Measurement of the photon energy spectrum in inclusive radiative B meson decays using the hadronic-tagging method

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von
Henrikas Svidras
aus Vilnius (Litauen)

Hamburg
2023

Abstract

This doctoral thesis presents an inclusive analysis of the photon energy spectrum in radiative B meson decays ($B \to X_s \gamma$) that involve strange hadrons and a photon. The data set consists of $\Upsilon(4S)$ mesons that primarily decay into pairs of B mesons. The partner B meson is fully reconstructed in a large number of hadronic decay channels using multivariate classifiers. This full reconstruction allows for the inference of the momentum and energy of the signal B meson through a technique referred to as hadronic-tagging. The data set analysed corresponds to an integrated luminosity of $189 \, \text{fb}^{-1}$ of electron-positron collisions at the $\Upsilon(4S)$ resonance energy provided by the SuperKEKB accelerator. This is the first measurement of the $B \to X_s \gamma$ decay with the Belle II experiment. Only the high energy photon from the signal B meson decay is reconstructed in order to achieve an unbiased inclusive sample of final states involving strange hadrons. The hadronic-tagging provides direct access to the signal B meson rest frame, leading to the photon energy spectrum without the additional kinematic smearing observed in the $\Upsilon(4S)$ rest frame. The partial branching fractions of the $B \to X_s \gamma$ decay are measured in eight intervals of photon energy in the signal B meson rest frame between $1.8 - 2.7$ GeV. The obtained signal yield for this photon energy range is 343 ± 122 events. The integrated branching fraction in this region is found to be $(3.54 \pm 0.78 \, (\text{stat.}) \pm 0.83 \, (\text{sys.})) \cdot 10^{-4}$. Additionally, the first and second moments of the photon energy spectrum are calculated for several photon energy thresholds. For photon energies above 1.8 GeV, they are determined to be $(2.284 \pm 0.065 \, (\text{stat.}) \pm 0.071 \, (\text{syst.}))$ GeV and $(0.0502 \pm 0.0157 \, (\text{stat.}) \pm 0.0176 \, (\text{syst.}))$ GeV^2, respectively. The results show excellent agreement with the Standard Model predictions and serve as a proof-of-concept for future hadronic-tagged radiative measurements at Belle II.

Zusammenfassung

Diese Doktorarbeit beschreibt eine inklusive Analyse des Photonen-Energiespektrums bei radiativen B-Meson-Zerfällen ($B \rightarrow X_s\gamma$), an denen Strange-Hadronen und ein Photon beteiligt sind. Der Datensatz besteht aus $\Upsilon(4S)$-Mesonen, die hauptsächlich in Paare von B-Mesonen zerfallen. Das Partner-B-Meson wird in einer großen Anzahl von hadronischen Zerfallskanälen mithilfe von multivariaten Klassifikatoren vollständig rekonstruiert. Diese vollständige Rekonstruktion ermöglicht den Rückschluss auf den Impuls und die Energie des Signal-B-Mesons durch eine Technik, die als hadronisches Tagging bezeichnet wird. Der analysierte Datensatz entspricht einer integrierten Luminosität von 189 fb^{-1} von Elektron-Positron-Kollisionen bei der $\Upsilon(4S)$-Resonanzenergie des SuperKEKB-Beschleunigers. Dies ist die erste Messung des $B \rightarrow X_s\gamma$-Zerfalls am Belle II Experiment. Nur das hochenergetische Photon aus dem Zerfall des Signal-B-Mesons wird rekonstruiert, um ein unverzerrtes, umfassendes Sample von Endzuständen zu erhalten, an denen Strange-Hadronen beteiligt sind. Die kinematische Einschränkung durch das hadronische Tagging ermöglicht einen direkten Zugang zum Signal-B-Meson-Ruhesystem, was zu einem Photonenenergiespektrum führt, ohne die zusätzliche kinematische Verschmierung, die im $\Upsilon(4S)$-Ruhesystem beobachtet wird. Das partielle Verzweigungsverhältnis des $B \rightarrow X_s\gamma$-Zerfalls wird im Ruhesystem des Signal-B-Mesons in acht Photonenergieintervallen innerhalb von $1.8 - 2.7$ GeV gemessen. Die erhaltene Anzahl von Signalkandidaten für diesen Photonenenergiebereich beträgt 343 ± 122. Das integrierte Verzweigungsverhältnis in diesem Bereich beträgt $(3.54 \pm 0.78\,(\text{stat.}) \pm 0.83\,(\text{sys.})) \cdot 10^{-4}$. Zusätzlich werden die ersten und zweiten Momente des Photonenenergiespektrums für mehrere Schwellenwerte der Photonenenergie berechnet. Für Photonenenergien oberhalb von 1.8 GeV werden sie zu $(2.284 \pm 0.065\,(\text{stat.}) \pm 0.071\,(\text{syst.}))$ GeV bzw. $(0.0502 \pm 0.0157\,(\text{stat.}) \pm 0.0176\,(\text{syst.}))$ GeV^2 bestimmt. Die Ergebnisse zeigen eine hervorragende Übereinstimmung mit den Vorhersagen des Standardmodells und dienen als Proof-of-Concept für zukünftige Messungen von radiativen $B \rightarrow X_s\gamma$ Zerfällen mit hadronischem Tagging am Belle II Experiment.

No book can ever be finished. While working on it we learn just enough to find it immature the moment we turn away from it. – KARL POPPER

Contents

Chapter 1

Introduction

The Standard Model is a theory describing the fundamental particles and their interactions that constitute the universe. The predictive power and precision of the Standard Model are arguably unmatched by any other theory in science. The Higgs boson was proposed already in 1964 [1], more than 50 years before its discovery [2], [3]. The top quark was discovered [4], [5] two decades after the existence of the third generation of fermions was inferred [6].

However, some experiments do show so far unexplained tensions with the predictions of the Standard Model. A notable example is the measurement of the anomalous magnetic moment of the muon [7]. Although the Standard Model can provide an astounding 8 significant digit agreement with the experiment, there is a disagreement at higher precision. Similarly, in the case of the anomalous magnetic moment of the electron, there is a tension between different experimental approaches [8]–[10]. The Standard Model has also not been able to provide a clear candidate for the excess matter that is inferred in the Universe (dark matter). Furthermore, the observed matter-antimatter asymmetry is larger than what is currently predicted in the Standard Model. The mathematical framework of the Standard Model, while incredibly powerful, appears to be arbitrary, with numerous free parameters. For example, the existence of exactly 3 generations of fermions with a mass hierarchy is not theoretically well-grounded but seems too organised to be accidental. The strong force conserves the charge-parity symmetry, although it is not required by the Standard Model framework (strong CP problem). There are many other experimental and theoretical inconsistencies (see Ref. [11] for a review) that cannot be adequately addressed by the Standard Model

These theoretical and experimental challenges motivate the search for discoveries that could lead to new developments in one of the most successful theories of science. While numerous extensions to the Standard Model have been proposed by the particle physics theory community, there is no clear evidence that any of them is the correct description of our Universe. At the moment, it seems that hints about a clear direction can only come from more precise measurements that will test the Standard Model particle properties, interactions and process rates.

This thesis contributes such a measurement in the radiative decay of the beauty quark, where it transitions to a strange quark and a photon. The measurement will be performed by studying data of B mesons decaying into a meson system originating from an s quark, X_s, and an energetic photon γ. The data that is analysed was collected in 2019-2020 by the Belle II

detector, which is located in Tsukuba, Japan at the KEK laboratory. Belle II collects data of electron-positron collisions, provided by the SuperKEKB collider. The colliding electron and positron beams create $\Upsilon(4S)$ resonances which decay into pairs of B mesons. The measurement presented here uses a technique called hadronic-tagging, which also reconstructs the partnering B meson, yielding a purer final measured data sample.

The standard notation of the radiative B meson decay processes is adopted to label it as $B \to X_s\gamma$. This decay channel is sensitive to particles not included within the Standard Model. Furthermore, the parameters of the photon energy spectrum are important inputs to other precision measurements of the Standard Model.

The previously mentioned points are discussed in detail in this thesis. It is therefore split up into 8 chapters, which introduce the theoretical background, experimental machinery and status of $B \to X_s\gamma$, data analysis tools and discusses the measurement itself. The current Chapter and Chapter 8 are dedicated to introduction and summary, respectively. The remaining Chapters are as follows:

- Chapter 2 introduces the main concepts of the Standard Model that are necessary to understand the basis of $B \to X_s\gamma$ decays, their importance and possible future developments. It is not a thorough mathematical overview but only serves as a synthesis of important concepts and conclusions. A wide array of literature is referenced for a deeper dive into the theoretical aspects of the analysis.

- Chapter 3 introduces the experimental status of radiative B decays which is intended as an introductory chapter for experimental $B \to X_s\gamma$ analysis. The information provided in the Chapter is only a summary, and not part of the original work in this thesis.

- Chapter 4 presents the experimental machinery (the SuperKEKB accelerator and the Belle II detector) and software used to collect and process the data for this analysis.

- Chapter 5 introduces important statistical and data science concepts related to parameter estimation and multivariate analysis which are employed in the analysis of the Belle II data.

- Chapter 6 presents the strategy and the analysis of the Belle II data leading to the measurement of $B \to X_s\gamma$ decay properties with the hadronic tagging technique. It presents a full overview of the evaluation of statistical and systematic uncertainties related to the analysis.

- Chapter 7 provides a comprehensive discussion of the results of Chapter 6 and compares them with experimental world averages. It also overviews the prospects and the impact of the result presented in this thesis to the theoretical status of the Standard Model.

Additional information, which supports or further explains the material presented in this thesis, is provided in the Appendices A to K.

Chapter 2

Theoretical overview

In this Chapter, the theoretical overview of the Standard Model, flavour physics and $B \rightarrow X_s \gamma$ is presented. Section 2.1 defines the main particle physics terminology that will be used throughout the thesis by a synthesis of information of well-established concepts about the Standard Model which can be found detailed, e.g., in Refs. [12]–[14]. Section 2.2 introduces the concept of flavour physics in the Standard Model and some main challenges related to it. Section 2.3 introduces the main points of interest in the study of B meson decays concerning flavour physics. Finally, the rest of the Chapter in Sections 2.4 to 2.6 provides the reader with an introduction of the theoretical foundation and the status of inclusive radiative decays.

2.1 The Standard Model of Particle Physics

The Standard Model (SM) is a quantum field theory based on a local gauge invariance given by the symmetry group $SU(3) \times SU(2) \times U(1)$. In quantum field theory, all particles are described as excitations of fields, $\psi(x)$, where $x = \{\vec{x}, t\}$ is a space-time coordinate. The Lagrangian, $\mathcal{L} = \mathcal{L}(\psi, \partial_\mu \psi)$, describes the dynamics originating from the excitations of a field ψ, where $\partial_\mu = \{\frac{\partial}{\partial \vec{x}}, \frac{\partial}{\partial t}\}$. In particular, it encodes the interactions and the free propagation of particles which are visually represented using Feynman diagrams.

The theoretical framework of the SM describes the electroweak and strong force interactions between elementary particles that constitute the world as we know it. The strong interaction between particles of the SM is described by quantum chromodynamics (QCD), whereas electromagnetic and weak interaction are unified by the electroweak theory. The elementary particles included in the SM have half-integer spin (fermions) or integer spin (bosons).

The spin-1 bosons are the mediators of the electromagnetic (photon), weak ($W^\pm$ and Z) and strong interactions (gluons). The spin-0 Higgs boson couples to all massive gauge bosons of the SM via the Higgs mechanism [1] and fermions via the Yukawa couplings [15]. The fermions are further split into two additional groups: quarks and leptons, which, respectively, can and cannot interact through the strong force. With the possible exception of neutrinos, all fermion fields have a left- and right-handed component, as the weak interaction only interacts with left-handed particles or right-handed antiparticles. Their spin, charge, mass and notation are summarised in Figure 2.1.
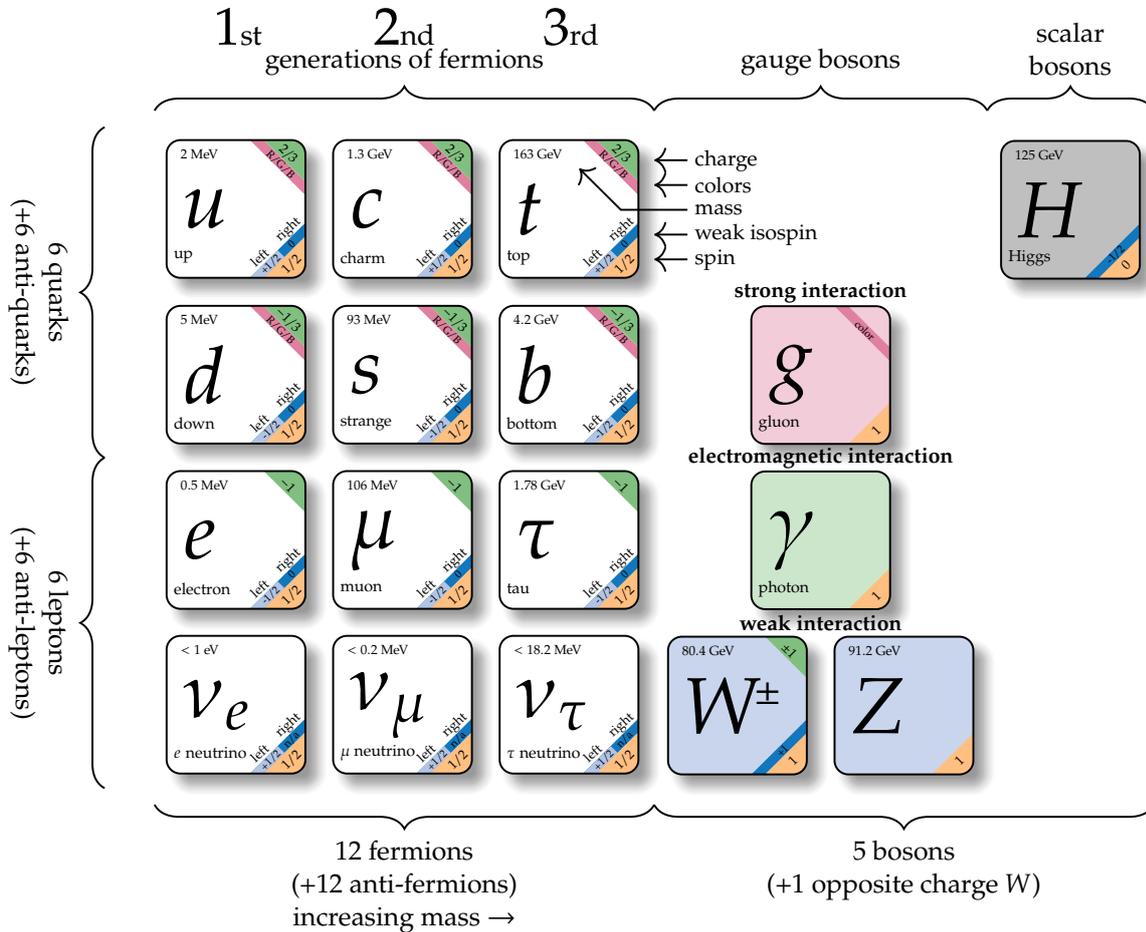
Figure 2.1: All the particles of the Standard Model of Particle Physics. Their mass, charge, spin, weak isospin and colours are listed. The gauge boson interactions are colour coded. The antiparticles can be obtained by inverting the signs of the weak isospin, charge, and swapping the left/right-handed chirality values. All the masses are approximate even if more precise measurements are available. They are based on the current knowledge of particle physics, as summarised in Ref. [16]. Note that quark masses are not well-defined as they cannot be observed in Nature independently. This provides their masses in the $\overline{\text{MS}}$ renormalisation scheme. See Ref. [16] for more information on the values. This is a modified version of the diagram of Ref. [17].

Quarks and leptons are organised in three *generations*, where each generation contains a charged lepton, a neutral lepton, an up-type quark, and a down-type quark. Each species of fermion has a distinct *flavour*, resulting in six quark and six lepton flavours in the SM. The strong and electromagnetic interactions do not change the flavour of particles, however, this can be achieved via the weak interaction. On the other hand, a transition between two generations is not possible for charged leptons [1].

