions between the two MC configurations. As a consequence, the BDT signal probability score evaluated on run-dependent samples differs from that on run-independent samples, indicating reduced generalizability of the model. Ideally, the BDT should be trained exclusively on run-dependent signal and generic MC samples, as they

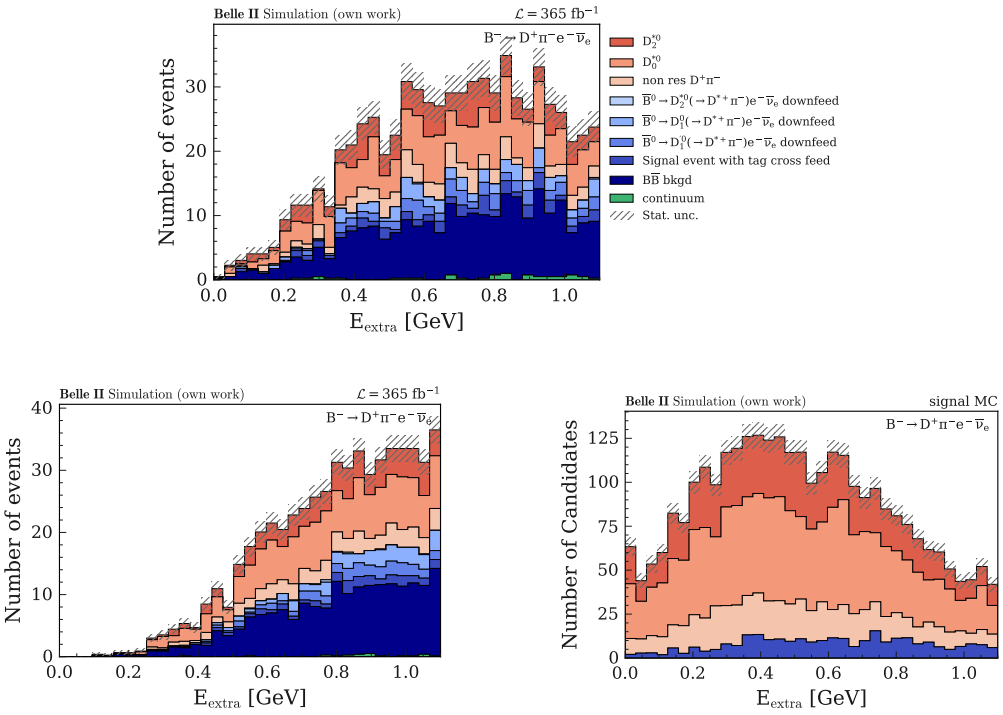


Figure 6.1: Distributions of E_{extra} for the MC15rd (upper plot), MC15rd (lower left) and ri signal MC (lower right) samples, exemplary shown for the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ decay mode. Note that the blue component in the signal MC distribution corresponds to signal events with TCF.

better reflect realistic background conditions relevant for data. However, generating sufficiently large run-dependent samples for this purpose would be computationally demanding.
Given the purpose of this thesis, which primarily aims to assess the feasibility of a SL tagged analysis of $B \to D^{(*)}\pi\ell\nu$ decays, we proceed with the current mixed training set strategy. To mitigate the aforementioned inconsistencies, we thus include both generic run-dependent and run-independent samples, as well as the run-independent MC sample, in the training set and apply additional reweighting to improve robustness. Moreover, as in the BDT training for the untagged analysis, we employ k-fold cross-validation to monitor overfitting and ensure stable performance across the full dataset. The combined dataset is partitioned into 80% for training and 20% for the test set.
For the further course of this analysis, we use the MC15rd generic sample while acknowledging that its performance may be slightly optimistic due to partial overlap with the training dataset.

For the BDT input variables we use slightly different sets of variables for every individual B decay mode. Distribution of these for every MC sample can be found in the appendix in Appendix D.2.

The binary classification of the BDT distinguishes between signal and background events. The signal category comprises all semi-leptonic decays of the B meson into D^{**} and D^* states, encompassing both resonant and non-resonant contributions. All D^{**} resonances are treated collectively as a single signal class.
It was found that the inclusion of down-feed and up-feed events in the signal category during BDT training significantly enhances the classifier's performance, leading to a stronger separation between

signal and background in the signal probability output. As a result, these events are retained in the signal class for the time being. A reassignment to their corresponding true decay modes will be addressed later in the migration procedure described in Section 7.1.

6.2 BDT input variables on real data

Additionally to the already discussed BDT training procedure, we apply correction weights to the generic MC15rd sample, to ensure that the BDT score and overall performance remain robust and consistent when applied to real data. To assess the effectiveness of this correction and the impact of our applied selection cuts in the previous analysis step, we further compare the distributions of BDT input variables from the correction-weighted, run-dependent MC samples with those from real Belle II data. In the following section, we first briefly describe the correction weights applied before comparing the MC simulation with Belle II data.

6.2.1 Corrections for simulated samples

Due to imperfections in detector and interaction modeling, the reconstruction efficiency and purity can differ between MC simulations and experimental data. To account for these discrepancies, correction weights, provided by the Belle II performance group, are applied to MC events. The applied corrections depend on the particle type, its kinematics, and the specific decay mode reconstructed.

Slow Pion Efficiency Correction Due to the small mass difference between D^* and D, the pion emitted in $D^* \to D\pi$ decays has low momentum, typically below 200 MeV/c. This results in reconstruction efficiency differences between MC and data [68]. Correction weights, provided in three bins of slow pion momentum, are applied with associated statistical and systematic uncertainties.

Lepton ID Efficiency Correction The lepton identification efficiency correction is derived from the two-photon process $e^+e^- \to e^+e^-\ell^+\ell^-$, and validated using inclusive $B \to XJ/\psi(\to \ell^+\ell^-)$ decays to account for hadronic effects. Correction factors are provided in bins of lepton momentum, polar angle, and identification criteria, and are applied to correctly reconstructed leptons.

Lepton Fake Rate Correction The probability for a lepton fake rate is defined as a hadron track to be wrongly identified as a lepton track. The misidentification rates for pions are studied using control channels such as $K^0_S \to \pi^+\pi^-$ and $e^+e^- \to \tau^\pm(1p)\tau^\mp(3p)$. Kaon misidentification rates are determined from $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$. Correction weights are provided as a function of the lepton momentum and polar angle in the laboratory frame, binned in two dimensions with corresponding central values and uncertainties.

Hadron PID Efficiency and Fake Rate Efficiencies and misidentification rates for charged kaons and pions are corrected using the decay $D^* \to D^0(\to K^-\pi^+)\pi^+$. Correction tables are binned in momentum and polar angle.

Additional correction weights, such as those related to the SL FEI calibration, are available and may be necessary for the further course of this analysis when using experimental data. However, since this

study is currently based solely on MC simulation, we stop here. The impact of the remaining correction weights is expected to be subdominant.
Therefore, we now can compare the distributions of the BDT input variables in data and MC simulation to verify that the selection criteria perform similarly in both. This validation is particularly important, as the BDT signal probability distribution obtained from MC is later used in a fit to extract the signal yields. Ensuring good agreement between data and MC in the input variables increases confidence in the reliability of the BDT model and the robustness of the signal extraction procedure. This consistency check also reinforces the overall credibility and validity of the analysis presented in this thesis. Figure 6.2, 6.3, 6.4 and 6.5 shows the BDT input variable distributions comparing the MC simulation with Belle II data for each B decay mode On the whole, the agreement between data and MC simulation looks quite good and there are no worrying outliers. This therefore emphasises the reliability of this work.

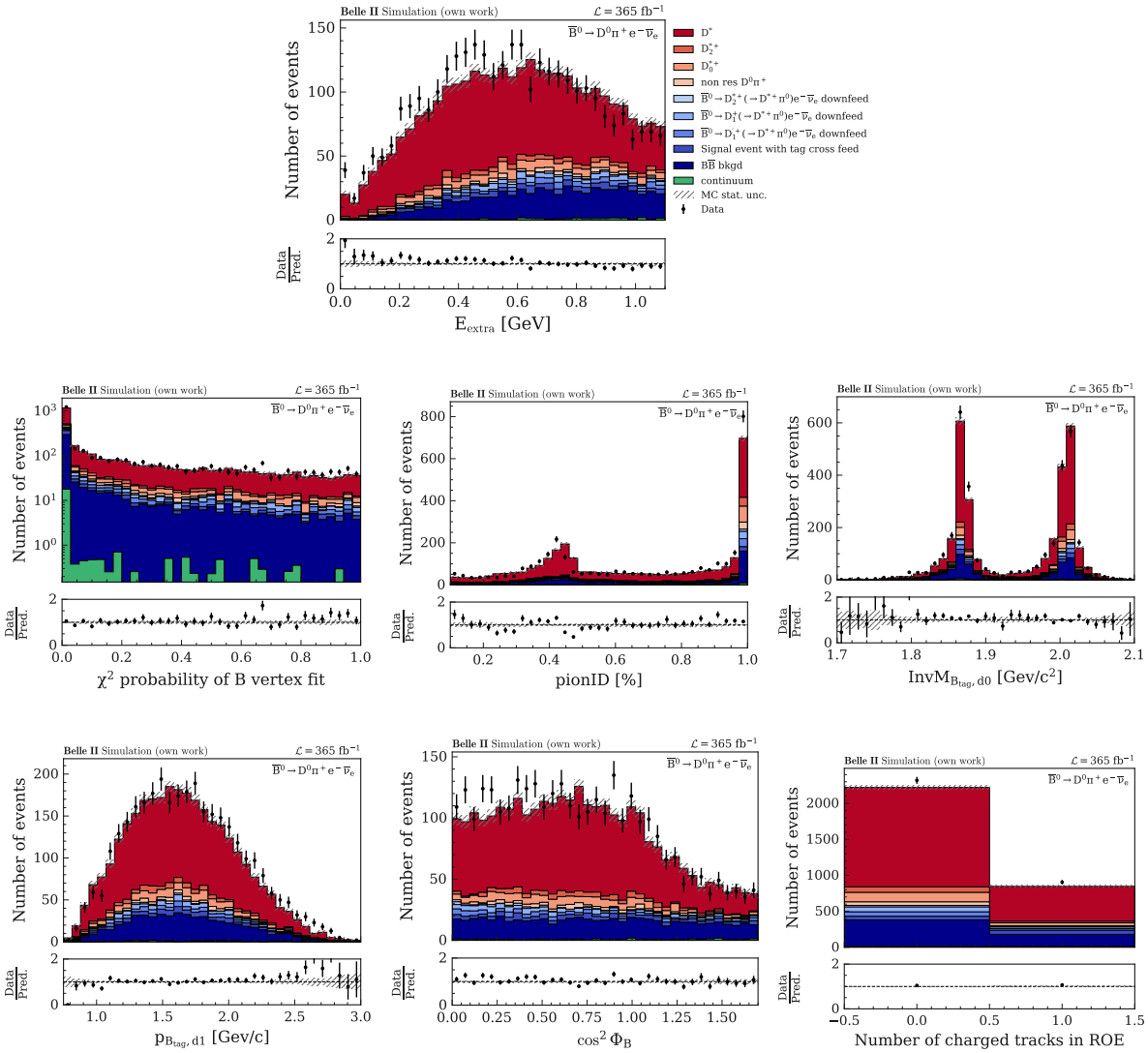


Figure 6.2: BDT input variables for the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode comparing MC simulation with real Belle II data.

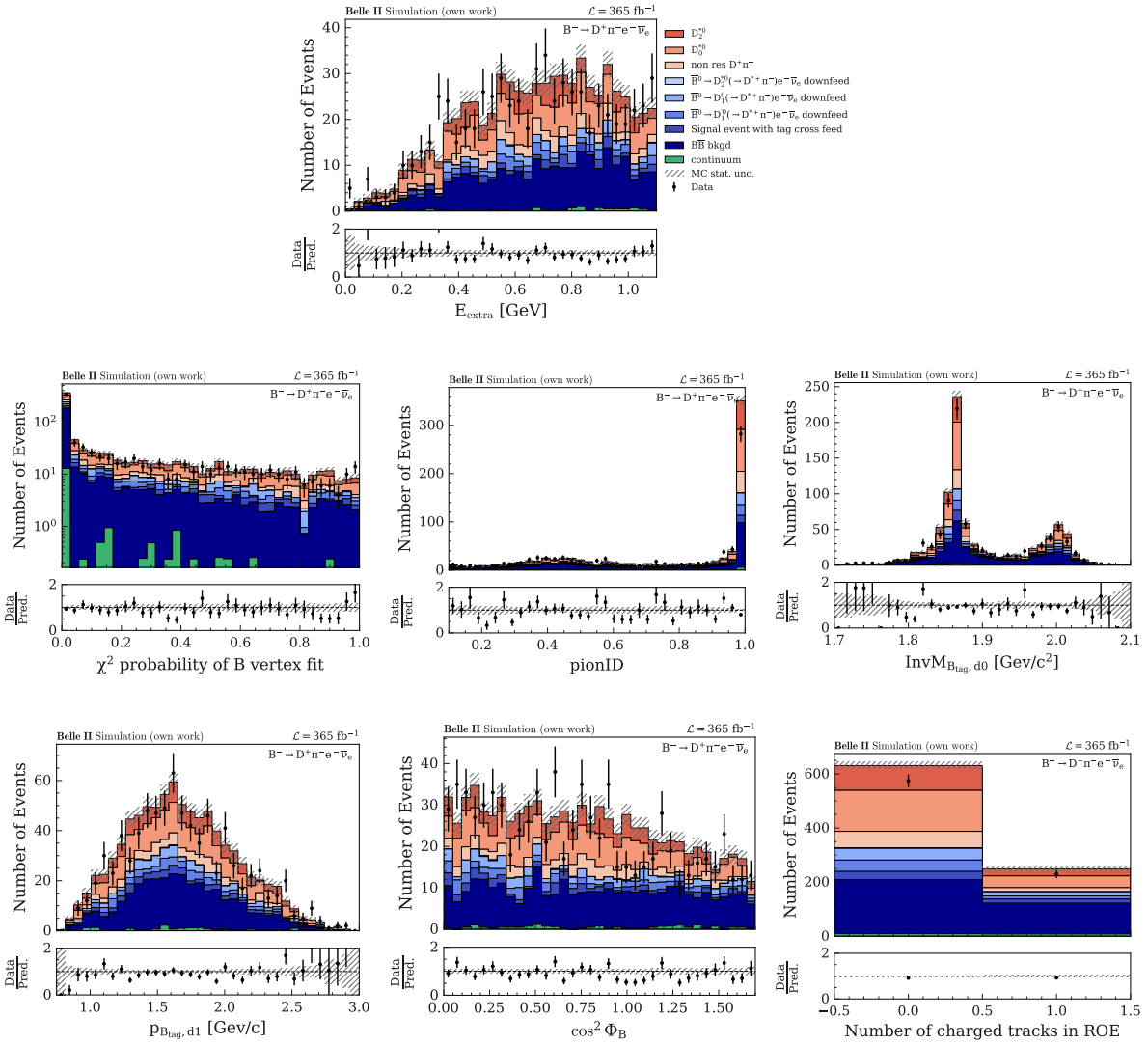


Figure 6.3: BDT input variables for the $B^- \to D^+\pi^- e^- \overline{\nu}_e$ decay mode comparing MC simulation with real Belle II data.

6.3 BDT performance and signal probability score

The performance of the trained BDT model is evaluated by examining the signal probability distributions and ROC curves. A comparison between the test and training datasets is shown in Fig. 6.6 to assess potential overtraining. For the $D\pi$ modes, the ROC curves indicate consistent performance between the training and test sets, suggesting that the model generalizes well.
In contrast, for the $D^*\pi$ modes, a noticeable discrepancy is observed and the training set achieves higher classification performance than the test set. This effect could be explained due to the limited statistics available in these modes, leading to relatively larger statistical uncertainties in the ROC evaluation (i.e. potentially larger error bands around the curve). When accounting for these uncertainties, the model's

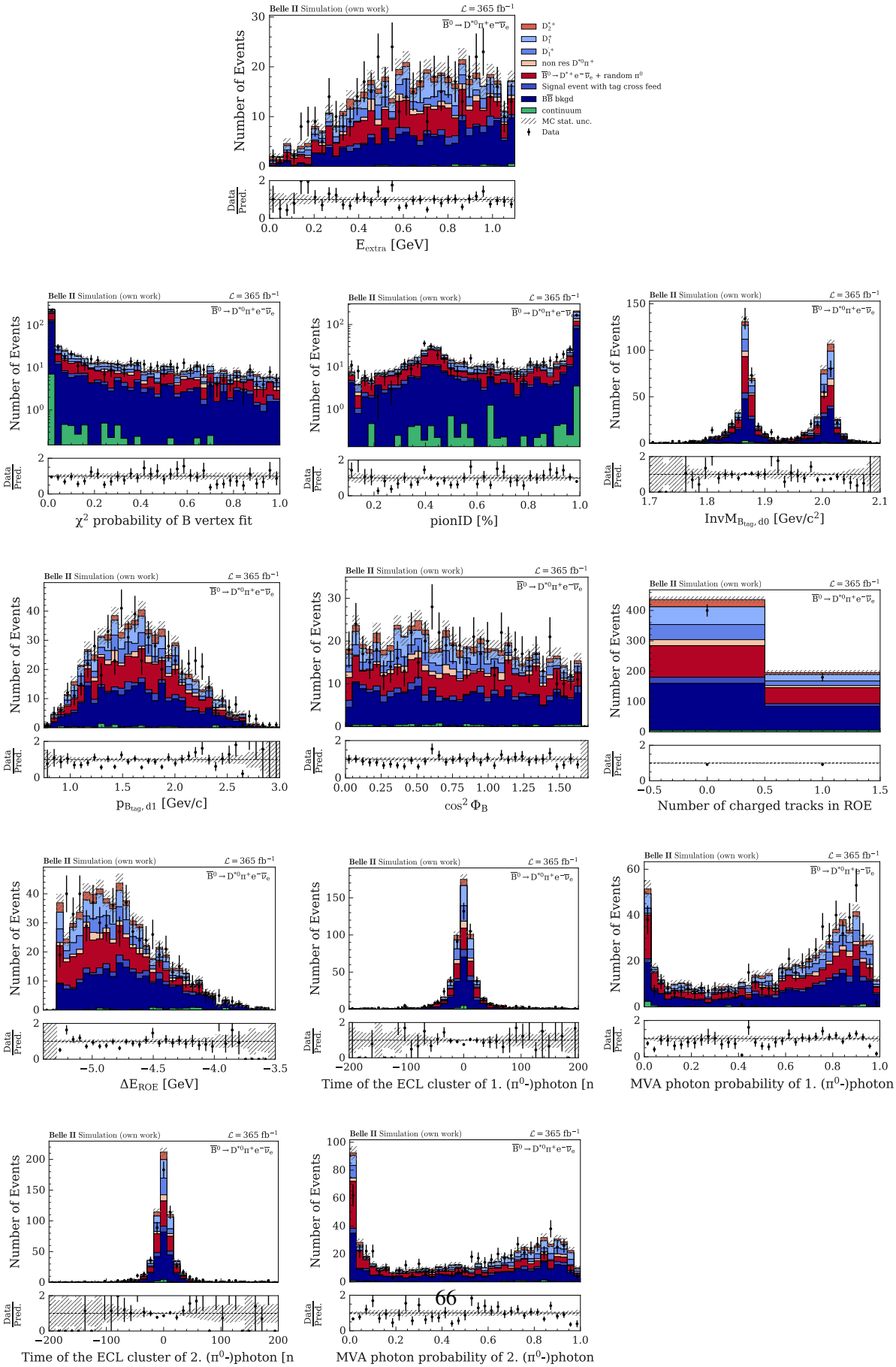


Figure 6.4: BDT input variables for the $\overline{B}^0 \to D^{*0}\pi^+e^-\overline{\nu}_e$ decay mode comparing MC simulation with real Belle II data.

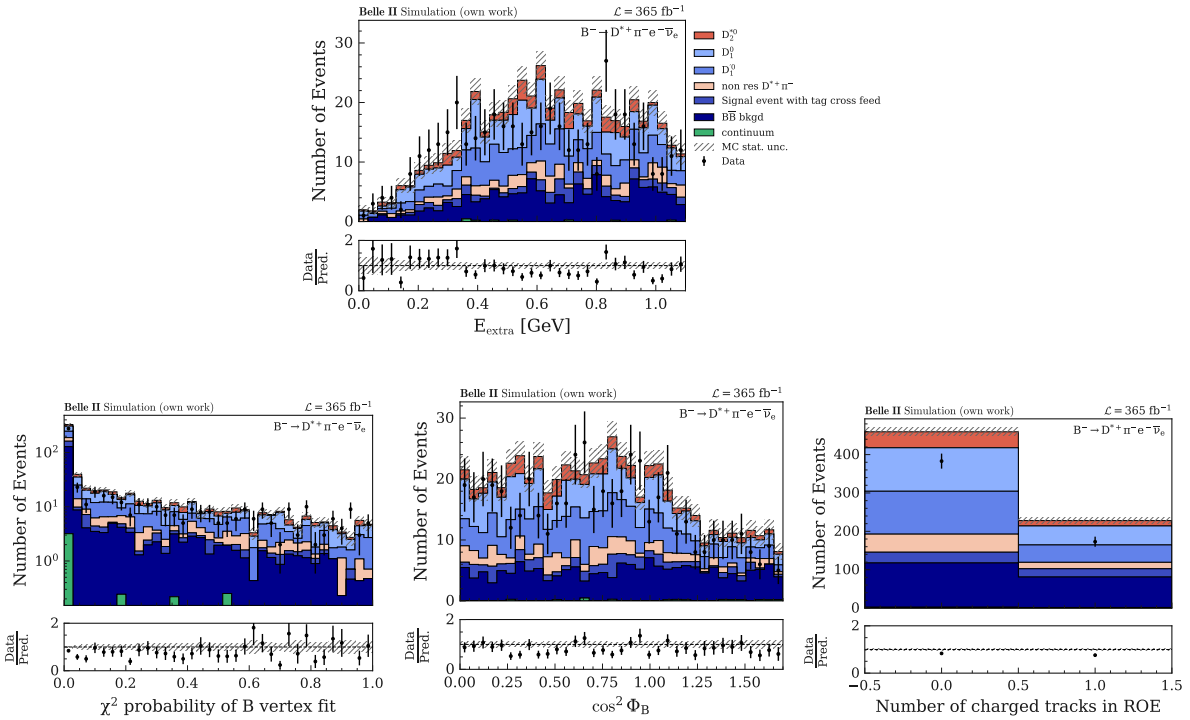


Figure 6.5: BDT input variables for the $B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$ decay mode comparing MC simulation with real Belle II data.

performance remain acceptable.
This is further supported by the signal probability distributions shown in **??**, where larger statistical fluctuations are visible in the test sample, particularly for the low-statistics modes, but the overall agreement between training and test sets remains reasonable. Ideally, the signal probability output from the BDT classifier should yield values close to 1 for signal events and close to 0 for background events, resulting in well-separated distributions. However, in our case, similar to the untagged analysis, this ideal separation is not achieved, and the classifier exhibits limited discriminative power between signal and background events, even in the SL-tagged analysis.
Nonetheless, for the purpose of signal extraction, the absolute separation between signal and background is less critical than the distinguishability and parameterizability of their respective BDT score distributions. As long as the signal and background templates can be reliably modeled and fitted to data, the BDT score remains a useful variable for performing statistical signal extraction.
Regarding this, the BDT score distributions show a reasonable level of separation between signal and background across most decay channels. The main exception is the $\bar{B}^0 \to D^{*0}\pi^+e^-\bar{\nu}_e$ channel, where the signal and background distributions exhibit significant overlap, limiting the effectiveness of a the signal extraction fit.
Furthermore, a look at the feature importance of the BDT input variables in Fig. 6.8 shows also for the SL tagged analysis that the variable with the highest information gain is the χ^2 probability of the B vertex fit. While further improvements may be possible, such as refining more hyperparameter optimization, incorporating additional input variables, or exploring alternative machine learning models, no dominant

background source could be clearly identified by a look into the invariant $D^{(*)}\pi$ mass spectra (Fig. 5.7) or to MC truth information. Given these considerations, the current BDT configuration represents the most effective solution within the available constraints. Therefore, we proceed with this setup in the subsequent stages of the analysis and will now turn our attention to the signal extraction fit in the next step.

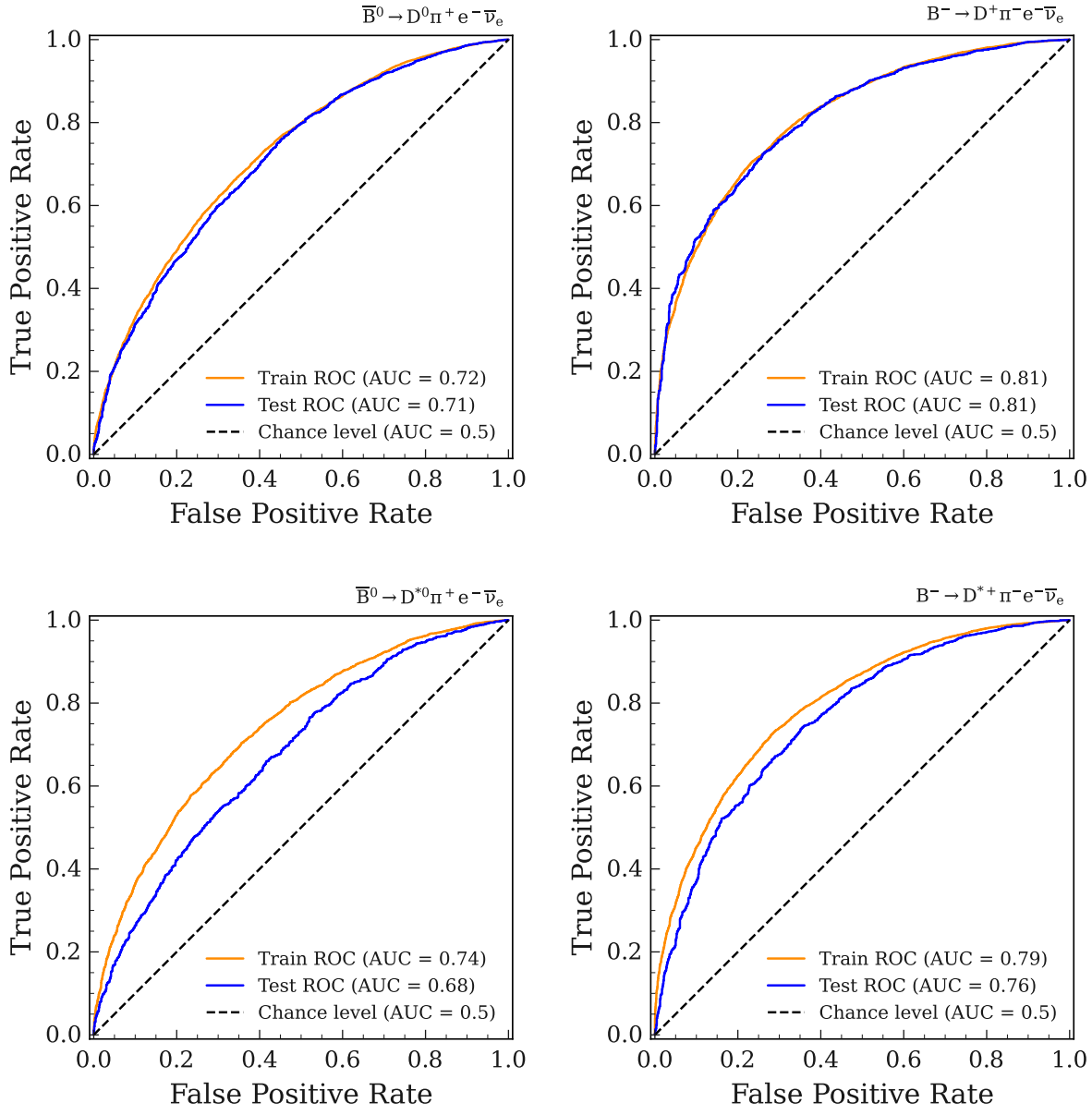


Figure 6.6: ROC curve for all four B decay modes.

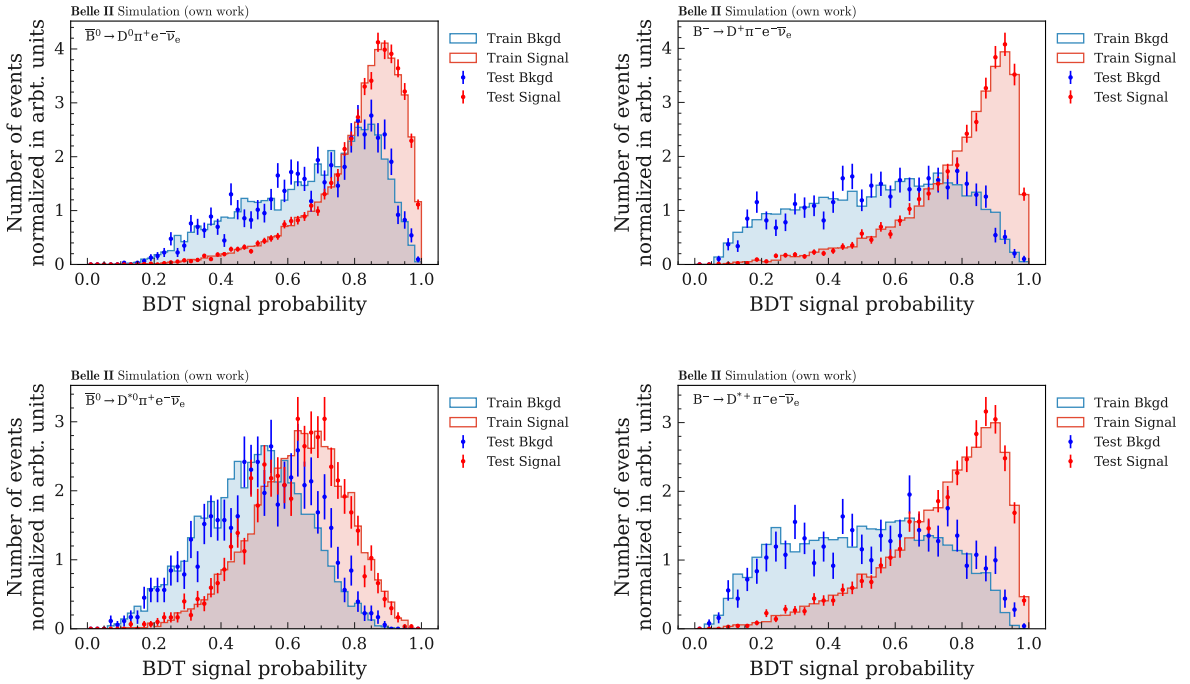


Figure 6.7: BDT signal probability for all four B decay modes.

6.4 Splot

In order to extract signal distributions from data in the presence of significant background contamination, we employ the *sPlot* technique. This method allows for the statistical disentanglement of overlapping signal and background components in a given dataset, based on a discriminating variable that exhibits distinguishable probability density functions (PDFs) for signal and background by applying event-by-event weights based on a maximum likelihood fit to the discriminating variable. The following section will explain the basics of this technique which are taken from [69]

6.4.1 Theoretical background

The sPlot technique operates with two categories of variables: the discriminating variable y, which provides good separation between signal and background with well-known Probability Density Functions (PDF) from simulation, and the control variable x, which represents the variable of interest for which we seek to extract the signal distributions. The approach involves fitting parametric templates for signal and background, then applying event-by-event weights to reconstruct the corrspodning distributions. The critical requirement is that the control variable x must be statistically independent of the discriminating variable y, which means that the joint probability density function factorizes as

$$f_i(x, y) = M_i(x) \cdot f_i(y) \,, \tag{6.1}$$

where $M_i(x)$ represents the PDF of the control variable for species i, and $f_i(y)$ denotes the PDF of the discriminating variable for the same species.