The generations only differ by their masses, where a mass hierarchy is observed:

$$m(\text{lepton neutrino}) \ll m(\text{charged lepton}) \lessgtr m(\text{down type quark}) < m(\text{up type quark}).$$

For quarks and charged leptons, only the first generation is stable and constitutes the absolute majority of the stable matter observed in the universe. On the other hand, all three generations of neutrinos appear to be stable. While leptons exist as free particles in nature, quarks are not observed free, due to the confinement effect, but rather as combinations of two or more quarks, collectively known as *hadrons*. The combinations of a quark and an anti-quark pair are called *mesons*, while combinations of three (anti-)quarks are called *baryons*. Some hadrons that will be commonly mentioned in this thesis are shown in Table 2.1.

Table 2.1: Examples of common mesons and baryons, with a focus on those mentioned in the thesis often. Their mass and lifetime values are approximate, even if more precise measurements are available [16].

Particle	Quark content	Mass [MeV/c^2]	Mean lifetime [ps]
Mesons			
π^+	$u\bar{d}$	140	2.6×10^4
π^0	$1/\sqrt{2}(u\bar{u} - d\bar{d})$	135	8.4×10^{-5}
η	$1/\sqrt{2}(u\bar{u} + d\bar{d} - 2s\bar{s})$	548	5×10^{-7}
K^+	$u\bar{s}$	494	1.2×10^4
K^0	$d\bar{s}$	498	-
K^0_S	$1/\sqrt{2}(d\bar{s} + s\bar{d})$	-	89
K^0_L	$1/\sqrt{2}(d\bar{s} - s\bar{d})$	-	5.1×10^4
D^+	$c\bar{d}$	1870	1.0
D^0	$c\bar{u}$	1865	0.4
B^+	$u\bar{b}$	5279	1.6
B^0	$d\bar{b}$	5279	1.5
$\Upsilon(4S)$	$b\bar{b}$	10579	1.2×10^{-8}
Baryons			
p	$u\,u\,d$	938	stable
n	$u\,d\,d$	940	800×10^{12}

The decay of particles is described by their lifetime τ (the time after which $1/e$ of the original particles remain, on average), which is the inverse of the total decay rate:

$$\Gamma = \frac{1}{\tau}, \tag{2.1}$$

---

[1] It is possible for neutrinos only if the SM is extended to include massive neutrinos.

which can be calculated by Fermi's golden rule. The decays of unstable particles are probabilistic processes and, particularly for heavy particles, occur via multiple different modes, often called *decay channels*. The total decay rate is the sum of individual decay rates:

$$\Gamma = \sum_j \Gamma_j, \tag{2.2}$$

where Γ_j corresponds to the decay rate of a specific decay channel, j. The *branching ratio* characterises Γ_j as a relative fraction of the total decay rate:

$$\mathcal{B}(j) = \frac{\Gamma_j}{\Gamma}. \tag{2.3}$$

2.2 Flavour physics

In the SM, each particle has a partner particle, called an *antiparticle*. For a particle ψ, an antiparticle can be obtained by performing the charge conjugation transformation:

$$C \ket{\psi} \rightarrow \ket{\bar{\psi}}, \tag{2.4}$$

which inverts the sign of all charges, lepton and baryon numbers corresponding to ψ. As a result, neutral bosons are their own antiparticle. All quarks (u, d, c, s, b, t) have a corresponding oppositely-charged anti-quark ($\bar{u}, \bar{d}, \bar{c}, \bar{s}, \bar{b}, \bar{t}$). Likewise, each charged lepton (e^-, μ^-, τ^-) has a positively charged particle (e^+, μ^+, τ^+). The antiparticles of neutrinos (ν_e, ν_μ, ν_τ) are denoted as ($\bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau$) [2].

Consider a parity transformation, which transforms a left-handed component into a right-handed component:

$$\mathcal{P} \ket{\psi(\vec{x}, t)} = \ket{\psi(-\vec{x}, t)}. \tag{2.5}$$

It has been experimentally observed that the weak interactions violate the $\mathcal{P}$ symmetry maximally (but not electromagnetic or strong).

After the violation of charge (C) and parity ($\mathcal{P}$) symmetries by the weak interaction became clear [19], it was initially assumed that their combination, $C\mathcal{P}$ symmetry, is conserved. In 1964, it was observed that the weak interaction can violate the combined charge-parity symmetry [20]. To incorporate $C\mathcal{P}$ violation into the SM, M. Kobayashi and T. Maskawa [21] built upon the predecessor work by N. Cabibbo [22], introducing the quark-mixing model based on the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The model introduces a difference between the d, s, b quark states when they propagate freely (mass eigenstate) and their d', s', b' states when they interact via the weak interaction (weak eigenstate). Specifically, they show that weak eigenstates of the quarks are linear combinations of the quark mass eigenstates:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.6}$$

---

[2]Whether neutrinos are their own antiparticle is an unanswered question in neutrino physics. The Majorana versus Dirac fermion problem is discussed widely in literature, e.g. Ref. [18].

where V_{ij} are the elements of a 3×3 unitary matrix known as the CKM matrix. Cabibbo initially introduced a real and unitary 2×2 matrix in order to universally explain the weak interactions of the then-known two generations of leptons, up, down and strange quarks. Kobayashi's and Maskawa's work extended this and showed that at least three generations of quarks are required (the existence of b and t was not known at the time) to incorporate the observed $\mathcal{CP}$ violation. In such a way, the $\mathcal{CP}$ violation is accounted for through complex diagonal elements ($V_{ij} \neq V_{ji}^*$).

The unitarity constraint implies:

$$\sum_i V_{ij} V_{ik}^* = \delta_{jk},$$
$$\sum_j V_{ij} V_{kj}^* = \delta_{ik}, \tag{2.7}$$

for any generation k. As a 3×3 unitary matrix, it can be fully described by 9 parameters, however, five quark fields can absorb a complex phase related to the matrix, which leaves only a single global complex phase. Therefore, the CKM matrix can be expressed in terms of the three quark-mixing angles and a $\mathcal{CP}$-violating complex phase. The standard representation of the CKM matrix is:

$$
\begin{aligned}
V_{\text{CKM}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{-13}e^{-i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} c_{12}c_{13} & s_{12}s_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}.
\end{aligned} \tag{2.8}
$$

Here $s_{ij} = \sin \theta_{ij}$ and $c_{ij} = \cos \theta_{ij}$, with three mixing angles θ_{ij}. The global phase responsible for $\mathcal{CP}$ in the flavour-changing processes is denoted by δ. These parameters are unknown in the SM and have to be measured experimentally. As of writing this thesis, the experimental magnitudes (omitting uncertainties) of the CKM matrix are [16]:

$$(|V_{\text{CKM}}|) \approx \begin{pmatrix} 0.974 & 0.224 & 0.004 \\ 0.221 & 0.975 & 0.042 \\ 0.009 & 0.042 & 1.014 \end{pmatrix}. \tag{2.9}$$

The diagonal values of the CKM matrix are the largest, which represents that flavour changes between the same generations are favoured. On the other hand, the off-diagonal elements are much smaller, indicating that transitions between different generations are highly suppressed.

Precise measurements of the CKM matrix elements have been performed to test the validity of the SM. One way to summarise the experimental status of flavour physics is through the so-called unitarity triangle. There are six vanishing unitarity conditions imposed through Equation (2.7), and they can be represented as triangles in the complex plane. The most common choice is taking

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0, \tag{2.10}$$

Figure 2.2: The constraints on the unitarity triangle in the $\bar{\rho}, \bar{\eta}$ plane. Shaded areas represent 95% confidence limits from different measurements constraining various angles of the triangle. Credit to Ref. [16].

and dividing it by a precisely known value of $V_{cd}V_{cb}^*$, which yields:

$$1 + \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = 0. \tag{2.11}$$

The vertices of this unitarity triangle are at $(0,0)$ and $(1,0)$ in the plane of $(\bar{\rho}, \bar{\eta})$, defined as $(\bar{\rho} + i\bar{\eta}) = V_{ud}V_{ub}^*/V_{cd}V_{cb}^*$, where the other two sides are given by $|V_{ud}V_{ub}^*/V_{cd}V_{cb}^*|$ and $|V_{td}V_{tb}^*/V_{cd}V_{cb}^*|$. Many independent measurements evaluating the CKM matrix elements can be combined to test the consistency of the SM with the experimental observations. As can be seen in Figure 2.2, experimental results show incredible consistency with the SM prediction.

2.3   The case for B physics

The B meson is the lightest hadron that contains a b quark. The decay of the unstable b quark always contains a flavour-changing process via weak interaction. Therefore, the study of their decays opens the possibility of measuring a wide range of parameters of the SM, including the CKM matrix, $\mathcal{CP}$ violation measurements, lepton universality measurements etc. Furthermore, significant inconsistencies from the theoretical expectations may lead to breakthroughs in the theoretical understanding of Nature. Some of the questions that are pursued through the study of B mesons (B physics) are [23]:

- What are additional sources of $\mathcal{CP}$ violation?

- Can there be additional Higgs bosons in Nature?

- How does dark matter fit in the Standard Model?

- Can the left-right symmetry of Nature, broken in weak interactions, be restored?

- Can charged lepton flavour-changing processes exist in Nature?

To exploit the properties of B meson decays, a special type of experiments, known as *B factories*, are designed to study B mesons. They will be introduced in Chapter 4.

2.4 The decay rate of $B \rightarrow X_{s/d}\gamma$

In the SM, $b \rightarrow s$ transitions are mediated by so-called flavour-changing neutral currents. These processes cannot occur directly and only happen through loops mediated by virtual particles. One of the decay channels used to study these transitions is the so-called rare radiative $b \rightarrow s\gamma$ and $b \rightarrow d\gamma$ decays. Due to the confinement effect and the hadronisation of the quarks, in reality, they manifest as $B \rightarrow X_s\gamma$ or $B \rightarrow X_d\gamma$, where X_s and X_d denote any meson state originating from the s or d quark hadronisation. The leading SM diagrams for $B \rightarrow X_{s/d}\gamma$ (used to collectively identify X_s and X_d states) processes are shown in Figure 2.3a.
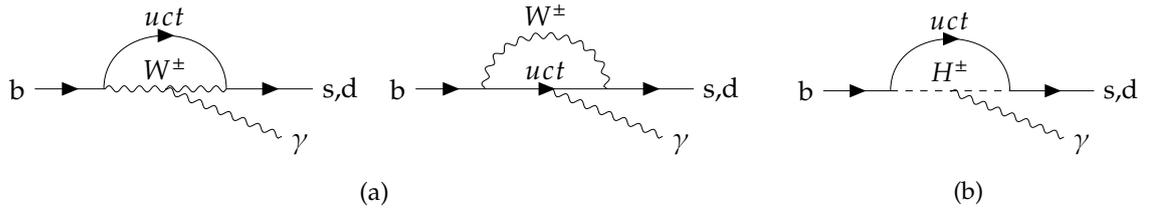
Figure 2.3: The Feynman diagrams for radiative $b \rightarrow s(d)$ transitions. Figure 2.3a shows the leading-order SM diagrams, where the $b \rightarrow s(d)$ transition occurs via electroweak loops. One possible beyond-SM scenario where this transition is mediated by a charged Higgs boson particle is shown in Figure 2.3b.

Since $B \rightarrow X_{s/d}\gamma$ decays proceed via $b \rightarrow s(d)\gamma$ transitions, they are sensitive probes for beyond-Standard-Model (BSM) particles. The leading contributions in the SM only happen via one-loop diagrams, as seen in Figure 2.3. The electroweak loop can get contributions from u, c, t quarks but they are suppressed as followed by the Glashow-Iliopoulos-Maiani mechanism [24]. Therefore, the loop is dominated by the much heavier top quark [25]. The masses of possible BSM weakly-interacting particles that can appear in the loops may be as high as $O(100 \text{ TeV})$ [26] which makes studying these decays appealing.

The decay rate of $B \rightarrow X_{s/d}\gamma$ involves both weak and strong interactions. The typical energy scale of these interactions is much lower than the electroweak scale: $\sim O(m_W)$. This motivates approximating the interactions mediated by heavy Z and $W^\pm$ bosons by an effective point-like vertex. An effective Lagrangian describing the $b \rightarrow s(d)\gamma$ can be written as [27], [28]:

$$\mathcal{L}_{\text{eff}} = \frac{4G_F}{\sqrt{2}} V_{tq}^* V_{tb} \left[\sum_{i=1}^{8} C_i(\mu) O_i(\mu) + \frac{V_{uq}^* V_{ub}}{V_{tq}^* V_{tb}} \sum_{i=1}^{2} C_i(\mu)(O_i(\mu) - O_i^u(\mu)) \right], \tag{2.12}$$

where G_F is the Fermi constant, $q \in \{s, d\}$, and V_{ij} are the appropriate CKM matrix elements. The factors C_i, known as *Wilson coefficients*, encode the high-energy weak contributions and can be calculated perturbatively. The operators O_i describe the effective point-like interactions vertices in the effective field theory. The renormalisation scale, μ, needs to be chosen close to the typical energies of the studied process. For the calculations of the total decay rate of radiative B meson decays it is conventionally set to be of the order of the b quark mass: $\mu \sim m_b$.

The exact expressions of operators O_i, relevant for $b \rightarrow s\gamma$ transitions, are given in Appendix A, while the sketches of effective processes that they represent are given in Figure 2.4. Coefficients C_{3-6} have been calculated and shown to be small, therefore the most important

$$\mathcal{L}_{\text{eff}} \propto C_7 \times \left[\begin{array}{c} \text{\Large Y} \\ \gamma \end{array} \right] + C_8 \times \left[\begin{array}{c} \text{\Large Y} \\ g \end{array} \right] + \sum_{i}^{1,\dots,6} C_i \times \left[\begin{array}{c} + \end{array} \right]$$

Figure 2.4:   Schematic representation of the SM effective $\mathcal{L}$ governing to $B \rightarrow X_s\gamma$. The effective interactions are normalised by the Wilson coefficients, C_i, and described by operators O_i (see Appendix A) whose sketches are shown here. They correspond to the effective Lagrangian in Equation (2.12).

contributions arise from $O_{1,2,7,8}$ [26], [29]. Furthermore, the ratio $V_{uq}^* V_{ub}/V_{tq}^* V_{tb}$ is small for the case of $q = s$ [30], and the terms including O_i^u are relevant only as higher-order corrections of the total decay rate [28]. The latter point does not hold for the $b \rightarrow d$ case, where the $V_{uq}^* V_{ub}/V_{tq}^* V_{tb}$ multiplied term is not numerically small and contributes already at the leading-order. However, for the rest of this Section, $b \rightarrow s\gamma$ is the main focus unless explicitly stated otherwise.

The total decay rate of the inclusive $B \rightarrow X_s\gamma$ is modelled as the rate of the parton decay, taking advantage of the quark-hadron duality and *local operator product expansion* [12], [31], [32]. In particular, additional non-perturbative components need to be considered for the SM calculation of the total decay rate [28]:

$$\Gamma(B \rightarrow X_s\gamma) = \Gamma(b \rightarrow s\gamma) + \Delta\Gamma_{\text{non-p.}}, \tag{2.13}$$

$\Gamma(b \rightarrow s\gamma)$ is the perturbatively calculable rate of b quarks decaying into charmless partons, and $\Delta\Gamma_{\text{non-p.}}$ is the non-perturbative contribution arising outside of local operator product expansion when accounting for the fact that the b quark is not stationary inside the bound B meson state. As long as an appropriately low photon energy threshold (in the decaying B meson rest frame) is chosen, the non-perturbative effects in Equation (2.13) can be considered smaller as they 'average out' over the spectrum. Due to contributions from $c\bar{c}$ resonances at low-E_γ, the threshold is conventionally chosen at $E_\gamma > 1.6$ GeV [33]. Until recently, a 5% uncertainty was associated with this assumption [34]. However, recent developments in understanding the non-perturbative effects [35] led to an improved treatment of these uncertainties [26]. One notable example of non-perturbative effects that are retained in total branching fraction calculations is the so-called resolved photon contributions. These contributions are a result of the photon coupling directly to partons instead of the effective weak interaction

vertex. The origin of the non-perturbative contribution will be further discussed with the differential decay rate in Section 2.5. It will also be seen that the threshold for E_γ is motivated also from the experimental side, as a large background process contamination is present in the low-E_γ region.

Using operator product expansion, one can calculate the matrix element $|\langle s\gamma | \mathcal{L}_{\text{eff}} | b \rangle|^2$ and integrate it from a chosen energy threshold, $E_\gamma > E_0$. Then the perturbative decay rate can be written as [26]:

$$\Gamma(b \to s\gamma) = \frac{G_F^2 m_b^5 \alpha_{\text{em}}}{32\pi^4} |V_{ts}^* V_{tb}|^2 \sum_{i,j=1}^{8} C_i^{\text{eff}}(\mu_b) C_j^{\text{eff}}(\mu_b) \hat{G}_{ij}(E_0, \mu_b), \tag{2.14}$$

where m_b is the b quark pole mass, and α_{em} is the fine-structure constant. The $C_i^{\text{eff}}(\mu_b)$ coefficients represent effective Wilson coefficients [36] that are independent of the renormalisation scheme and defined through linear combinations of the Wilson coefficients of Equation (2.12). Finally, the functions $\hat{G}_{ij}$ encapsulate the terms describing the interference between the operators $O_{i,j}$, which arise in the squared matrix element. As already mentioned before, the dominant terms arise from a handful of operators O_i, and the combined efforts to calculate them have brought $B \to X_s\gamma$ theoretical estimates to the next-to-next-to-leading-order precision. The dominant functions $\hat{G}_{77}$ [37], $\hat{G}_{78}$ [38], and $\hat{G}_{(1,2)7}$ [26], [39] have been evaluated. Contributions from $\hat{G}_{(1,2,8)8}$ [40], [41] have also been calculated. As the calculations have already reached next-to-next-to-leading-order precision, even contributions involving O_i^u are accounted for [42]. The mixing of $O_{1-6} \to O_8$ has been also described [43]. Some of the next-to-next-to-leading-order corrections depend on the mass of the charm quark. However, as the evaluation of such corrections at the physical mass of the charm quark is complicated, the mass is interpolated between $m_c = 0$ and $m_c \gg m_b$ [44].

Equation (2.14) contains a 5th power dependence on the ill-defined pole mass of the b quark. Furthermore, uncertainties arising from the CKM matrix element determination also directly enter the calculation. To minimise the uncertainties related to these values, the calculation of $b \to s\gamma$ decay rate is usually normalised to the semi-leptonic decay rate $b \to u\ell\nu$. An alternative choice is the experimentally more attainable $b \to c\ell\nu$, however, using the charmless decay rate allows separating the m_c determination problem from the problem of calculating higher-order corrections. In doing so, the decay rate ratio is expressed as [45]:

$$\frac{\Gamma(b \to s\gamma)_{E_\gamma > E_0}}{|V_{cb}/V_{ub}|^2 \Gamma(b \to u\ell\bar{\nu})} = \left| \frac{V_{ts}^* V_{tb}}{V_{cb}} \right|^2 \frac{6\alpha_{\text{em}}}{\pi} P(E_0). \tag{2.15}$$

Here, $P(E_0)$ denotes the perturbatively calculable contribution. Replacing $b \to B$, as seen with Equation (2.13), required the introduction of a non-perturbative correction, $N(E_0)$:

$$\frac{\Gamma(B \to X_s\gamma)_{E_\gamma > E_0}}{\Gamma(B \to X_c\ell\nu_\ell)} = \frac{\mathcal{B}(B \to X_s\gamma)_{E_\gamma > E_0}}{\mathcal{B}(B \to X_c\ell\nu_\ell)} = \left| \frac{V_{ts}^* V_{tb}}{V_{cb}} \right|^2 \frac{6\alpha_{\text{em}}}{\pi C} [P(E_0) + N(E_0)]. \tag{2.16}$$

The additional semileptonic phase-space factor:

$$C = \left| \frac{V_{ub}}{V_{cb}} \right|^2 \frac{\Gamma(B \to X_c\ell\bar{\nu})}{\Gamma(B \to X_u\ell\bar{\nu})}, \tag{2.17}$$

accounts for the choice to use the $B \rightarrow X_c \ell \bar{\nu}$ as a normalisation channel for the total decay rate. It is determined using the experimental value of the branching ratio of $B \rightarrow X_c \ell \bar{\nu}$, which is known to a high precision [16], [46]. This choice is preferable compared to the $|V_{ub}|$-suppressed and more model-dependant $B \rightarrow X_u \ell \bar{\nu}$.

At the leading-order, where non-perturbative effects are disregarded, effects from $C_{2,7,8}$ are the most important and this can be expressed in a compact form [36]:

$$\frac{\mathcal{B}(B \rightarrow X_s \gamma)}{\mathcal{B}(B \rightarrow X_c \ell \nu_\ell)} = \left| \frac{V_{ts}^* V_{tb}}{V_{cb}} \right|^2 \frac{6\alpha_{\text{em}}}{\pi C} |C_7^{(0)\text{eff}}(\mu)|^2. \tag{2.18}$$

Here $C_7^{(0)\text{eff}}(\mu)$ is the effective leading-order Wilson coefficient which was briefly introduced before. At leading-order, it takes the form:

$$C_7^{(0)\text{eff}}(\mu) = \eta^{16/23} C_7^{(0)}(\mu) + \frac{8}{3} \left(\eta^{14/23} - \eta^{16/23} \right) C_8^{(0)}(\mu) + C_2^{(0)}(\mu) \sum_i^8 h_i \eta^{a_i}, \tag{2.19}$$

where the SM coefficients $C_{2,7,8}^{(0)}$ are expressed in terms of the ratio of $W^\pm$ and top-quark masses, $x = (m_t/m_W)^2$:

$$C_2^{(0)}(\mu) = 1;$$
$$C_7^{(0)}(\mu) = F_7^{(1)}(x) \equiv \frac{3x^3 - 2x^2}{4(x-1)^4} \ln x + \frac{-8x^3 - 5x^2 + 7x}{24(x-1)^3}; \tag{2.20}$$
$$C_8^{(0)}(\mu) = F_8^{(1)}(x) \equiv \frac{-3x^2}{4(x-1)^4} \ln x + \frac{-x^3 + 5x^2 + 2x}{8(x-1)^3}.$$

The coefficient η is expressed as $\eta = \alpha_s(m_W)/\alpha_s(\mu)$, whereas h_i and a_i are the eigenvalues of a scheme-independent matrix $\gamma^{(0)\text{eff}}$, which is defined in Appendix A of Ref. [36]. The matrix $\gamma^{(0)\text{eff}}$ governs the leading-order strong interaction corrections to $b \rightarrow s\gamma$, and its elements appear in the renormalisation group equation of the effective Wilson coefficients.

Considering the effective interactions in Figure 2.4, it may seem peculiar that $O_{2,8}$ are important at the leading-order. However, because $B \rightarrow X_s \gamma$ receives Glashow-Iliopoulos-Maiani suppression terms ($\sim m_q^2/m_W^2$), two-loop contributions involving a gluon exchange ($\sim \ln m_t^2/m_b^2$) are important [47]. In the absence of QCD, one has $\eta = 1$ and $B \rightarrow X_s \gamma$ decays are governed solely by the photonic dipole exchange. Consider evaluating Equation (2.19) explicitly for $m_t = 170$ GeV/c^2, $\mu = 5$ GeV/c^2, $\alpha_s(m_Z) = 0.118$ [32]:

$$C_7^{(0)\text{eff}}(\mu_b) \approx 0.695 \cdot C_7^{(0)}(\mu) + 0.085 \cdot C_8^{(0)}(\mu) - 0.158 \cdot C_2^{(0)}(\mu)$$
$$\approx 0.695 \cdot (-0.193) + 0.085 \cdot (-0.096) - 0.158 \tag{2.21}$$
$$\approx -0.300.$$

The additive QCD two-loop contributions including terms with C_8 and C_2 enhance the decay rate significantly. On the other hand, multiplicative QCD correction in the first term suppresses the decay rate.

Beyond the leading-order approximation, one has to take into account the full form of Equation (2.16). A sketch of the next-to-leading-order can be found in Ref. [45], whereas the most up-to-date total decay rate calculation is described in Ref. [26], and references therein. A detailed description of the higher-order calculation of the total $B \to X_s\gamma$ decay rate is beyond the scope of this thesis, but such calculations can be followed up in the references introduced in this Chapter, such as Ref. [26] that includes the most up-to-date estimation of the total decay rate of $B \to X_s\gamma$ in the SM:

$$\mathcal{B}(B \to X_s\gamma) = (3.40 \pm 0.17) \times 10^{-4}. \tag{2.22}$$

For $B \to X_d\gamma$, as evident from Equation (2.12), the result is suppressed by a factor $|V_{td}/V_{ts}|^2 \approx 4.2\%$ with respect to the $B \to X_s\gamma$ result. Evaluating the decay with the exchanged CKM factors and other previously discussed caveats [28]:

$$\mathcal{B}(B \to X_d\gamma) = (0.173 \pm^{0.12}_{0.22}) \times 10^{-4}. \tag{2.23}$$

For $B \to X_s\gamma$, the uncertainties in Equation (2.22) arise from unevaluated higher-order effects, the necessity to perform an interpolation in m_c, and a parametric uncertainty that also encodes the non-perturbative effects. The first two amount to 3% each, whereas the last is considered at 2.5%. This amounts to a total uncertainty of approximately 5%. For significant accuracy improvements in the future, higher-order calculations will not be sufficient. It is necessary to remove the dependence on m_c interpolation and improve the treatment of the parametric uncertainties (non-perturbative effects) to go below the $\sqrt{3^2 + 2.5^2} \approx 3.9\%$ uncertainty.

2.5 The photon energy spectrum of $B \to X_{s/d}\gamma$

It was already briefly discussed (when introducing Equation (2.13)) that theoretical and experimental evaluations of the total $B \to X_s\gamma$ decay rate employ a photon energy threshold. The non-perturbative effects that occur at different energy scales compared to the effective $b \to s\gamma$ transition can be partially integrated out in a treatment of the total decay rate. However, the choice of a lower threshold spoils this approximation and introduces non-perturbative corrections. These effects directly manifest as differences in the shape of the photon energy spectrum.

At the leading-order, $b \to s(d)\gamma$ is a two-body decay, which means that the differential decay rate peaks near the half of b quark mass, $m_b/2 \sim E_\gamma$. It is also clear that photon energies larger than half of the B meson mass, $E_\gamma > m_B/2$ are not allowed. Higher-order effects, such as gluonstrahlung and the Fermi motion of the b quark within the B meson, smear the distribution. The necessity to account for all of these effects occurring at different energy scales motivates the use of soft-collinear effective theory to describe the $B \to X_s\gamma$ photon energy spectrum [48], [49]. It allows factorising different contributions to the decay rate, $d\Gamma$, into terms originating from effective hard-interaction vertices, collinear particles and soft particles:

$$d\Gamma \propto \mathcal{H} \times \mathcal{J} \otimes \mathcal{S}. \tag{2.24}$$

$\mathcal{H}, \mathcal{J}, \mathcal{S}$ represent the hard, jet and hadronic-soft functions, respectively, with $\otimes$ symbolising convolution between the two terms. It is further assumed that the $\mathcal{S}$ can be factorised into a partonic soft function, $\mathcal{S}_{\text{partonic}}$, and a non-perturbative *shape function*, $\mathcal{F}$:

$$\mathcal{S} = \mathcal{S}_{\text{partonic}} \otimes \mathcal{F}, \tag{2.25}$$

which means that the application of soft-collinear effective theory is fully capable to separate perturbative and non-perturbative contributions in the differential decay rate of $B \to X_s \gamma$. This is shown graphically in Figure 2.5, which sketches out the leading-order perturbative and non-perturbative effects in describing the $B \to X_s \gamma$ spectrum.

Due to the larger $B \to X_s \gamma$ decay rate and an overall smaller background process rate, experimental values have the highest precision around the peak region. However, from the theory side, the peak part is mostly governed by non-perturbative effects, encoded within the shape function [49]. The shape function encodes the b quark residual momentum distribution within the B meson. Therefore, a reliable theoretical description of the shape function is critical to make sensible experimental and theoretical comparisons of the photon energy spectrum.
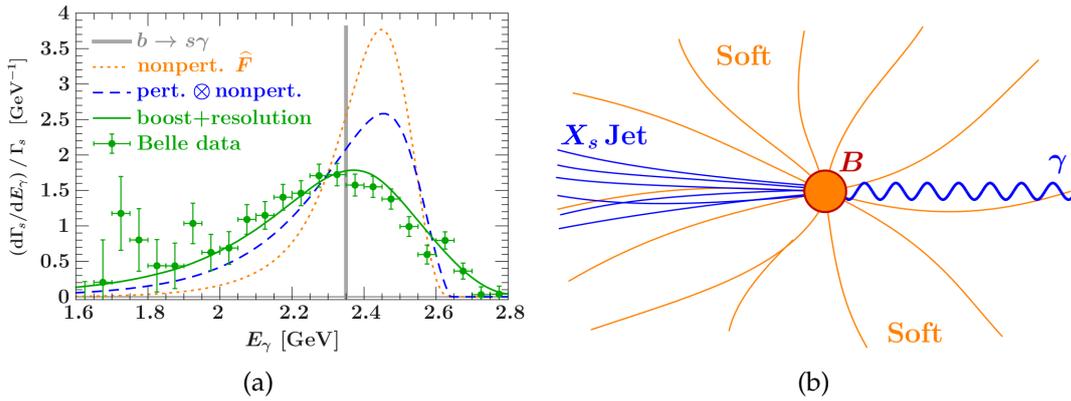
Figure 2.5: The schematic representation of the theoretical $B \to X_s \gamma$ spectrum components and the comparison of them in data. Note that the Figures are presented only for illustration purposes and do not represent a highly-accurate depiction. In (a), the leading-order δ function spectrum, the non-perturbative shape function effects, and the convolution of these effects with the perturbatively calculable ones are represented. It also includes the experimental effects, such as finite detector resolution. In (b), the origin of perturbative and non-perturbative effects in a $B \to X_s \gamma$ decay are illustrated. Credit to Dr. Frank Tackmann for the Figures. The data points in (a) correspond to the Belle measurement in [50].

An additional motivation to study the shape function is the fact that it is a universal property of the B meson at the leading-order in $1/m_b$ [51], [52]. This allows extracting the functional form by a precise experimental determination of the $B \to X_s \gamma$ spectrum and using it to improve the precision of other measurements. For example, the measurement of $|V_{ub}|$ uses $B \to X_u \ell \bar{\nu}_\ell$ decays, however, it suffers orders-of-magnitude larger backgrounds due to the presence of $B \to X_c \ell \bar{\nu}_\ell$ in most phase-space regions. In the regions where the $b \to c$ is kinematically-forbidden, the theoretical predictions for $B \to X_u \ell \bar{\nu}_\ell$ are dependent on the

non-perturbative shape functions. The extracted precise inputs from $B \to X_s\gamma$ could then be used to predict the $B \to X_u\ell\bar{\nu}_\ell$ spectra. This is an important relationship which could lead to a model-independent evaluation of the $|V_{ub}|$ element [51].

Let's consider Equations (2.24) and (2.25) in a bit more detail. Following the treatment of Ref. [49], the differential decay rate takes the form:

$$\frac{d\Gamma}{dE_\gamma} = 2\frac{G_F^2\alpha_{\text{em}}m_b^5}{32\pi^4}|V_{ts}V_{tb}^*|^2\mathcal{H}(x,\mu) \times \int dk \mathcal{P}(m_b, x-k, \mu_i)\mathcal{F}(k) + O\left(\frac{\Lambda_{\text{QCD}}}{m_b}\right), \qquad (2.26)$$

where $x = m_B - 2E_\gamma$, $\mathcal{P}$ is used to denote the perturbatively calculable $\mathcal{J} \otimes \mathcal{S}_{\text{partonic}}$ and the symbolic convolution has now been replaced by an integration over a dummy momentum k. The higher-order corrections introduce additional shape functions but are suppressed by a factor of $1/m_b$ [53]. Generally, $\mathcal{H}$ and $\mathcal{P}$ are calculable in perturbation theory (see Appendix A of [49]). $\mathcal{H}$ is approximately expressed in terms of the effective Wilson coefficient C_7^{eff}. Moreover, at the leading-order in perturbation theory, $\mathcal{P}$ is expressed as a Dirac delta distribution:

$$\mathcal{P}(m_b, k, \mu) = \delta(k) + O(\alpha_s), \qquad (2.27)$$

and therefore integrating out the momentum k:

$$\frac{d\Gamma}{dE_\gamma} \propto |C_7^{\text{eff}}|^2\mathcal{F}(x). \qquad (2.28)$$

This shows, as stated before, that the peak region (close to $m_b/2$) is governed by the shape function. In this case, $|C_7^{\text{eff}}|$ is the normalisation of the spectrum. The shape function can be formally expanded in its moments as [49], [54]:

$$\mathcal{F}(x) = \sum_n \frac{(-1)^n}{n!}A_n\frac{d^n\delta(x)}{dx^n}, \qquad (2.29)$$

where

$$A_n = \int dk k^n\mathcal{F}(k). \qquad (2.30)$$

Equation (2.30) is the expression for the n-th moment of the shape function, where $A_0 = 1$ is fixed because the shape function is chosen to be normalised. Note that if one naively neglects $n > 1$ terms, the integral over the total E_γ of Equation (2.28) would take the form $\Gamma \sim |C_7^{\text{eff}}|^2$, which is consistent with Equation (2.18). Therefore, for total decay rate calculations (such as the ones sketched in Section 2.4) it is enough to know the first moments of the shape function, as the more delicate $\mathcal{F}(k)$ dependence is suppressed at leading-orders.