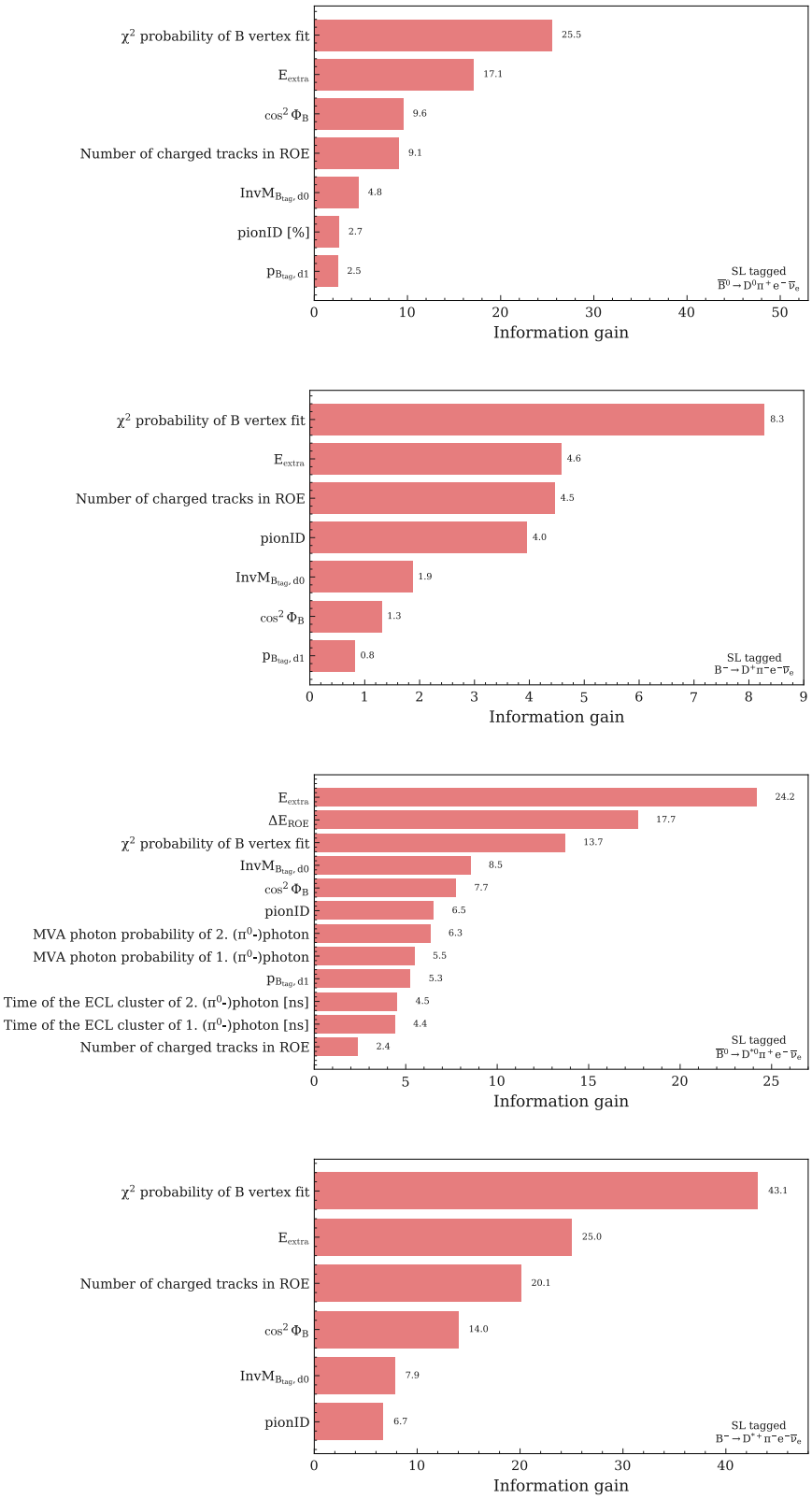


Figure 6.8: Feature importance ranking for all four B decay modes.

The mathematical foundation rests on an extended maximum likelihood fit. For a data sample containing N events with N_s different species, the log-likelihood function is defined as

$$\mathcal{L} = \sum_{e=1}^{N} \ln\left(\sum_{i=1}^{N_s} N_i f_i(y_e)\right) - \sum_{i=1}^{N_s} N_i, , \tag{6.2}$$

where N_i represents the expected number of events for species i, and $f_i(y_e)$ is the PDF value for species i evaluated at the discriminating variable values of event e. The control variable x does not appear in this likelihood expression, which is fundamental to the technique's validity.
The covariance matrix of the fitted yields is obtained from the second derivatives of the negative log-likelihood:

$$V_{ij}^{-1} = \frac{\partial^2(-\mathcal{L})}{\partial N_n \, \partial N_j} = \sum_{e=1}^{N} \frac{f_n(y_e) f_j(y_e)}{\left(\sum_{k=1}^{N_s} N_k f_k(y_e)\right)^2} \, . \tag{6.3}$$

This matrix encodes the statistical correlations between the fitted yields and provides the mathematical link between the likelihood fit results and the subsequent distribution reconstruction.

The core concept lies in the calculation of event-specific weights, so called *sWeights*, which determine the contribution of each event to the reconstructed distribution of each species. For event e and species n, the sWeight is calculated according to

$${}_s\mathcal{P}_n(y_e) = \frac{\sum_{j=1}^{N_s} V_{nj} f_j(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}, . \tag{6.4}$$

Each event receives a weight that depends on the covariance matrix elements V_{nj}, the PDF values $f_j(y_e)$ for all species evaluated at the event's discriminating variable values, and the normalization factor in the denominator. The sWeights can take negative values, particularly in regions where discrimination between species is poor.

6.4.2 Correlation check

As mentioned in the previous section, the sPlot technique requires the discriminant variable to be statistically independent, and therefore also uncorrelated, to the variable on which we want to perform the background subtraction. In our case this means that the BDT signal probability should be uncorrelated to the kinematic variables $M(D^{(*)}\pi)$, q^2 and $\cos\theta_V$ (for now). To evaluate these correlations we use two different statistical measures. First, we compute the Pearson correlation coefficients, which quantify linear correlations between the variables. These values are summarized in Table 6.1. Second, we evaluate the Kendall τ rank correlation coefficients, which assess the presence of monotonic (not necessarily linear) relationships between variables. The corresponding results are provided in Table 6.2. In both tests, we observe that the condition of statistical independence between the BDT output and the variables of interest is not always fulfilled. This is particularly evident for the correlation with the invariant mass variable $M(D^{(*)}\pi)$, where the strongest deviations from independence are observed. Notably, the $D^{*+}\pi^-$ decay mode exhibits the smallest correlation, indicating the best compliance with the assumption.
Alternatively, we could require orthogonality to these variables in BDT training, but this comes at the cost of less discriminating power, which is already critical for us. In the BDT training, we have made

Table 6.1: Pearson correlation coefficients ρ of BDT signal probability with parameter of interests.

	Signal			Background		
	$\rho_{M(D\pi)}$	ρ_{q^2}	$\rho_{\cos\theta_V}$	$\rho_{M(D\pi)}$	ρ_{q^2}	$\rho_{\cos\theta_V}$
$D^0\pi^+e^-\bar{\nu}_e$	-0.05 ± 0.01	-0.01 ± 0.01	0.04 ± 0.01	-0.11 ± 0.02	-0.02 ± 0.02	0.03 ± 0.02
$D^+\pi^-e^-\bar{\nu}_e$	0.12 ± 0.02	-0.13 ± 0.02	-0.02 ± 0.02	0.03 ± 0.03	-0.05 ± 0.03	-0.06 ± 0.03
$D^{*0}\pi^+e\bar{\nu}_e$	-0.08 ± 0.02	-0.002 ± 0.02	0.04 ± 0.02	-0.08 ± 0.02	-0.004 ± 0.02	-0.01 ± 0.02
$D^{*+}\pi^-e^-\bar{\nu}_e$	0.01 ± 0.02	-0.06 ± 0.02	-0.02 ± 0.02	0.004 ± 0.02	-0.0005 ± 0.02	0.02 ± 0.02

Table 6.2: Kendall τ correlation coefficients ρ of BDT signal probability with parameters of interest.

	Signal			Background		
	$\rho_{M(D\pi)}$	ρ_{q^2}	$\rho_{\cos\theta_V}$	$\rho_{M(D\pi)}$	ρ_{q^2}	$\rho_{\cos\theta_V}$
$D^0\pi^+e^-\bar{\nu}_e$	-0.048 ± 0.007	0.029 ± 0.007	0.002 ± 0.007	-0.088 ± 0.013	0.013 ± 0.013	-0.017 ± 0.013
$D^+\pi^-e^-\bar{\nu}_e$	0.073 ± 0.015	0.018 ± 0.015	-0.095 ± 0.015	0.023 ± 0.018	-0.041 ± 0.018	-0.037 ± 0.018
$D^{*0}\pi^+e^-\bar{\nu}_e$	-0.050 ± 0.012	0.027 ± 0.012	0.002 ± 0.012	-0.084 ± 0.013	-0.006 ± 0.013	0.008 ± 0.013
$D^{*+}\pi^-e^-\bar{\nu}_e$	0.005 ± 0.011	0.014 ± 0.011	-0.036 ± 0.011	0.016 ± 0.014	0.011 ± 0.014	-0.007 ± 0.014

sure that all input variables are already uncorrelated to the variables of interest, but we could continue to invest there and optimise the BDT training in this respect.
Another possibility that takes correlations into account would be the COW [70] method. However, this requires a polynomial parametric fit, which proves to be difficult with our signal probability distribution. Despite the observed deviations from ideal statistical independence, we proceed with the current setup to assess the applicability and robustness of the sPlot technique in our analysis. These limitations are noted and will be addressed in future optimisations. The next step focuses on the parameterisation of the BDT signal probability distributions for both signal and background components.

6.4.3 Fit to signal and background distribution

Finding a suitable model that describes the background and signal distribution of the BDT signal probability proved to be not that trivial. One method that is suitable for such a case is Kernel Density Estimation (KDE), which provides a way to estimate the probability density function from sampled data. The details are discussed in the following and are taken from [71].

Kernel Density Estimator

KDE constructs the estimate by placing a kernel function centered on each data point and aggregating their contributions.
Given a set of n sample points $\{x_1, \ldots, x_n\}$, the KDE is defined as

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \tag{6.5}$$

where $K(x)$ denotes the kernel function, $h > 0$ is the bandwidth parameter, and $x \in \mathbb{R}$ is the evaluation point. The kernel function specifies the shape and influence of each sample point, while the bandwidth

controls the width or spread of the kernel function and defines the range of influence. A commonly used kernel is the Gaussian kernel, given by

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \tag{6.6}$$

Increasing the bandwidth results in a smoother estimate by extending the influence of individual data points.
In this analysis, the bandwidth is determined automatically using the default adaptive method implemented in the `zfit` [72] library. This method first applies a global rule-of-thumb (e.g., Scott's [73] or Silverman's rule [74]) to obtain an initial pilot estimate of the density. The local bandwidths are then scaled inversely with this estimate, resulting in narrower kernels in regions with dense data and wider kernels in sparse areas. The goal of this procedure is to minimize the mean integrated squared error (MISE) between the estimated and true density.
This work tried both approaches, an exact and grid-based variants of KDE. The exact KDE computes the full sum described in Equation 6.5 without approximation, preserving all information from the sample. However, its computational cost scales linearly with the number of sample and evaluation points, making it more suitable for smaller datasets.
To address the scalability issue, the grid-based KDE introduces a binning procedure that reduces the number of data points by aggregating data points onto a grid. Let $\{g_l\}_{l=1}^{N}$ be a set of uniformly spaced grid points with associated weights $\{c_l\}_{l=1}^{N}$. The Grid KDE takes the form

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{l=1}^{N} c_l \, K\left(\frac{x - g_l}{h}\right), \tag{6.7}$$

which lowers computational complexity by evaluating the kernel sum over grid points instead of individual data points. This approximation is better suited for large datasets.
Since both approaches yield the same results, if it comes to fit stability and the quality of the fit, we continue with the grid-based KDE.

The KDEs are constructed using `zfit.pdf.KDE1DimGrid` from the `zfit` library over the observable space SigProb $\in [0, 1]$, representing the BDT output score. The corresponding KDEs are then extended to incorporate expected event yields via `set_yield`, where the initial values for the signal and background yields are taken from the number of signal- and background events in the MC sample.
The full model is constructed as a sum of the extended signal and background PDFs:

$$\mathcal{P}_{\text{tot}} = N_{\text{sig}} \cdot \mathcal{P}_{\text{sig}}(\text{SigProb}) + N_{\text{bkgd}} \cdot \mathcal{P}_{\text{bkgd}}(\text{SigProb}), \tag{6.8}$$

where N_{sig} and N_{bkgd} are the yields and $\mathcal{P}_{\text{sig}}$ and $\mathcal{P}_{\text{bkgd}}$ are the signal and background KDEs, respectively. An extended unbinned maximum likelihood fit is then performed, where the fit is applied to the BDT score of asimov data. The minimization is carried out using the `iMinuit` minimizer [63]. The result yields the best-fit signal and background contributions, which are then used in subsequent sPlot weight calculations. Finally the sWeights are then computed from the fitted KDE model and the Asimov data using the `compute_sweights(model, data)` function, provided by the `hepstats.splot` package. These weights allow us to statistically disentangle the signal and background contributions for our

Table 6.3: χ^2_{red} of total KDE fit.

	χ^2_{red}	p-value
$\bar{B}^0 \to D^0 \pi^+ e^- \bar{\nu}_e$	1.33	0.063
$B^- \to D^+ \pi^- e^- \bar{\nu}_e$	1.24	0.121
$\bar{B}^0 \to D^{*0} \pi^+ e^- \bar{\nu}_e$	0.45	0.999
$B^- \to D^{*+} \pi^- e^- \bar{\nu}_e$	1.67	0.0025

variables of interest.

For technical reasons, the calculation of sWeights requires an unweighted dataset as input. Therefore, the KDE fit and subsequent signal extraction were carried out on the unweighted MC15rd sample. This technical limitation naturally disappears when applying the method to real data, which is unweighted by nature. Nevertheless, it has been verified that the BDT score distribution changes only marginally when applying correction weights, and therefore the influence on the KDE model and sWeight calculation should be also minimal. After computing the sWeights, they are then multiplied again by the correction weights applied per event discussed in Section 6.2.1. This ensures that the further analysed spectrum is more closely resemble real data.

6.4.4 Results

First we discuss the fit results of the KDEs estimating the BDT signal probability distribution. Figure 6.9, 6.10, 6.11 and 6.10 show the results of the KDE fit to the Asimov data for each B decay mode. Additionally, the KDEs of the signal and background distributions used for the fit, as described in Eq. (6.8), are shown. At first glance, it looks promising. A quantitative evaluation of the goodness-of-fit using the reduced chi-squared χ^2_{red} and the corresponding p-value, as summarized in Table 6.3, shows that high p-values and χ^2_{red} values close to or below unity are expected when using Asimov data. This behavior is observed only in the $\bar{B}^0 \to D^{*0} \pi^+ e^- \bar{\nu}_e$ channel. For the remaining decay modes, the fit quality deteriorates primarily due to a poor description of the last bin in the BDT score distribution, which is characterized by small statistical uncertainties and a sharper drop in event counts than predicted by the KDE model. This discrepancy likely accounts for the higher χ^2 values. Given the context and limitations of the model, this deviation remains within acceptable bounds for the present analysis.
Now, we examine the sPlot-weighted distributions shown in Fig. 6.13, 6.14 and 6.15, taking into account that especially the variable $M(D^{(*)}\pi)$ exhibits a relatively strong correlation with the BDT signal probability. This correlation suggests that the assumption of statistical independence required by the sPlot technique may have been violated, and this must be considered when interpreting the resulting distributions. The plots look fine in this regard. However, in the lower $D^{(*)}\pi$ mass regions, noticeable discrepancies between the sPlot-weighted distribution and the expected signal shape become apparent, indicating limitations in the accuracy of the background subtraction in these regions. In conclusion, while the sPlot technique is functional and yields meaningful results, the current signal extraction set up requires further refinements and needs to be slightly adjusted.

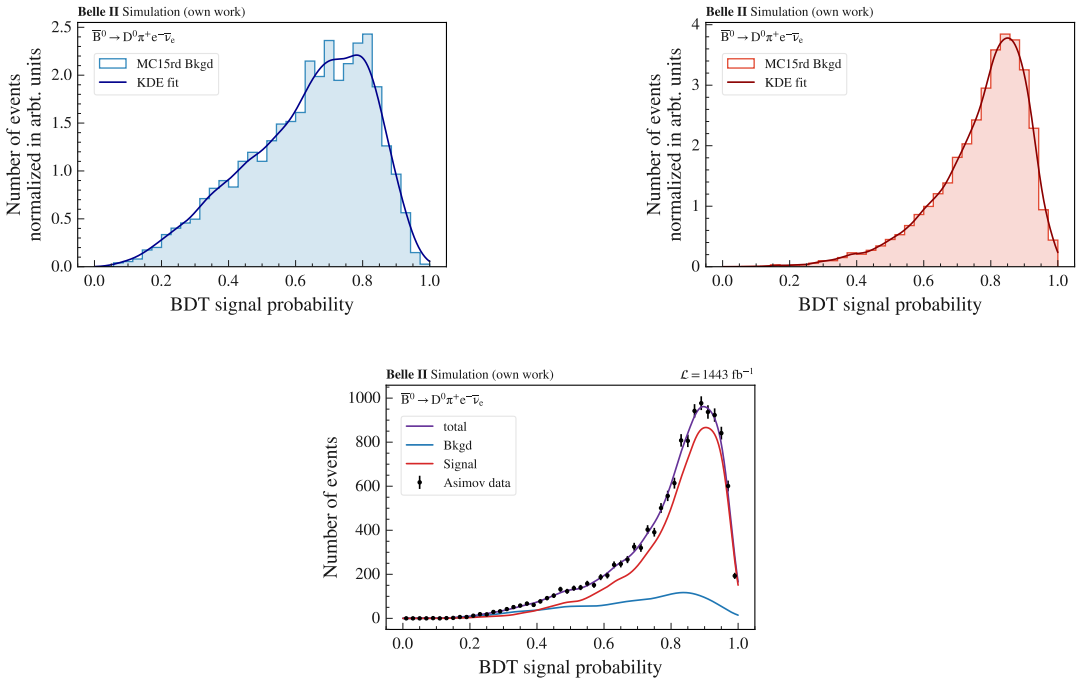


Figure 6.9: KDE fit on signal- and background distributions of the BDT score for the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ decay channel.

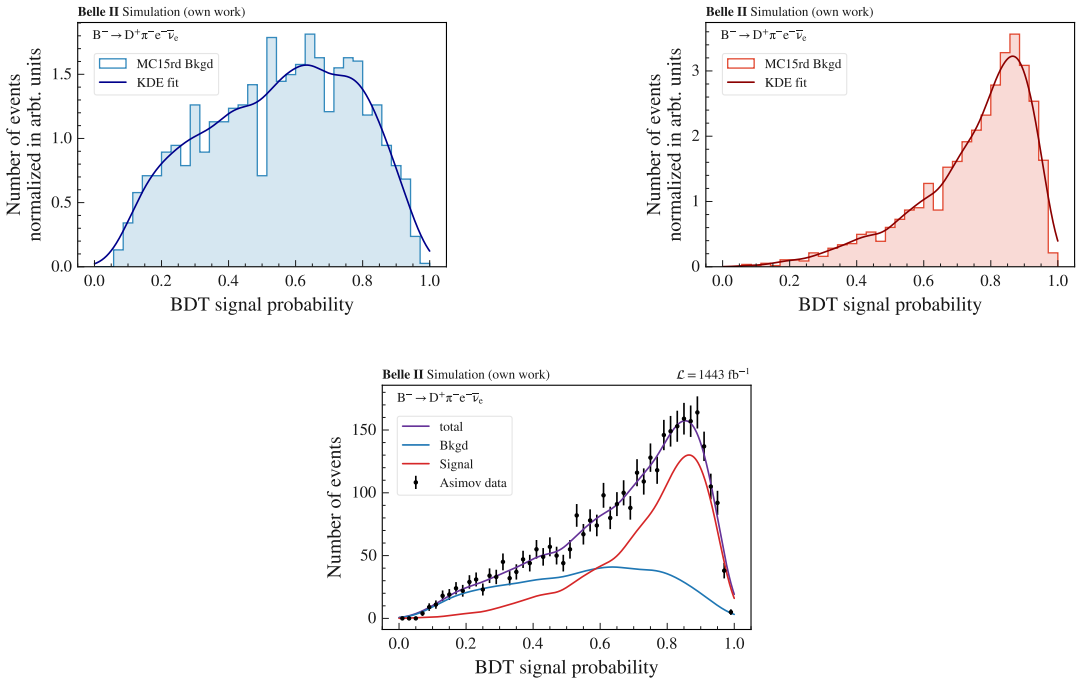


Figure 6.10: KDE fit on signal- and background distributions of the BDT score for the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ decay channel.

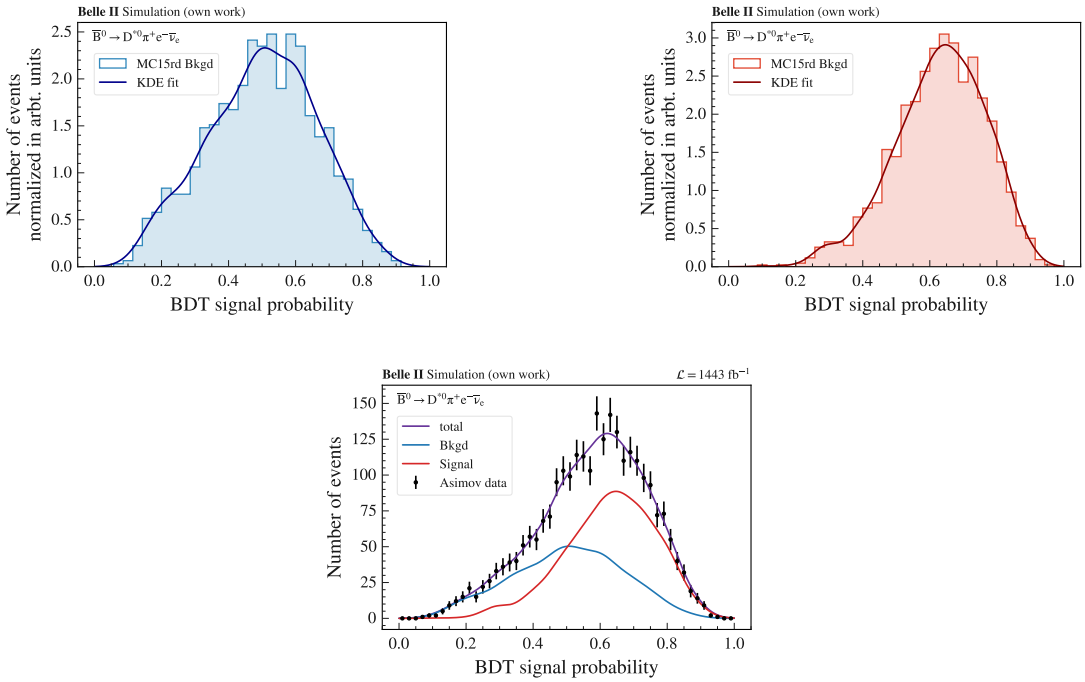


Figure 6.11: KDE fit on signal- and background distributions of the BDT score for the $\bar{B}^0 \to D^{*0}\pi^+e^-\bar{\nu}_e$ decay channel.

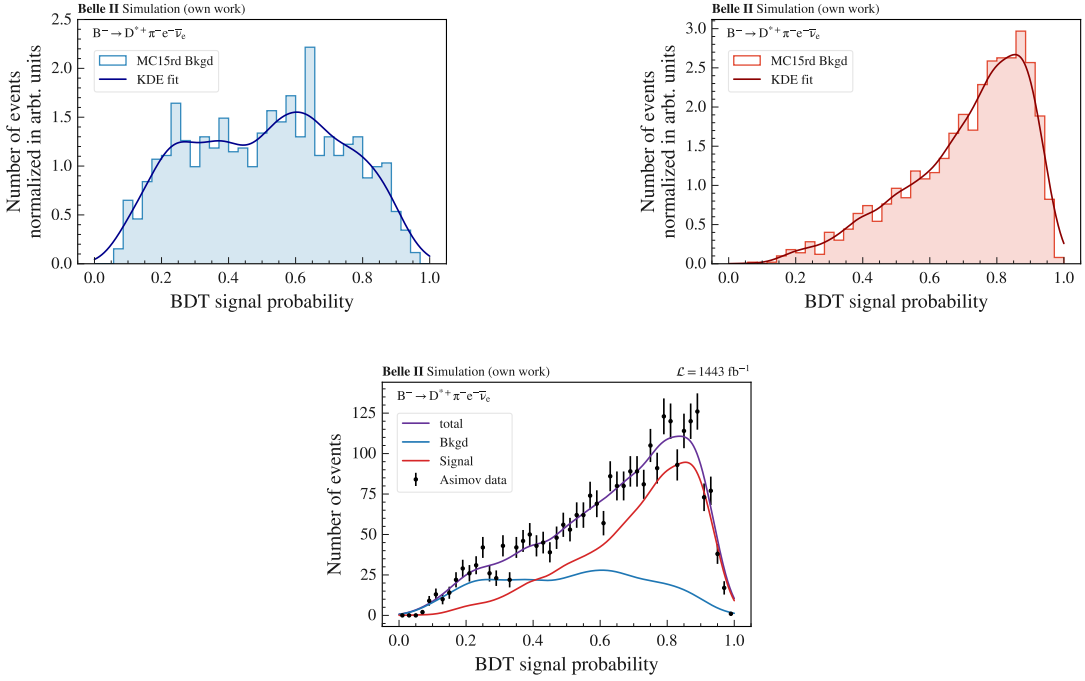


Figure 6.12: KDE fit on signal- and background distributions of the BDT score for the $B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$ decay channel.

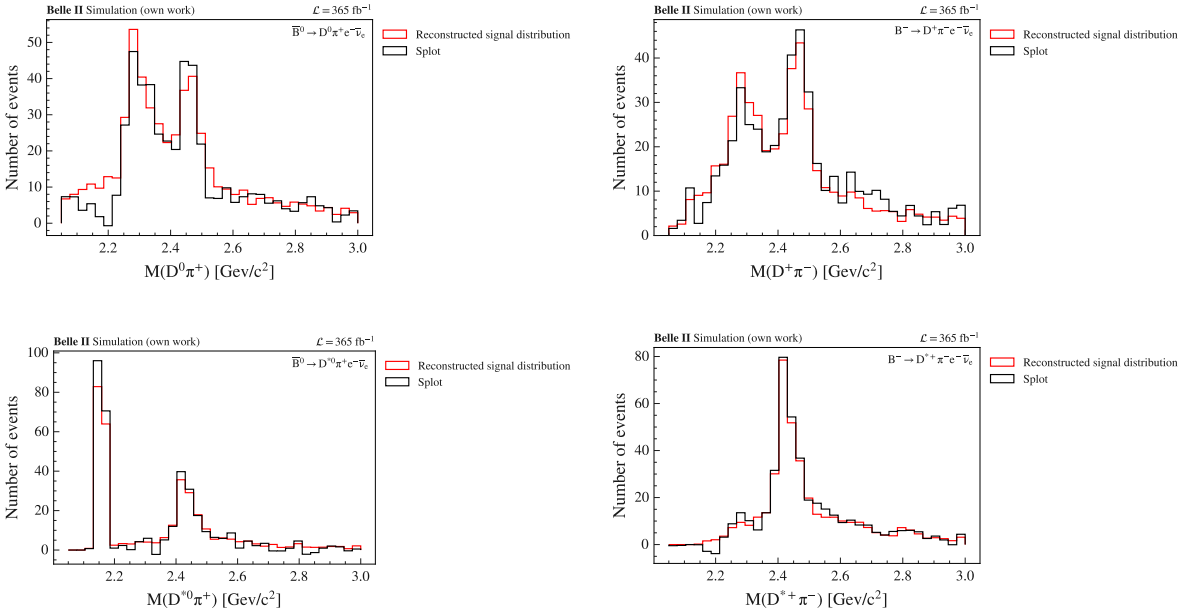


Figure 6.13: SPlots of $M(D^{(*)}\pi)$ variable for all four B decay modes.

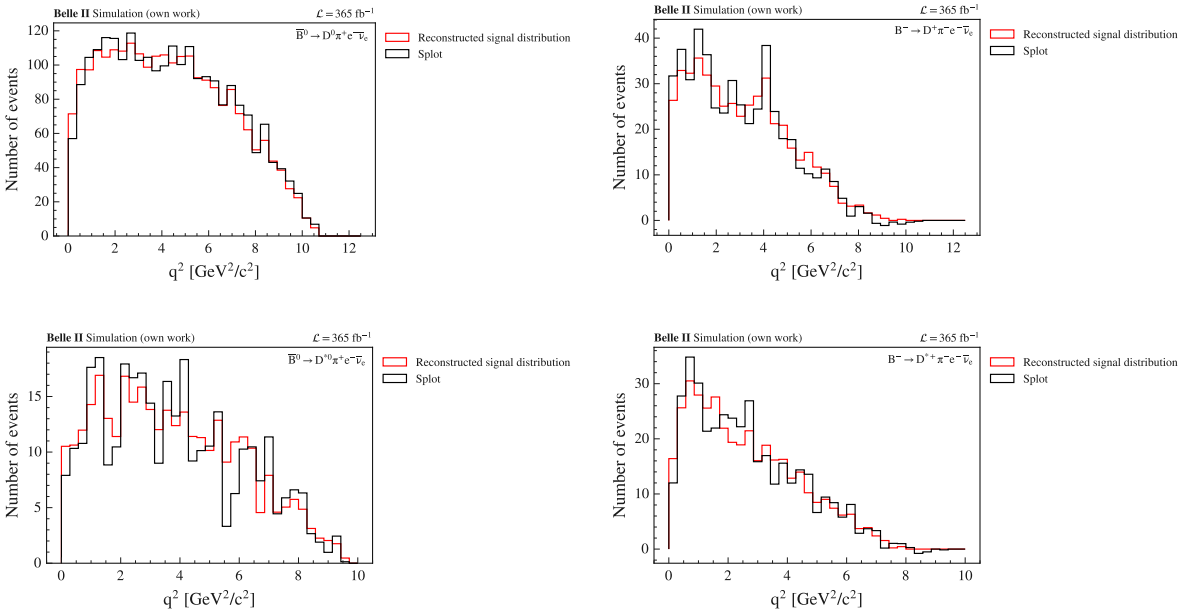


Figure 6.14: SPlots of q^2 variable for all four B decay modes.

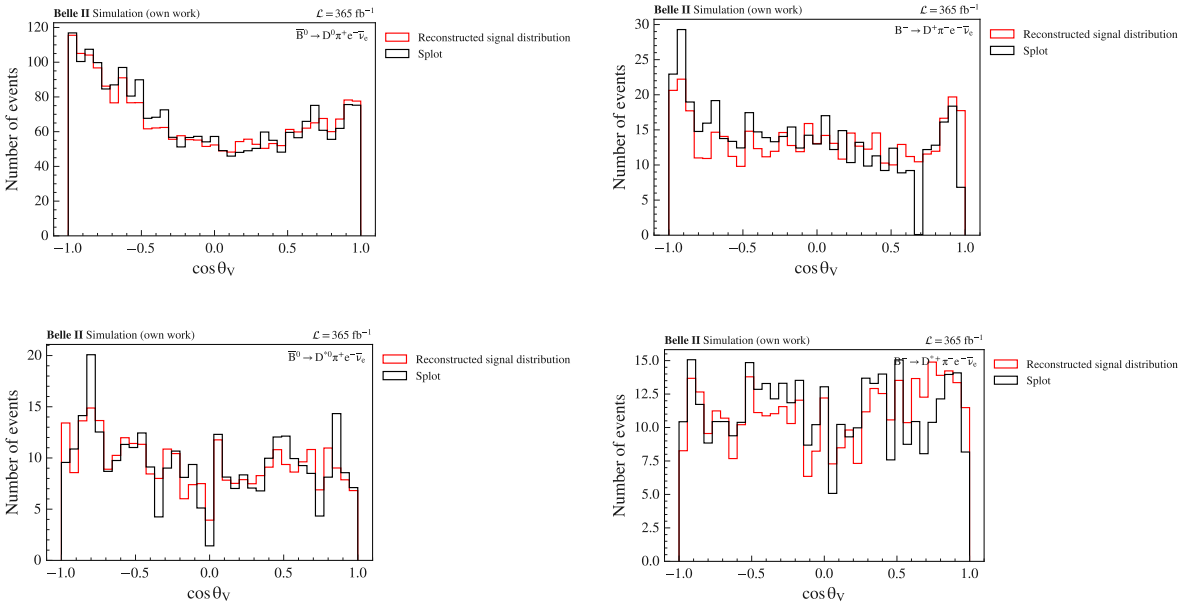


Figure 6.15: SPlots of $\cos\theta_V$ variable for all four B decay modes.

CHAPTER 7

Migration and efficiency corrections

The next step in the analysis strategy involves accounting for detector effects through migration corrections and evaluating selection efficiencies. This chapter provides a preliminary investigation of these aspects and is limited in scope, serving primarily as a first-step feasibility study.
To address the impact of detector resolution and event misreconstruction, a migration matrix approach is employed. In particular, the method is extended to allow for migration between different B decay modes, providing a framework to handle down-feed contributions. This modified migration scheme is applied, as a test case, to the invariant mass distribution of the $D^{(*)}\pi$ system.
In addition, we present the distributions of selection efficiencies for each of variables of interest, $M(D^{(*)}\pi)$, q^2 and $\cos\theta_V$, studied in this work. While no full efficiency corrections are applied at this stage, the information offers valuable insight into the selection behavior and lays the groundwork for potential future differential corrections.

7.1 Migration matrix

In experimental measurements, detector resolution and reconstruction inefficiencies can lead to deviations between the true physical quantities and their reconstructed counterparts. As a result, an event reconstructed in bin i may have originated from a true value corresponding to a different bin j. This effect is commonly described by a *migration matrix* $\mathcal{M}_{ij}$, which encodes the conditional probability of reconstructing an event in bin i given that its true value lies in bin j:

$$\mathcal{M}_{ij} = P(\text{reco in bin } i \,|\, \text{true in bin } j) \tag{7.1}$$

To recover an estimate of the true distribution from the observed (reconstructed) distribution, an *unfolding* procedure is applied. The simplest form of unfolding consists in inverting the migration matrix:

$$\mathbf{x}_{\text{true}} = \mathcal{M}^{-1}\,\mathbf{x}_{\text{reco}} \tag{7.2}$$

where $\mathbf{x}_{\text{true}}$ denotes the binned yields of the true spectrum (of variable x) the and $\mathbf{x}_{\text{reco}}$ the corresponding reconstructed yields. Many other approaches exist to mitigate the impact of increased statistical variance, particularly relevant in cases of large bin-to-bin migrations, where a single true bin may receive significant weighted contributions from multiple reconstructed bins. These methods also aim to reduce dependence

on the MC prior, which can influence the unfolding result. While this issue is less critical in the current study, techniques such as iterative Bayesian unfolding can help propagate the shape observed in data rather than relying solely on MC predictions, thereby reducing model bias. However, such unfolding techniques and regularization strategies are beyond the scope of this thesis and analysis (for now).

In our case, as discussed in Section 5.4 we find down feed components in the $D\pi$ channels originating from a semi-leptonic $B \to D^*\pi$ decay. In the migration procedure we therefore try not only to migrate between the bins in one channel but also in between the channels and thus migrate down feed components in the $D\pi$ channel back to the corresponding true value in $D^*\pi$. In the context of this thesis, we explore unfolding, as rarely used in other analyses, as a conceptually approach to handle down-feed contributions. However, this method can result in highly non-trivial migration matrices with significant off-diagonal elements, deviating substantially from a near-diagonal structure and potentially complicating the inversion.

Now we only focus on the charged B decay mode for the migration procedure. For the migration procedure, the neutral B modes turned out to be somewhat more difficult and are therefore still being worked on. In particular, the dominant D^* resonance in the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ channel complicates the normalization of the migration matrix, making it more sensitive to statistical fluctuations. Especially also since we have up-feed events in the $D^{0*}\pi^+$ channel, which we try to migrate to the $D^0\pi^+$ channel, where up-feed events have much fewer counts compared to the correct D^* peak in the $D^0\pi^+$ mode. An initial attempt to apply a Bayesian unfolding approach did not yield a significant improvement. Furthermore, a technical limitation arises from incomplete MC truth information, especially for SCF events (and also to a lesser extent for down-feed events), true kinematic values were often unavailable. This issue prevented the inclusion of such events in the migration process. While this effect is negligible in the charged B modes due to the relatively small SCF contribution (see Table C.1), it becomes critical in the $\bar{B}^0 \to D^{*0}\pi^+e^-\bar{\nu}_e$ channel, where SCF constitutes approximately 50% of the sample. As a result, the reliability of the migration matrix in this channel is reduced. Additionally, the overall low event statistics in the semi-leptonic $\bar{B}^0 \to D^{*0}\pi^+$ channel amplify the instability in the unfolding procedure.

For the charged B modes, the $D^{(*)}\pi$ invariant mass spectrum is shown again in Fig. 7.1 to provide a reference for the unfolding procedure and to illustrate the structure and magnitude of the down-feed contributions. The corresponding migration matrix can be found in Fig. 7.2, where one can clearly

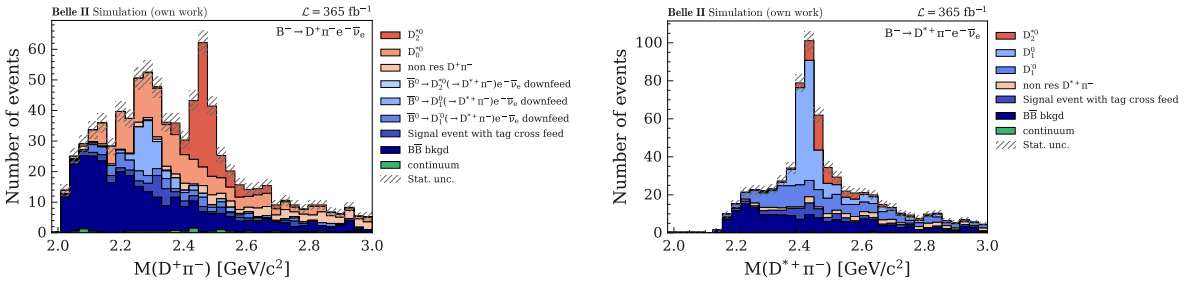


Figure 7.1: $M(D^{(*)}\pi)$ distributions of the charged B modes, to illustrate the proportion of down feed components once again.

see the influence of the down feed components in the non-diagonal elements in the migration matrix. Applying the migration matrix to unfold the invariant mass spectrum, on which the matrix itself was

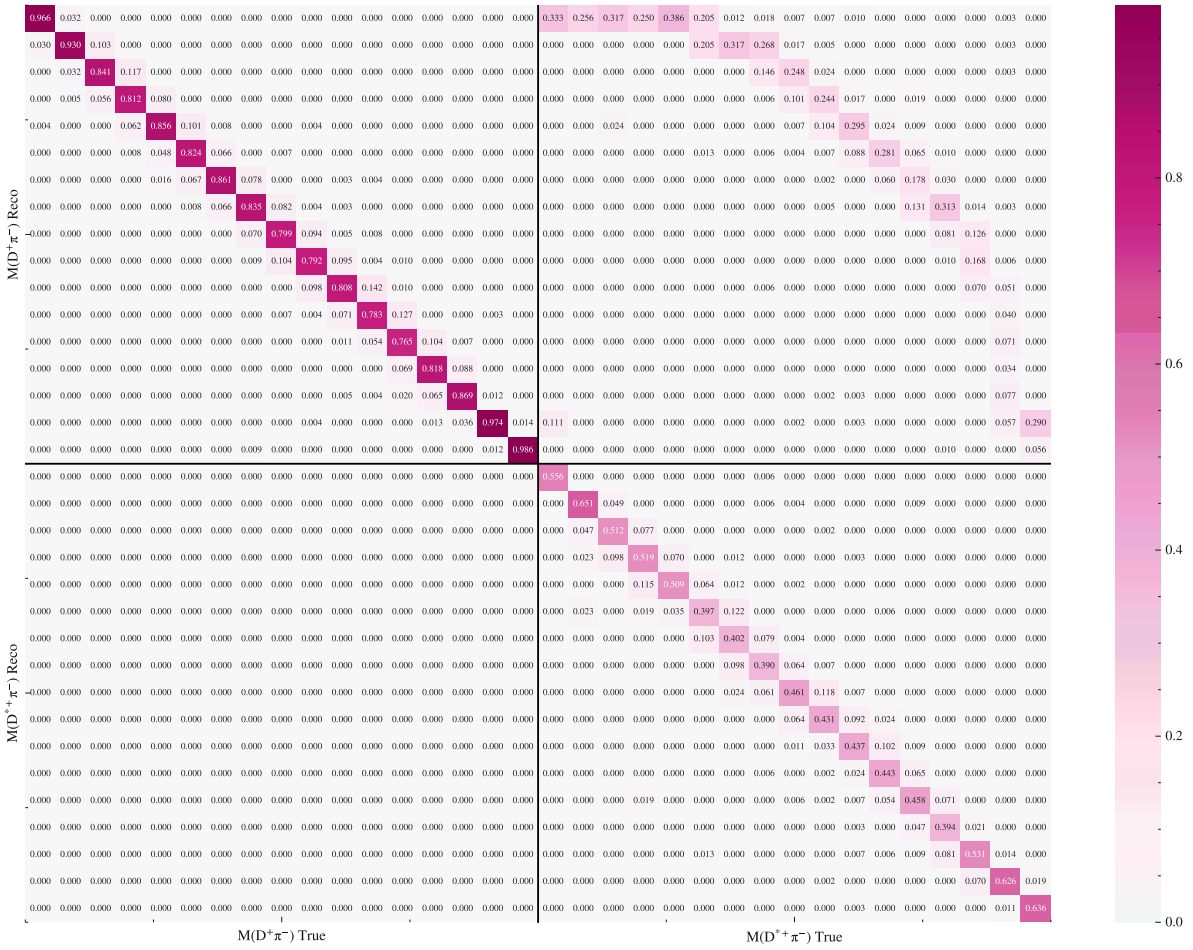


Figure 7.2: Migration matrix for the two charged B modes, where in this approach we also try do migrate the down-feed events back to the corresponding $D^*\pi$ channel.

constructed, serves as a valuable consistency check, since by construction, the unfolded distribution should exactly reproduce the true underlying spectrum. This one can see in Fig. 7.3, where it closes up exactly. The statistical uncertainties shown in the unfolded spectrum were obtained using a bootstrapping procedure. Specifically, Equation Eq. (7.2) was evaluated for 1000 statistically fluctuated realizations of $\mathbf{x}_{\text{reco}}$. The corresponding uncertainties were then extracted from the resulting the covariance matrix (and its diagonal elements) of the unfolded distributions.

Now we can also unfold our invariant $D^*\pi$ mass spectrum, which we have obtained using the sPlot method, as background subtracted mass spectrum. Since both the migration matrix and the fit (with Asimov data) are derived from the same MC sample, the resulting differences are minimal and primarily reflect statistical fluctuations of the signal yields in the invariant mass distribution. The unfolded $D^{(*)}\pi$ mass spectrum after background subtraction using the sPlot technique is shown in Fig. 7.4. The result demonstrates reasonable agreement within statistical uncertainties, although some deviations suggest that further refinement of the procedure may be beneficial[1].

[1] Furthermore, isospin corrections have not yet been applied yet in this context. These arise from the fact that the down-feed

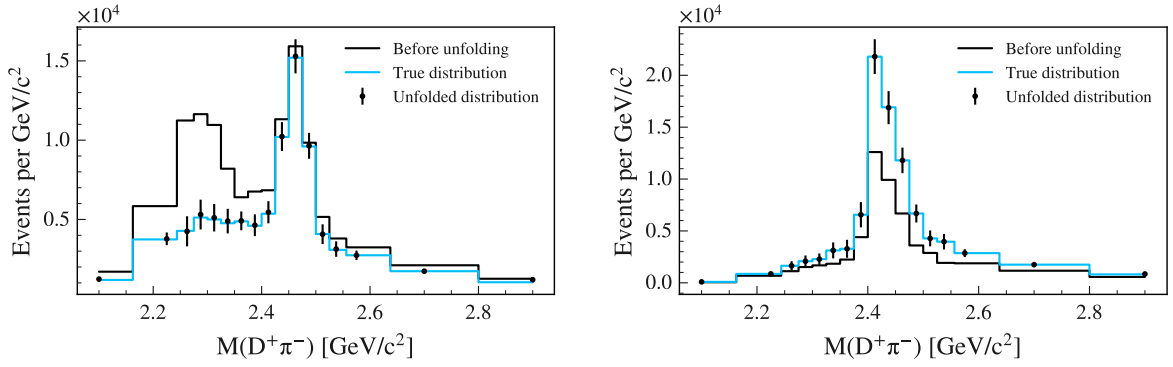


Figure 7.3: Unfolded $M(D^{(*)}\pi)$ spectrum on the MC sample where we build up the migration matrix on.

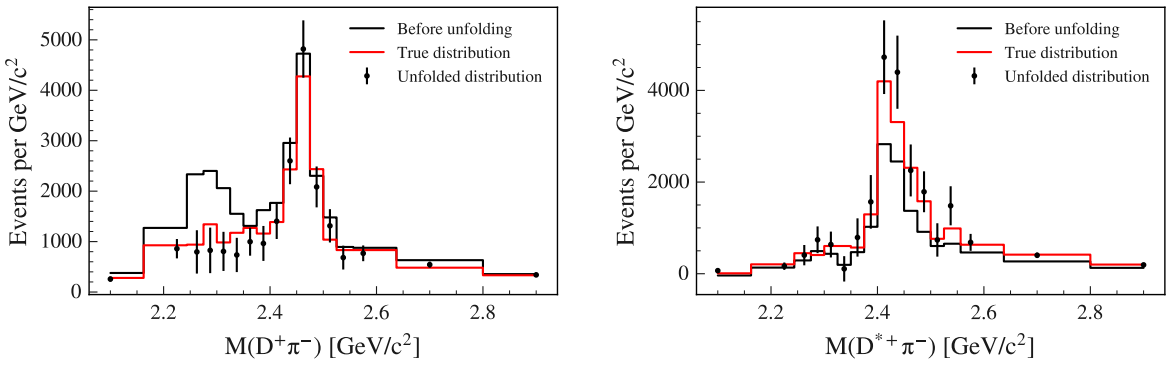


Figure 7.4: Unfolded $M(D^{(*)}\pi)$ spectrum on the background subtracted mass spectrum, we achieved with the sPlot technique. Thus, this spectrum already encounters primarily statistical fluctuation arising from the fit.

7.2 Efficiency distributions

After correcting for migration effects, selection efficiencies must also be taken into account. In this thesis, we restrict ourselves to examining the efficiency distributions of the observables of interest, in order to identify potential biases introduced by selection cuts. Even if the efficiency is not uniform across the kinematic range, differential corrections can, in principle, be applied. While this is not a major concern for the current study, it is of methodological interest. A full correction for selection efficiency, required for instance to extract differential decay rates such as $d\Gamma/dq^2$, constitutes a significant next step in the analysis, but lies beyond the scope of this work. The respective efficiency distributions for every B mode individually are given in Fig. 7.5, 7.6, 7.7 and 7.8. In addition, the plots also include the true distributions of the $M(D^{(*)}\pi)$, q^2, and $\cos\theta_V$ variables in respect to the MC modeling. The figures also show the true distributions corresponding to the reconstructed spectrum, illustrating the impact of selection inefficiencies and detector effects. Additionally, the generator-level distributions, which

contribution originates from decays such as $B^- \to D^{*+}(\to D^+\pi^0)\pi^- e^- \bar{\nu}_e$, while the reconstructed signal mode corresponds to $B^- \to D^{*+}(\to D^0\pi^+)\pi^- e^- \bar{\nu}_e$. Accounting for isospin symmetry, these events should be weighted by a factor of 2. However, given their limited impact on the overall evaluation of the unfolding method, they are expected to play a minor role in the evalualtion of this analysis.

include all simulated events prior to any reconstruction or selection, are shown for reference. These comparisons help quantify losses due to reconstruction failures and selection cuts. Roughly speaking, the efficiency looks stable for almost all B modes and in all variables of interest.

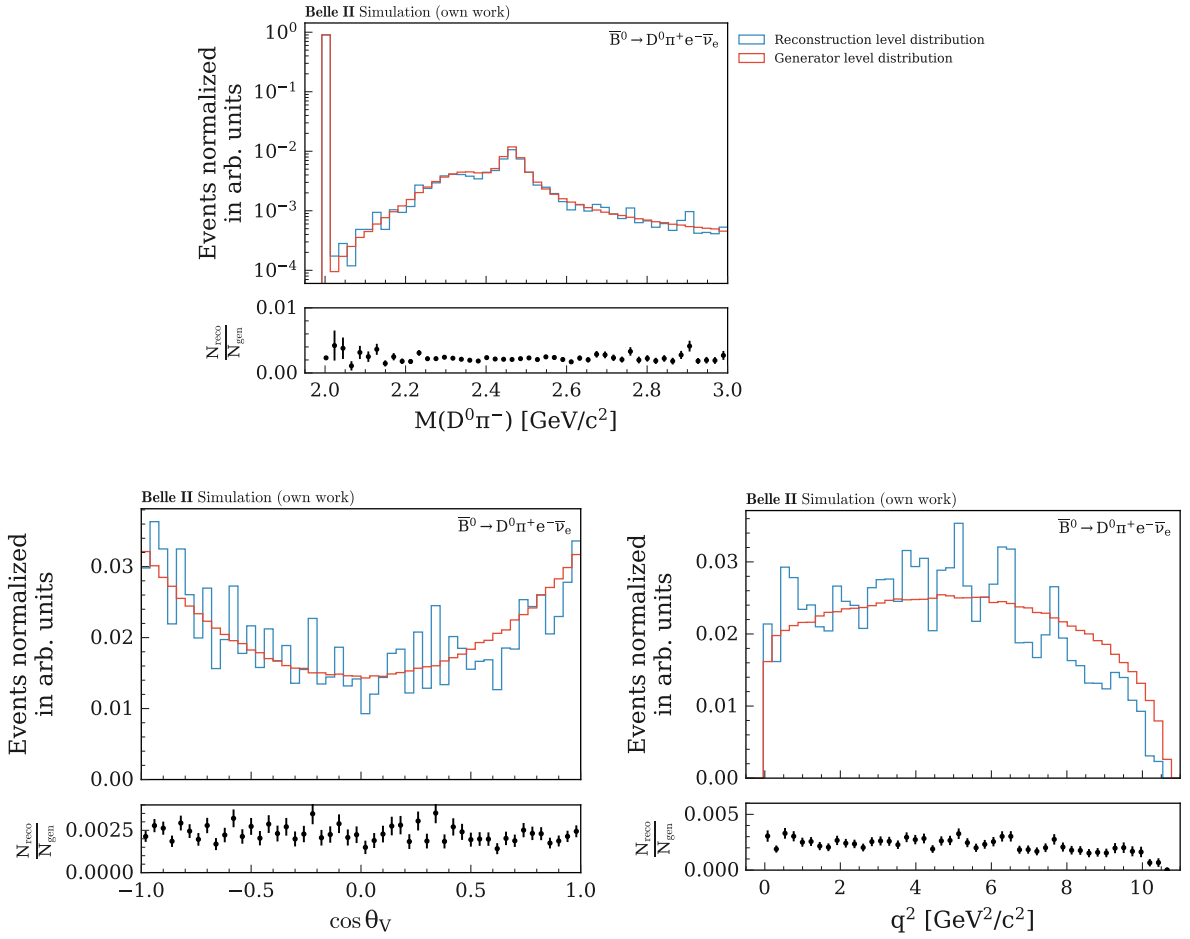


Figure 7.5: Efficiency distributions of the parameter of interest for the $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ decay mode. Since the D^* resonance dominates this decay mode, we provide additional plots for this channel in which the efficiency is shown separately for all contributing resonant states as well as for the non-resonant component. This allows for a more differential view on how the detector and analysis selection affects each component individually. The figures can be found in Appendix E.

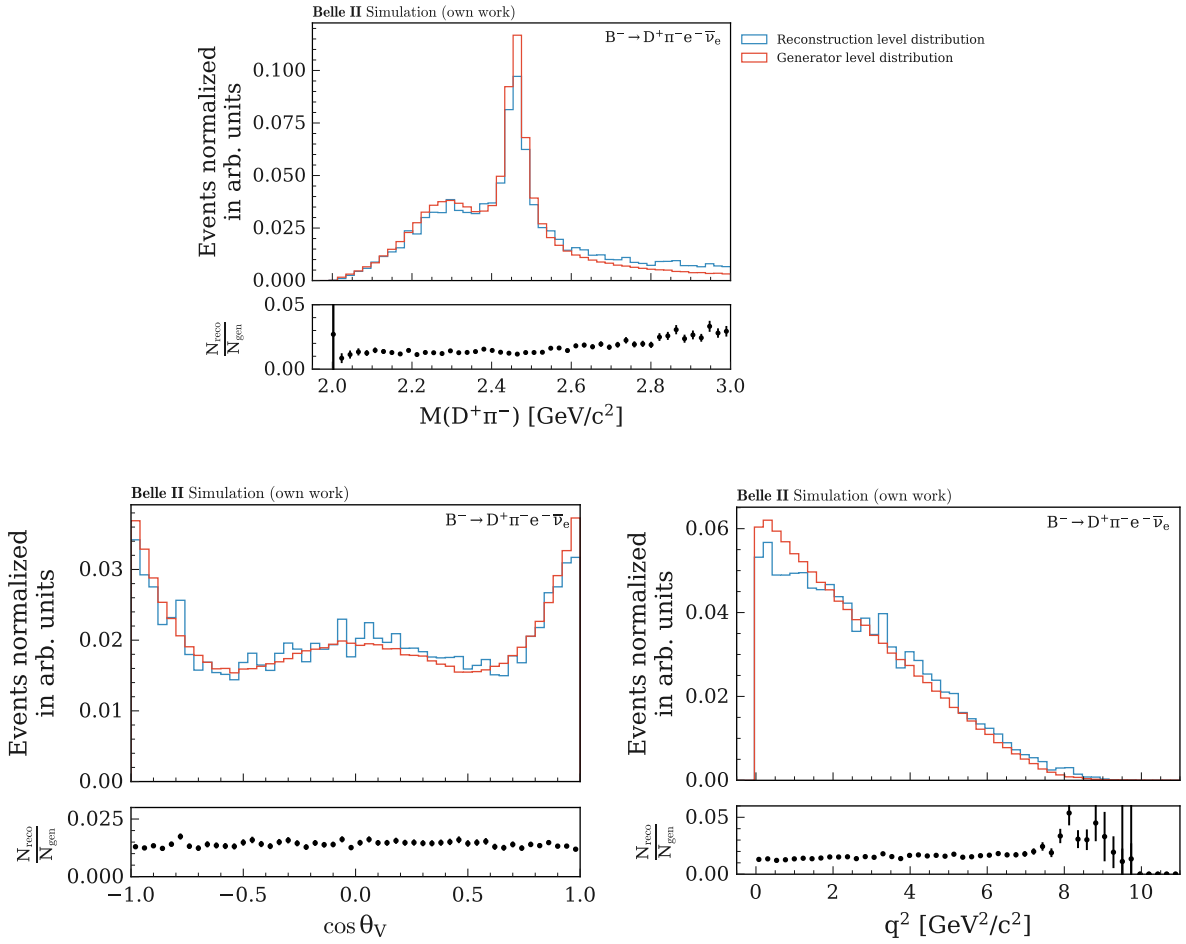


Figure 7.6: Efficiency distributions of the parameter of interest for the $B^- \to D^+\pi^- e^- \bar{\nu}_e$ decay mode.

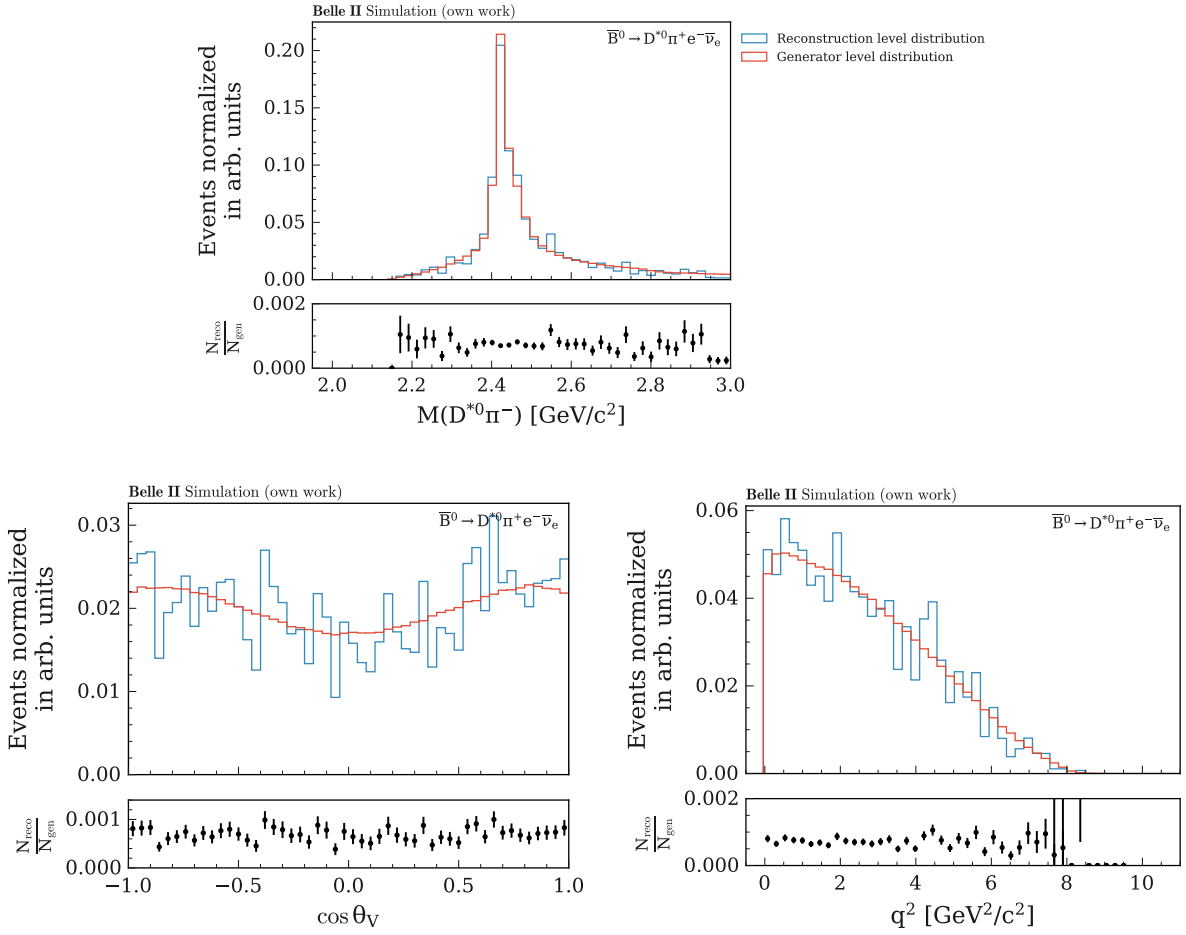


Figure 7.7: Efficiency distributions of the parameter of interest for the $\bar{B}^0 \to D^{*0}\pi^+e^-\bar{\nu}_e$ decay mode.

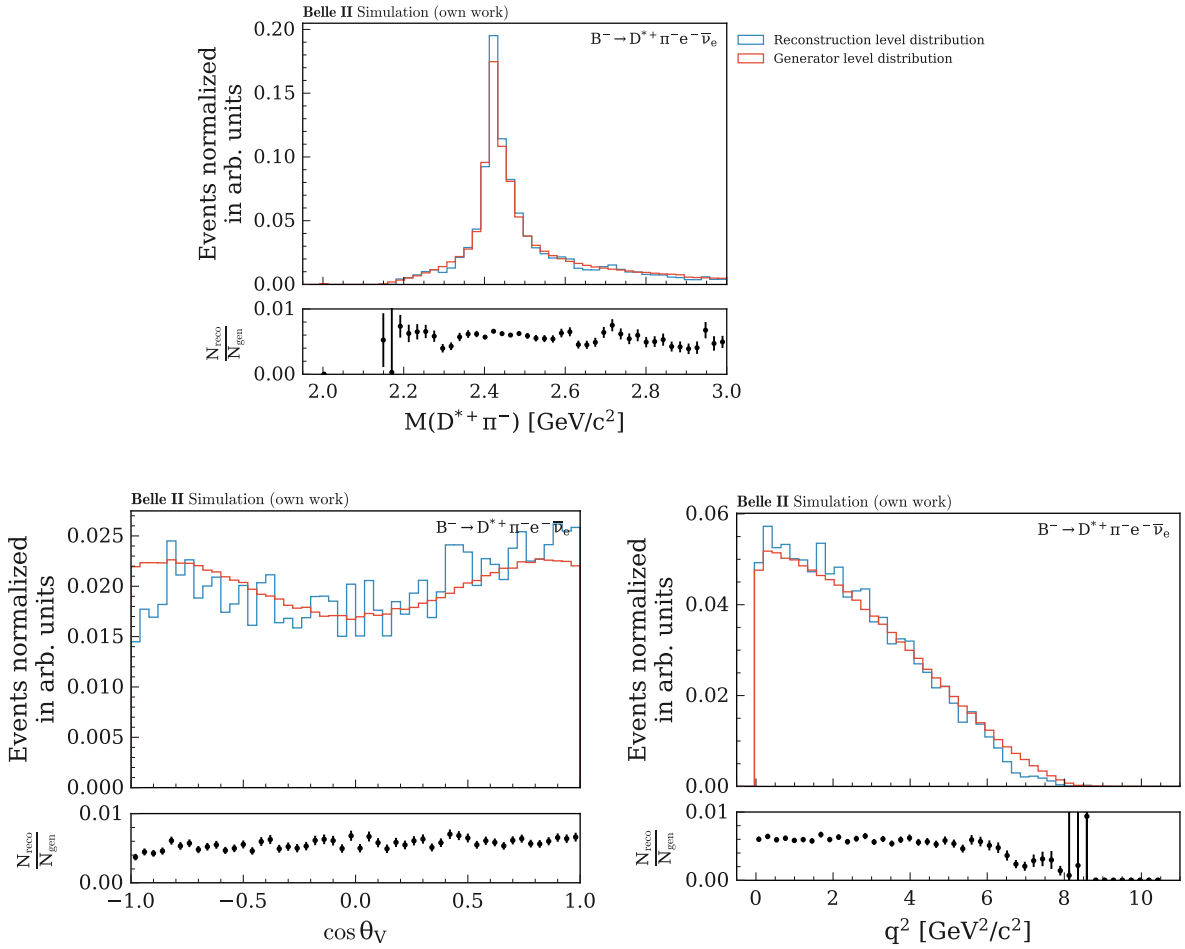


Figure 7.8: Efficiency distributions of the parameter of interest for the $B^- \to D^{*+}\pi^- e^- \bar{\nu}_e$ decay mode.