For example, consider taking only terms up to the first order. In such case, using the definition of an average,

$$\langle E_\gamma \rangle = \frac{\int dE_\gamma E_\gamma \frac{d\Gamma}{dE_\gamma}}{\int dE_\gamma \frac{d\Gamma}{dE_\gamma}}, \qquad (2.31)$$

the shape function is directly related to the average of the photon energy spectrum. Similarly, the variance (and higher-order moments) contribute if more terms from Equation (2.29) are considered. As shown in Refs. [55]–[57] (and others) two important relations emerge:

$$\langle E_\gamma \rangle \sim m_b/2 + O(\Lambda_{\text{QCD}}); \quad \langle E_\gamma{}^2 \rangle - \langle E_\gamma \rangle^2 \sim \lambda_1/12 + O(\Lambda_{\text{QCD}}), \qquad (2.32)$$

with the b quark pole mass, m_b, the b quark kinetic energy parameter, λ_1, and Λ_{QCD} is the chosen strong interaction energy scale.

The form of the shape function is not well-known and is non-perturbative, unlike the Wilson coefficients. Therefore, theoretical and experimental comparisons always lead to modelling-related uncertainties. There are many works which propose different ways to describe it [57]–[62]. For example, Ref. [59] describes it based on a technique called dressed gluon exponentiation. The result of Ref. [58] provides several functional forms of the shape function, based on the leading-order shape function that can be extracted from $B \to X_s \gamma$ decays. An analysis by the SIMBA collaboration [62] describes a model-independent treatment of the shape function based on Equation (2.26). In particular, any chosen shape function is expanded in a complete set of orthonormal basis functions, and can therefore be extracted, together with the normalisation, from a global fit to the available experimental results. The strength of this approach is a consistent method to combine several experimental inputs for the extraction of the shape function and its moments.

One of the most commonly chosen inclusive $B \to X_s \gamma$ models for experimental analyses is known as the Kagan-Neubert model [63]. It provides a next-to-leading-order description of the inclusive $B \to X_s \gamma$ spectrum. In this model, the shape function takes a simple form:

$$\mathcal{F}(x) = N \left(1 - \frac{x}{m_B - m_b} \right)^a \exp\left\{ (1 + a) \frac{x}{m_B - m_b} \right\}, \qquad (2.33)$$

The shape function satisfies the necessary moment constraints through these relations: $A_0 = 1, A_1 = 0, A_2 = -\lambda_1/3$. The parameter a is also related to $A_2 = (m_B - m_b)/(1 + a)$. Therefore, the shape function can be fully described by two free parameters m_b and λ_1. This functional expression is easy to implement and interpret. Furthermore, the Kagan-Neubert model is readily available in the `EvtGen` Monte Carlo generator, commonly used by B-factories [64]. The $B \to X_s \gamma$ inclusive decay is implemented as the `BTOXSGAMMA` model within the generator. These points made it the conventional choice for an experimental description of the inclusive E_γ spectrum in many past measurements. The spectra, generated with varied values of λ_1 and m_b, using the kinetic scheme, are shown in Figure 2.6, otherwise using the default `BTOXSGAMMA` setup.

The previous argument may also be used inversely: an accurate experimental measurement of the $B \to X_s \gamma$ spectrum can be used to precisely determine the parameters m_b, λ_1 and higher moments of the shape function. One such example is the aforementioned SIMBA collaboration result [62]. The evaluated values of m_b and λ_1 by the SIMBA collaboration from fitting available $B \to X_s \gamma$ experimental results are:

$$m_b^{1S} = 4.750 \pm 0.043 \text{ GeV}/c^2; \quad \lambda_1^{\text{inv}} = -0.210 \pm 0.083 \text{ GeV}^2/c^4, \qquad (2.34)$$

where fitting, theoretical and parametric uncertainties are combined. These results originate directly from experimental data fits and have slightly larger than world average uncertainties. The superscripts $1S$ and 'inv' indicate a renormalisation scheme that is chosen by the authors.
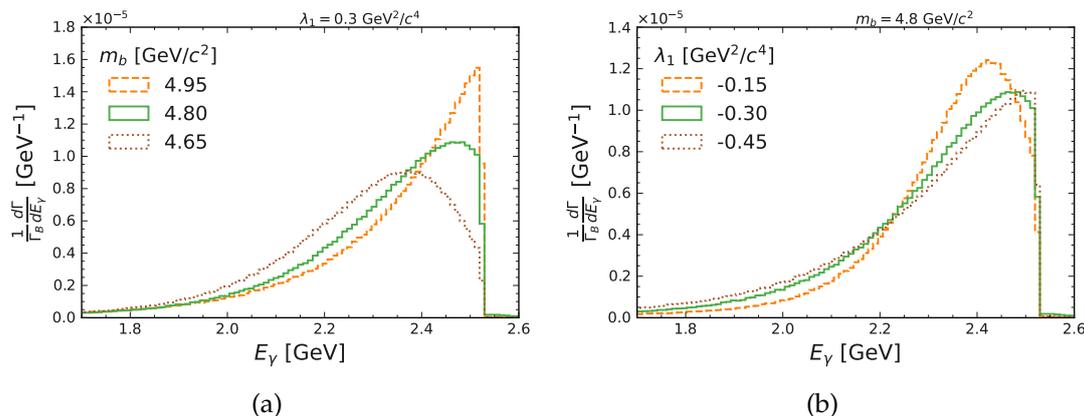
Figure 2.6: The $B \to X_s\gamma$ spectrum predicted by the Kagan-Neubert model, generated using the `EvtGen` generator's `BTOXSGAMMA` model. In (a), variations of m_b affecting the spectrum while keeping λ_1 constant are shown. In (b), the inverse scenario is depicted. The values here are provided in the kinetic scheme and correspond to those used by the authors of Ref. [63]. The generated spectra correspond to 10^6 events of both B^+ and B^0 decay modes. The full line corresponda to the default setup used by `BTOXSGAMMA`. The model generates the E_γ spectrum for $m_{X_s} > 1.1$ GeV/c^2, which corresponds to $E_\gamma \lessgtr 2.52$ GeV (see Equation (3.2)).

The works presented in this thesis will use the Kagan-Neubert model to generate the in-clusive photon energy spectrum. The m_b and λ_1 values measured by the SIMBA collaboration will be used, as they are extracted from all available experimental evidence, and contain slightly larger uncertainties making it a conservative estimate. Generally, these parameters are heavily dependent on the renormalisation scheme that is chosen. The relations provided in [49] will be used to transform the '1S/inv' scheme to the kinetic scheme at the precision of $O(\alpha_s^2)$, which can be used in the Kagan-Neubert model.

Lastly, it is also important to stress that the Kagan-Neubert (or any inclusive) model cannot generate resonant structures that the m_{X_s} exhibits. As a two body decay, m_{X_s} and E_γ are directly related:

$$m_{X_s}^2 = m_B^2 - 2m_B E_\gamma. \tag{2.35}$$

Therefore, the theoretical inclusive photon energy spectrum must always be interpreted in the picture of quark-hadron duality. In the low-E_γ region non-resonant and resonant decays are effectively indistinguishable as there are numerous available kinematic states. The inclusive model describes the spectrum as an average of all the states, and this is further enhanced by resolution effects in experimental data (see Section 5.3). In the high-E_γ region, however, the spectrum is dominated by several resonances, most notably the well-separated $B \to K^*(892)\gamma$. A study by the authors of the Kagan-Neubert model [63] shows that the inclusive model for $m_{X_s} \gtrsim 1.1$ GeV/c^2 should be supplemented by the resonant decays. This approach is followed when implementing a model for the analysis in Section 6.2.3

2.6   New-physics opportunities in $B \to X_s \gamma$ decays

The last Section introduced that $B \to X_s \gamma$ is important to constrain the SM parameters, such as m_b, through the description of the B meson shape function and its moments. However, one of the main motivations to study $B \to X_s \gamma$ decays is their sensitivity to BSM models. Considering the SM diagrams in Figure 2.3, any non-SM particles that couple to quarks and/or photons could contribute.

If light particles that couple to quarks would exist, they would generally be expected to have been observed by now. Therefore, the new degrees of freedom introduced by BSM models are expected to be heavy. In effective field theory terms:

$$C_i^{\text{SM}} \to C_i^{\text{SM}} + \Delta C_i^{\text{BSM}}. \tag{2.36}$$

That is, BSM physics can manifest by modifying any Wilson coefficient contributing to the $b \to s \gamma$ Lagrangian (Equation (2.12)). New Wilson coefficients may also be introduced by certain theories. Some of the models which could exhibit these effects for $B \to X_s \gamma$ include two-Higgs-doublet models (2HDM) [65]–[67], minimal supersymmetric models with minimal flavour violation [66], [68]–[70] and left-right symmetric models [66]. More exotic models include general minimal-supersymmetric theories [71], models with extra dimensions [72]–[75], the littlest Higgs models [76], [77], and so-called 331 models [78].

One of the most compelling SM extensions is the set of the 2HDM models. A thorough overview of different types of 2HDM models is presented in Ref. [79]. The model is attractive because it can be incorporated into many theories that provide insight into the long-standing issues of the SM. In supersymmetric theories, a second Higgs doublet is required to give masses to both u- and d-type quarks. Furthermore, having a second Higgs doublet could allow solving the strong $\mathcal{CP}$ problem. Finally, it may also provide answers to the baryon asymmetry observed in the Universe. An example of a charged Higgs boson, a prediction of (but not exclusively) 2HDM models, contributing to the $b \to s \gamma$ transition is shown in Figure 2.3b.

Two main types of 2HDM models contribute in the quark sector with vanishing tree-level flavour-changing neutral current contributions: so-called type-I 2HDM, where all quarks can couple to only one of the Higgs doublets, or type-II 2HDM, where u-type quarks couple to one doublet and d-type couple to another. In the SM, a Lagrangian for a charged Higgs boson interaction with quarks, following these requirements, is [67]:

$$\mathcal{L}_{H^+} = \frac{1}{\sqrt{\sqrt{8}G_F}} \sum_{i,j=1} \bar{u}_i (A_u m_{u_i} V_{ij} P_L - A_d m_{d_j} V_{ij} P_R) d_j H^+ + \text{h.c.} \tag{2.37}$$

Here, A_q are couplings, related to either u- or d-type quarks, with masses m_u and m_d. The sum runs over quark flavours i and j, and V_{ij} is the appropriate element of the CKM matrix. P_L and P_R are the chiral projection operators, and h.c. implies an additional hermitian conjugate term. In the type-I 2HDM, the couplings to different u- and d-type quarks are:

$$A_u = A_d = \frac{1}{\tan \beta}, \tag{2.38}$$

whereas for type-II 2HDM:

$$A_u = -\frac{1}{A_d} = \frac{1}{\tan\beta}. \tag{2.39}$$

These couplings are expressed in terms of $\tan\beta$, which is the ratio of Higgs vacuum expectation values of the two doublets in these theories. At the leading-order, the interactions are fully described by $\tan\beta$ and the mass of the charged Higgs, m_{H^+}, but these receive corrections at higher-orders related to other parameters in 2HDM or super-symmetric theories [80].

The decay rate calculation for the 2HDM model in the above-described cases proceeds similarly as described in Section 2.4. By modifying the effective Wilson coefficients based on Equation (2.36) and calculating appropriate corrections [81]:

$$\Delta C_i^{(0)2HDM} = \begin{cases} 0, & i = 1, ..., 6; \\ \frac{A_u^2}{3} F_i^{(1)}(y) - A_u A_d F_i^{(2)}(y), & i = 7, 8; \end{cases} \tag{2.40}$$

where $F_i^{(1)}(y)$ have already been defined in Equation (2.20), and

$$
\begin{aligned}
F_7^{(2)}(y) &= \frac{3y^2 - 2y}{6(y-1)^3} \ln y + \frac{-5y^2 + 3y}{12(y-1)^2}, \\
F_8^{(2)}(y) &= \frac{-y}{2(y-1)^3} \ln y + \frac{-y^2 + 3y}{4(y-1^2)},
\end{aligned}
\tag{2.41}
$$

with $y = m_t^2/m_{H^+}^2$. Using Equation (2.38), in the case of type-I 2HDM, the charged Higgs contribution takes the form $A\cot^2\beta - B\cot^2\beta$ which interferes with the SM value destructively. On the other hand, type-II 2HDM (Equation (2.38)) will take the form $A\cot^2\beta + B$ which is always constructive [82].

Higher-order corrections for 2HDM are provided in Refs. [67], [81] and have reached the next-to-next-to-leading-order precision. Based on Equation (2.40) and higher-order corrections, one can quantify the effect such models would have on the $B \to X_s\gamma$ branching fraction. The dependence of the $B \to X_s\gamma$ decay rate on m_{H^+} is shown in Figure 2.7 for two particular cases of $\tan\beta$ (see later Figure 2.8).

The destructive (constructive) interference predicted by the type-I (type-II) 2HDMs is apparent in Figure 2.7. Comparing the experimental and theoretical results seen in the Figures allows ($\tan\beta$, m_{H^+}) parameter space ranges to be determined. In both cases, as the charged Higgs mass increases, the predicted total decay rate approaches the SM values. This is because, in the $m_H^+ \to \infty$ regime, the charged Higgs is effectively decoupled from the SM. The bounds imposed by next-to-next-to-leading-order $B \to X_s\gamma$ results on the (m_{H^+}, $\tan\beta$) space are shown in Figure 2.8.

It can be seen that ($\tan\beta$, m_{H^+}) space is well-bounded for type-I and type-II 2HDM results. The strongest bounds arise for type-I 2HDM results in the range of $\tan\beta \in (0.4, 2)$. On the other hand, type-II 2HDM saturates at the high $\tan\beta$ limit. For values outside these ranges other experimental bounds are necessary [82]. Using the methods laid out in this Section, Ref. [26] rules out type-II 2HDM models with a charged Higgs lighter than 800 GeV at 2σ. A compatible result is also obtained in Ref. [83].

(a)   (b)
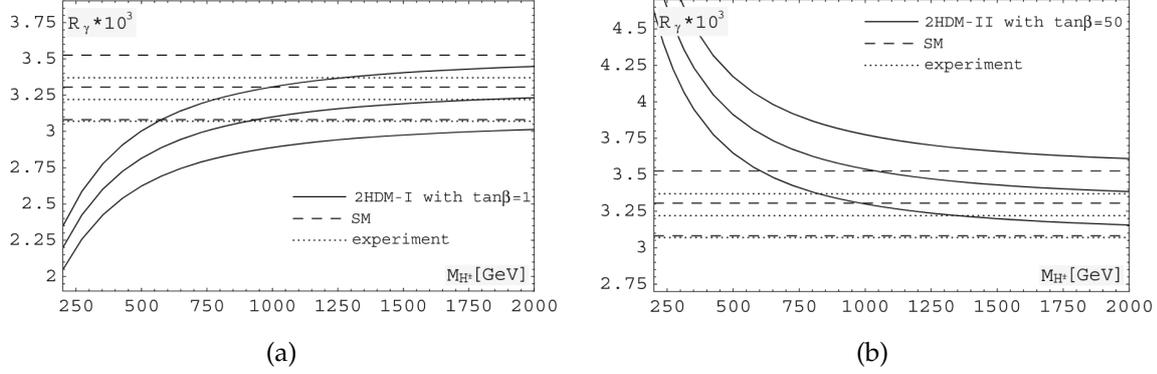
Figure 2.7: Variations of the $B \to X_s \gamma$ total decay rate (normalised by $B \to X_c \ell \bar{\nu}$ decay rate) as a function of the mass of the charged Higgs boson. In (a), the dependence for a type-I 2HDM model is shown, whereas (b) shows that for type-II 2HDM. The predictions are made assuming $\tan \beta = 1$ and $\tan \beta = 50$ for both types, respectively. The experimental bounds, as well as theoretical ones (not the most up-to-date), are shown as well. The Figures are taken from Ref. [82].
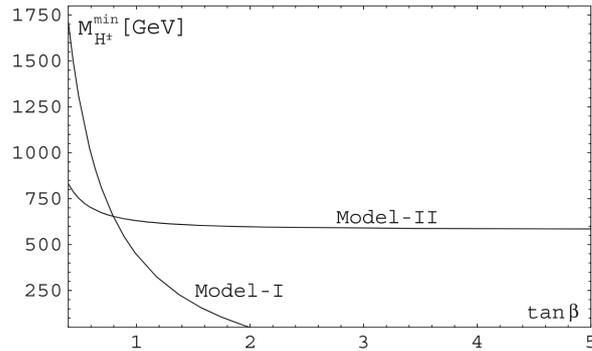
Figure 2.8: Excluded $(\tan \beta, m_{H^+})$ parameter space (95% confidence level lower bound) for type-I and type-II 2HDM models, based on Figure 2.7. The Figure begins at $\tan \beta > 0.4$ as better alternative bounds exist below that range. The Figure is taken from Ref. [82].