CHAPTER 8

Conclusion

This thesis presented a first feasibility study of the semileptonic decay $B \to D^{**}\ell\nu$ at Belle II, using both SL B tagging and an untagged reconstruction strategy. The aim was to evaluate whether these approaches offer promising prospects for further investigation of these challenging decay modes.

In the untagged analysis, it was found that the background conditions are particularly demanding, primarily due to the high rate of events, being misreconstructed because of confusions with the tag side B. In many cases, the pion was incorrectly assigned from the tag side, complicating background suppression. A BDT was trained to improve the signal-to-background discrimination, but the resulting separation power remained insufficient. Visual inspections of vertex and kinematic variables also did not yield substantial improvements. Future directions for the untagged approach could involve more sophisticated BDT tuning, additional input variables, alternative machine learning algorithms, or a detailed re-examination of vertex fitting configurations.

In contrast, the SL tagged analysis showed significantly more favorable background conditions. A combination of carefully chosen selection criteria was sufficient to suppress most of the background without requiring a multivariate classifier. In particular, the variable $\cos^2\phi_B$ demonstrated strong discriminative power. Further background reduction was achieved by combining tag-side particles with the signal side and applying a D^* mass veto. A comparison of our signal sensitivity with the hadronically tagged measurement performed by Belle [24] shows that our analysis achieves comparable, and in several cases even superior, performance across most B decay modes. Notably, higher signal yields and improved signal-to-background ratios are observed in the majority of channels. An exception is the $\bar{B}^0 \to D^{*0}(\to D^0\pi^0)\pi$ decay mode, where our analysis yields fewer signal events and a less favorable signal-to-background ratio, likely due to the challenges associated with reconstructing neutral particles and the dominant background contributions in this channel.
Following background suppression, signal extraction was carried out using the *sPlot* technique. This method requires both a discriminative variable and statistical independence between that variable and the parameters of interest ($M(D^{(*)}\pi)$, q^2, and $\cos\theta_V$). As no single variable with high discriminative power remained after event selection, a BDT was trained specifically for the purpose of providing a signal probability score to be used in the fit. Although the input variables were chosen to be uncorrelated with the mass spectrum, the resulting BDT output still exhibited correlations up to approximately 10%, formally violating the sPlot assumption of independence. Nevertheless, the technique was applied, and the resulting sWeighted distributions appeared reasonable, though discrepancies were observed in the

lower mass regions. This suggests that further refinement of the signal extraction procedure is necessary for more robust results.
The last component of the analysis focused on correcting for detector resolution and selection biases. This was explored through a basic implementation of unfolding via a migration matrix, as well as an inspection of selection efficiencies. The migration matrix was extended to allow for reconstruction-mode transitions between B decay channels, thereby providing a strategy to re-incorporate down-feed events (e.g., $D^*\pi \to D\pi$ misreconstruction). While the migration showed potential for the charged B modes, it remained problematic for the neutral modes, where SCF contributions and dominant resonances complicate the procedure and therefore it is still under development. Regarding selection efficiency, distributions of the efficiency across the key observables $M(D^{(*)}\pi)$, q^2, $\cos\theta_V$ were studied and found to be sufficiently stable to support future unfolding and differential measurements.

In summary, this work demonstrates that the SL tagged analysis offers a promising path forward for studying $B \to D^{**}\ell\nu$ decays at Belle II, particularly given its manageable background environment. However, improvements are still needed in signal extraction and unfolding methods. The findings of this thesis provide a solid foundation for further refinement and pave the way for a more detailed measurement of these decays with Belle II.

APPENDIX A

More variable distributions for untagged analysis

A.1 More vertex variables

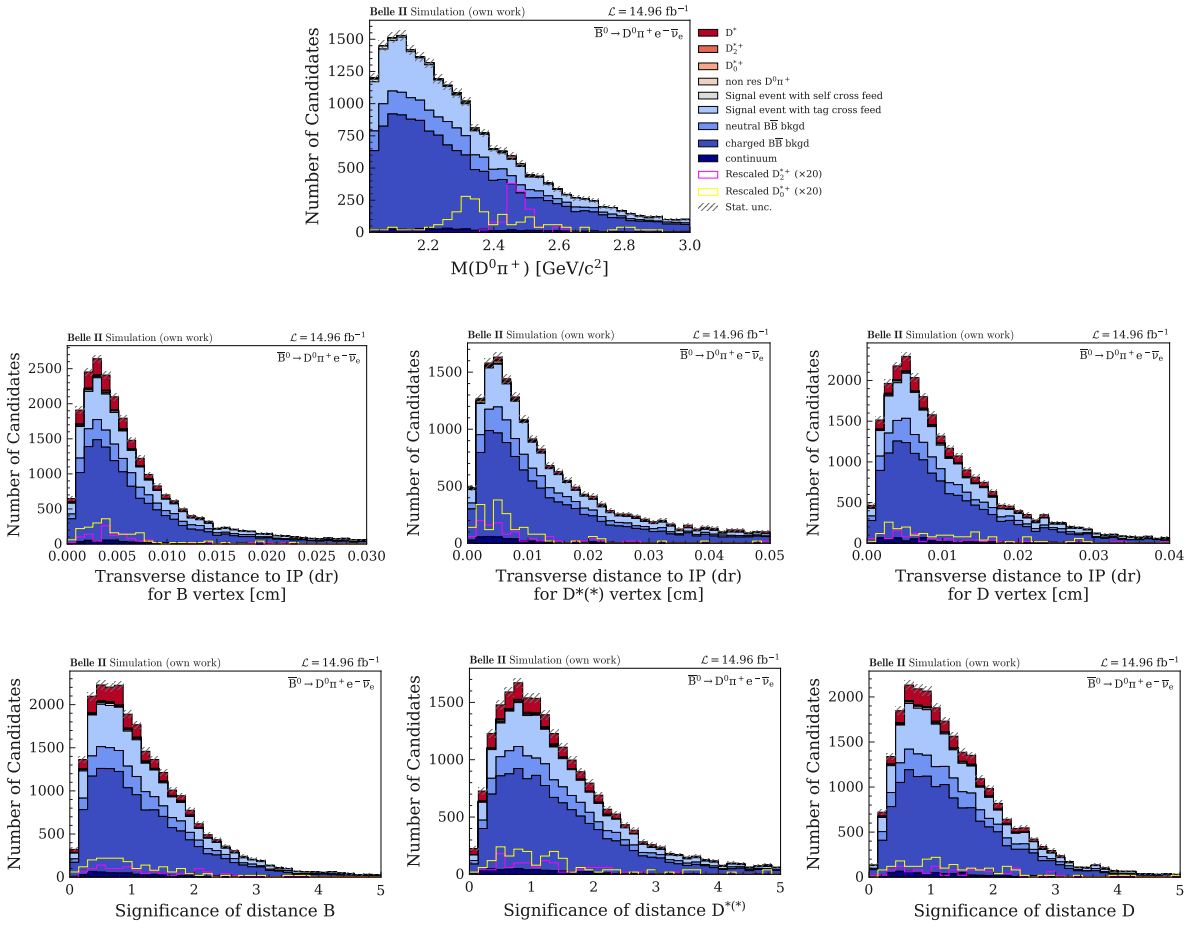


Figure A.1: More vertex variable distributions for the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode.

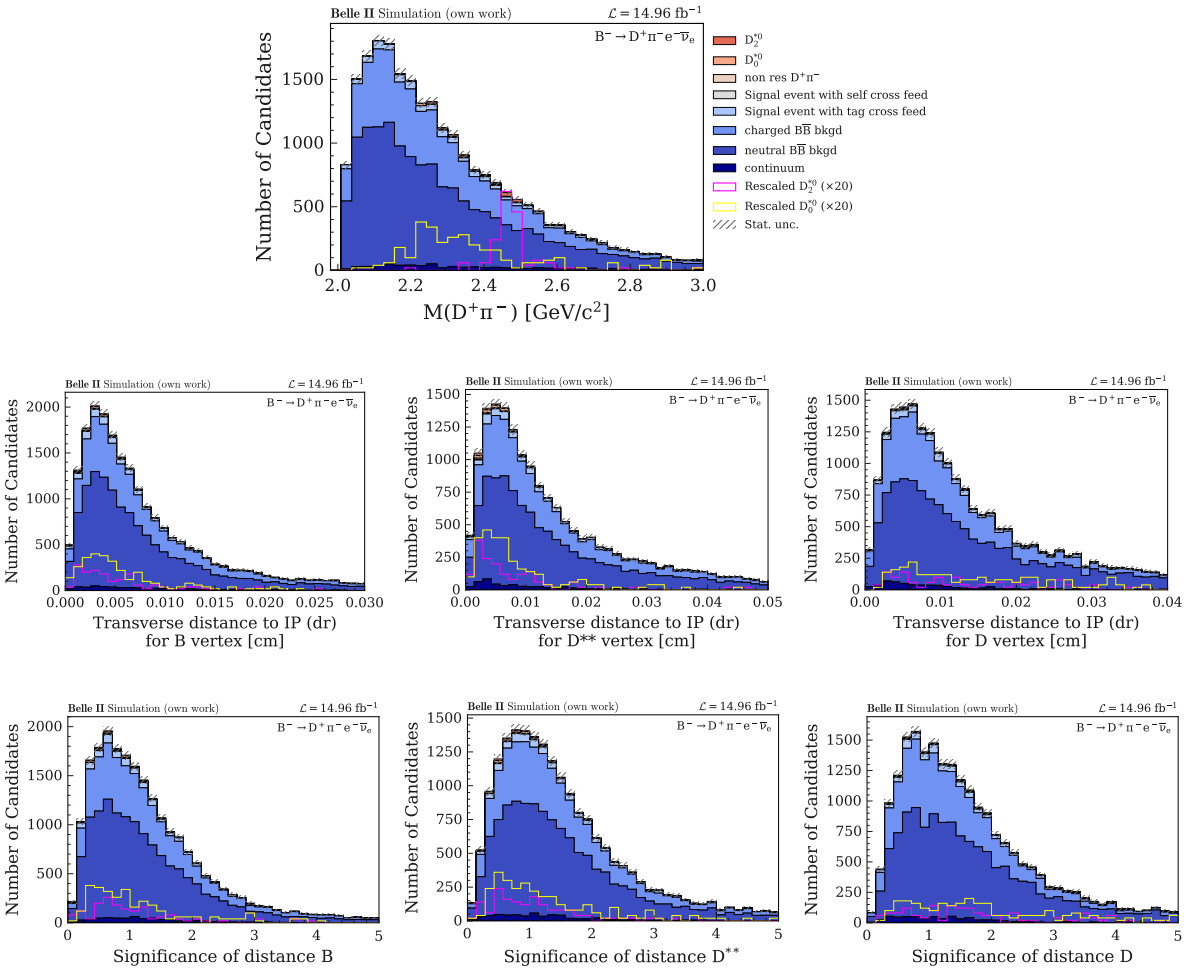


Figure A.2: More vertex variable distributions for the $B^- \to D^+\pi^-\ell^-\overline{\nu}_\ell$ decay mode.

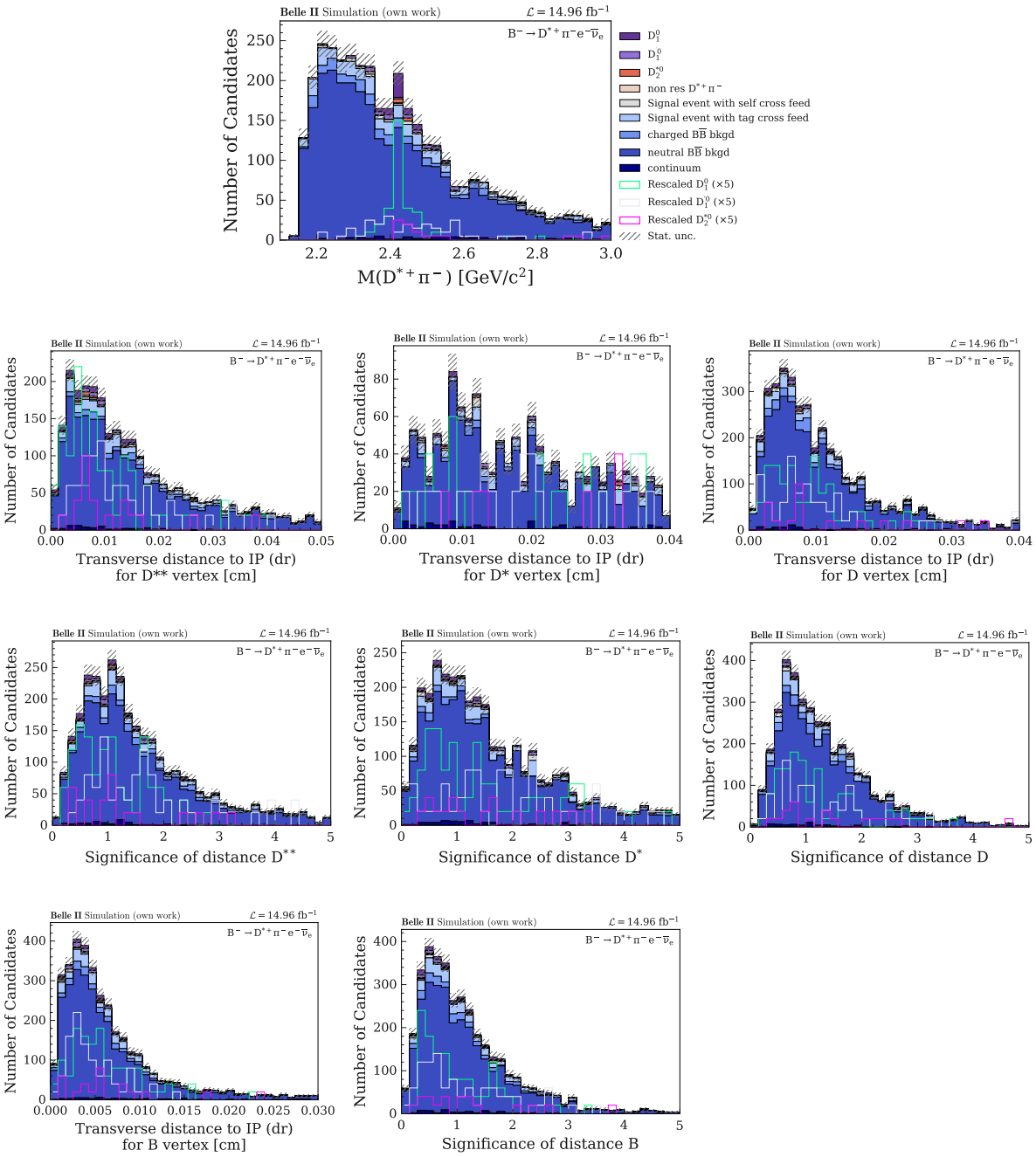


Figure A.3: More vertex variable distributions for the $B^- \to D^{*+}\pi^-\ell^-\overline{\nu}_\ell$ decay mode.

A.2 More variables

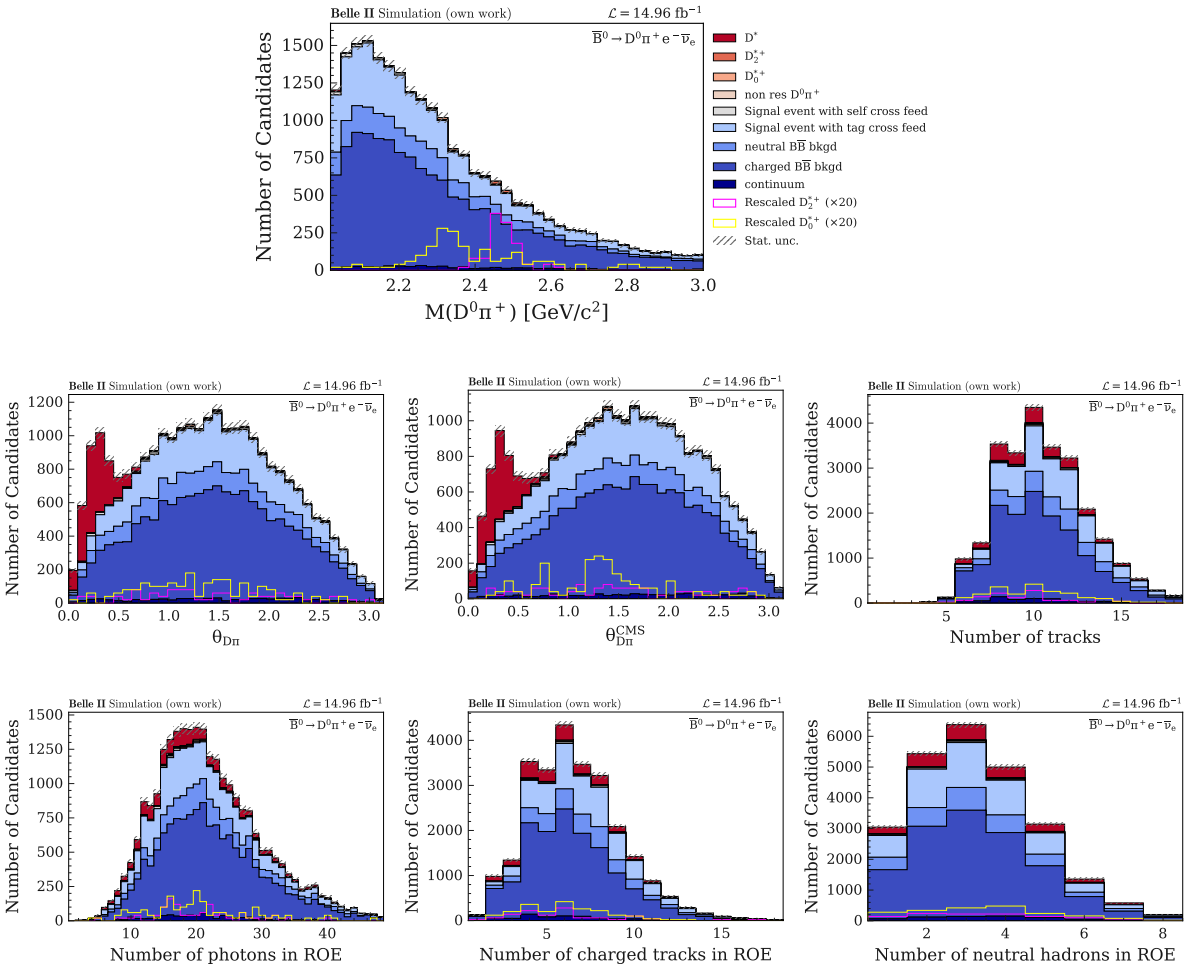


Figure A.4: More variable distributions for the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode.

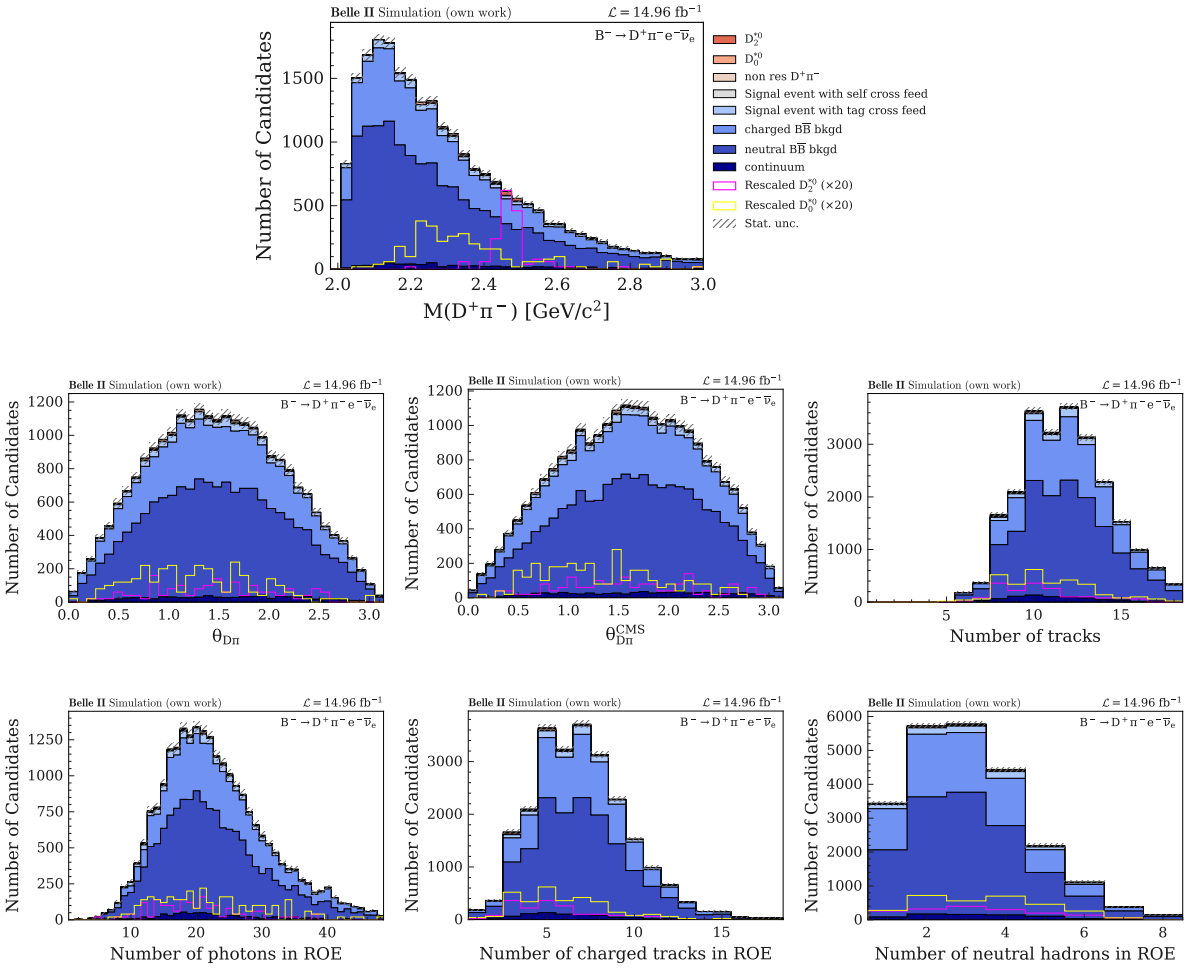


Figure A.5: More variable distributions for the $B^- \to D^+\pi^-\ell^-\bar{\nu}_\ell$ decay mode.

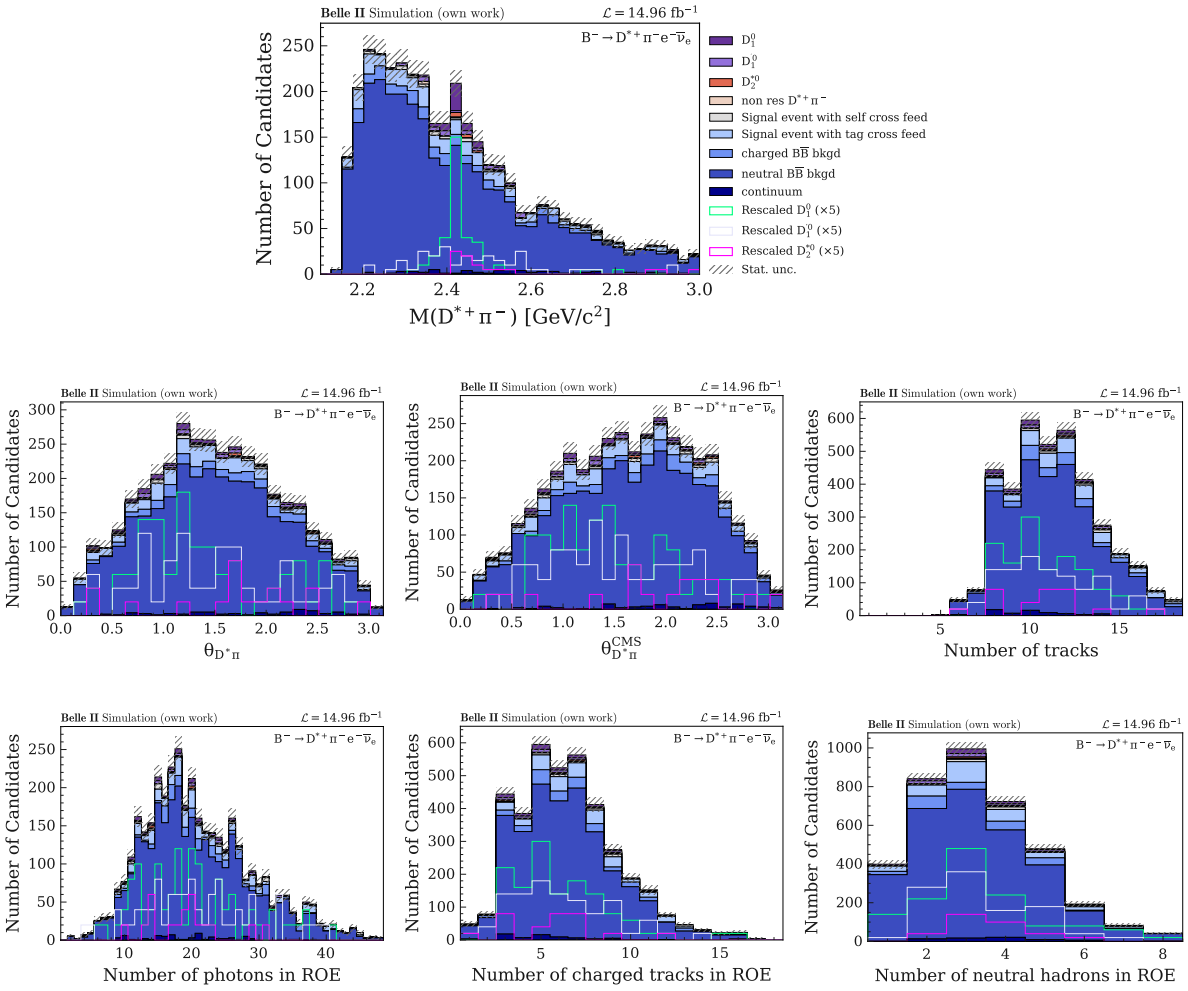


Figure A.6: More variable distributions for the $B^- \to D^{*+}\pi^-\ell^-\bar{\nu}_\ell$ decay mode.

APPENDIX B

Further BDT training plots for untagged analysis

B.1 Input variables for Signal MC for the untagged approach

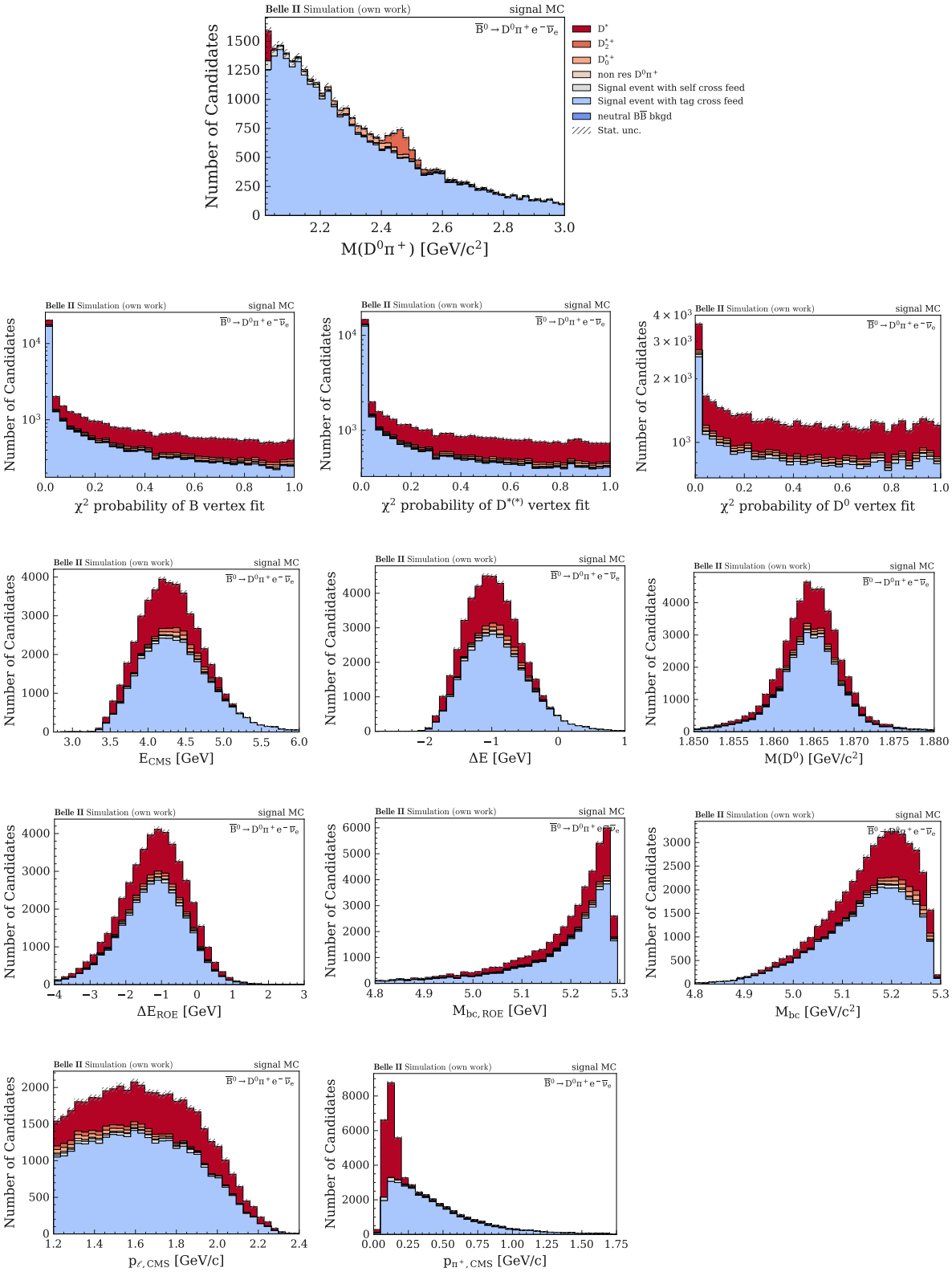


Figure B.1: BDT input variables for the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

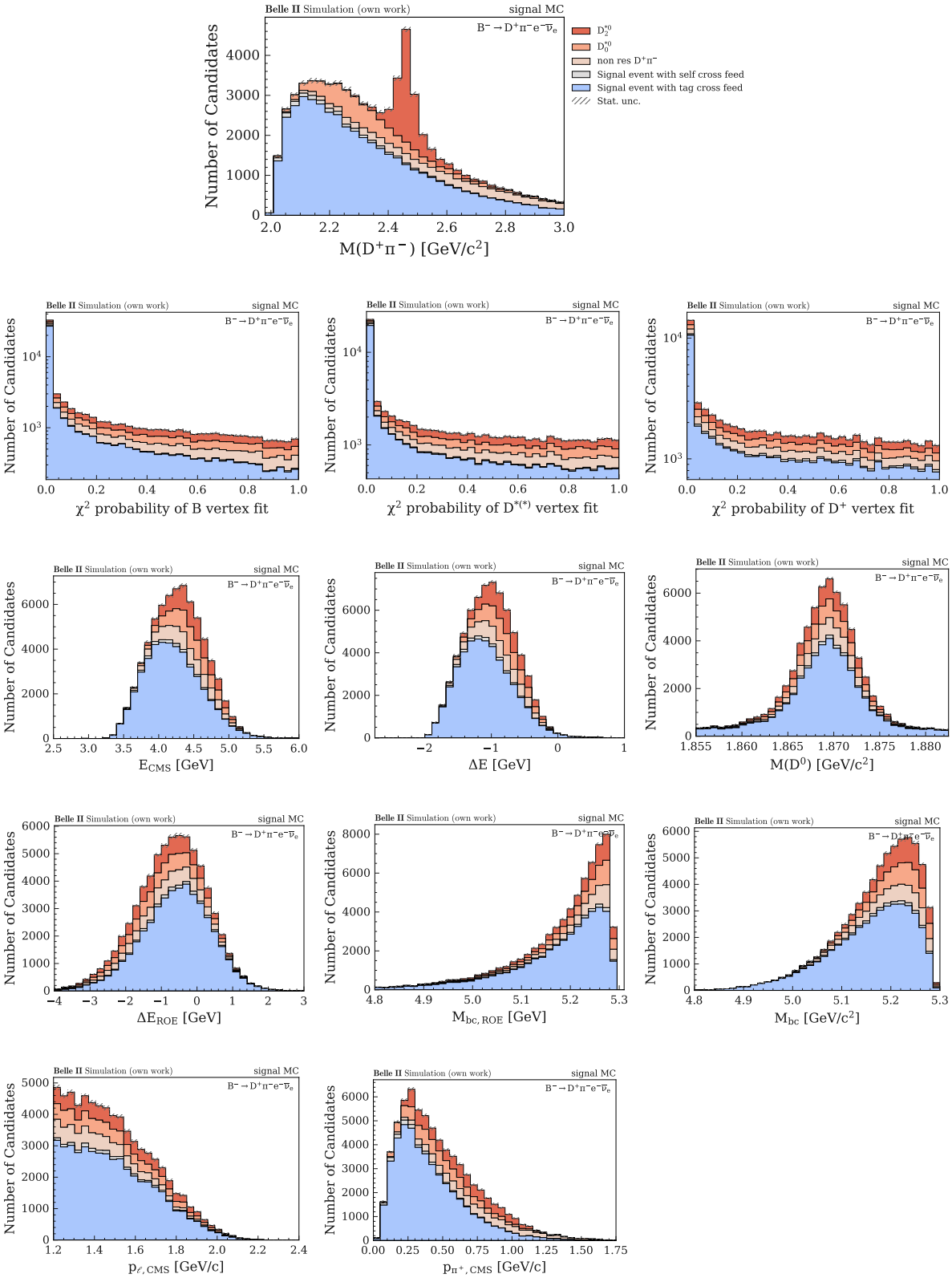


Figure B.2: BDT input variables for the $B^- \to D^+\pi^-\ell^-\bar{\nu}_\ell$ decay mode. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

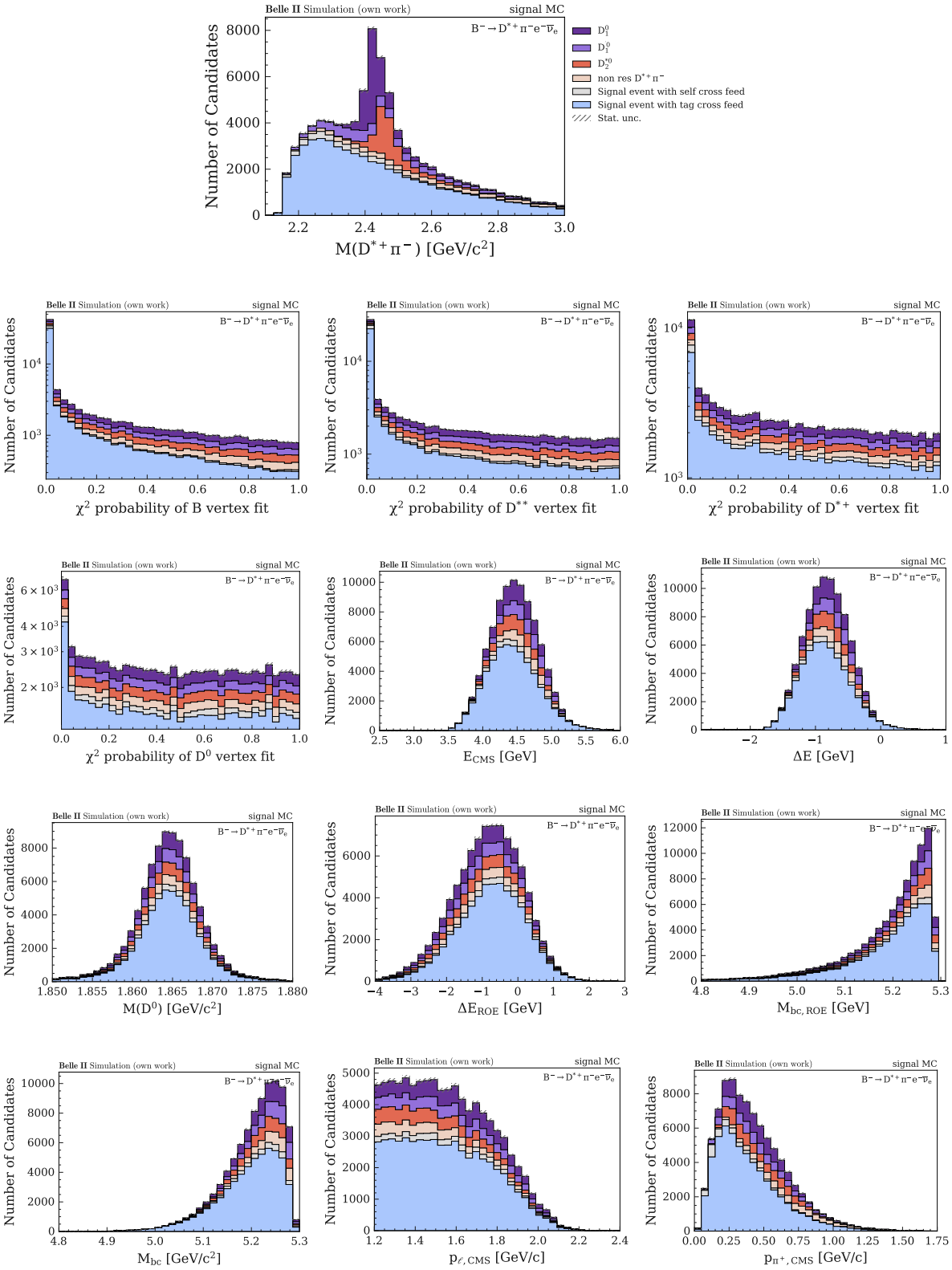


Figure B.3: BDT input variables for the $B^- \to D^{*+}\pi^-\ell^-\bar{\nu}_\ell$ decay mode. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

B.2 Different BDT training approaches for $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ channel.

B.2.1 Including the D^* resonance in the BDT training

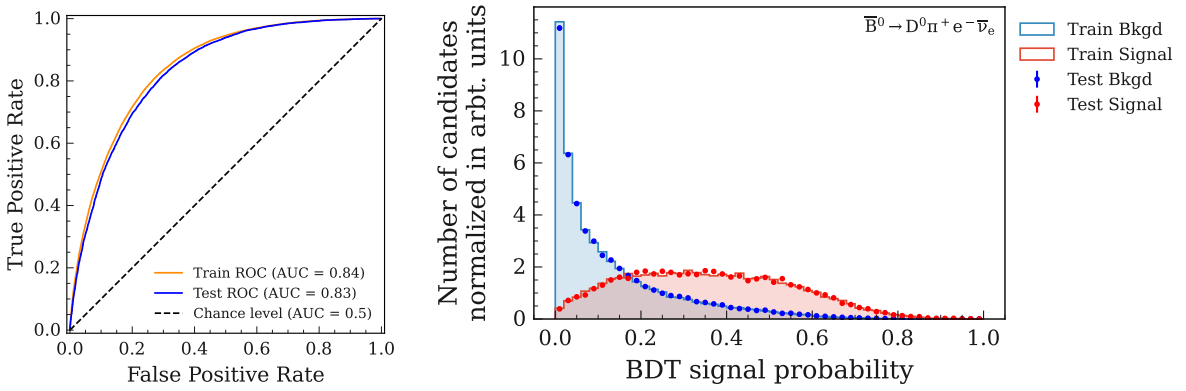


Figure B.4: (Left) ROC curve and (right) BDT signal probability for the BDT training of the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode, for the case that the D^* resonance is included.

B.2.2 Multi-classification approach

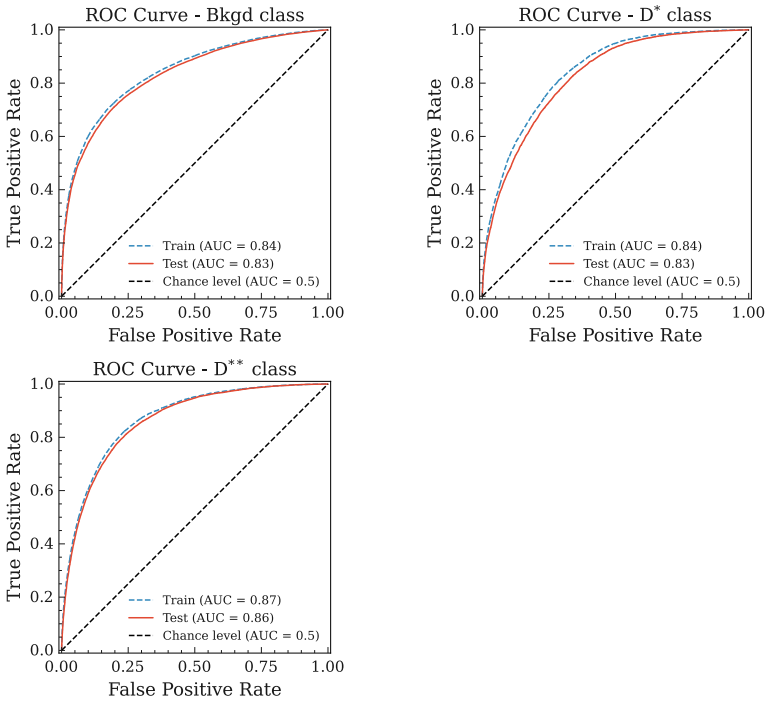


Figure B.5: ROC curves for the multi-class BDT training approach for the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ channel.

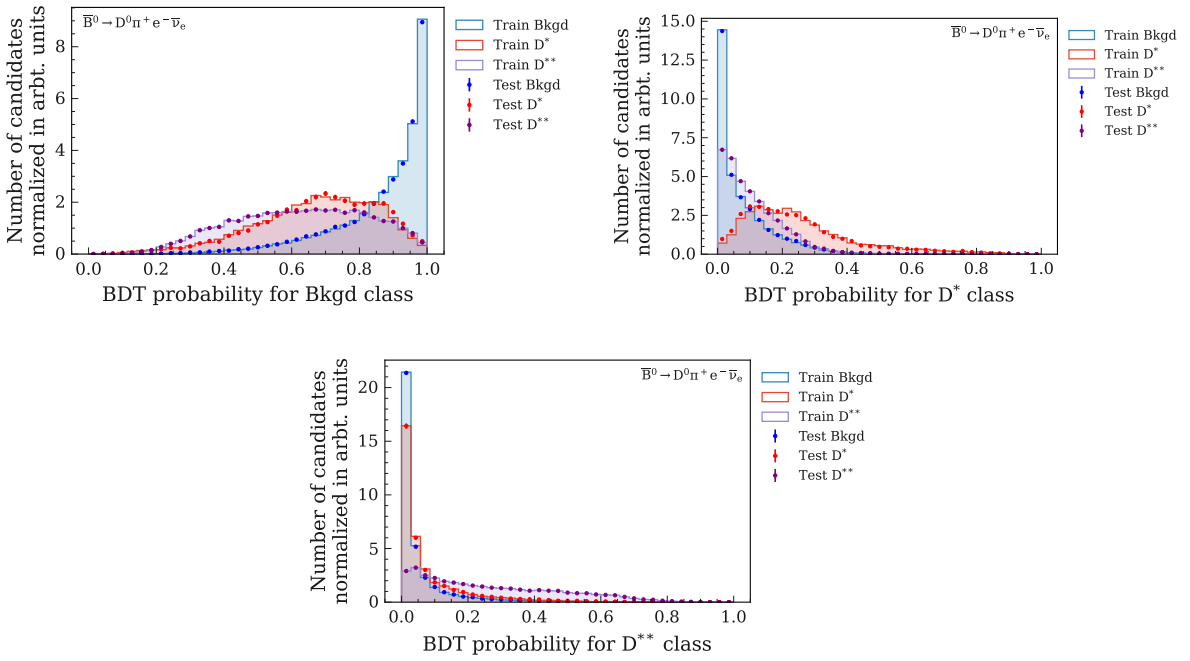


Figure B.6: BDT multi-class probabilities for the train and test set for the untagged analysis for the $\overline{B}^0 \to D^0 \pi^+ \ell^- \overline{\nu}_\ell$ channel.

APPENDIX C

Supplemental material to the SL tagged analysis regarding the event reconstruction and selection

C.1 Plots of online cut variables

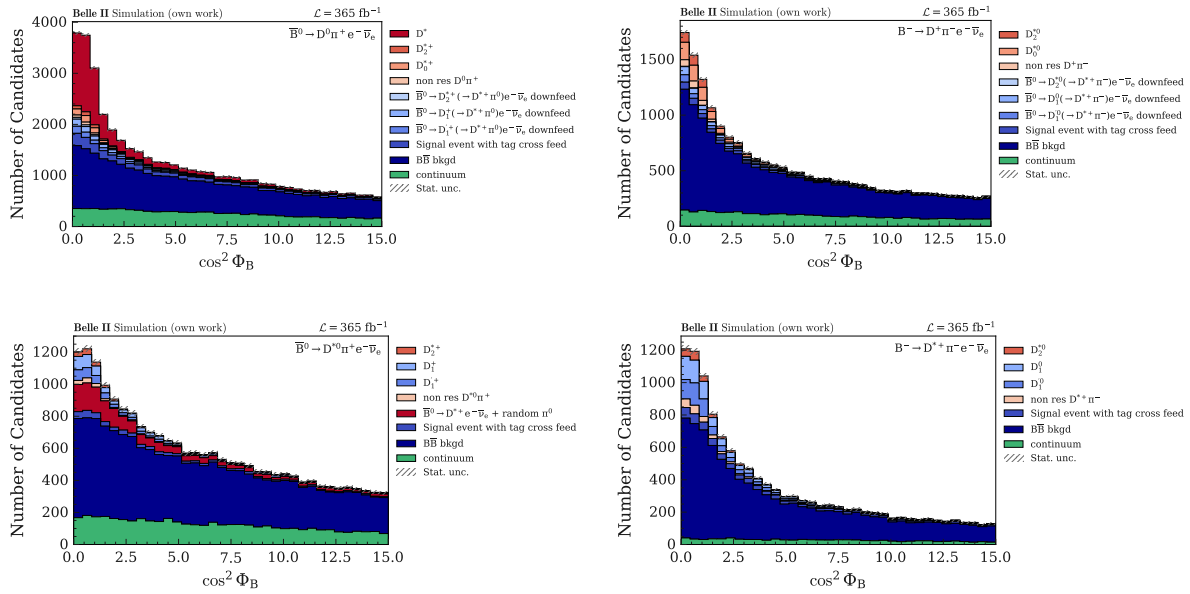


Figure C.1: Distribution of $\cos^2 \phi_B$ right after the reconstruction, where only online selection cuts are applied.

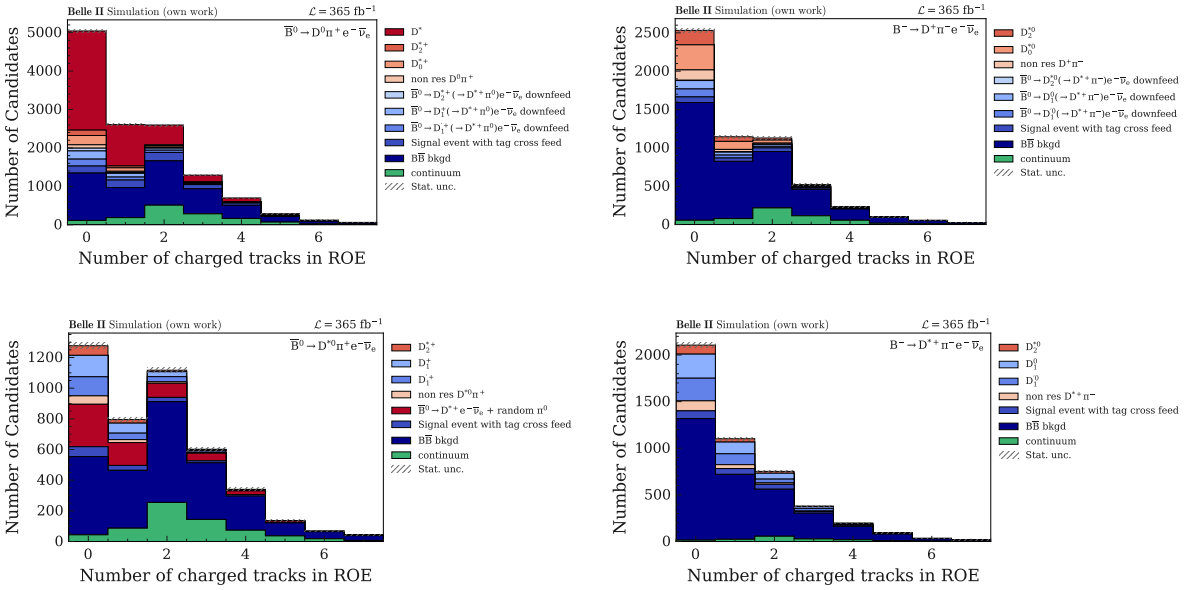


Figure C.2: Distribution of number of charged tracks in ROE after the $\cos^2 \phi_B$ cut is already applied.

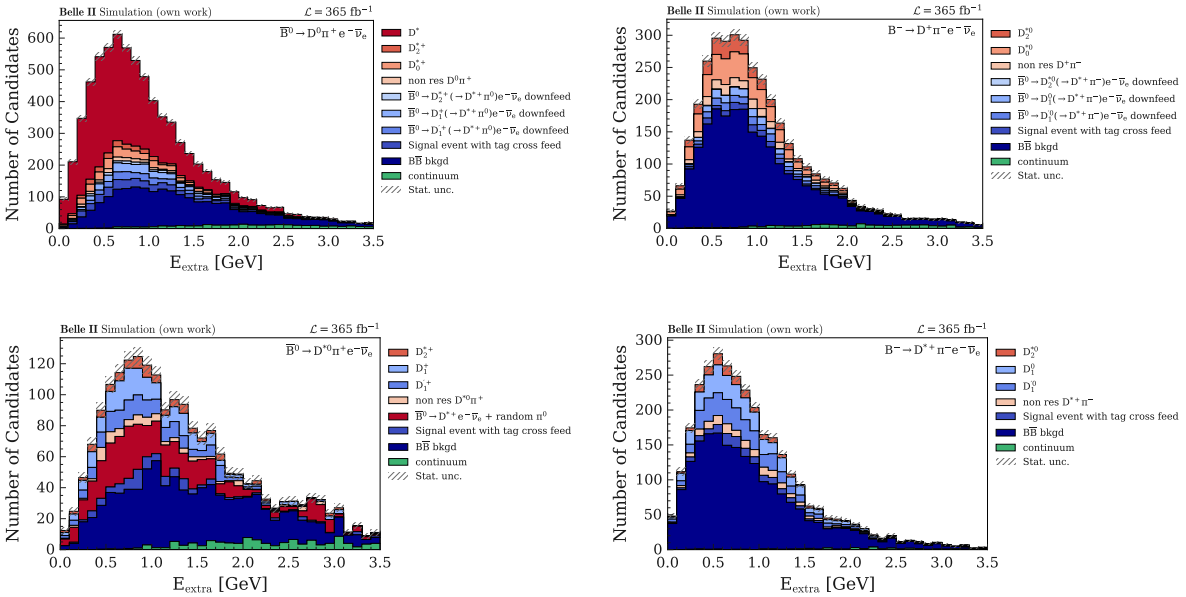


Figure C.3: Distribution of $E_{\text{extra}}^{\text{ECL}}$ after the $\cos^2 \phi_B$ and the number of charged tracks in ROE cut is already applied.

C.2 Self cross feed distributions

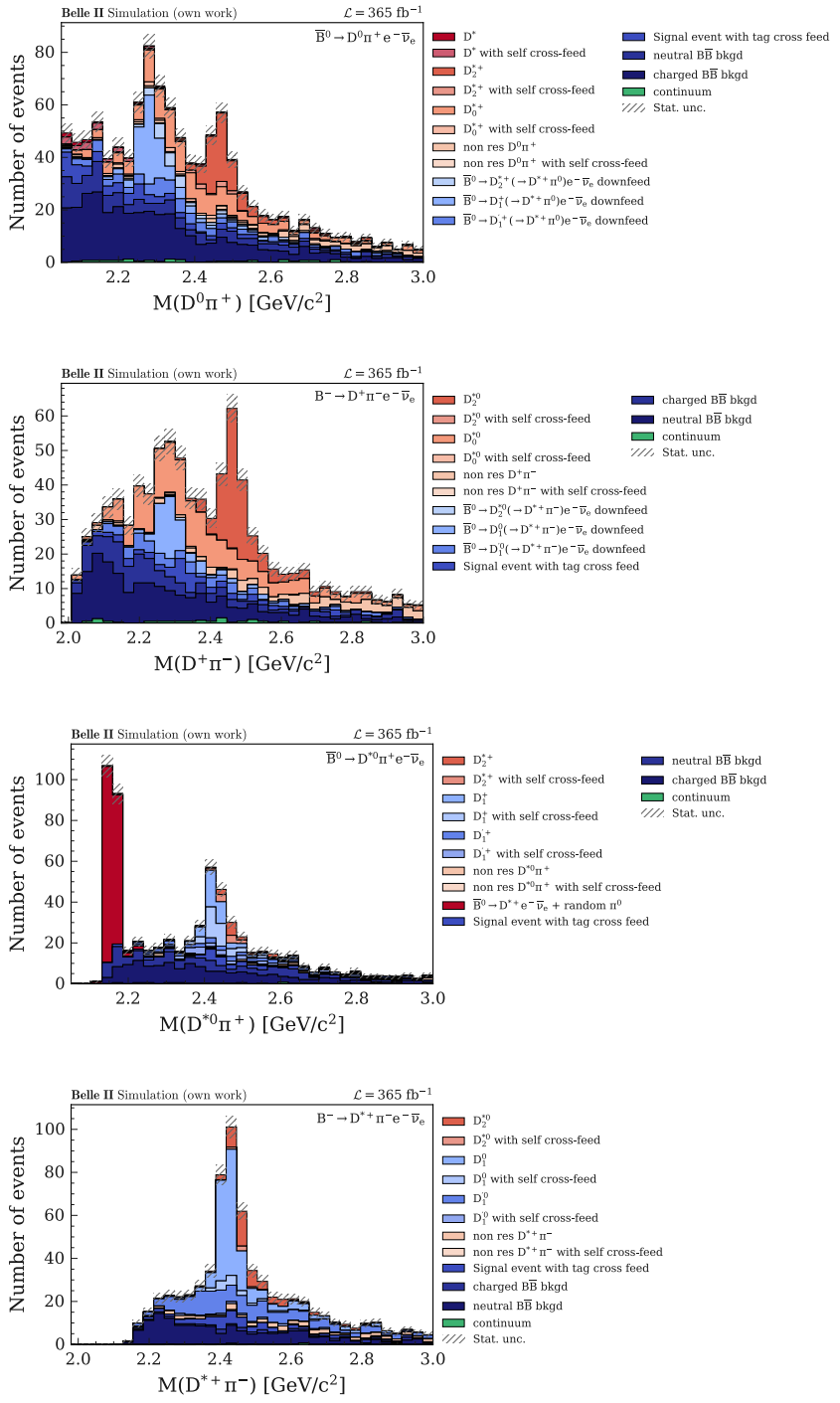


Figure C.4: $M(D^{(*)}\pi)$ spectrum after the BCS is applied. In these distributions, the SCF events are shown separately and are not included in the corresponding signal categories (as usual).

Table C.1: Fraction of SCF events to all signal events for all four B decay modes.

	Fraction of SCF events in %
$\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$	9.36 ± 0.65
$B^- \to D^+\pi^-e^-\bar{\nu}_e$	1.11 ± 0.53
$\bar{B}^0 \to D^{*0}\pi^+e^-\bar{\nu}_e$	50.28 ± 5.74
$B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$	9.46 ± 1.44

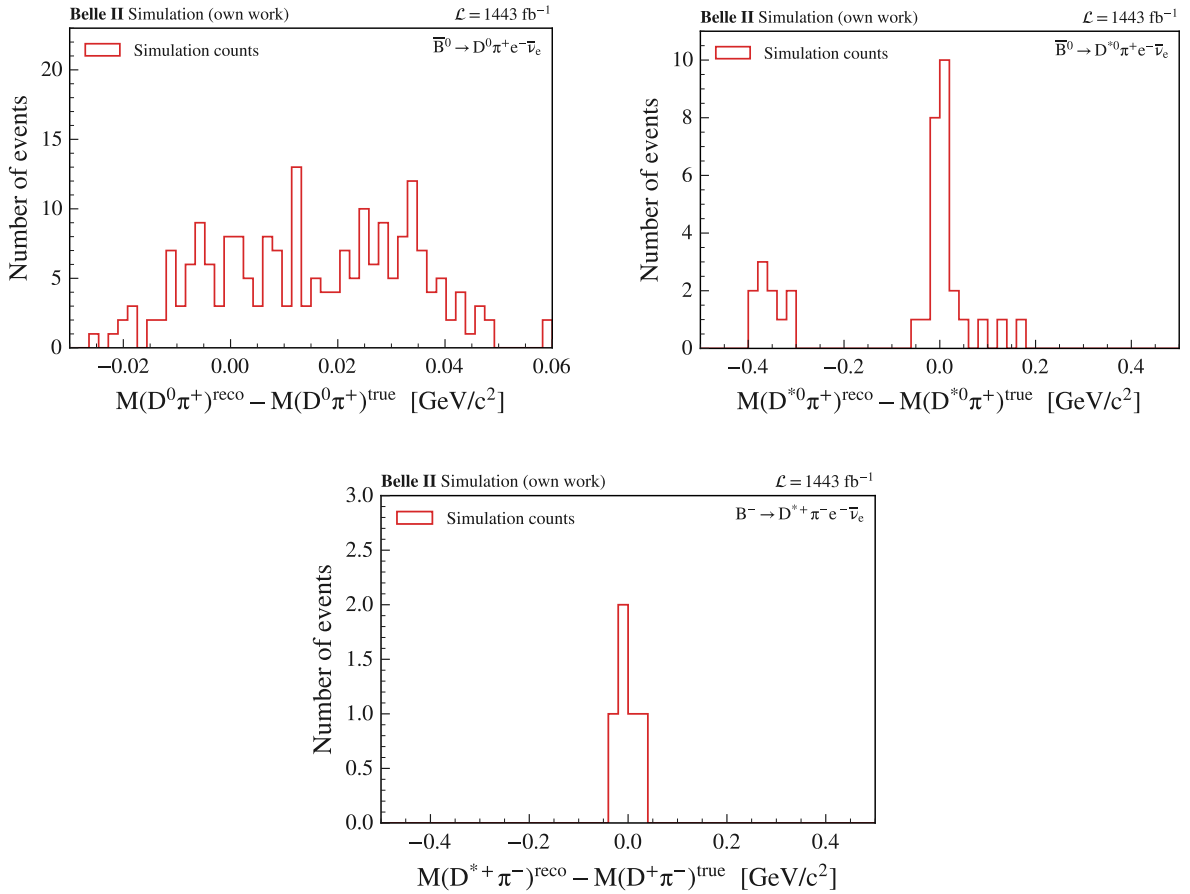


Figure C.5: $M(D^{(*)}\pi$ distributions of residuals for SCF events. Note that the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ channel is excluded, since it does more or less not have any SCF events. It is important to note that a considerable fraction of events is lost due to the limited availability of MC truth information. In particular, the MC matching frequently fails in the presence of a reconstructed SCF event.

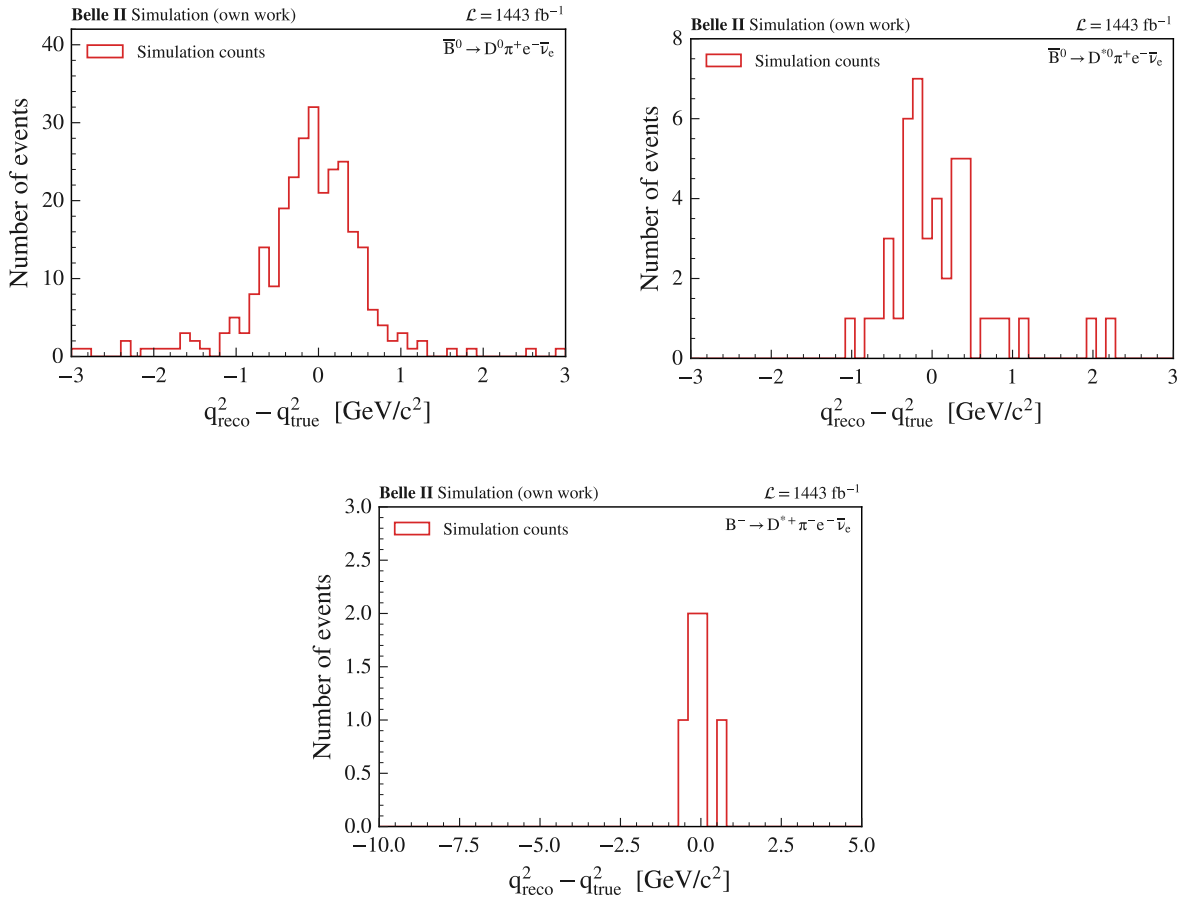


Figure C.6: q^2 distributions of residuals for SCF events. Note that the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ channel is excluded, since it does more or less not have any SCF events. It is important to note that a considerable fraction of events is lost due to the limited availability of MC truth information. In particular, the MC matching frequently fails in the presence of a reconstructed SCF event.