Chapter 3

Experimental overview of rare radiative decays

There are two main types of analyses which retain the meaning from the theoretical description presented in Chapter 2: *inclusive* analyses and *exclusive* analyses. Another method, which is a mixture of the two methods, is known as the *sum-of-exclusive* method. Theoretical motivations for inclusive measurements of $B \rightarrow X_s\gamma$ have already been discussed in Sections 2.4 and 2.5. Similar motivations hold for other electroweak decay channels, such as $B \rightarrow X_s\ell\bar{\ell}$ ($\ell \in \{\mu^-, e^-, \nu\}$). However, compared to the latter, $B \rightarrow X_s\gamma$ contains one fewer interaction vertex and is therefore enhanced by two orders of magnitude, which makes it experimentally accessible with smaller data sets. In this Chapter, the main methods of performing an inclusive measurement are introduced while focusing predominantly on $B \rightarrow X_{s/d}\gamma$, although, generally, these methods are applicable to other decay channels as well.

3.1 Past measurements of $B \rightarrow X_{s/d}\gamma$ decays

Inclusive measurements target a wide selection of decay products; in the case of $X_{s/d}$, the sum of all states that originate in $b \rightarrow s$ or $b \rightarrow d$ transitions. Such states include resonant and non-resonant particles. A notable resonant state is $B \rightarrow K^*(892)\gamma$. Experimentally, it is highly accessible due to its narrow and isolated peak near the $b \rightarrow s\gamma$ two-body decay kinematic limit. A similar state for $b \rightarrow d\gamma$ is $B \rightarrow \rho(770)\gamma$. Non-resonant states include combinations of one or more pions and, in the case of $b \rightarrow s$, kaons. While the goal of an inclusive measurement is to measure the energy of photons from all these decays simultaneously, an exclusive measurement attempts to select one or several particular states from the spectrum.

The world average values of inclusive radiative B decay measurements are [16], [84]:

$$\mathcal{B}(B \rightarrow X_s\gamma) = (3.49 \pm 0.19) \times 10^{-4},$$
$$\mathcal{B}(B \rightarrow X_d\gamma) = (0.09 \pm 0.03) \times 10^{-4}. \tag{3.1}$$

The values of Equation (3.1) can be compared to the theoretical prediction in Equations (2.22) and (2.23). $B \rightarrow X_s\gamma$ results show an excellent agreement between theory and experiment. For $B \rightarrow X_d\gamma$, values agree within 2σ, however, the dominant source of experimental uncertainty is statistical.

Table 3.1 provides the experimental status of the most prominent observed $B \rightarrow X_{s/d}\gamma$ decay channels as of writing this. Due to the final state similarity and overlap between various resonant and non-resonant decay modes, only the most prominent resonant decays have been measured. However, even in the case of a relatively isolated decay channel, such as $B \rightarrow K^*(892)\gamma$, the overall precision of inclusive measurements is higher (compare to Equation (3.1)).

Table 3.1: Branching fractions of $B \rightarrow X_s\gamma$ modes for charged and neutral modes. The Table only includes decay modes that have been observed and (for $B \rightarrow X_s\gamma$ only) have a branching fraction $\gtrsim 10^{-5}$. The B^+ decays are ordered in terms of the experimental precision $\mathcal{B}/\Delta\mathcal{B}$, whereas B^0 are ordered in relation to B^+, where applicable. The values correspond to the averages of experimental measurements given in Refs. [16], [84].

$B^+ \rightarrow X_s\gamma$ exclusive modes

Decay mode	Branching fraction ($\times 10^{-4}$)
Two decay products	
$B^+ \rightarrow K^*(892)^+\gamma$	0.392 ± 0.022
$B^+ \rightarrow K_1(1270)^+\gamma$	$0.438 \pm^{0.071}_{0.063}$
$B^+ \rightarrow K_2^*(1430)^+\gamma$	0.138 ± 0.040
$B^+ \rightarrow K^*(1410)^+\gamma$	$0.271 \pm^{0.080}_{0.061}$
$B^+ \rightarrow K^*(1680)^+\gamma$	$0.670 \pm^{0.170}_{0.140}$
$B^+ \rightarrow K_1(1400)^+\gamma$	$0.097 \pm^{0.054}_{0.038}$
Three or more decay products	
$B^+ \rightarrow K^*(892)^0\pi^+\gamma$	0.233 ± 0.012
$B^+ \rightarrow K^+\pi^+\pi^-\gamma$	0.258 ± 0.015
$B^+ \rightarrow K^0\pi^+\pi^0\gamma$	0.456 ± 0.052
$B^+ \rightarrow K^+\pi^+\pi^-\gamma$ (non-resonant)	$0.099 \pm^{0.017}_{0.020}$

$B^0 \rightarrow X_s\gamma$ exclusive modes

Decay mode	Branching fraction ($\times 10^{-4}$)
Two decay products	
$B^0 \rightarrow K^*(892)^0\gamma$	0.418 ± 0.025
-	-
$B^0 \rightarrow K_2^*(1430)^0\gamma$	0.124 ± 0.024
-	-
-	-
-	-
Three or more decay products	
-	-
$B^0 \rightarrow K^+\pi^-\pi^0\gamma$	0.407 ± 0.038
$B^0 \rightarrow K^0\pi^+\pi^-\gamma$	0.199 ± 0.018
-	-

$B \rightarrow X_d\gamma$ exclusive modes

Decay mode	Branching fraction ($\times 10^{-4}$)
$B^+ \rightarrow \rho^+(770)\gamma$	$0.0098 \pm^{0.0025}_{0.0024}$
$B^0 \rightarrow \rho^0(770)\gamma$	0.0086 ± 0.0015
$B^0 \rightarrow \omega(782)\gamma$	$0.0044 \pm^{0.0018}_{0.0016}$

Currently, inclusive measurements are only attainable at B-factories. The relatively low-background environment offered by e^+e^- collisions enables treating the $X_{s/d}$ system as a 'missing-momentum' system with few explicit requirements. Although the process of event reconstruction may introduce biases to the inclusive system, this is expected to be a much smaller effect than other experimental factors, such as finite resolution or statistical fluctuations. Conversely, at hadron colliders, large proton-proton collision background makes it complicated to select an unbiased and model-independent inclusive sample. This is further complicated by multiple proton pairs interacting in a collision event and creating large multijet backgrounds, usually referred to as *pileup*. At the time of writing this, no inclusive B measurement has been performed outside e^+e^- collision experiments.

Even in the case of exclusive radiative measurements, such as $B \rightarrow K^*\gamma$, B-factories have historically outperformed hadron collider experiments (such as LHCb) due to the cleaner e^+e^- collision environment. Final states that include neutral particles and photons are prob-

lematic to measure accurately in hadron colliders. Therefore, with several exceptions (e.g. Ref. [85]), the field of rare radiative *B* decay measurements is dominantly probed by the *B* factory experiments.

3.2 Techniques for inclusive $B \to X_s\gamma$ measurements

Historically, three different techniques were applied for inclusive $B \to X_s\gamma$ analyses at *B* factories: sum-of-exclusive measurements, untagged inclusive measurements and tagged inclusive measurements. These experiment techniques are explained in more detail in this Section. The primary focus will be given to the hadronic-tagged technique, which is applied for the measurement described in this thesis. The summary of the most precise measurements that are used in the experimental average of Equation (3.1) is given in Table 3.2.

Table 3.2: Different experiments and their most precise results using various techniques of measuring $B \to X_s\gamma$. These results are included in the total $B \to X_s\gamma$ world average (Equation (3.1)) [16], [84]. The thresholds of the photon energy in the decaying *B* meson rest frame (E_γ^B), quoted in the corresponding papers, are also provided. The branching fractions are extrapolated to 1.6 GeV, using extrapolation factors calculated in Ref. [86]. The Belle[†] measurement was not published or used in the averages but is included here as the lepton-tagged measurement with the largest data sample.

Year	Experiment	Technique	Data used	Energy threshold	$\mathcal{B}(B \to X_s\gamma) \times 10^{-4}$ [$E_\gamma^B > 1.6$ GeV]
2001	CLEO [87]	Untagged	9.1 fb^{-1}	$E_\gamma^B > 2.0$ GeV	$3.29 \pm 0.44 \pm 0.29$
2007	BaBar [88]	Hadronic-tagged	210 fb^{-1}	$E_\gamma^B > 1.9$ GeV	$3.90 \pm 0.91 \pm 0.64$
2009	Belle [50]	Untagged/Lepton-tagged	605 fb^{-1}	$E_\gamma^B > 1.7$ GeV	$3.47 \pm 0.15 \pm 0.40$
2012	BaBar [89]	Lepton-tagged	347 fb^{-1}	$E_\gamma^B > 1.7$ GeV	$3.32 \pm 0.16 \pm 0.31$
2012	BaBar [90]	Sum-of-exclusive	429 fb^{-1}	$E_\gamma^B > 1.7$ GeV	$3.52 \pm 0.20 \pm 0.51$
2014	Belle [91]	Sum-of-exclusive	711 fb^{-1}	$E_\gamma^B > 1.7$ GeV	$3.75 \pm 0.18 \pm 0.35$
2016	Belle[†] [92]	Lepton-tagged	711 fb^{-1}	$E_\gamma^B > 1.6$ GeV	$3.12 \pm 0.10 \pm 0.21$

3.2.1 Sum-of-exclusive technique

The sum-of-exclusive measurement technique embodies the idea of reconstructing all X_s states separately and summing them up into an inclusive spectrum. In practice, this is, of course, impossible; previous BaBar and Belle analyses ([90], [91]) reconstruct a sum of 38 exclusive channels that amount to roughly 70% of the total $B \to X_s\gamma$ decay width. The selected final states include various combinations of one or multiple $K^\pm$, K_S^0, $\pi^\pm$, π^0, η. Modes with up to three kaons, four pions and one η-meson are considered.

A significant challenge of the method is the proper treatment of $B \to X_s\gamma$ events that have been incorrectly reconstructed in one of the 38 final states. Photons originating in non-$B \to X_s\gamma$ decay chains, particularly decays of the type $B \to D^{(*)}\rho^+$ and non-*B* events, also contribute significantly as background. Much more than in the case of inclusive measurements, this technique strongly depends on the X_s fragmentation modelling, which has to be carefully tuned and calibrated to represent the experimental data. Finally, a measure-

ment performed this way is only 'pseudo-inclusive', meaning that additional uncertainties for unaccounted decay phase space are incurred.

The main advantages of this method are due to the precise knowledge of the X_s system, which gives additional information about the B meson, and a higher degree of control of background. The knowledge of the charge and flavour of the decaying B meson enables the measurement of the $C\mathcal{P}$ and isospin asymmetries [93]. Furthermore, it is the only inclusive measurement technique that has been able to distinguish X_s and X_d states experimentally [94]. Direct reconstruction of the mass of the X_s system, m_{X_s}, and decaying B meson mass, m_B, allows expressing the photon energy directly in the signal B meson rest frame:

$$E_\gamma^B = \frac{m_B^2 - m_{X_s}^2}{2m_B},$$ (3.2)

which is otherwise only directly obtainable by hadronic-tagged inclusive measurements (see Section 3.2.3). Furthermore, the full reconstruction of the candidate B meson allows using the well-defined initial state of the e^+e^- collision for additional background suppression. Two observables can be defined:

$$\Delta E \equiv E_B^* - \sqrt{s}/2,$$ (3.3)

known as the energy difference, expressed in terms of the energy of the B meson in the collision centre-of-mass frame, E_B^*, and

$$M_{bc} \equiv \sqrt{(\sqrt{s}/2)^2 - (p_B^*)^2},$$ (3.4)

known as the beam-constrained mass, expressed in terms of the momentum of the B meson in the collision centre-of-mass frame, p_B^*. From Equations (3.3) and (3.4) it is clear that B candidates which are reconstructed correctly tend to have a resonant behaviour in M_{bc} and ΔE, with their distributions peaking at the nominal B mass ≈ 5.28 GeV/c^2 and 0, respectively. The backgrounds tend to have broader or even non-peaking shapes. An M_{bc} distribution is shown in Figure 3.1, as measured by the Belle sum-of-exclusive measurement [91]. In the past, these analysis techniques achieved an average signal reconstruction efficiency of 3.5% (larger for greater values of E_γ^B) after correcting for missing X_s modes [91].

3.2.2   Untagged technique

The untagged inclusive measurements are the simplest conceptually as they only require a high-energy photon in the final state, without any explicit assumptions about the X_s or partner B meson decay. Such a simple requirement guarantees that photons from all $B \to X_s\gamma$ decays are included in the selected data sample. Because any partner B meson state is accepted, they are sometimes also called *inclusive-tagged* or *fully inclusive*. For the rest of the thesis *untagged* is used to stress the difference between the signal B decay and the partner B.

Although the concept of this approach is simple, the measurement is highly challenging experimentally. In particle collisions and subsequent decay processes, high-energy photons can originate through numerous ways, such as initial e^+e^- state radiation, $e^+e^- \to q\bar{q}$ processes, B decays etc. In the previously introduced sum-of-exclusive technique this problem is solved by using the information of decays of X_s and using observables such as M_{bc} and
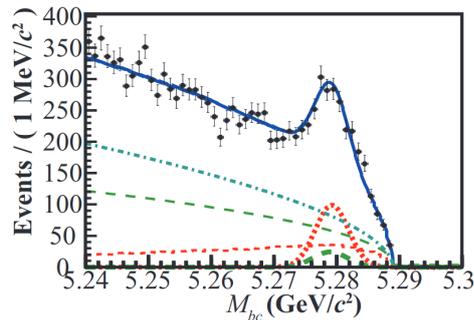
Figure 3.1: The distribution of M_{bc}, as seen in the $1.9 > M_{X_s} > 1.8$ GeV/c^2 interval by Ref. [91] in the sum-of-exclusive state analysis. The data points are fitted in an unbinned maximum likelihood fit with a combination of fit functions for signal events (red, thick and short dashed), cross-feed (red, thin and short dashed), peaking $B\bar{B}$ (green, thick and long dashed), non-peaking $B\bar{B}$ (green, thin and long dashed) and $q\bar{q}$ background events (blue, dash-dotted).

ΔE to suppress incorrect photon candidates. Conversely, in the inclusive case, X_s is treated as unreconstructed, and therefore a background suppression procedure is performed such that no selection bias is introduced to the X_s system. An example spectrum based on Belle II simulation is shown in Figure 3.2a, which highlights the signal-to-background difference.

The predominant background originates from events where no B mesons are created, for example, in $e^+e^- \to q\bar{q}$ events ($q \in \{u, d, s, c\}$), where a π^0 is created through hadronisation or decays of other hadrons. The π^0 then decays asymmetrically into two photons, which mimics the isolated high-energy photon of $B \to X_s\gamma$ decays. In particular, $q = u, c$ processes contribute strongly, since they tend to produce energetic π^0 or charm mesons (whose decay chains include π^0). Semileptonic and hadronic B decays that produce π^0, η and ρ mesons, are also among the main sources of background contribution.

The usual way to perform this analysis involves using boosted decision trees or other multivariate methods (see Section 5.2) to suppress highly prominent backgrounds. For example, the technique applied in Refs. [50], [87] combines high-energy photons with all the other photons in the event and vetoes those that are compatible with π^0 or similar decays. The contributions from $e^+e^- \to q\bar{q}$ are suppressed by parametrising the different decay topologies that are observed for $B\bar{B}$ and $q\bar{q}$ events (see Section 6.5). Data samples collected below the $\Upsilon(4S)$ resonance, containing only $e^+e^- \to q\bar{q}$ events, are used to subtract the non-$B\bar{B}$ contributions remaining after the background treatment. The $B\bar{B}$ background contribution is usually removed using simulation. An example result of the extracted $B \to X_s\gamma$ spectrum from Belle II data using 63.1 fb^{-1}, which is a work that I was involved in during the doctoral research [95], is shown in Figure 3.2b.

The strength of the untagged technique is the 'truly' inclusive approach, which ensures that all X_s states are selected, as well as a large signal reconstruction efficiency. Previous analyses (e.g. Ref. [50]) report an average selection efficiency in their final sample of $\sim 10\%$ (increasing with photon energy), which is several times higher than the sum-of-exclusive approach. On the other hand, the missing kinematic information of the X_s system yields a

complicated and inefficient background suppression process, which means that such measurements have a small signal-to-background ratio. Moreover, the kinematic information of the B decay cannot be accessed, which only allows a measurement of the photon energy in the e^+e^- collision frame. To reach the theoretically more desirable B meson decay frame, additional modelling uncertainties have to be introduced.

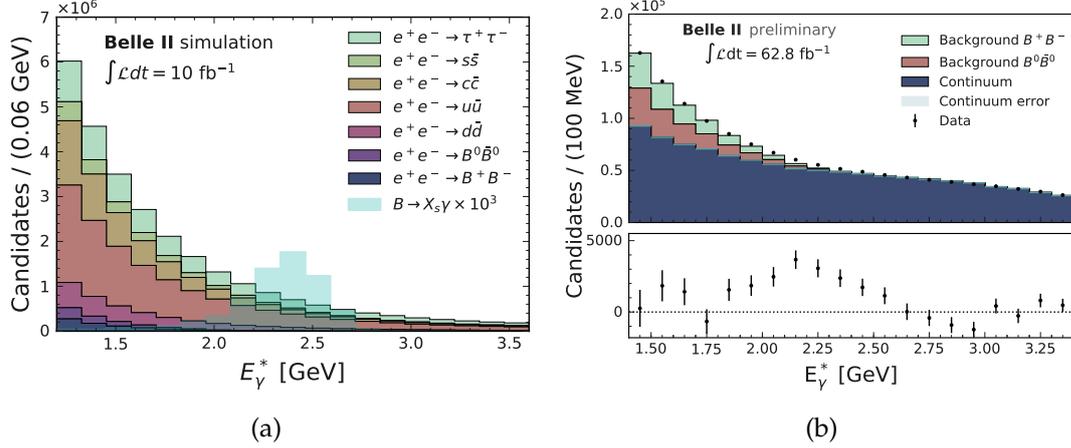(a)   (b)

Figure 3.2:   Photon energy spectra in $B \rightarrow X_s\gamma$ decays before (a) and after (b) background suppression. Before the background suppression is performed, the signal fraction is roughly 1000 times smaller than the background. After background suppression, subtracting the remaining continuum ($e^+e^- \rightarrow q\bar{q}$ decays) and $B\bar{B}$ background yields the $B \rightarrow X_s\gamma$ spectrum (bottom panel). Both Figures show Belle II simulated data. In (b), official Belle II data from Ref. [95] is shown. Figure 3.2a is produced for illustrative purposes only.

3.2.3   Tagged techniques

To overcome multiple of the issues that come with untagged measurements presented in Section 3.2.2 while still selecting an inclusive X_s sample (conversely to Section 3.2.1), additional information about the second B meson from the $\Upsilon(4S)$ decay can be used. The naming stems from the similarity to the *tag-and-probe* methods. In this approach, the partnering B meson is fully reconstructed, or some of its decay products are identified. This B meson will henceforth be referred to as the *tag-B* meson. The application of kinematic constraints on the event arising from the tag-B is called *tagging*. The schematic idea of tagging is shown in Figure 3.3a. Three main tagging techniques have been used in the past at B-factories:

- *lepton-tagging*, where a lepton originating from tag-B decays is reconstructed;

- *semileptonic-tagging*, where the tag-B is reconstructed as a semileptonic B decay of the form $B \rightarrow D^{(*)}\ell\bar{\nu}$;

- *hadronic-tagging*, where the tag-B is reconstructed as a decay that involves the final states $B \rightarrow$ hadrons, such as $B \rightarrow K\pi$.
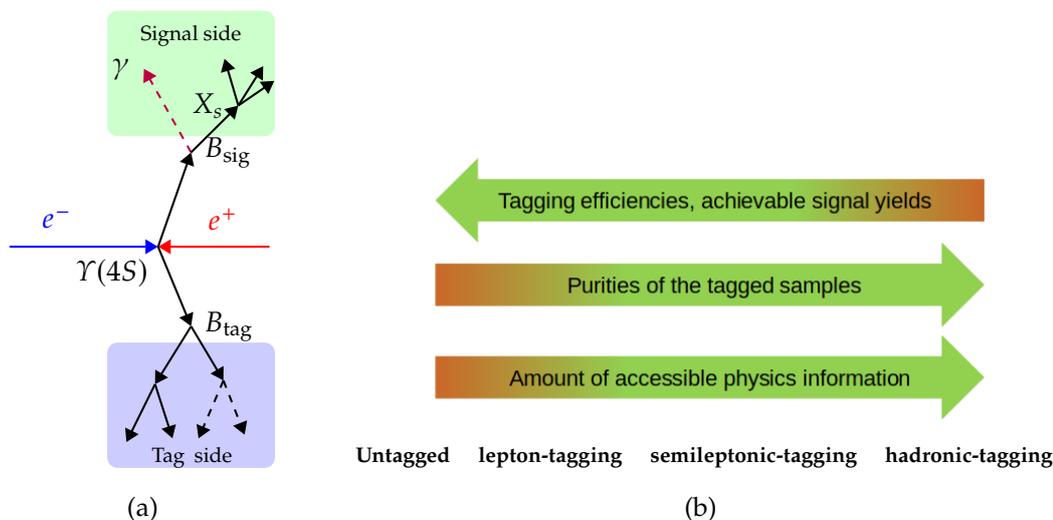
Figure 3.3: Schematic representation of tagging is shown in (a). The idea of tagging is using the tag-side B decay products to apply kinematic constraints on the signal-side B decay ($B \rightarrow X_s \gamma$ in the example). In (b), the advantages and disadvantages related to using different B decay products for tagging at B-factories are highlighted. A detailed discussion of these techniques is given in the text. Credit to Dr. Markus Röhrken for Figure 3.3b.

The main advantages of these techniques are summarised in Figure 3.3b.

Leptonic tagging has been used as a 'successor' method for $B \rightarrow X_s \gamma$ untagged analyses by BaBar and Belle [50], [89], [92], providing a higher degree of background control while still retaining a larger efficiency. In the past, lepton-tagged analyses achieved an average signal reconstruction efficiency of up to 3% (increasing with photon energy). Going a step further and reconstructing the charm meson and the lepton from the semileptonic B decay gives an even higher degree of background control. Measuring the angle between the reconstructed $D\ell$ system and the decaying B meson, for example, provides excellent background suppression [96]. A major complication is the fact that the reconstruction of the semileptonic decay chain necessarily reduces the overall efficiency. B mesons have large semileptonic decay rates, meaning that the lower efficiency due to D reconstruction is partially offset by the large statistical samples. However, the presence of a neutrino in the final state complicates the technique further. On average, this technique is at least an order of magnitude less efficient than the untagged approach and several times less efficient than the lepton-tagged method [23]. Despite successful application for missing energy modes, e.g., $B \rightarrow K^+ \nu\nu$ [97], semileptonic tagging was never used for $B \rightarrow X_s \gamma$.

Of particular interest is hadronic tagging, which, in the context of $B \rightarrow X_s \gamma$, has been performed only once by BaBar [88], using roughly 50% of their total data set. Compared to other tagged and untagged inclusive techniques, this is the only method which fully reconstructs the kinematics of the tag-B due to the absence of neutrinos in the final state. As a result, with the beam constraint requirements, one can calculate quantities such as M_{bc} and ΔE (see Equations (3.3) and (3.4)) for the tag-side B meson. The ability to rely on these distributions and, in particular, perform a signal extraction fit, similar to the one given in Figure 3.1, allows

suppressing previously dominant $e^+e^- \rightarrow q\bar{q}$ components to negligible levels.

Because both B mesons at B-factories are created from decays of $\Upsilon(4S)$, the full knowledge of the tag-B properties allows inferring the charge, momentum and flavour of the signal-B meson, and consequentially, measuring the desired observables in the decaying-B rest frame. A mathematical description of a Lorentz transformation using tag-B constraints is provided in Appendix C. Therefore, one regains all the benefits that the sum-of-exclusive technique offers, while still ensuring that no selection requirements are imposed on the X_s system. However, a complication that follows is the fact that hadrons can have thousands of decay chains and an efficient reconstruction of a statistically significant tag-B sample is complicated. Compared to semileptonic tagging, the hadronic tagging technique has a several times lower efficiency, although this is compensated by a very high purity of the tagged data sample [23]. The BaBar analysis achieved a signal efficiency of ~0.2% depending on the E_γ^B interval (increasing with photon energy).

Hadronic-tagged measurements of $B \rightarrow X_s\gamma$ have uncertainties that are mainly related to the modelling of $B\bar{B}$ background, M_{bc} fitting and correlation between tag-side decays and signal-side reconstruction. The previous hadronic-tagged BaBar measurement [88] achieved a 16% systematic uncertainty and a 23% statistical uncertainty of the branching fraction. Both uncertainties are expected to be improved with larger data samples. As evident from Table 3.2, historically, sum-of-exclusive and lepton-tagged methods have been the most precise measurements of the $B \rightarrow X_s\gamma$ spectrum. The uncertainty of the hadronic-tagged measurement is higher but comparable to that of untagged, leptonic-tagged and sum-of-exclusive measurements, even though the hadronic-tagged analysis has been performed with only half of the available BaBar data. The hadronic-tagged technique can therefore provide one of the world's most accurate measurements with an increased data sample [98].

Moreover, because the hadronic-tagged analysis incurs different systematic uncertainties and relies less on simulation, it is a powerful cross-check of the other tagged analyses. It is also important to note that these techniques produce samples that are not highly correlated due to different backgrounds specific to the analysis procedure (see e.g. Ref. [50]). Therefore, different tagging techniques *complement* but do not *compete with* each other.

Chapter 4

Experimental setup

A common strategy to study B meson decays is using dedicated colliders, known as B factories. B-factory experiments operate by producing large quantities of $B\bar{B}$ pairs, through the creation of $\Upsilon(4S)$ mesons that primarily decay to a B meson pair ($B\bar{B}$) [16]. Historically, two B-factory experiments operated:

- BaBar at the PEP-II accelerator at SLAC, USA [99];

- Belle at the KEKB accelerator at KEK, Japan [100].

Another general flavour physics experiment is LHCb [101] with the LHC accelerator at CERN, collecting B meson data produced in proton-proton collisions. It has been operating since 2008.

Belle and BaBar are sometimes also referred to as first-generation asymmetric-energy B factories, as they are the first that employed asymmetric beam collisions (see Section 4.1). CLEO [102] and ARGUS [103] experiments also collected significant B meson data sets at symmetric electron-positron (e^+e^-) collision energies and are therefore considered predecessors to B-factories. An upgraded version of Belle, known as Belle II, began collecting data in 2018. Its main purpose is the collection of e^+e^- collision data at the centre of mass energies ($\sqrt{s}$) at or near the $\Upsilon(4S)$ meson mass. The colliding beams are provided by the SuperKEKB collider. This Chapter provides an overview of Belle II and introduces the main concepts of the SuperKEKB accelerator.

4.1 The SuperKEKB accelerator

The SuperKEKB accelerator, discussed in detail in Ref. [104], is a double-ring electron-positron collider. It is an upgraded version of the KEKB collider [105] that operated with the Belle experiment. The SuperKEKB accelerator complex is schematically shown in Figure 4.1. A photo-cathode radio-frequency gun produces two electron beams. The first beam is subsequently accelerated to 7 GeV by a linear accelerator into the electron ring. On the other hand, positrons are created by directing the second electron beam to a tungsten target. The positrons are singled out using a magnetic field and accelerated to 1.1 GeV, injected into

the damping ring and, finally, accelerated again by the linear accelerator to 4 GeV before entering the main positron ring. The subsequent collision occurs in the interaction region inside the Belle II detector, where the electron and positron rings meet (see Section 4.2).
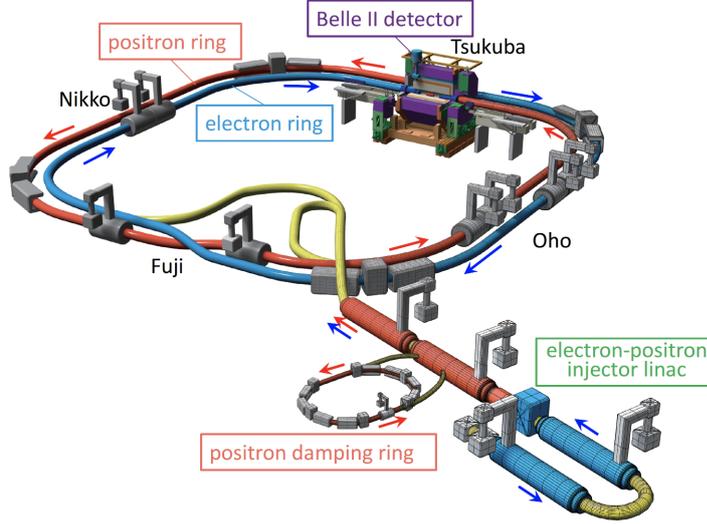

Figure 4.1: The schematic visualisation of the SuperKEKB accelerator complex. The main components that contribute to the acceleration of electrons and positrons are shown. The four straight sections are named after Japanese cities. Credit to Ref. [104].

The beam asymmetry (7 GeV for e^- and 4 GeV for e^+) is an important design characteristic of SuperKEKB, which allows separating B meson decay vertices by $\mathcal{O}(\mu m)$, necessary for measurements such as time-dependent $\mathcal{CP}$ violation [96]. On the other hand, the exact collision energy is chosen to operate at $\sqrt{s} = \sqrt{(7+4)^2 - (7-4)^2}$ GeV ≈ 10.58 GeV, which corresponds to the $m_{\Upsilon(4S)}$, hence fulfilling the requirements of a B-factory experiment.

SuperKEKB also collects data at different $\sqrt{s}$. For example, in the setup when the collision energy is lowered by 60 MeV, $e^+e^- \to \Upsilon(4S)$ events are not produced. Such data, containing no $\Upsilon(4S)$ events, is called *off-resonance* data. Conversely, the conventional previously mentioned setup is referred to as *on-resonance* data.

It is important to emphasise that an e^+e^- collision at 10.58 GeV can produce more than just the $\Upsilon(4S)$. Many other processes occur, such as $e^+e^- \to \ell^+\ell^-$ or $e^+e^- \to q\bar{q}$, and the production cross-section of all these processes depends on $\sqrt{s}$. This is shown for $\sqrt{s} \approx 10.58$ GeV in Figure 4.2, with more details about the exact values of the cross-sections provided in Appendix B.

Although by far the largest cross-sections are related to the *low-multiplicity* processes, such as $e^+e^- \to e^+e^-$ and $e^+e^- \to \mu^+\mu^-$ (see Appendix B), they differ largely from typical $\Upsilon(4S) \to B\bar{B}$ events. On the other hand, the *continuum* ($e^+e^- \to q\bar{q}$ and $e^+e^- \to \tau^+\tau^-$) events are a significant background process for many analyses aiming to measure B meson decays [1].