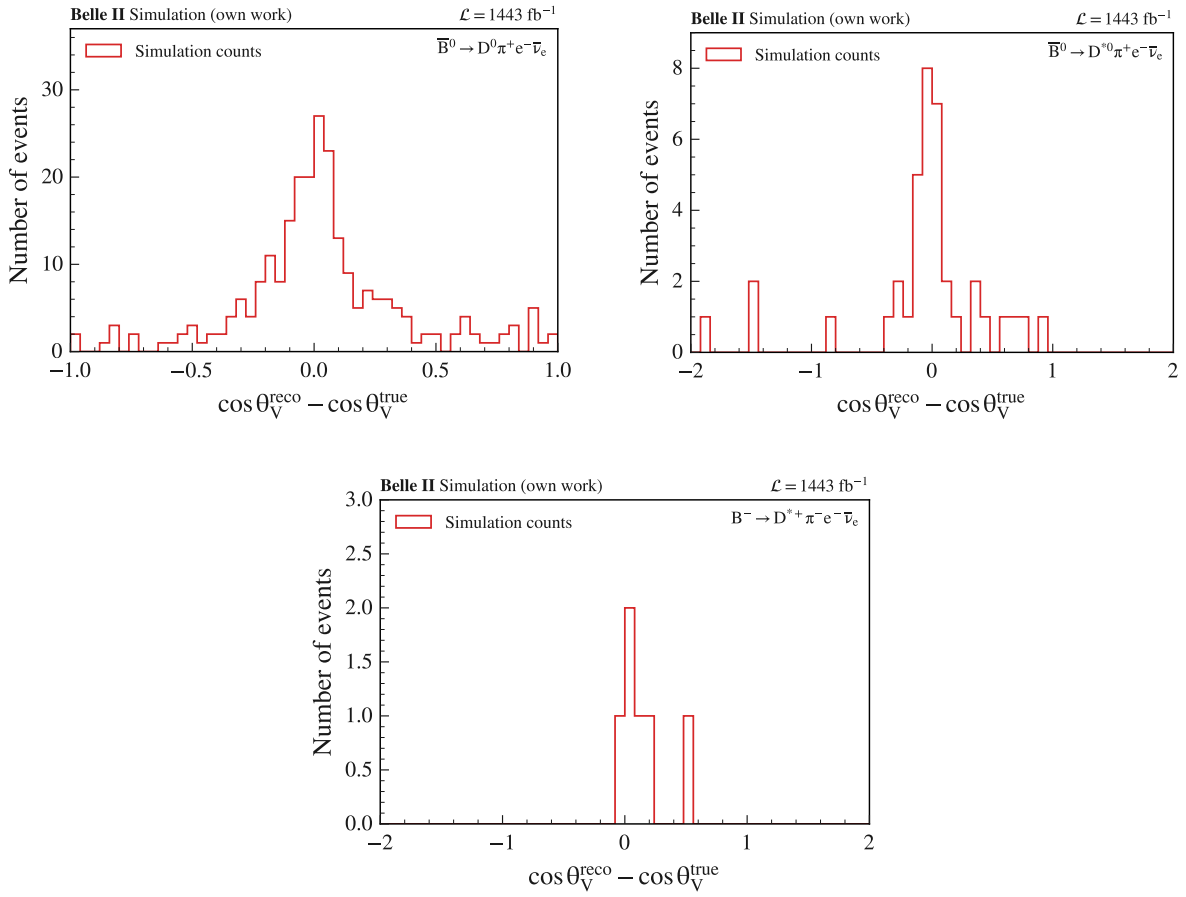


Figure C.7: $\cos\theta_V$ distributions of residuals for SCF events. Note that the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ channel is excluded, since it does more or less not have any SCF events. It is important to note that a considerable fraction of events is lost due to the limited availability of MC truth information. In particular, the MC matching frequently fails in the presence of a reconstructed SCF event.

C.3 BCS variable

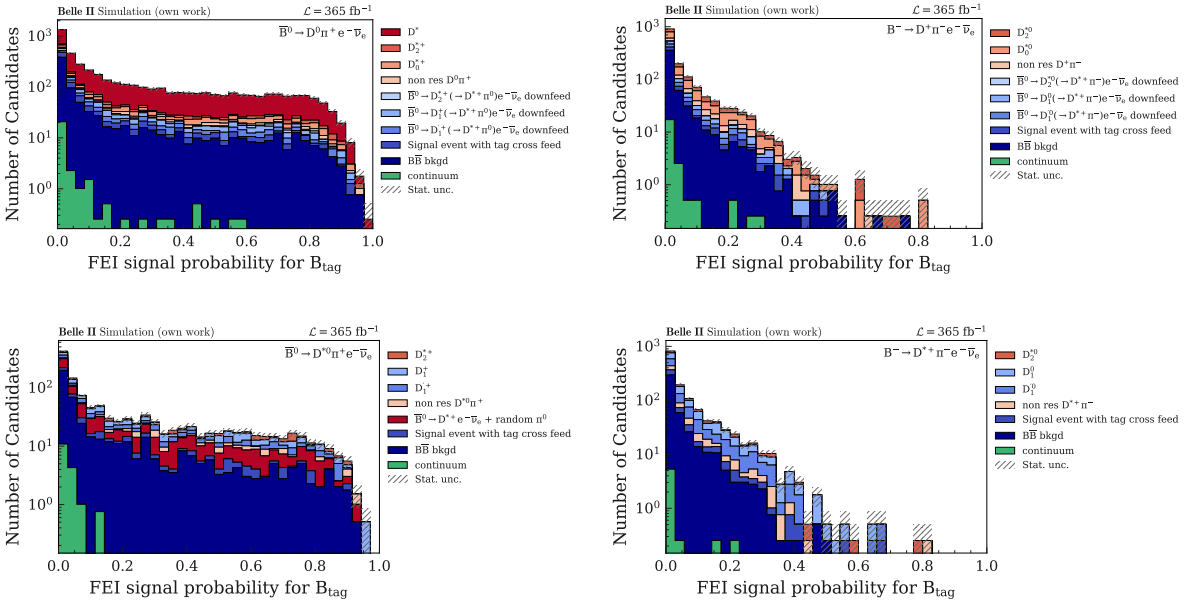


Figure C.8: Signal probability distribution of SL FEI. This variable is used to perform the BCS on. Note that the log-scale on the y-axis is on.

C.4 q^2 and $\cos\theta_V$ after BCS

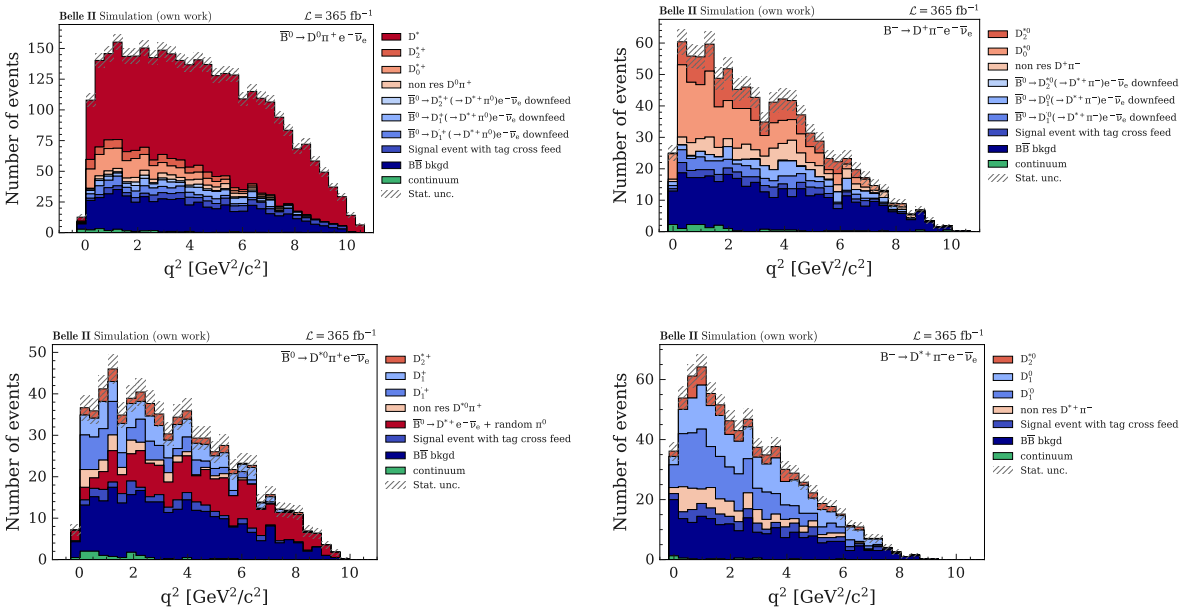


Figure C.9: q^2 distributions after the BCS is applied for all four B decay modes.

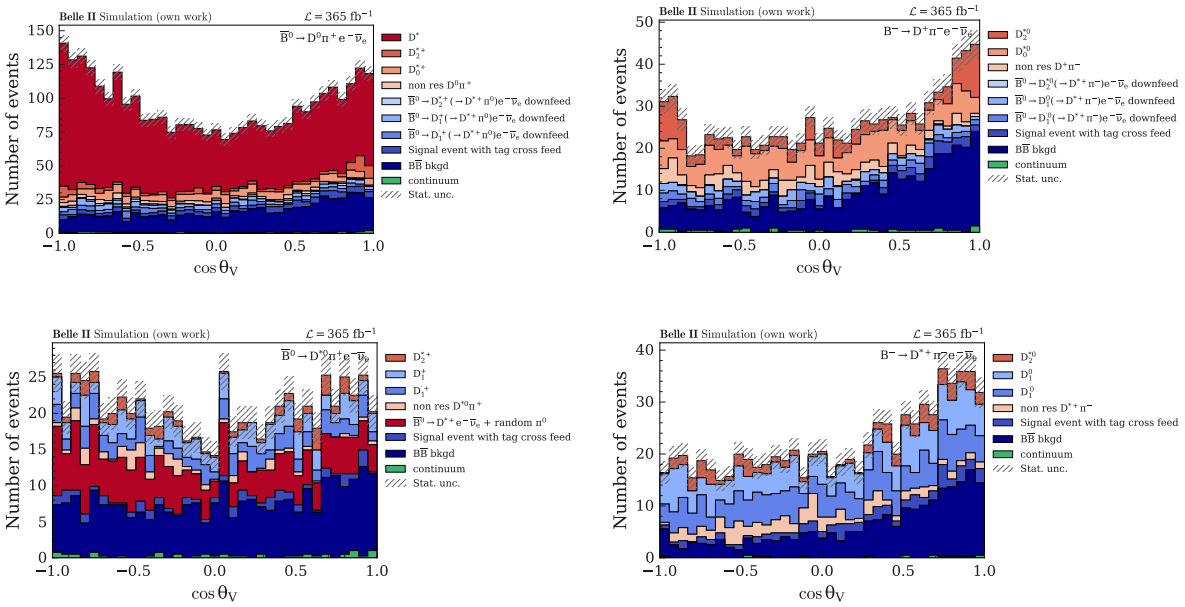


Figure C.10: $\cos\theta_V$ distributions after the BCS is applied for all four B decay modes.

APPENDIX D

Supplemental material to the SL tagged analysis regarding the signal extraction

D.1 $\cos\theta_{BY}$

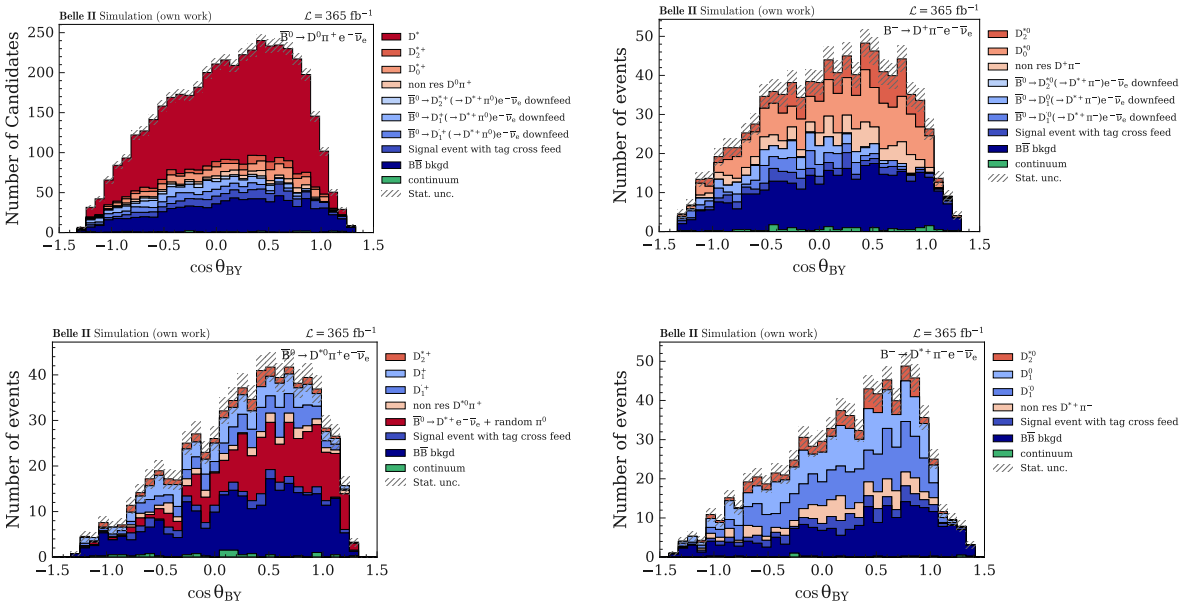


Figure D.1: $\cos\theta_{BY}^{\text{sig}}$ distributions for all four B decay modes.

D.2 BDT input variable distributions

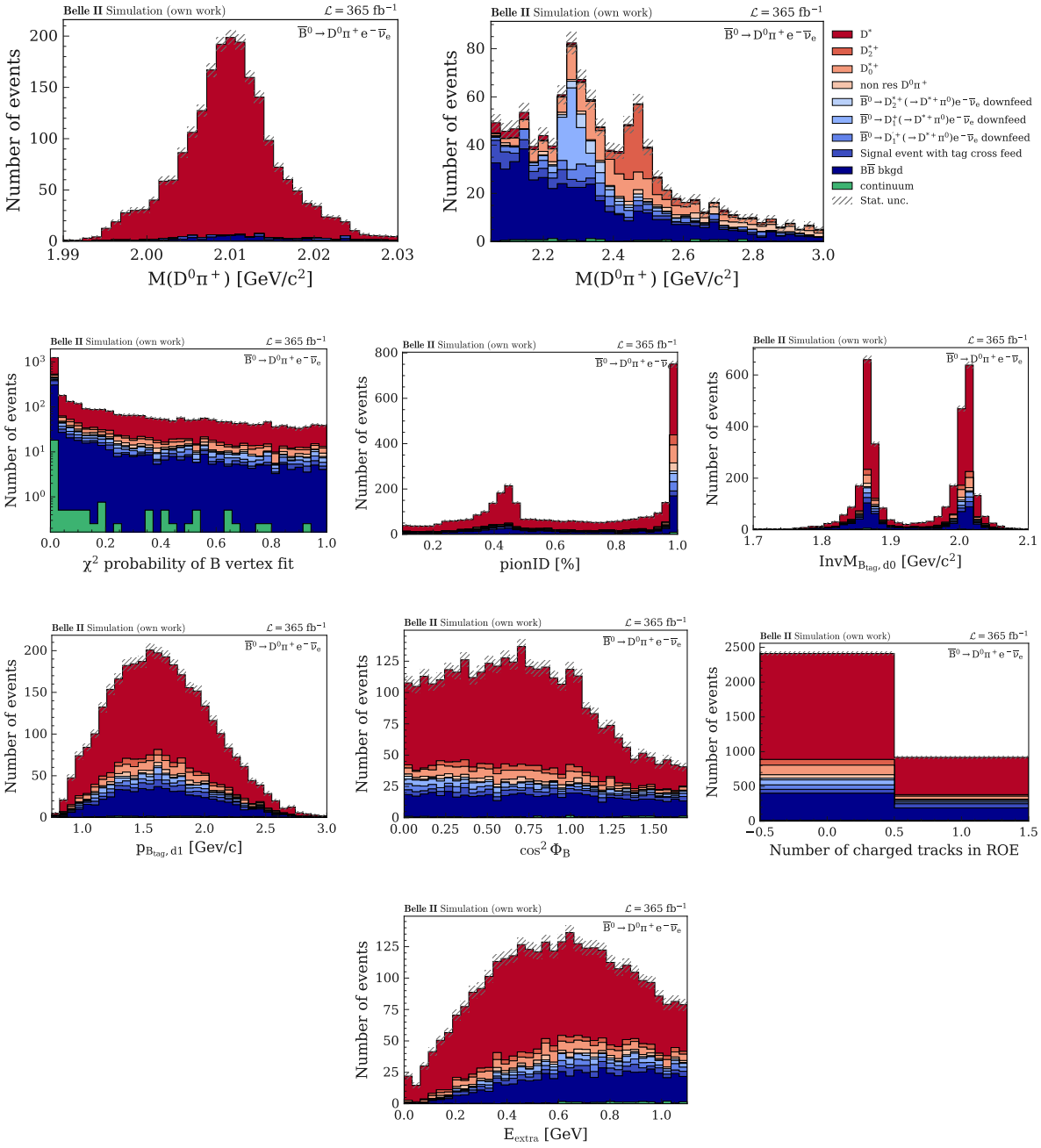


Figure D.2: BDT input variables for the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode for the MC15rd sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

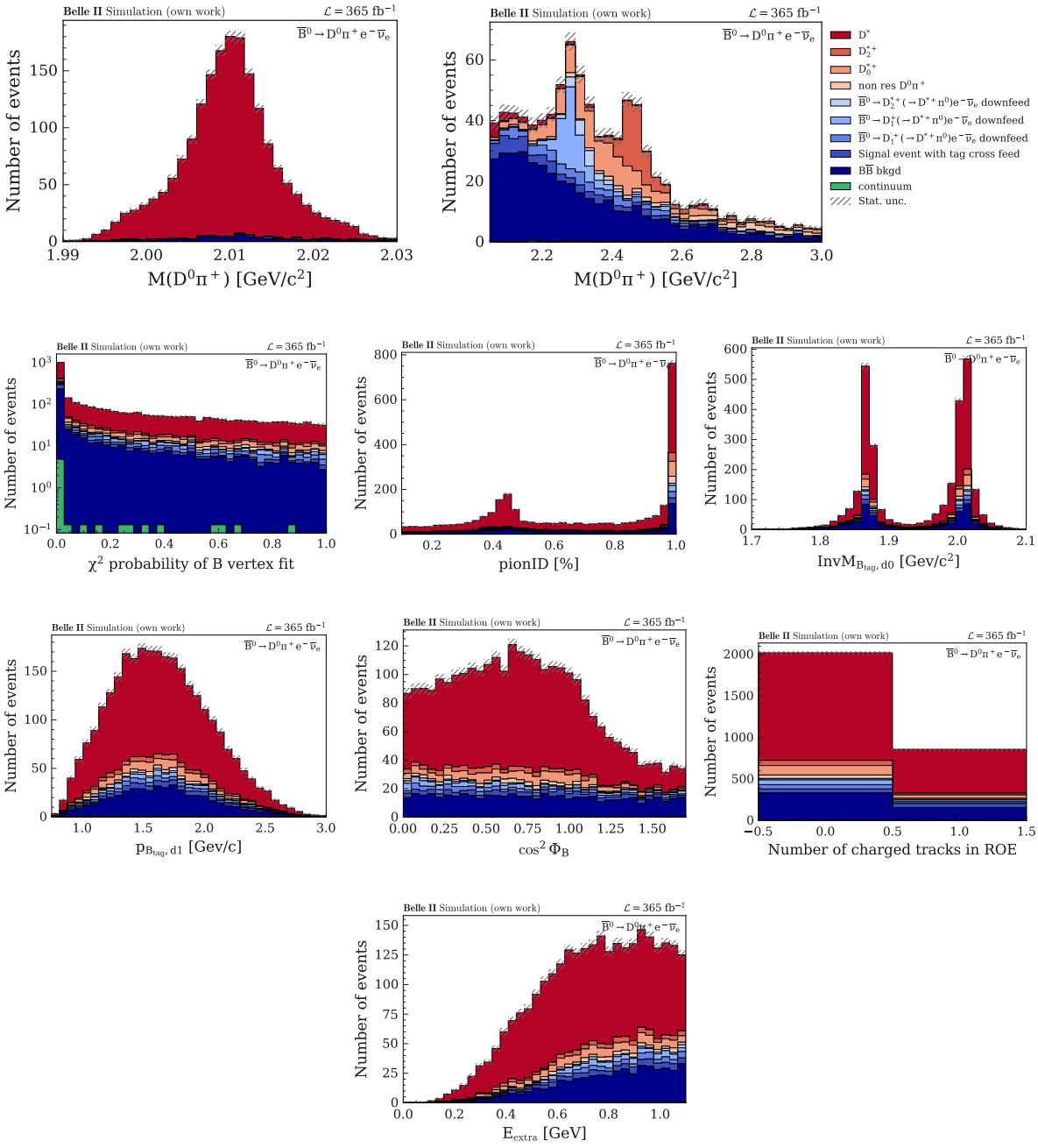


Figure D.3: BDT input variables for the $\overline{B}^0 \to D^0\pi^+\ell^-\overline{\nu}_\ell$ decay mode for the MC15ri sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

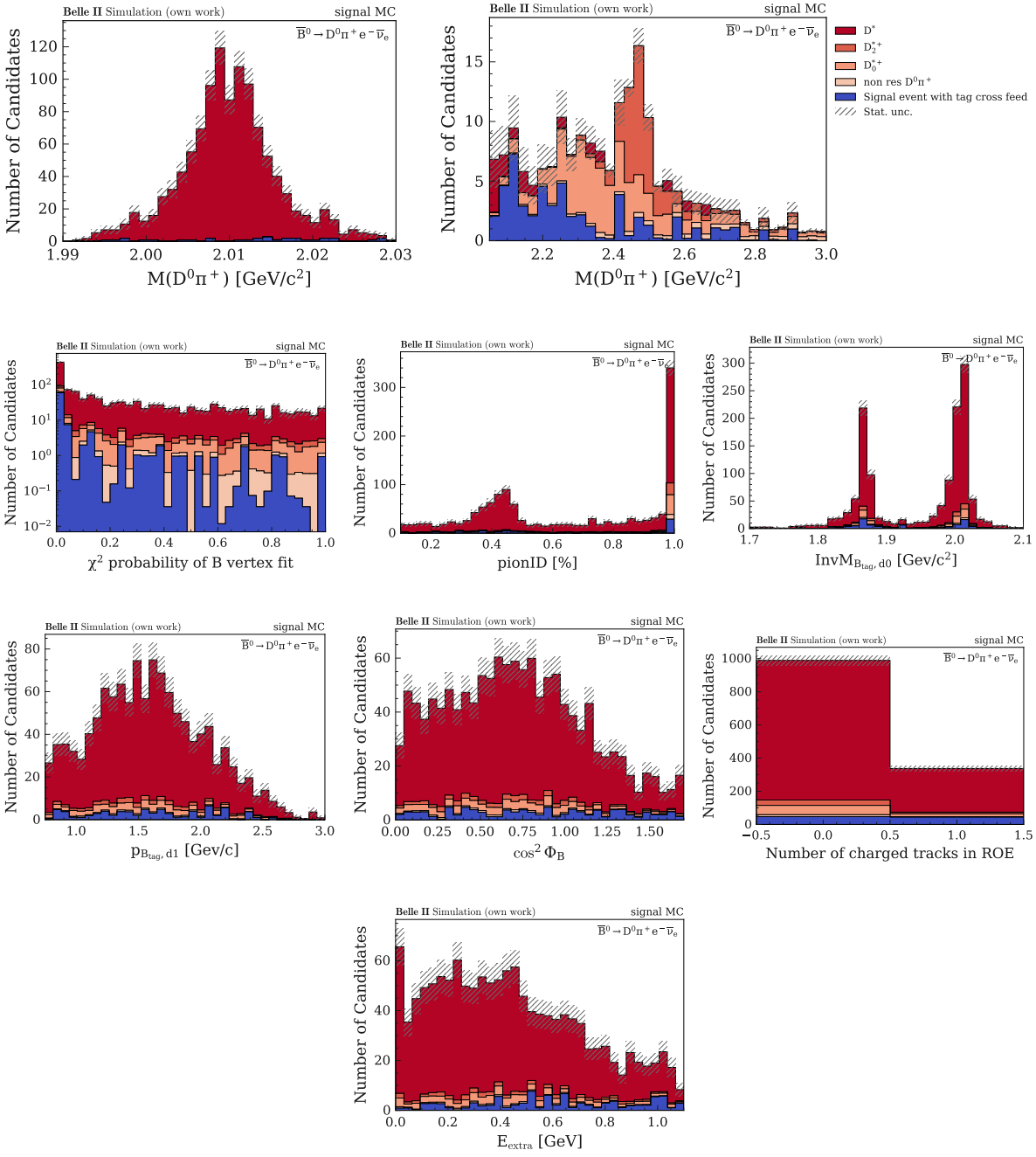


Figure D.4: BDT input variables for the $\overline{B}^0 \to D^0\pi^+e^-\overline{\nu}_e$ decay mode for the ri signal MC sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

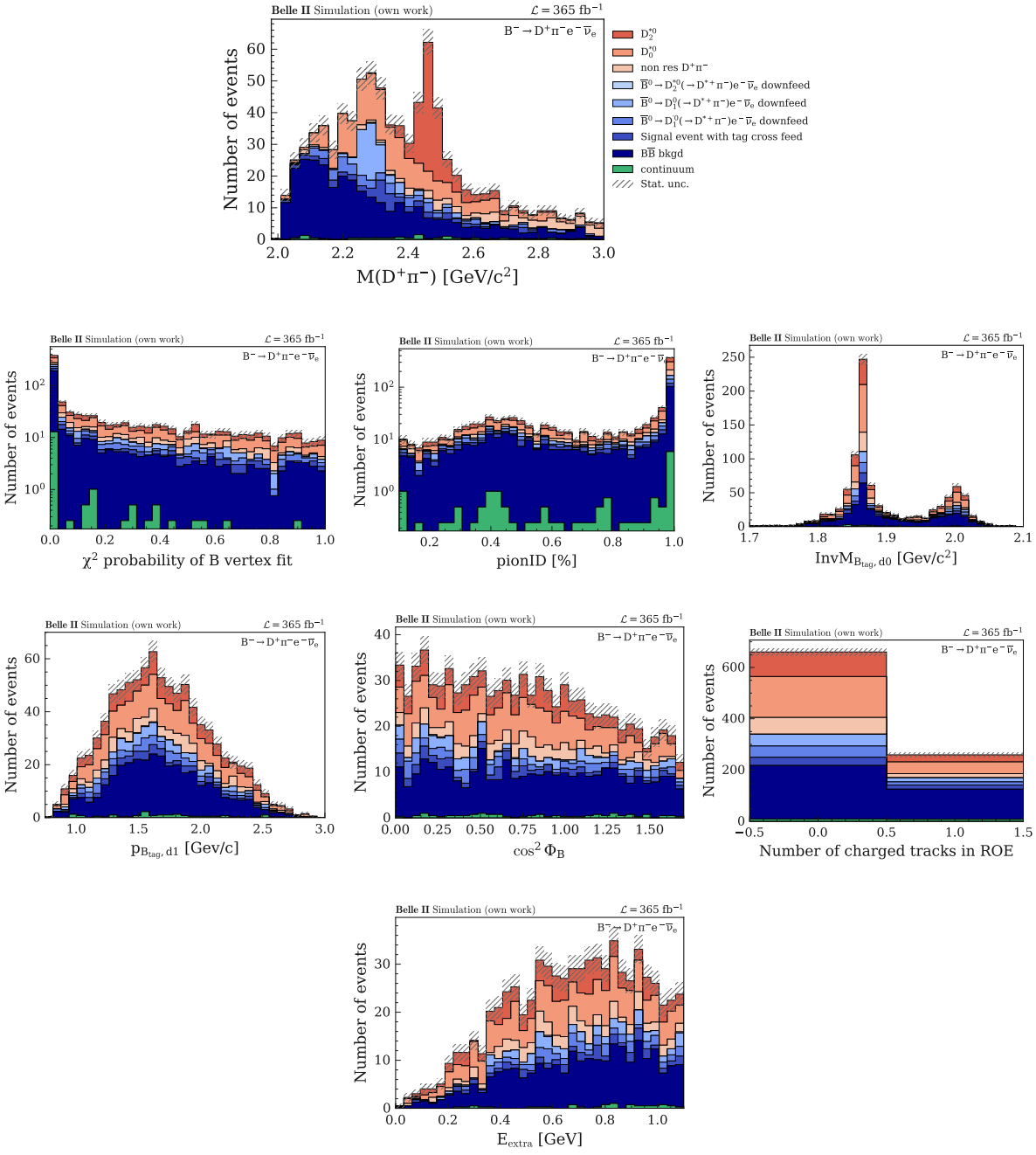


Figure D.5: BDT input variables for the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ decay mode for the MC15rd sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