SuperKEKB has a design luminosity of $L = 8 \times 10^{35}$ cm^{-2}s^{-1}, which is 40 times higher

---

[1]Because of the large $e^+e^- \to q\bar{q}$ and $e^+e^- \to \tau^+\tau^-$ production cross-sections, B-factories are also used to study charm mesons or $\tau^\pm$ and similar decays. In such cases, the continuum events are the events of interest.
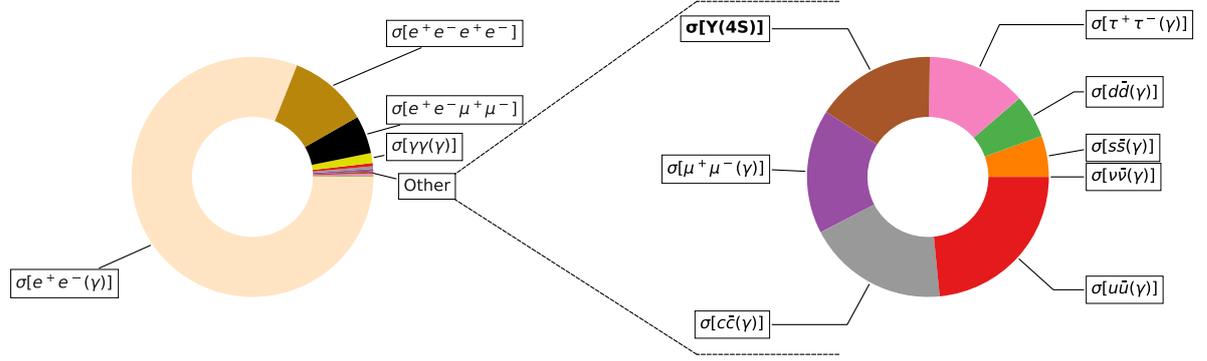
Figure 4.2: Relative comparison of the largest $e^+e^- \rightarrow X$ production cross-sections at B-factories. The absolute scale of the cross-section is $\mathcal{O}(\text{nb}^{-1})$. The exact numbers composing the charts are listed in Appendix B and taken from [23].

than the maximum achieved by KEKB [106]. Currently, SuperKEKB holds the instantaneous luminosity world record, which at the time of writing this thesis is $L = 4.65 \times 10^{34} \text{ cm}^{-2} s^{-1}$. This is enabled by the increased beam intensity, upgraded beam focusing technique (nano-beam scheme) and other improvements which are beyond the scope of this thesis.

4.2 The Belle II experiment

The Belle II detector, discussed in detail in Ref. [107], is designed to reconstruct the final states of e^+e^- collisions. Belle II operates since 2018 and has collected 364 fb^{-1} of on-resonance and 42 fb^{-1} of off-resonance data by the time of writing this thesis. The colliding beams are supplied by the SuperKEKB accelerator as discussed in Section 4.1. The experiment is designed to collect 50 ab^{-1} of collision data, which will be nearly 50 times that of Belle. Belle II consists of several detector subsystems (Sections 4.2.1 to 4.2.7) arranged cylindrically around the interaction region. The visual representation of the Belle II detector is given in Figure 4.3 and shows the subdetectors with their acronyms.

The coordinate system of Belle II is defined as follows: the x axis is defined to be horizontal and points to the outside of the tunnel with respect to the accelerator's main rings, the y axis is vertically upward, and the z axis is defined in the direction of the electron beam. The azimuthal angle, ϕ, and the polar angle, θ, are defined with respect to the z axis. Three regions in the detector are defined based on θ:

- forward endcap ($12° < \theta < 31°$),

- barrel ($32° < \theta < 129°$),

- backward endcap ($131° < \theta < 155°$).

Superconducting Solenoid

Pixel Detector (PXD)

Silicon Vertex Detector (SVD)

Central Drift Chamber (CDC)

TOP counter (TOP)

Aerogel RICH counter (ARICH)

Electromagnetic Calorimeter (ECL)
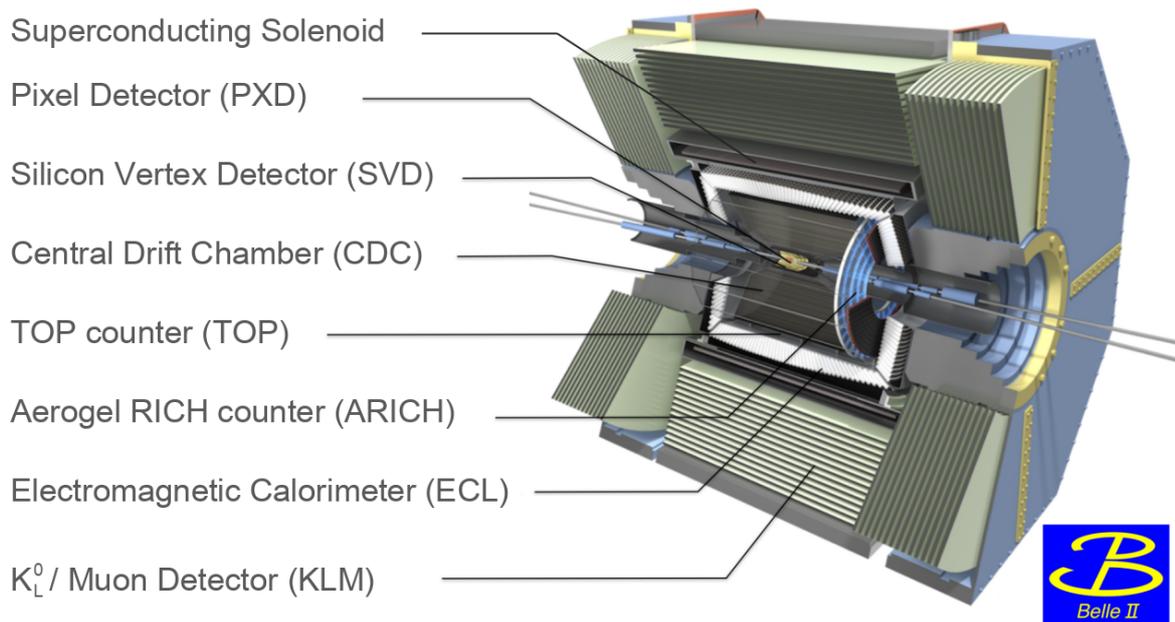
K_L^0 / Muon Detector (KLM)

Figure 4.3:   The schematic representation of the Belle II detector. The Belle II cylinder is approximately 7 meters in diameter and 7.5 meters long. The description of all sub-systems is provided in Sections 4.2.1 to 4.2.7. The Figure is adapted from Ref. [108]. Credit to the Belle II collaboration.

4.2.1 Pixel vertex detector

Closest to the interaction region is the pixel vertex detector (PXD) [107]. Its main purpose is the high-precision measurement of short-lived particle decay vertices, such as that of B mesons.

The PXD contains two layers of depleted p-channel field-effect transistors (DEPFET) to detect charged particles that pass through it [109]. The first layer is located at a 14 mm radius from the interaction region, whereas the second is at 22 mm. Each layer consists of 16 and 24 sensor modules, which are glued in pairs along the short edge to make 8 and 12 ladders[2], respectively. Each module contains 768×250 DEPFET pixels.

The PXD is designed to operate in harsh radiation conditions due to the proximity to the interaction region, while maintaining high-precision vertex reconstruction and a low material budget. The sensors can withstand a 20 Mrad radiation dose and have a ~ 0.2% radiation length per layer [107], while maintaining an average spatial resolution of approximately 15 μm and a hit efficiency of 98% after 4 years of data taking [110]. The PXD is shown in Figure 4.4a, whereas its schematic location in the Belle II detector is depicted in Figure 4.5.

(a)   (b)
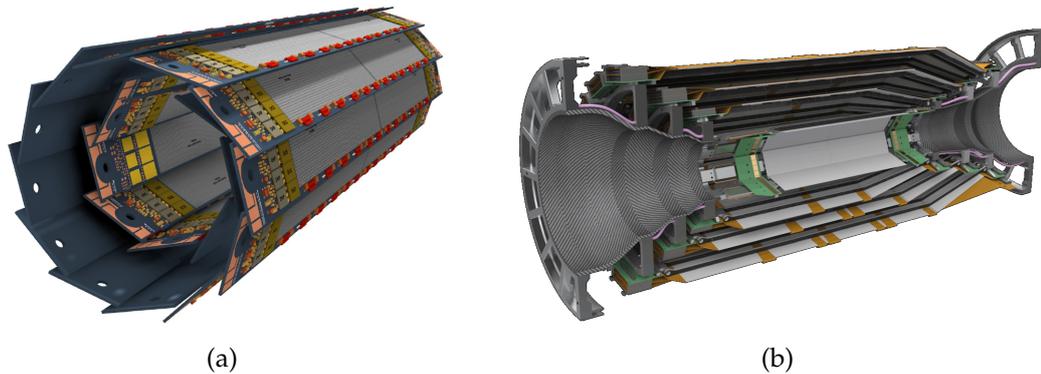
Figure 4.4: The pixel (a) and silicon (b) vertex detectors. They are installed in the Belle II detector as shown in Figure 4.5. The PXD is arranged into two layers that are designed to contain 16 modules in the first layer and 24 in the second. The SVD is composed of four layers around the PXD, with a total of 172 double-sided silicon strip sensors. Credit to Belle II PXD and SVD groups.

4.2.2 Silicon vertex detector

The silicon vertex detector (SVD) [111] surrounds the PXD and is the second subsystem responsible for charged particle detection. Its main roles include the reconstruction of short-lived particle decay vertices together with the PXD, standalone charged particle trajectory reconstruction for low momentum particles, and particle identification through specific ionisation measurements.

The SVD contains four layers. The first layer is composed of 7 ladders with 2 sensors, the second of 10 ladders with 3 sensors, the third of 12 ladders with 4 sensors, and the final of

---

[2]Until the scheduled 2023 Belle II upgrade the outer layer contained only 2 ladders.

16 ladders with 5 sensors. The 172 double-sided silicon strip sensors are composed, in total, of 224 000 strips. Each sensor is based on an n-type bulk implanted with p and n-doped sensitive strips. The p and n strips are aligned perpendicularly and on opposite sides of the sensor. The visual representation of the SVD is shown in Figure 4.4b, whereas its schematic location in the Belle II detector is depicted in Figure 4.5.

As charged particles pass through the SVD sensors, the electrons and holes created through ionisation drift to p and n strips, respectively. The perpendicularity of the strips ensures that the spatial coordinates of the passing particle can be inferred.
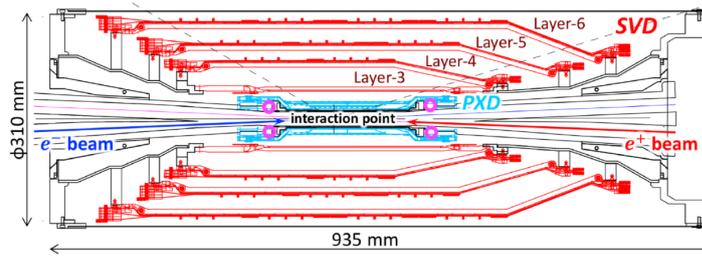

Figure 4.5: The Belle II pixel and silicon vertex detectors are shown inside the Belle II detector. The sizes of both vertex detection components and the interaction point location are noted. The Figure is taken from Ref. [111].

4.2.3 Central Drift Chamber

The central drift chamber (CDC) [112] is the central subsystem responsible for the reconstruction of charged particle trajectories inside the Belle II experiment. As such, its main objective is the measurement of particle momenta and charge. The CDC also provides particle identification information through specific ionisation measurements and participates in the decision to save the event information (*triggering*). It is a large volume drift chamber filled with a 50% helium and 50% ethane mixture. The CDC begins after the SVD at 160 mm and is contained within an outer cylinder radius of 1130 mm. It consists of 14336 readout wires distributed across 56 layers. Each readout wire is surrounded by 8 field wires that create an electric field in the chamber. As a charged particle passes through the chamber ionising the gas, the resulting electrons are accelerated in the electric field creating avalanches that are read out as signal by the wires. In order to obtain three-dimensional information about the particle trajectory, some layers in the CDC are skewed. The first 8 layers are axial, whereas the rest alternate between axial and skewed every 6 layers. The grouping of layers is shown in Figure 4.6a, whereas Figure 4.6b illustrates the difference between axial ad skewed layers.

The CDC covers $\theta \in (17, 150)°$ range and provides a highly accurate measurement of charged particle trajectories with a spatial resolution of $0.1 - 0.2$ cm and a transverse momentum resolution of 0.5% for the majority of particles resulting from B meson decays [113].

4.2.4 Particle identification systems

Belle II has two dedicated particle identification systems: an aerogel ring imaging Cherenkov counter (ARICH) in the forward endcap region and a time of propagation (TOP) chamber
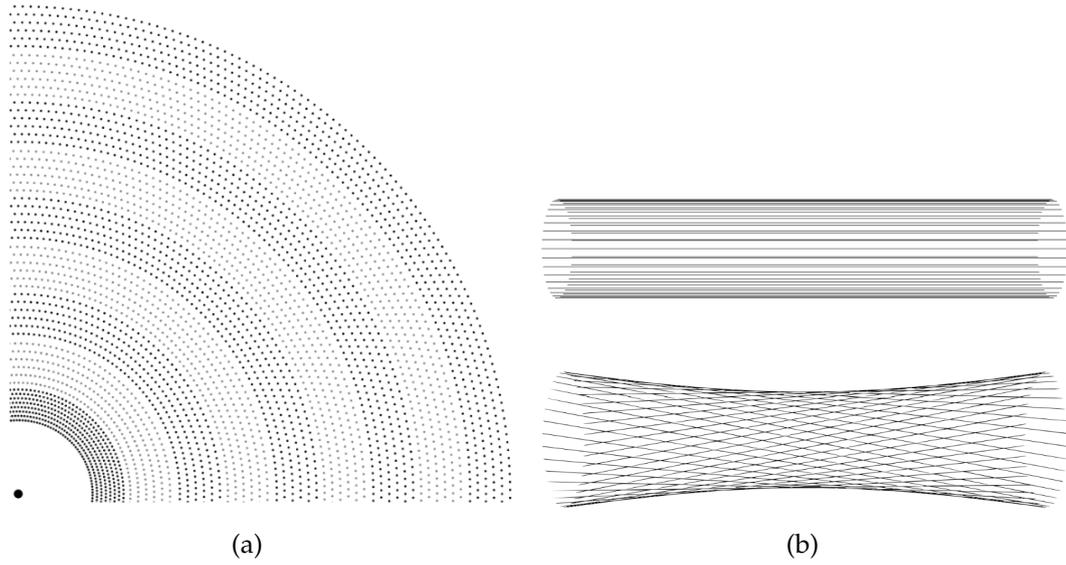
(a)   (b)

Figure 4.6: The Belle II central drift chamber schematic representation. In (a), a quadrant of the CDC in the r-ϕ plane is shown. Different axial and skewed layer groups are visible. In (b), the axial (upper) and skewed (lower) wires are shown. The skew is exaggerated for illustrative purposes. Credit to Ref. [114].

in the barrel region. Both detectors are located outside the CDC and are tasked with the distinction between particle species.

The ARICH detector [115] consists of an array of silica aerogel used as a radiator. As charged particles pass through the material at a speed greater than the phase velocity of light in that medium, they emit Cherenkov photons, which are detected by photon sensors. This working principle is depicted in Figure 4.7a. The angle of the emitted Cherenkov light, θ_C, can be used to calculate its velocity, β, given the refractive index of the radiator material, n:

$$\beta = \frac{1}{n \cdot \theta_C}. \tag{4.1}$$

The velocity information, combined with the knowledge of the particle's momentum from the CDC, SVD and PXD allows identifying the species of a particle through its mass. ARICH is designed to provide separation information for $\pi^\pm$ and $K^\pm$ in $(0.4, 4)$ GeV/c momentum range, and for $\pi^\pm, \mu^\pm, e^\pm$ below 1 GeV/c.

The TOP detector [116] consists of sixteen $270 \times 45 \times 2$ cm quartz radiator bars. The working principle is illustrated in Figure 4.7b. Due to the high refractive index of quartz, the Cherenkov light emitted by passing particles undergoes total internal reflection. One end of the quartz crystal contains a spherical mirror that reflects the light to the opposite end containing microchannel plate photomultiplier tube arrays. The photon time of arrival is the sum of the time of flight of the charged particle to the quartz radiator and the time of propagation in the crystal. Given a precisely known angle of the incoming particle, the time of propagation inside the chamber is a function of θ_C. The time of arrival of the photons is
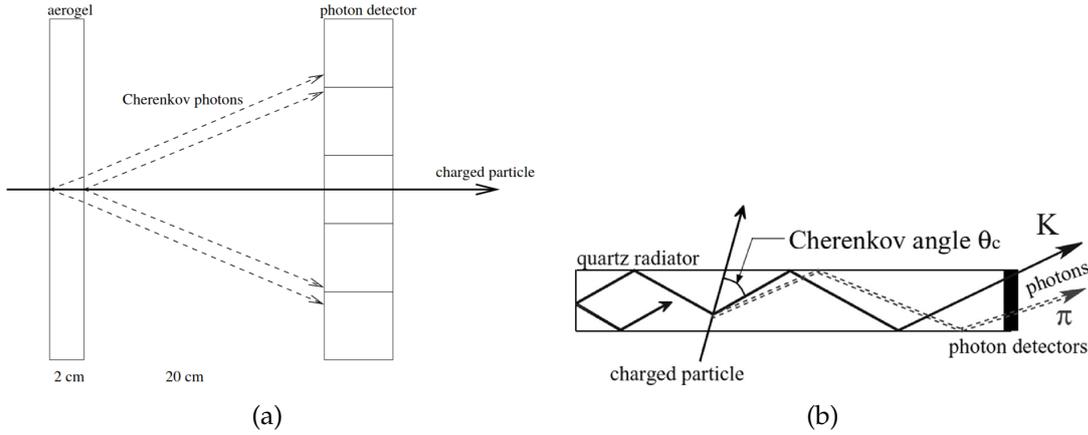
Figure 4.7:   The schematic working principle of the Belle II particle identification detectors: aerogel ring-imaging Cherenkov counter (a) and a time of propagation chamber (b). ARICH covers the forward endcap region, whereas TOP covers the barrel. Credit to Refs. [115] and [116], respectively.

compared to the expected distributions for different particle hypotheses ($e^{\pm}, \mu^{\pm}, \pi^{\pm}, K^{\pm}, p^{\pm}$) and corresponding likelihood values are computed for each (see Ref. [117] for details).