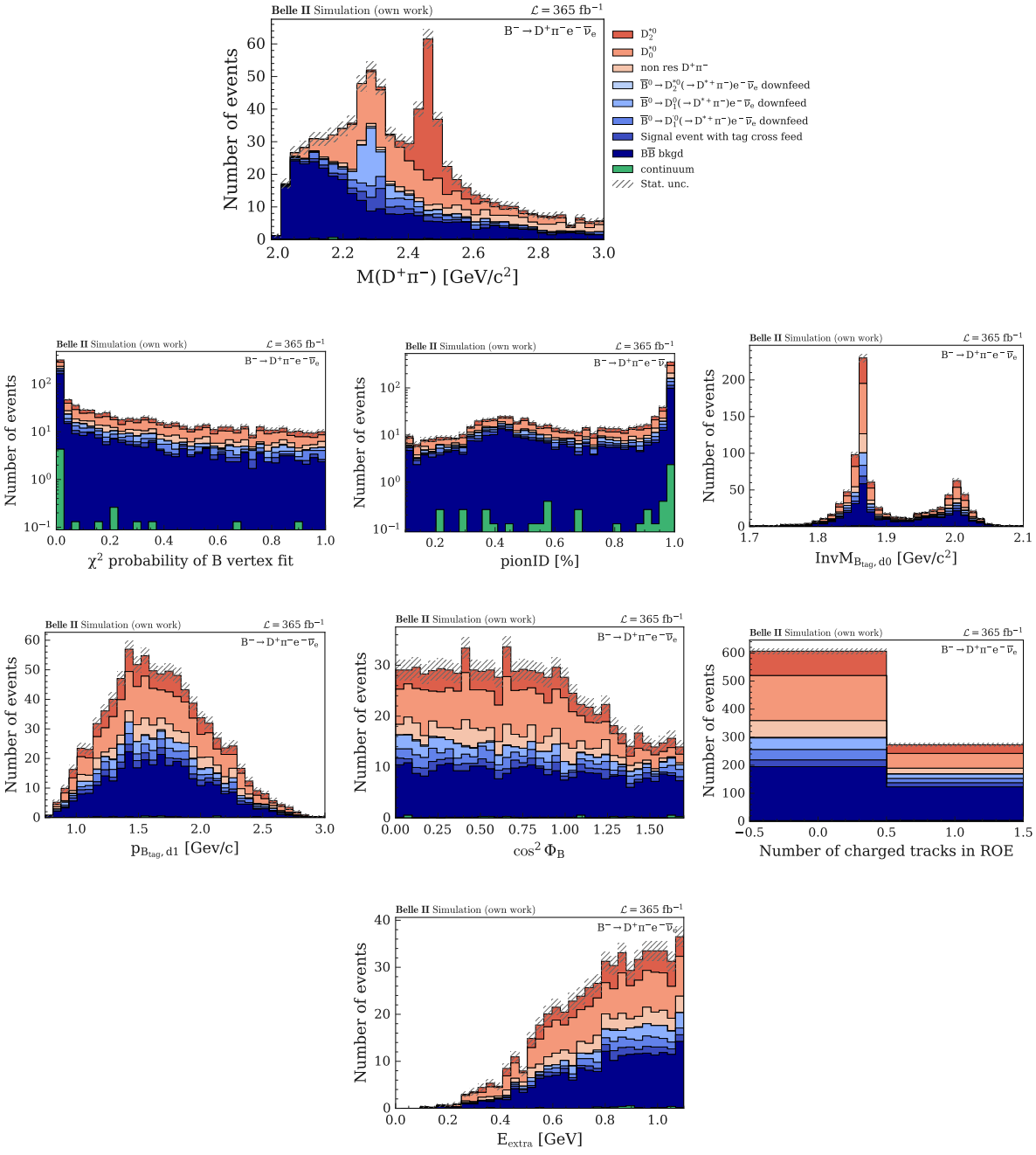


Figure D.6: BDT input variables for the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ decay mode for the MC15ri sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

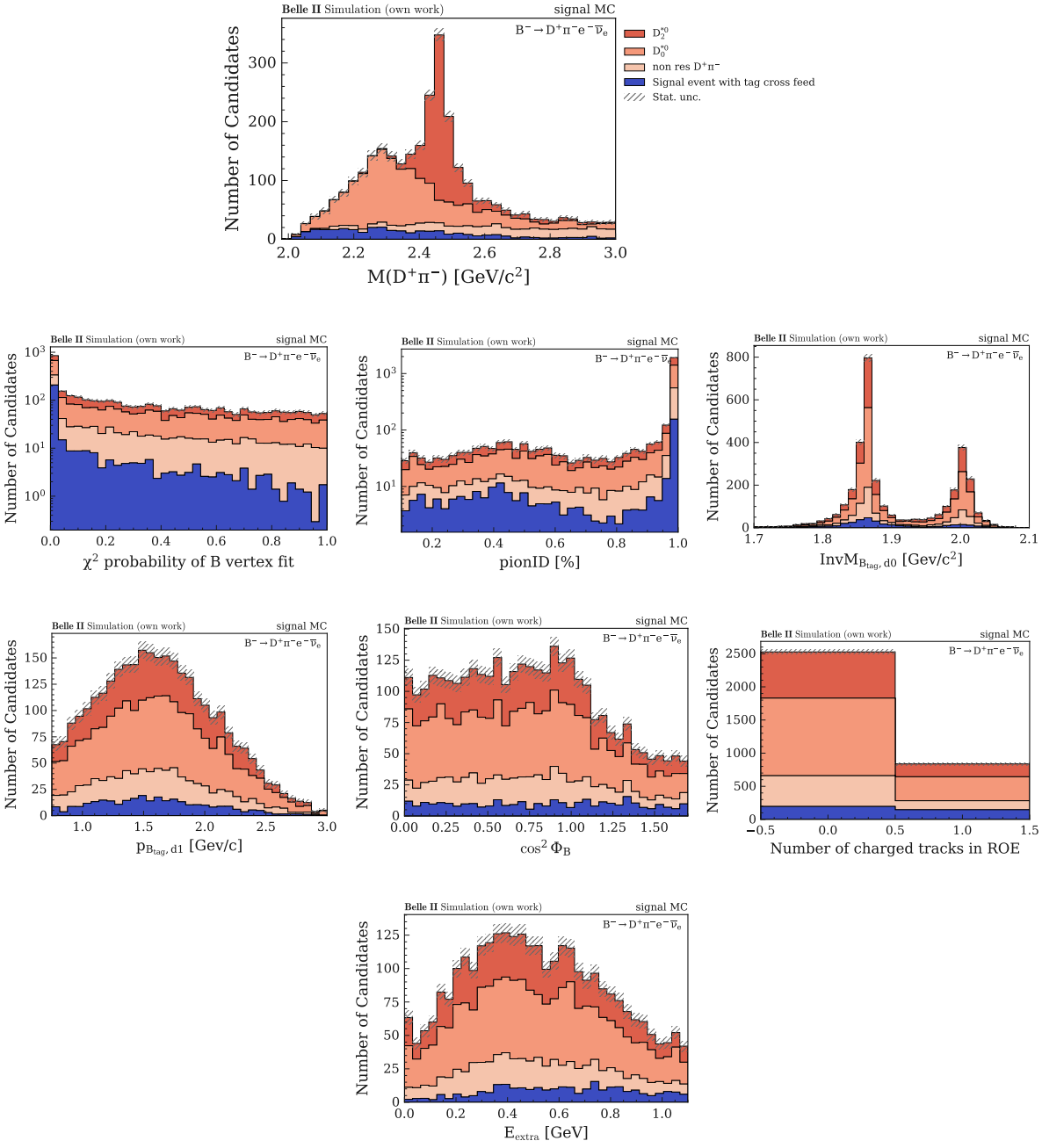


Figure D.7: BDT input variables for the $B^- \to D^+\pi^-e^-\bar{\nu}_e$ decay mode for the ri signal MC sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

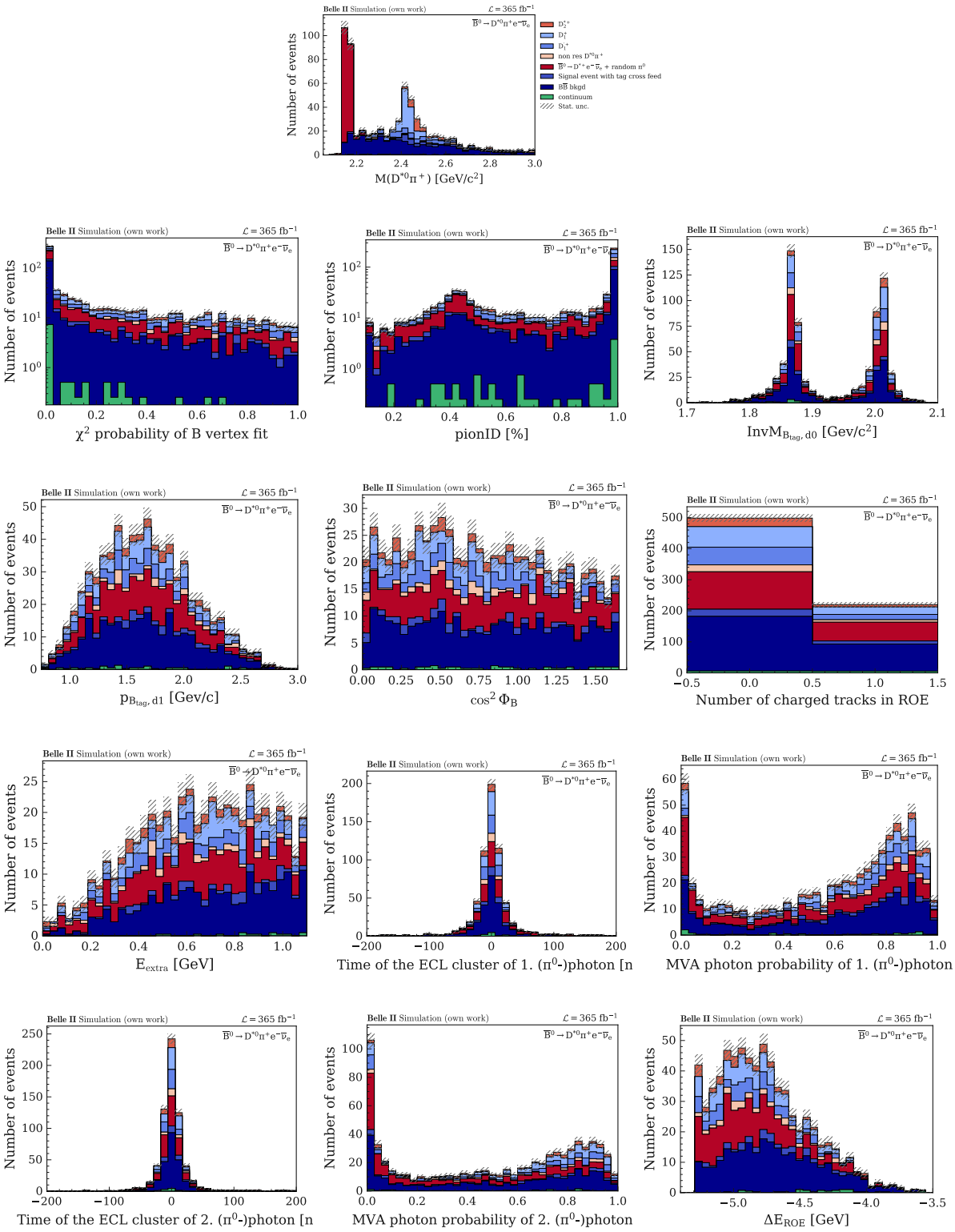


Figure D.8: BDT input variables for the $\overline{B}^0 \to D^{*0}\pi^+e^-\overline{\nu}_e$ decay mode for the MC15rd sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

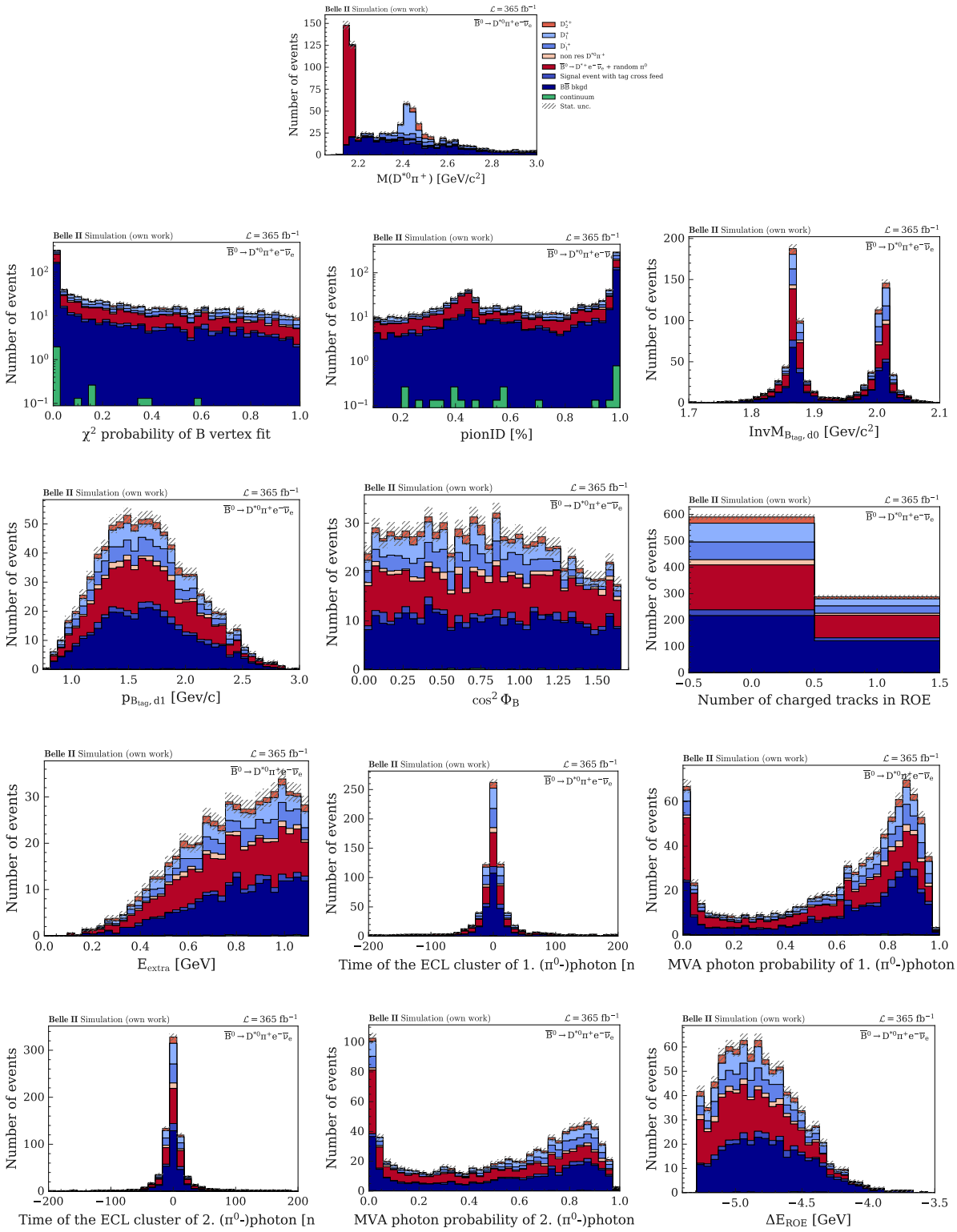


Figure D.9: BDT input variables for the $\overline{B}^0 \to D^{*0}\pi^+ e^- \overline{\nu}_e$ decay mode for the MC15ri sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

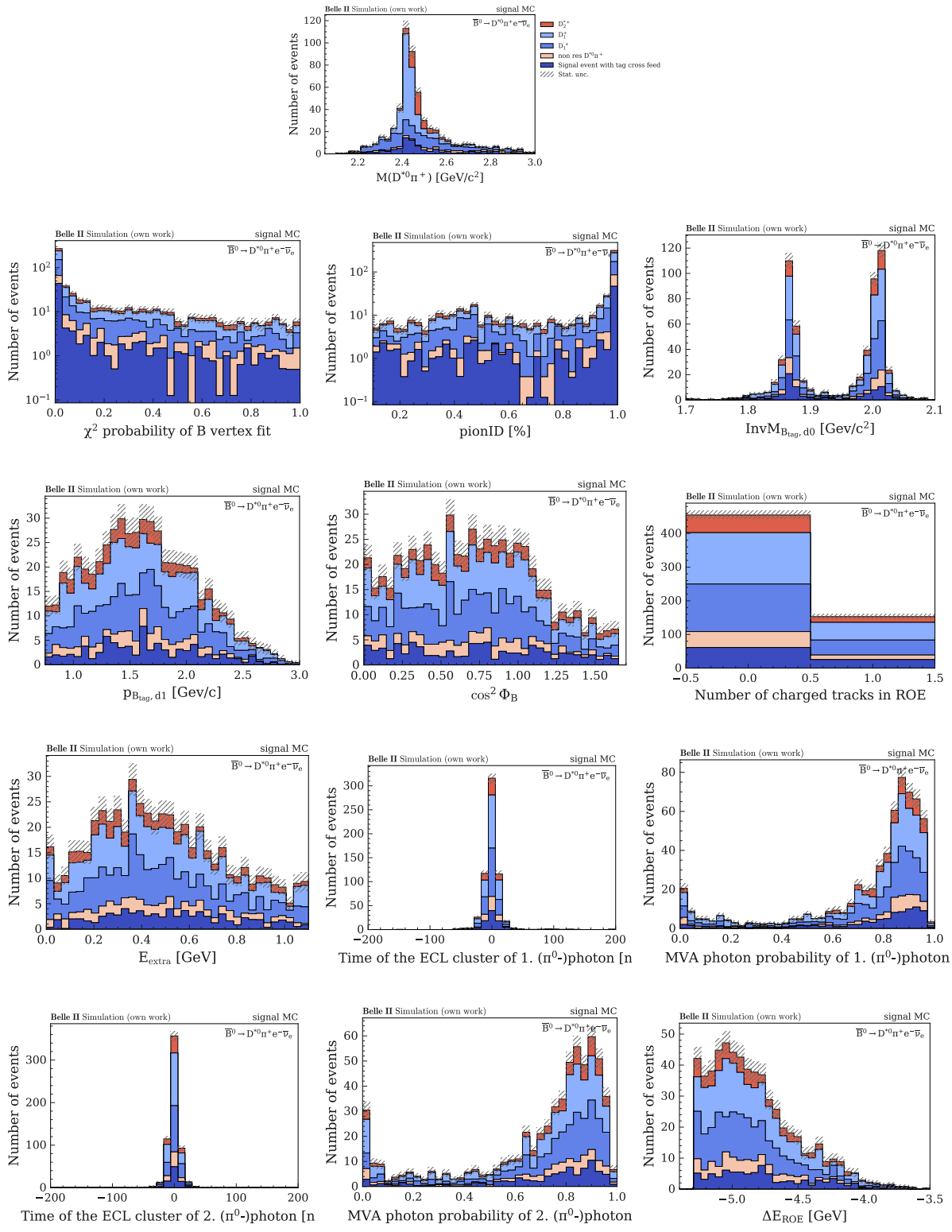


Figure D.10: BDT input variables for the $\overline{B}^0 \to D^{*0}\pi^+ e^- \overline{\nu}_e$ decay mode for the ri signal MC sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

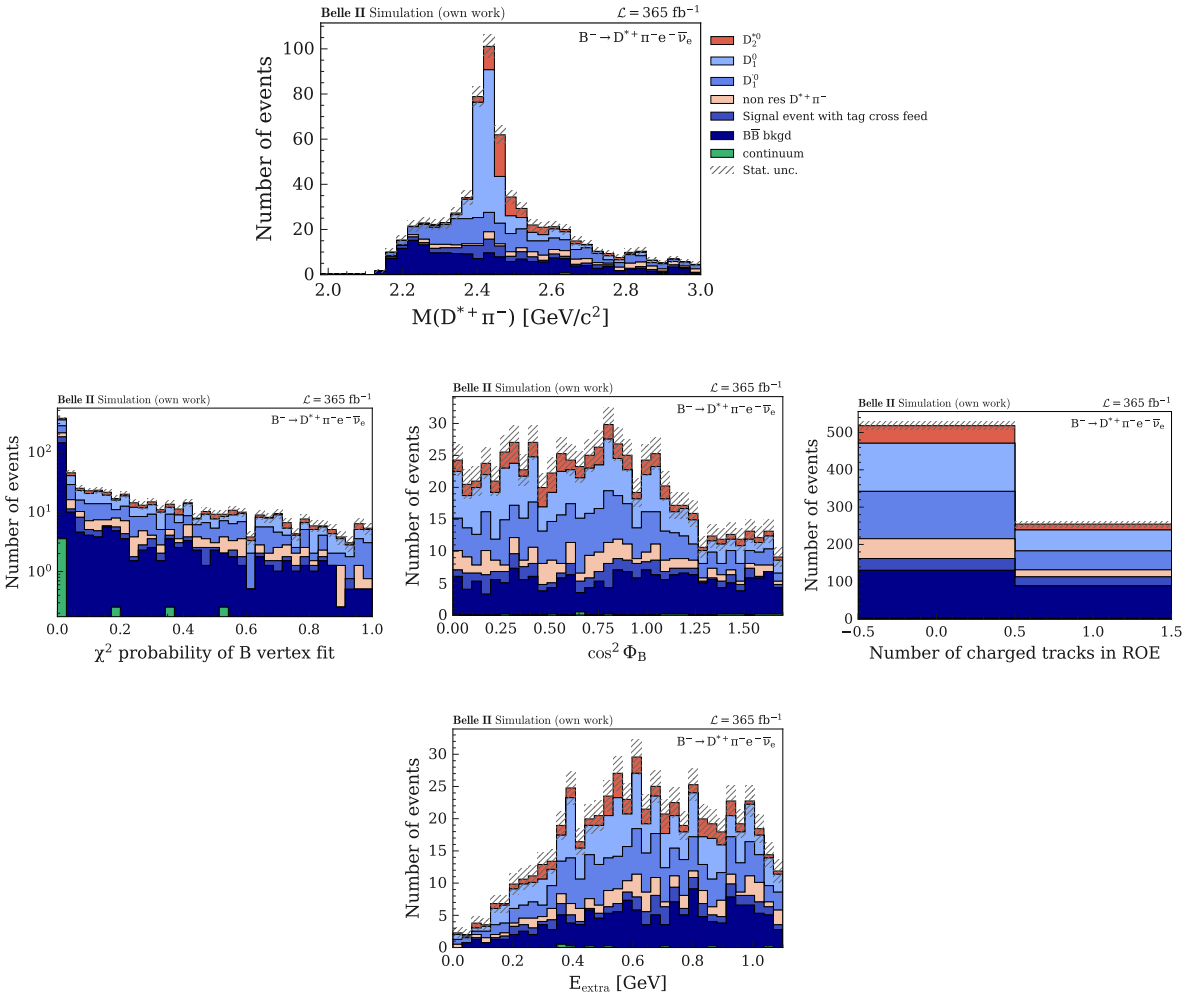


Figure D.11: BDT input variables for the $B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$ decay mode for the MC15rd sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

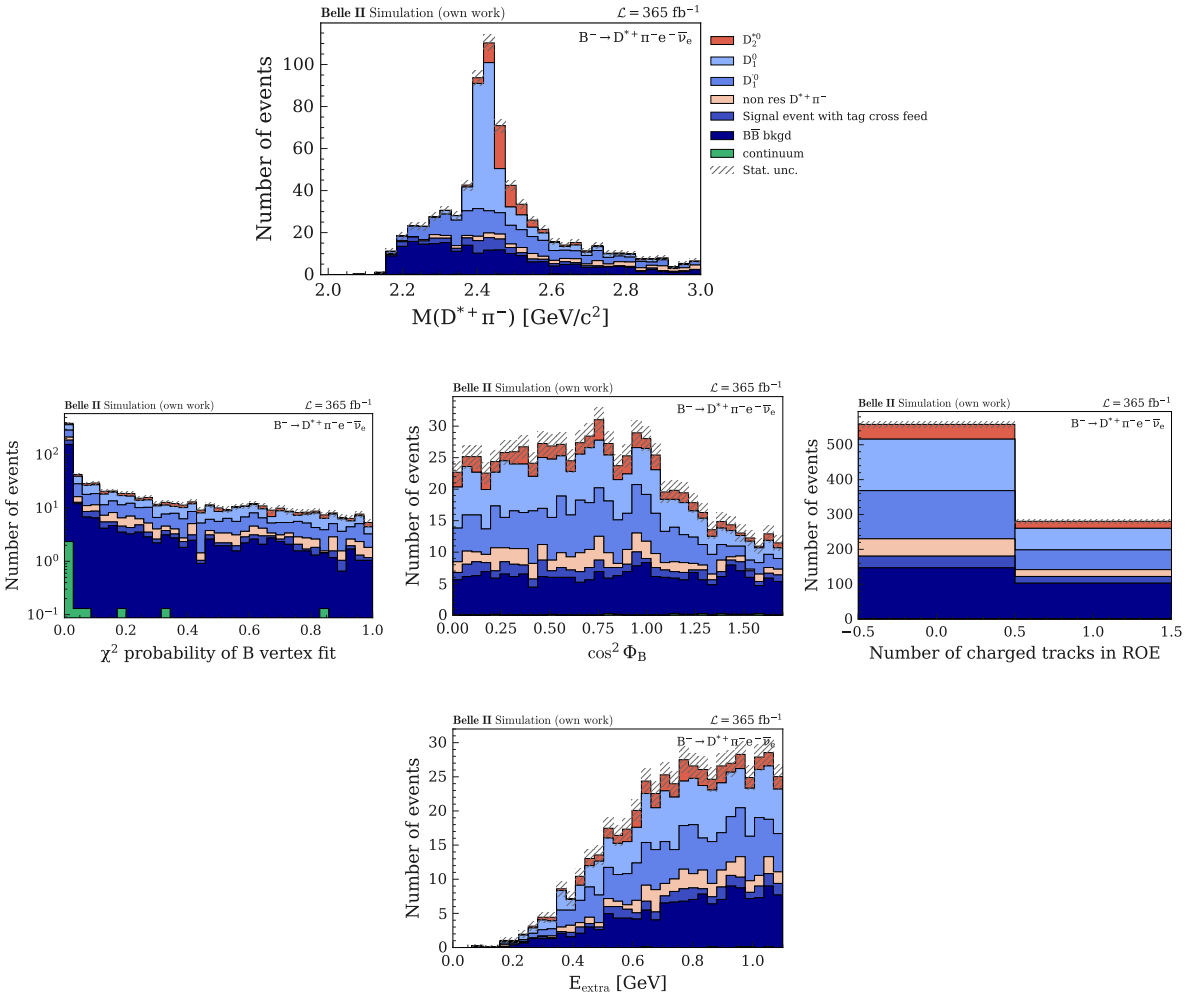


Figure D.12: BDT input variables for the $B^- \to D^{*+}\pi^-e^-\bar{\nu}_e$ decay mode for the MC15ri sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

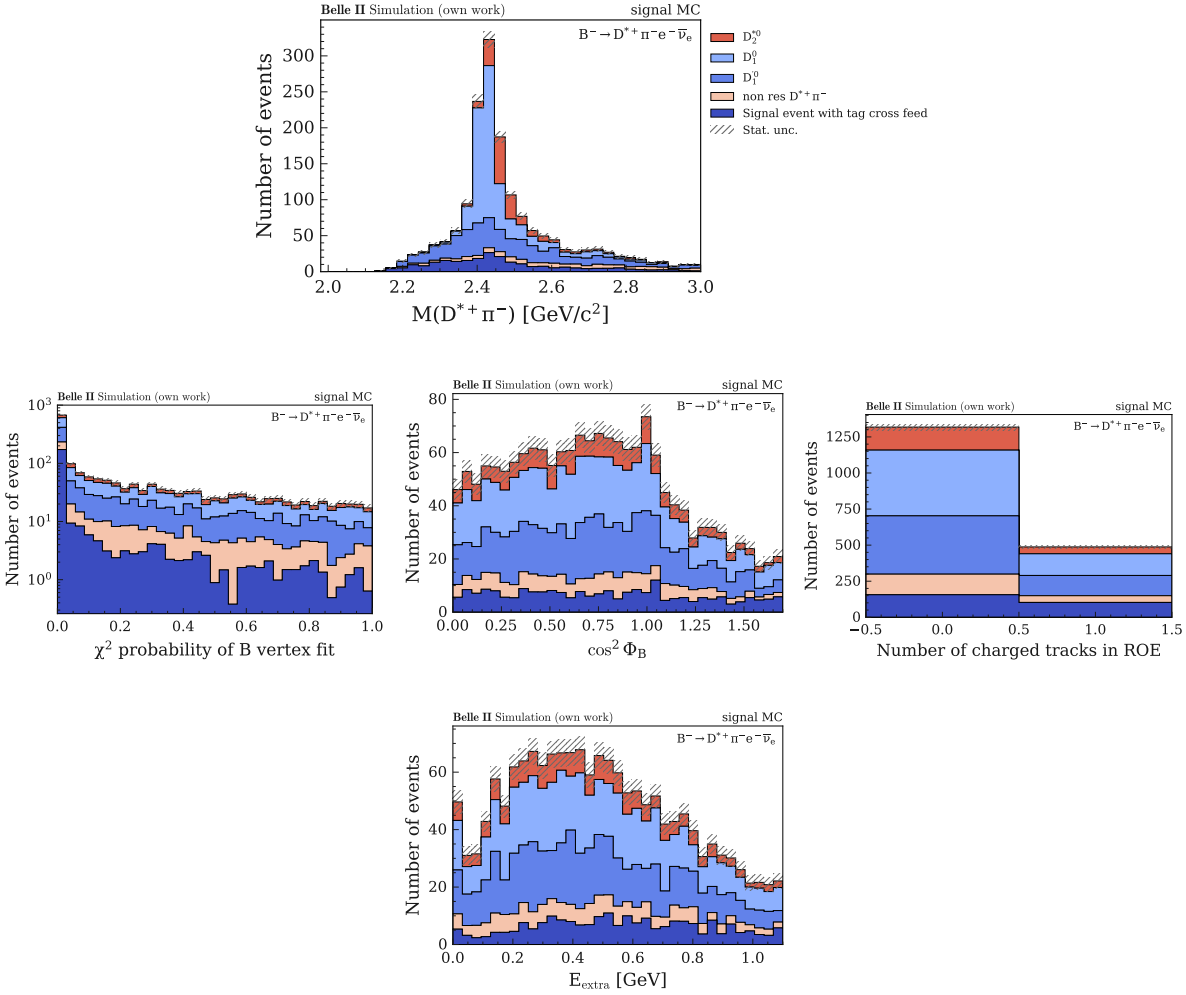


Figure D.13: BDT input variables for the $B^- \to D^{*+}\pi^-e^-\overline{\nu}_e$ decay mode for the ri signal MC sample. Note that the invariant mass spectrum of the $D\pi$ system is not used in the BDT training, but is included to this figure for illustrative purposes.

APPENDIX E

Efficiency distributions for the $\bar{B}^0 \to D^0\pi^+$ mode individually for every resonance

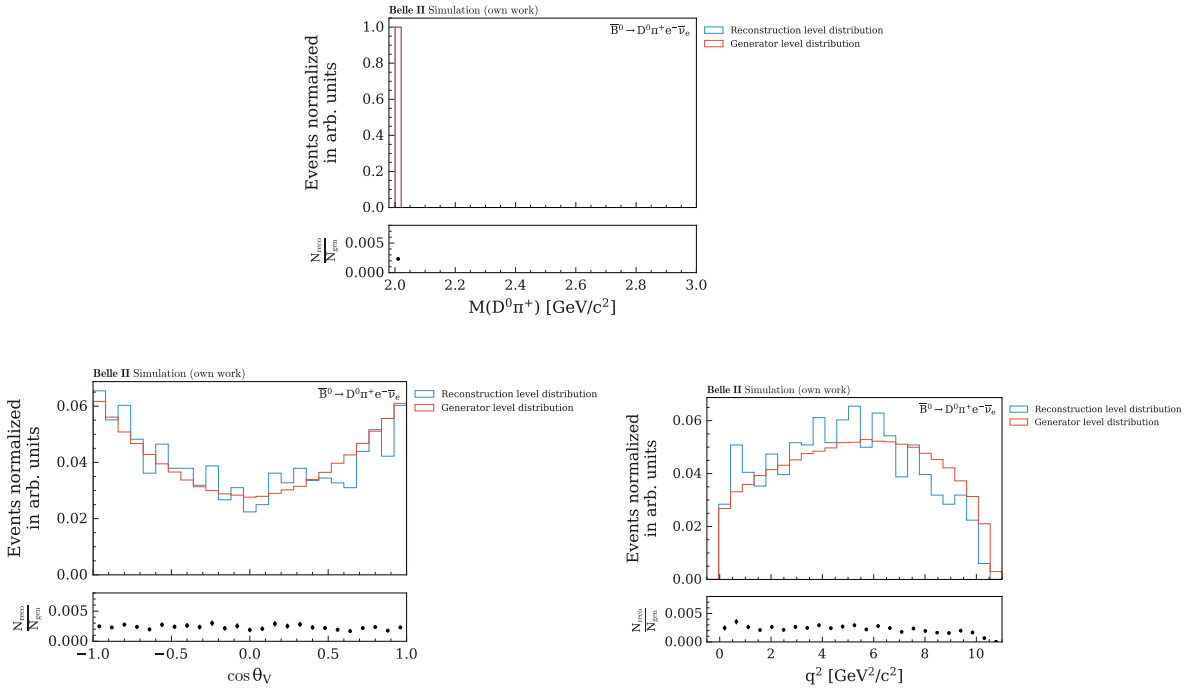


Figure E.1: Efficiency distributions of the parameter of interest for the $\bar{B}^0 \to D^{*+}(\to D^0\pi^+)e^-\bar{\nu}_e$ decay mode.

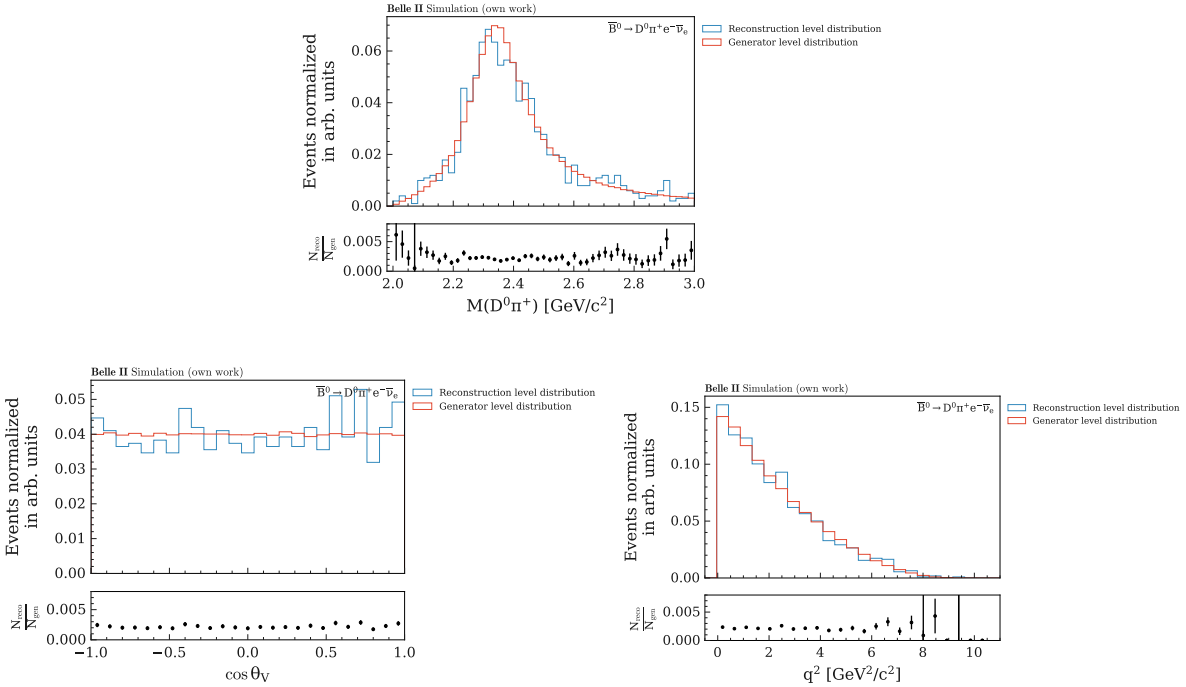


Figure E.2: Efficiency distributions of the parameter of interest for the $\bar{B}^0 \to D_0^{*+}(\to D^0\pi^+)e^-\bar{\nu}_e$ decay mode.

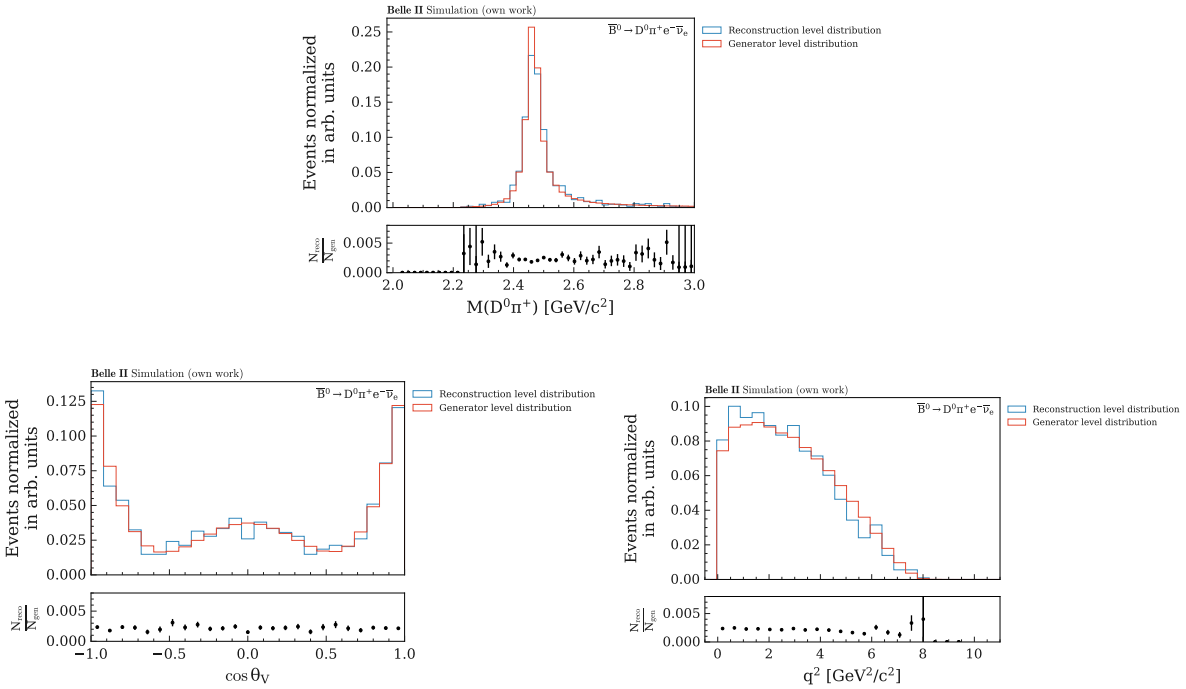


Figure E.3: Efficiency distributions of the parameter of interest for the $\bar{B}^0 \to D_2^{*+}(\to D^0\pi^+)e^-\bar{\nu}_e$ decay mode.

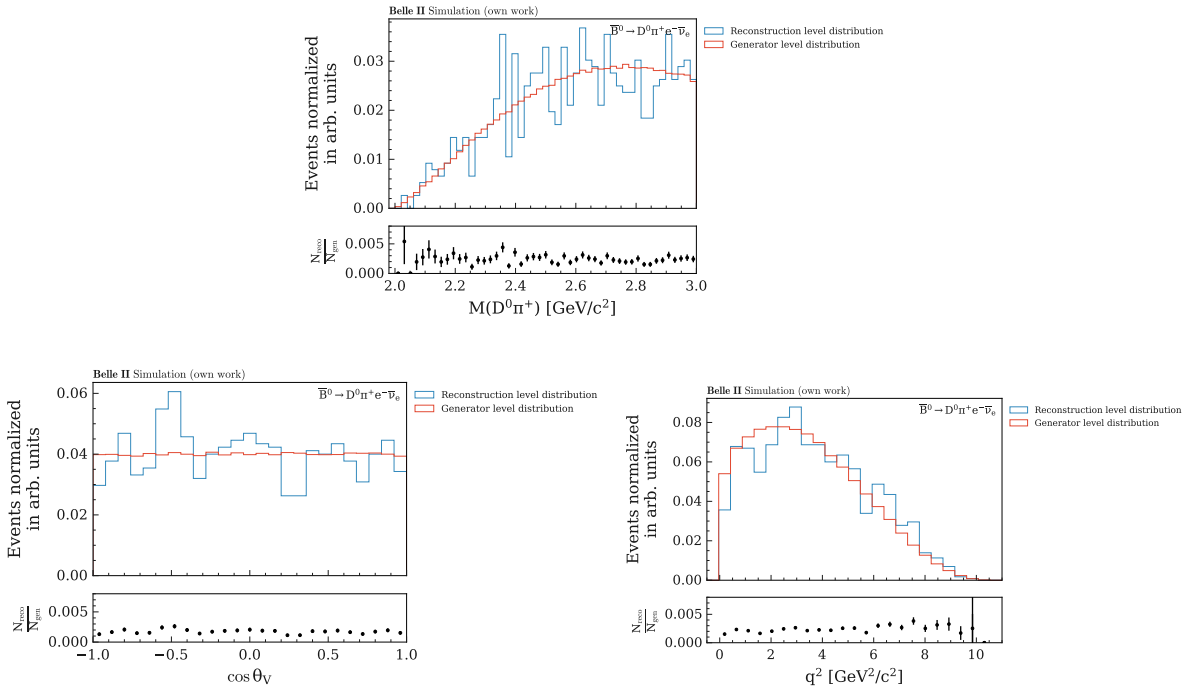


Figure E.4: Efficiency distributions of the parameter of interest for the non resonant $\bar{B}^0 \to D^0\pi^+e^-\bar{\nu}_e$ decay mode.

APPENDIX F

Comparison of this work with the theory model

In the following, we compare the $D\pi$ invariant mass spectrum obtained in this study with the fit results from a model-independent description of $B \to D\pi\ell\nu_\ell$ decays [1], briefly described in Section 2.3.1, where the theoretical model was fitted to hadronic tagged Belle data [24].
This study is based solely on MC simulations, where the D_0^* state is modeled using a single Breit-Wigner resonance. The comparison, shown in Fig. F.1, clearly indicates that this parametrization is insufficient and leads to an inaccurate prediction of the branching fraction for the $(D\pi)_S$-wave contribution.

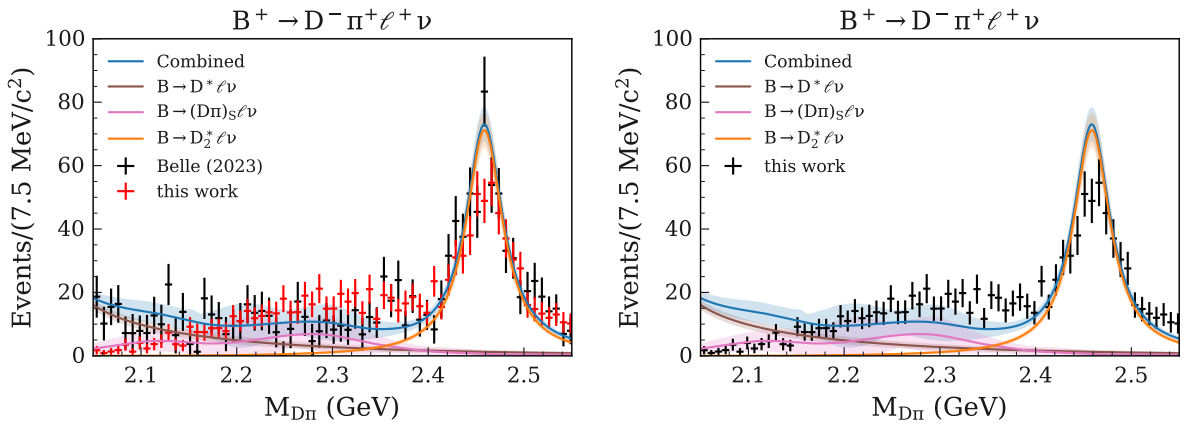


Figure F.1: Comparison of fitted invariant $D^+\pi^-$ mass spectrum from [1] to hadronic tagged Belle data [24] and the distribution of this work. (Left) In black the data points of the Belle measurement and in red the reconstructed signal MC distribution of this work scaled to the luminosity of Belle (711 fb^{-1}). (Right) The Belle data is removed for a better comparison of the fitted model with the results of this work.

Bibliography

[1] E. J. Gustafson, F. Herren, R. S. V. de Water, R. van Tonder and M. L. Wagman, *A model-independent description of $B \to D\pi\ell\nu$ decays*, Phys. Rev. D **110** (2024) L091502, arXiv: 2311.00864 [hep-ph], URL: https://arxiv.org/abs/2311.00864 (cit. on pp. iii, 1, 9–11, 125).

[2] 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** (7 2018) 075011, URL: https://link.aps.org/doi/10.1103/PhysRevD.97.075011 (cit. on pp. 1, 8, 18).

[3] D. J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) 152, BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/S0168900201000894 (cit. on pp. 1, 16).

[4] Particle Data Group and S. N. et al., *Review of Particle Physics*, Physical Review D **110** (2024) 030001 (cit. on pp. 1, 5, 7).

[5] C. A. Manzari and D. J. Robinson, *On-shell recursion and holomorphic HQET for heavy quark hadronic resonances*, (2025), arXiv: 2402.12460 [hep-ph], URL: https://arxiv.org/abs/2402.12460 (cit. on p. 1).

[6] M.-L. Du, F.-K. Guo, C. Hanhart, B. Kubis and U.-G. Meißner, *Where is the lightest charmed scalar meson?*, Phys. Rev. Lett. **126** (2021) 192001, arXiv: 2012.04599 [hep-ph], URL: https://arxiv.org/abs/2012.04599 (cit. on p. 1).

[7] M. Mai, U.-G. Meißner and C. Urbach, *Towards a theory of hadron resonances*, Physics Reports **1001** (2023) 1, ISSN: 0370-1573, URL: http://dx.doi.org/10.1016/j.physrep.2022.11.005 (cit. on p. 1).

[8] N. Isgur and M. B. Wise, *Weak decays of heavy mesons in the static quark approximation and Weak transition form factors between heavy mesons*, Phys. Lett. B **232, 237** (1989, 1990) 113, 527, Phys. Lett. B 232 (1989) 113; Phys. Lett. B 237 (1990) 527 (cit. on p. 3).

[9] J. M. Flynn and N. Isgur, *Heavy-quark symmetry: ideas and applications*, Journal of Physics G: Nuclear and Particle Physics **18** (1992) 1627, ISSN: 1361-6471, URL: http://dx.doi.org/10.1088/0954-3899/18/10/004 (cit. on p. 3).

[10] A. K. Leibovich, Z. Ligeti, I. W. Stewart and M. B. Wise,
Semileptonic B decays to excited charmed mesons, Physical Review D **57** (1998) 308,
ISSN: 1089-4918, URL: http://dx.doi.org/10.1103/PhysRevD.57.308 (cit. on pp. 3, 8).

[11] F. U. Bernlochner, Z. Ligeti and S. Turczyk,
A proposal to solve some puzzles in semileptonic B decays, Physical Review D **85** (2012),
ISSN: 1550-2368, URL: http://dx.doi.org/10.1103/PhysRevD.85.094033
(cit. on pp. 3, 4).

[12] R. Eden, P. Landshoff, D. Olive and J. Polkinghorne, *The Analytic S-Matrix*,
Cambridge University Press, 1966 (cit. on p. 5).

[13] C. Hanhart, *The role of hadron-hadron interactions in hadron spectroscopy*,
Eur. Phys. J. A **49** (2013) 82, arXiv: 1311.6627 [hep-ph] (cit. on p. 5).

[14] G. Breit and E. Wigner, *Capture of Slow Neutrons*, Phys. Rev. **49** (1936) 519 (cit. on p. 5).

[15] Ulf-G. Meißner, *Two-pole structures in QCD: Facts, not fantasy!*, Symmetry **12** (2020) 981,
arXiv: 2005.06909 [hep-ph] (cit. on pp. 5, 6).

[16] M.-L. Du, F.-K. Guo, C. Hanhart, B. Kubis and U.-G. Meißner,
Where is the lightest charmed scalar meson?, Phys. Rev. Lett. **126** (2021) 192001,
arXiv: 2012.04599 [hep-ph] (cit. on pp. 5, 6).

[17] R. Aaij et al., *Amplitude analysis of $B^- \to D^+\pi^-\pi^-$ decays*, Phys. Rev. D **94** (2016) 072001,
arXiv: 1608.01289 [hep-ex] (cit. on pp. 5, 6).

[18] K. M. Watson, *The Effect of Final State Interactions on Reaction Cross Sections*,
Phys. Rev. **88** (1952) 1163 (cit. on p. 7).

[19] F. Metzner, *Preparation of a Measurement of $\mathcal{R}(D^{(*)})$ with Leptonic τ and Hadronic FEI Tag at the Belle Experiment*, PhD thesis: Karlsruhe Institute of Technology (KIT), 2022 (cit. on p. 7).

[20] F. U. Bernlochner, M. Franco Sevilla, D. J. Robinson and G. Wormser,
Semitauonic b-hadron decays: A lepton flavor universality laboratory,
Rev. Mod. Phys. **94** (1 2022) 015003,
URL: https://link.aps.org/doi/10.1103/RevModPhys.94.015003 (cit. on pp. 7, 8).

[21] C. G. Boyd, B. Grinstein and R. F. Lebed,
Constraints on Form Factors for Exclusive Semileptonic Heavy to Light Meson Decays,
Phys. Rev. Lett. **74** (23 1995) 4603,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.4603 (cit. on pp. 8, 18).

[22] S. Faller, T. Feldmann, A. Khodjamirian, T. Mannel and D. van Dyk,
Disentangling the Decay Observables in $B^- \to \pi^+\pi^-\ell^-\bar{\nu}_\ell$, Phys. Rev. D **89** (2014) 014015,
arXiv: 1310.6660 [hep-ph] (cit. on p. 9).

[23] X.-W. Kang, B. Kubis, C. Hanhart and U.-G. Meißner, *B_l4 decays and the extraction of $|V_{ub}|$*,
Phys. Rev. D **89** (2014) 053015, arXiv: 1312.1193 [hep-ph] (cit. on p. 9).

[24] F. Meier et al., *First observation of branching fractions with hadronic tagging at Belle*,
Physical Review D **107** (2023), ISSN: 2470-0029,
URL: http://dx.doi.org/10.1103/PhysRevD.107.092003
(cit. on pp. 11, 43, 54, 57, 87, 125).

[25] A. L. Yaouanc, J.-P. Leroy and P. Roudeau,
Model for nonleptonic and semileptonic decays by $B^0 \to D^{**}$ transitions with $BR(j = 1/2) \ll BR(j = 3/2)$ using the Leibovich–Ligeti–Stewart–Wise scheme,
Phys. Rev. D **105** (2022) 013004, arXiv: 2111.09535 [hep-ph] (cit. on p. 11).

[26] B. Aubert et al., *Measurements of the $B \to D^*$ form factors using the decay $\overline{B}^0 \to D^{*+} e^- \overline{\nu}_e$*,
Phys. Rev. D **74** (9 2006) 092004,
URL: https://link.aps.org/doi/10.1103/PhysRevD.74.092004 (cit. on pp. 12, 59).

[27] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth and N. Braun, *The Belle II Core Software*,
Computing and Software for Big Science **3** (2018) 1,
URL: https://doi.org/10.1007/s41781-018-0017-9 (cit. on p. 13).

[28] Belle II Collaboration, *The experiment*, https://software.belle2.org/release-08-01-03/sphinx/online_book/fundamentals/02-datataking.html, Accessed: 2025-06-14, 2024 (cit. on pp. 13, 15).

[29] T. Abe et al., *Belle II Technical Design Report*, 2010, arXiv: 1011.0352 [physics.ins-det],
URL: https://arxiv.org/abs/1011.0352 (cit. on pp. 13, 16).

[30] E. Kou et al., *The Belle II Physics Book*,
Progress of Theoretical and Experimental Physics **2019** (2019) 123C01, ISSN: 2050-3911,
eprint: https://academic.oup.com/ptep/article-pdf/2019/12/123C01/32693980/ptz106.pdf,
URL: https://doi.org/10.1093/ptep/ptz106 (cit. on pp. 13, 14, 16).

[31] I. H. de la Cruz, *The Belle II experiment: fundamental physics at the flavor frontier*,
Journal of Physics: Conference Series **761** (2016) 012017, ISSN: 1742-6596,
URL: http://dx.doi.org/10.1088/1742-6596/761/1/012017 (cit. on p. 14).

[32] D. Zhou, K. Ohmi, Y. Funakoshi and Y. Ohnishi, *Luminosity performance of SuperKEKB*, 2023,
arXiv: 2306.02692 [physics.acc-ph], URL: https://arxiv.org/abs/2306.02692
(cit. on p. 14).

[33] A. Abashian et al., *The Belle detector*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **479** (2002) 117, Detectors for Asymmetric B-factories,
ISSN: 0168-9002,
URL: https://www.sciencedirect.com/science/article/pii/S0168900201020137
(cit. on p. 14).

[34] Belle II Collaboration, *Belle II Archives*, Accessed: 2025-06-14, 2025,
URL: https://www.belle2.org/archives/ (cit. on p. 15).

[35] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*,
Computer Physics Communications **191** (2015) 159, ISSN: 0010-4655,
URL: https://www.sciencedirect.com/science/article/pii/S0010465515000442
(cit. on p. 16).

[36] S. Jadach, B. Ward and Z. Wąs,
The precision Monte Carlo event generator KK for two-fermion final states in e+e- collisions,
Computer Physics Communications **130** (2000) 260, ISSN: 0010-4655,
URL: https://www.sciencedirect.com/science/article/pii/S0010465500000485
(cit. on p. 16).

[37] E. Barberio, B. van Eijk and Z. Wąs,
Photos - a universal Monte Carlo for QED radiative corrections in decays,
Computer Physics Communications **66** (1991) 115, ISSN: 0010-4655,
URL: https://www.sciencedirect.com/science/article/pii/001046559190012A
(cit. on p. 16).

[38] S. Agostinelli et al., *Geant4 - a simulation toolkit*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002,
URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688
(cit. on p. 17).

[39] B. Brau and D. Greenwald, *Run-dependent Monte Carlo Production at Belle II*,
tech. rep. BELLE2-NOTE-TE-2024-001, Belle II Collaboration, 2024,
URL: https://indico.belle2.org/event/15670/contributions/94920/attachments/35072/51949/BELLE2-NOTE-TE-2024-001.pdf (cit. on p. 17).

[40] Belle II Collaboration, *Belle II Data Production Collection Summary*,
https://xwiki.desy.de/xwiki/bin/view/BI/Belle%20II%20Internal/Data%20Production%20WebHome/Collection%20summary/, Accessed: 2025-06-16, 2025
(cit. on p. 17).

[41] Belle II Collaboration, *Belle II MC main page*, https://xwiki.desy.de/xwiki/bin/view/BI/Belle%20II%20Internal/Data%20Production%20WebHome/MC%20main%20page/,
Internal documentation; accessed 2025-06-16, 2025 (cit. on p. 17).

[42] Belle II Collaboration, *Skimming — basf2 documentation*,
https://software.belle2.org/release-08-01-03/sphinx/online_book/basf2/skim.html, Accessed: 2025-06-16, 2025 (cit. on p. 17).

[43] I. Adachi et al., *Measurement of the integrated luminosity of data samples collected during 2019-2022 by the Belle II experiment**, Chinese Physics C **49** (2025) 013001, ISSN: 2058-6132,
URL: http://dx.doi.org/10.1088/1674-1137/ad806c (cit. on pp. 17, 18).

[44] R. Glattauer et al., *Measurement of the decay $B \to D\ell\nu_\ell$ in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{cb}|$*,
Phys. Rev. D **93** (3 2016) 032006,
URL: https://link.aps.org/doi/10.1103/PhysRevD.93.032006 (cit. on p. 18).

[45] D. Ferlewicz, P. Urquijo and E. Waheed, *Revisiting fits to $B^0 \to D^{*-}\ell^+\nu_\ell$ to measure $|V_{cb}|$ with novel methods and preliminary LQCD data at nonzero recoil*, Phys. Rev. D **103** (7 2021) 073005,
URL: https://link.aps.org/doi/10.1103/PhysRevD.103.073005 (cit. on p. 18).

[46] X. Zhou, S. Du, G. Li and C. Shen, *TopoAna: A generic tool for the event type analysis of inclusive Monte-Carlo samples in high energy physics experiments*, Computer Physics Communications **258** (2021) 107540, ISSN: 0010-4655, URL: https://www.sciencedirect.com/science/article/pii/S0010465520302575 (cit. on pp. 18, 40, 41).

[47] T. Keck et al., *The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment*, Computing and Software for Big Science **3** (2019) 6 (cit. on pp. 19, 20).

[48] F. Bernlochner, L. Cao, C. Lyu, C. Schwanda and B. I. Collaboration, *Determination of $|V_{cb}|$ using $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with Belle II*, Phys. Rev. D **108** (2023) 092013, BELLE2-PUB-PH-2023-024, Belle II Collaboration (cit. on pp. 21, 48).

[49] E. von Törne, *Advanced Laboratory Course physics 601E215: Particle-antiparticle oscillations at BELLE-II*, University of Bonn, Version 1.03, n.d. (Cit. on p. 22).

[50] F. Bernlochner, L. Cao, C. Lyu, C. Schwanda and D. Dorner, *Determination of $|V_{cb}|$ using $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with Belle II*, tech. rep. BELLE2-NOTE-PH-2022-001, Working group: Semileptonic and Missing Energy Decay: Belle II Collaboration, 2022, URL: https://docs.belle2.org/pub_data/documents/196/ (cit. on p. 23).

[51] M. T. Prim et al., "Design and Performance of the Belle II High Level Trigger", *Proceedings of Science (ICHEP2020)*, PoS ICHEP2020 (2021) 769, SISSA, 2021 769 (cit. on p. 23).

[52] Belle II Collaboration, *Belle II Software Online Book: Data Taking – Triggers and Filters*, Accessed: 2025-07-09, Belle II Software Group, 2025, URL: https://software.belle2.org/development/sphinx/online_book/fundamentals/02-datataking.html#triggers-and-filters (cit. on p. 23).

[53] Belle II Collaboration, *Belle II Charged PID Recommendations for Lepton ID – Moriond 2024*, DESY XWiki, Accessed: 2025-07-09, 2024, URL: https://xwiki.desy.de/xwiki/bin/view/BI/Belle%5C%20II%5C%20Internal/Physics%5C%20Performance%5C%20Webhome/Charged%5C%20PID%5C%20performance/Charged%5C%20PID%5C%20Recommendations/Recommendations%5C%20for%5C%20Lepton%5C%20ID%5C%20-%5C%20Moriond2024/ (cit. on p. 23).

[54] Belle II Collaboration, *How do we look for Bremsstrahlung photons?*, Accessed: 2025-07-09, Belle II Software Group, 2025, URL: https://software.belle2.org/release-08-01-03/sphinx/online_book/basf2/various_additions.html?highlight=bremscorrected#how-do-we-look-for-bremsstrahlung-photons (cit. on p. 24).

[55] Belle II Collaboration, *Belle II Charged Hadron PID Recommendations – Release 6 (Moriond 2023, 2024)*, DESY XWiki, Accessed: 2025-07-09, 2023, URL: https://xwiki.desy.de/xwiki/bin/view/BI/Belle%5C%20II%5C%20Internal/Physics%5C%20Performance%5C%20Webhome/Charged%5C%20PID%5C%20performance/Charged%5C%20PID%5C%

20Recommendations/Recommendations%5C%20for%5C%20Hadron%5C%20ID%5C%20-%5C%20release6%5C%20%5C%28Moriond2023%5C%2C%5C%20Moriond2024%5C%29/ (cit. on p. 24).

[56] Belle II Collaboration, *Rest of Event (ROE) — Belle II Software Documentation*, Accessed: 2025-07-10, 2024, URL: https://software.belle2.org/release-08-01-03/sphinx/online_book/basf2/roe.html#roe-masks (cit. on p. 25).

[57] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016 785 (cit. on p. 29).

[58] T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework", *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, 2019 2623 (cit. on p. 29).

[59] R. Tibshirani, *Regression Shrinkage and Selection via the Lasso*, Journal of the Royal Statistical Society: Series B (Methodological) **58** (1996) 267 (cit. on p. 29).

[60] A. E. Hoerl and R. W. Kennard, *Ridge Regression: Biased Estimation for Nonorthogonal Problems*, Technometrics **12** (1970) 55 (cit. on p. 29).

[61] XGBoost Development Team, *Booster.get_score*, XGBoost Python API documentation, https://xgboost.readthedocs.io/en/stable/python/python_api.html#xgboost.Booster.get_score, 2025 (cit. on p. 31).

[62] G. Punzi, *Sensitivity of searches for new signals and its optimization*, 2003, arXiv: physics/0308063 [physics.data-an], URL: https://arxiv.org/abs/physics/0308063 (cit. on p. 37).

[63] H. Dembinski et al., *scikit-hep/iminuit*, Zenodo (2024), URL: https://doi.org/10.5281/zenodo.13923658 (cit. on pp. 37, 73).

[64] A. Frey and N. Rauls, *Measurement of the branching ratio $B \to D^{**}\ell\nu_\ell$ with the hadronic Full Event Interpretation using Belle II data*, Belle II Internal Note BELLE2-NOTE-PH-2023-006, Georg-August-Universität Göttingen and Belle II Collaboration, 2023 (cit. on p. 44).

[65] B. I. Collaboration et al., *Test of lepton flavor universality with measurements of $R(D^+)$ and $R(D^{*+})$ using semileptonic B tagging at the Belle II experiment*, 2025, arXiv: 2504.11220 [hep-ex], URL: https://arxiv.org/abs/2504.11220 (cit. on pp. 45, 46).

[66] Belle II Software, *Semileptonic Full Event Interpretation (FEI) – Documentation (of release 08-01-03)*, https://software.belle2.org/release-08-01-03/sphinx/skim/doc/02-physics.html?highlight=sl%20fei#skim.WGs.fei.feiSL, Accessed July 2025, 2025 (cit. on p. 46).

[67] I. Adachi et al., *Measurement of the branching fractions of $\overline{B} \to D(*)K - K_{(S)}^{(*)0}$ and $\overline{B} \to D(*)D_s^-$ decays at Belle II*, Journal of High Energy Physics **2024** (2024), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP08(2024)206 (cit. on p. 47).

[68] D. Dorner, P. Rados and C. Schwanda,
Relative tracking efficiency study of charged, low momentum pions,
Belle II Internal Note BELLE2-NOTE-PL-2021-001, Internal Document:
Belle II Collaboration, 2021 (cit. on p. 63).

[69] M. Pivk and F. Le Diberder, *: A statistical tool to unfold data distributions*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **555** (2005) 356, ISSN: 0168-9002,
URL: http://dx.doi.org/10.1016/j.nima.2005.08.106 (cit. on p. 69).

[70] H. Dembinski, M. Kenzie, C. Langenbruch and M. Schmelling,
Custom Orthogonal Weight functions (COWs) for event classification,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1040** (2022) 167270, ISSN: 0168-9002,
URL: https://www.sciencedirect.com/science/article/pii/S0168900222006076
(cit. on p. 72).

[71] M. Steiner, *Performance of Univariate Kernel Density Estimation Methods in TensorFlow*,
Bachelor Thesis: University of Zurich, 2021,
URL: https://astroviking.github.io/ba-thesis/ (cit. on p. 72).

[72] J. Eschle, A. P. Navarro, R. S. Coutinho and N. Serra, *zfit: scalable pythonic fitting*,
SoftX (2019), arXiv:1910.13429 [phystat] (cit. on p. 73).

[73] D. W. Scott, "Multivariate Density Estimation and Visualization",
Handbook of Computational Statistics: Concepts and Methods,
ed. by J. E. Gentle, W. K. Härdle and Y. Mori,
Berlin, Heidelberg: Springer Berlin Heidelberg, 2012 549, ISBN: 978-3-642-21551-3,
URL: https://doi.org/10.1007/978-3-642-21551-3_19 (cit. on p. 73).

[74] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, 1st ed., Routledge, 1998,
URL: https://doi.org/10.1201/9781315140919 (cit. on p. 73).

List of Figures

List of Tables