The TOP provides an 85% identification efficiency for $K^{\pm}$ at a 10% $\pi^{\pm}$ misidentification rate [118]. The ARICH has a 94% efficiency for $K^{\pm}$ identification with a $\pi^{\pm}$ misidentification rate of 11% [117].

4.2.5   Electromagnetic calorimeter

Belle II reuses the calorimeter of Belle, with upgraded readout electronics [107]. The electromagnetic calorimeter (ECL) surrounds the previously mentioned detector systems and covers the barrel and both endcap regions. It is the main subdetector for photon detection and their energy measurements. Furthermore, the ECL provides information necessary to differentiate electrons from hadrons, participates in K_L^0 detection together with the KLM (see Section 4.2.7), supplies triggering information, and measures the luminosity collected by the detector.

The ECL covers $\theta \in (12.4, 155.1)°$ and is composed of 8736 thalium-doped caesium iodide crystals [119]. Each crystal is approximately 16 radiation lengths long [120]. The rear surfaces of the crystals contain glued photodiodes with preamplifiers. As electromagnetically interacting particles pass through the calorimeter, they induce cascades of particles through interaction with the dense detector material, called *electromagnetic showers*. The photodiodes read out the scintillation light of the shower and convert it to a digital signal. The sketch of a single ECL crystal is given in Figure 4.8.

The ECL has excellent performance: a photon energy resolution which varies from 4 % at 100 MeV [119] to 2 % at 5 GeV [120], a position resolution of $5 - 10$ mm [119], and a mass-resolution of 5 MeV/c^2 (12 MeV/c^2) for the composite π^0 (η) meson [119].
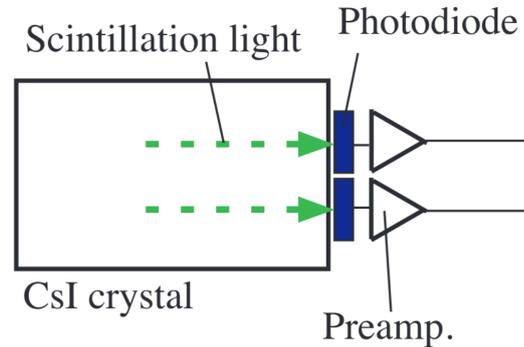
Figure 4.8: A schematic depiction of one of the crystals that comprise the Belle II electromagnetic calorimeter. A signal resulting from the shower of an electromagnetically interacting particle reaches the photodiode and is amplified by the preamplifier. Credit to Ref. [119].

4.2.6 Superconducting magnet

Surrounding the ECL [107], there is the superconducting solenoid. The coil is made from a niobium-titanium-copper alloy and is wound around an aluminium support cylinder. The cooling is performed using a liquid helium cryogenic system. It generates a 1.5 T magnetic field necessary to bend the trajectories of charged particles, enabling a transverse momentum measurement and charge separation. The magnetic field is directed along the z direction and was measured to be homogeneous and vary less than $O(1\%)$ in the entire volume [114].

4.2.7 K_L^0 and μ detector

The K_L^0 and μ detector (KLM) [121] is the outermost subsystem of Belle II. The KLM is composed of alternating layers of up to 4.7 cm thick iron plates and detector active parts. The iron plates decelerate the traversing particles and also act as a return yoke for the magnet.

The barrel and endcap regions differ by design [122]. The barrel region contains 14 iron layers and 15 detector layers, which are aligned parallel to the z axis. There, two innermost detector layers are instrumented with scintillator strips, whereas the remaining layers use resistive plate chambers. The endcap region contains 14 iron and detector layers each, which are aligned perpendicular to the z axis. Conversely, all 14 detector layers use plastic silicon strips with silicon photomultipliers.

The K_L^0 mesons interact with the nuclei in the iron plates and cause hadronic showers, which are read out by the silicon strip detectors or the resistive plate chambers. This process may already occur in the ECL, which is why it is also a part of K_L^0 detection. Minimum ionising particles, such as μ with momentum larger than 0.6 GeV/c, traverse the detector in a straight line, depositing only small amounts of energy in the system.

4.3 The Belle II software

The Belle II analysis software (`basf2`) [123] is an open-source software framework developed for collision event reconstruction, analysis and any other tasks necessary for the physical in-
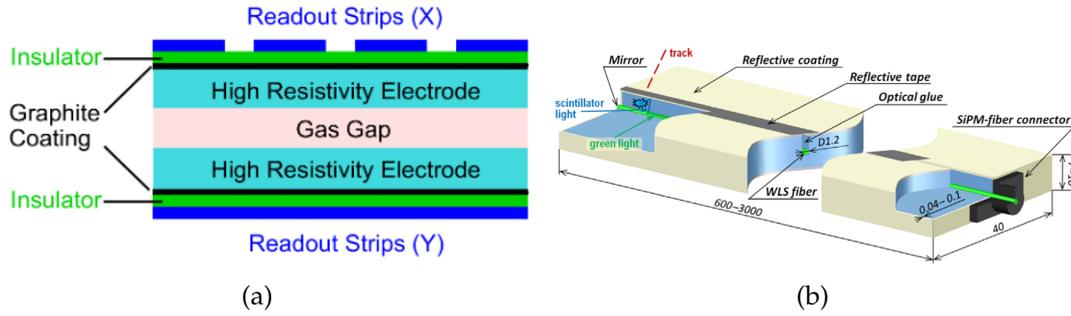
(a)   (b)

Figure 4.9: The Belle II K^0_L and μ detector active detector layers in the barrel (a) and the endcap regions (b). Iron plates are sandwiched between the detector layers. Credit to Refs. [122] and [121], respectively.

terpretation of the data recorded by the Belle II detector. The software is primarily based on Python and C++ programming languages. This Section briefly introduces the charged and neutral particle reconstruction strategies, which are implemented in basf2.

4.3.1 Charged particle reconstruction

Tracking refers to the charged particle trajectory reconstruction. The tracking process in Belle II is discussed broadly in Ref. [114]. Each particle trajectory is modelled by a helix with 5 parameters (track):

- d_0: the distance of the point of the closest approach to the z axis;

- ϕ_0: the angle between the transverse momentum and the x axis at the point of the closest approach;

- ω: the track curvature signed with the particle charge;

- z_0: the z coordinate at d_0;

- $\tan \lambda$: the tangent of the track dip angle (see Figure 4.10c).

These track parameters are visualised in Figure 4.10. The tracks are reconstructed by combining the information from CDC and/or SVD with information from PXD, if present.

4.3.2 Photon reconstruction

Particles showering in the ECL material deposit energy into the ECL crystals. The deposited energy and the time of each energy deposit are recorded by the calorimeter. Clusters are sets of energy deposits in the ECL that are associated with the interactions of a single particle. A graphical illustration of the cluster reconstruction in an experimental environment is shown in Figure 4.11.

The identification of photons exploits the fact that the energy deposited in the cluster by an incident photon has a cylindrical symmetry in the lateral direction with an exponentially decreasing energy deposition away from the incident axis. On the other hand, neutral
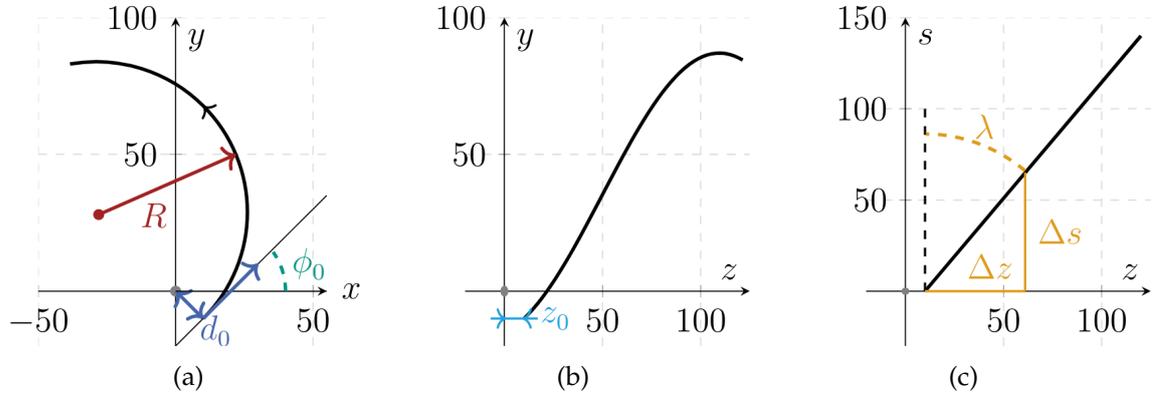
Figure 4.10: A schematic representation of a track in Belle II and the corresponding helix parameters that model it. All dimensions are in centimetres. In this context, s corresponds to the path length along the circular trajectory in the xy plane. The definitions of the helix parameters are given in Section 4.3.1. The Figures are taken from Ref. [114].
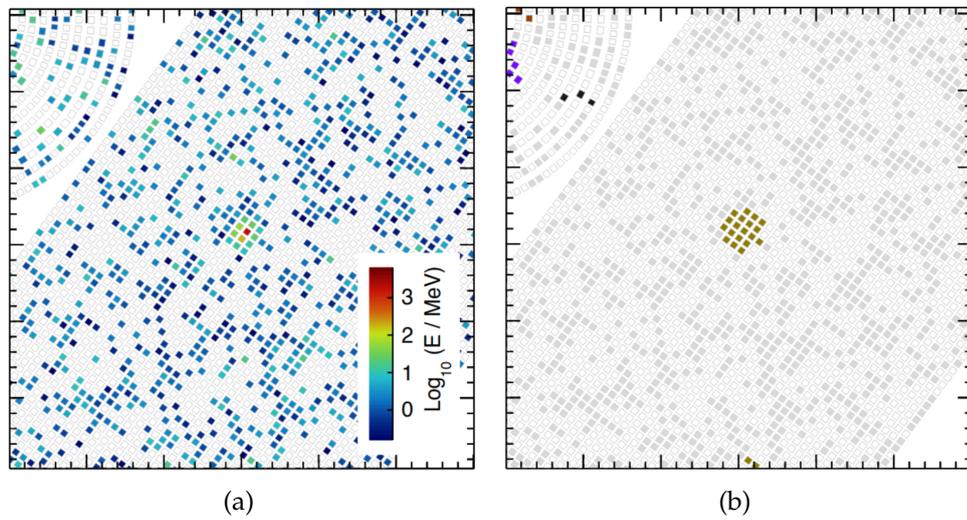
Figure 4.11: The Belle II calorimeter in the (θ, ϕ) plane is visualised, showcasing energy deposits in a single simulated photon event in the centre of the image. Each point corresponds to a single ECL crystal. The low-energy deposits resulting from beam background radiation are included in (a). The cluster reconstruction algorithm of basf2 singles out the cluster from the photon, rejecting background as seen in (b). Credit to Belle II Neutrals group.

or charged hadron interactions tend to produce less confined and more asymmetric shower shapes. To distinguish photons from charged particles (particularly electrons) tracks are extrapolated to the ECL and compared for consistency with reconstructed clusters.

Chapter 5

Data Analysis Techniques

5.1 Parameter estimation

The distributions of observables that are measured experimentally always contain a degree of statistical uncertainty. In physics, observables are often modelled with smooth distributions, which can be parametrised by a set of numbers. Parameter estimation, henceforth referred to as *fitting*, is a process of extraction of parameters from observed distributions. It is one of the key tasks of statistical analyses and, consequentially, particle physics measurements. Nearly every particle physics analysis features some kind of fitting for the result extraction. It is also present in particle trajectory reconstruction algorithms, calorimeter cluster shape parametrisation, calibration procedures, and others. One of the most common examples is the fitting of a particle species invariant mass distribution reconstructed from its decay products. It was, for example, used in some of the most precise of the Higgs boson mass measurements [124], [125], in the $H^0 \rightarrow ZZ^* \rightarrow 4\ell$ and $H^0 \rightarrow \gamma\gamma$ decay channels.

Fitting consists of two main steps: point estimation and uncertainty (confidence interval) estimation. In the former, the best estimate for a set of parameters is derived, which describes a given data set. The latter sets the confidence interval on each parameter estimate in the set.

This Section gives a brief overview of the most relevant method that was used in the work presented in this thesis. For a more in-depth consideration of statistical methods in data science and physics, the readers are referred to Refs. [126], [127] and for rigorous proofs of the underlying statistical concepts, Refs. [128]–[130]. The material presented here only summarises the details found in these books and articles.

5.1.1 Maximum-likelihood method

One of the most common and popular methods for parameter estimation is the maximum likelihood method. It is also used widely in this analysis, as described in Chapter 6. A measured data set can be defined as $x = (x_1, x_2, ..., x_N)$, with x_i being the i-th result of N independent measurement following an unknown probability density function (PDF), $f(x)$. The shape of $f(x)$ is often parametrised as $f(x; a)$, where $a = (a_1, ..., a_M)$ is an M-dimensional vector of unknown parameters. For a fitting procedure, one has to construct an *estimator*, which is a function of the observed data that can provide an estimated numerical value, $\hat{a}$,

corresponding to the parameter vector a.

A very common example is the maximum likelihood estimator, which uses the likelihood function. The likelihood function is built from one- or multidimensional PDF $f(x; a)$ of the measured values x:

$$\mathcal{L}(x; a) = \prod_{i=1}^{N} f(x_i; a). \tag{5.1}$$

In this case, the maximum likelihood estimate of the parameters a corresponds to $\hat{a}$ for which $\mathcal{L}(x; a)$ is globally maximised. Because the product of many components can vary over several orders of magnitude, in real-life applications it is more convenient to work with sums. Therefore, a log-likelihood function is used in practice:

$$l(x; a) \equiv \ln \mathcal{L}(x; a) = \sum_{i=1}^{N} \ln f(x_i; a). \tag{5.2}$$

Since a logarithm is a monotonic function, the maximum of a function is the same as the maximum of its logarithm. The maximum of the log-likelihood function satisfies the standard requirement for an extremum point:

$$\frac{\partial l(x; a)}{\partial a_j} = 0 \quad \text{for} \quad j = 1, ..., M. \tag{5.3}$$

The solutions of Equation (5.3) are maximum likelihood estimates of $\hat{a}$. Generally, it is impossible to find an extremum in a large parameter space using analytical methods. In practice, numerical procedures and dedicated software packages for optimisation are usually used. Many optimisers used in modern-day computers tend to minimise functions, rather than maximise them, therefore a *negative log-likelihood* function is often used. The maximum-likelihood method is unbiased and consistent as the number of measurements grows, i.e. $N \to \infty$. However, it requires a good assumption of the form of the PDF $f(x; a)$ that the data follows. If the assumed distribution is incorrect, the results can be unreliable.

5.1.2   Variance of the maximum likelihood method

Assuming Equation (5.3) is satisfied, the log-likelihood function can be expanded as a series at $a = \hat{a}$ and approximated as a parabola:

$$l(x; a_1, a_2, ..., a_N) = l(x; \hat{a}_1, \hat{a}_2, ..., \hat{a}_N) + \frac{1}{2} \sum_{i,k} \frac{\partial^2 l}{\partial a_i \partial a_k} (a_k - \hat{a}_k) + ... \tag{5.4}$$

The likelihood function generally has any non-Gaussian shape, however, it can be shown [129] that in the asymptotic limit, $N \to \infty$, any given $f(x; a)$ can have its likelihood function approximated with a multivariate Gaussian distribution:

$$\mathcal{L} \propto \exp\left\{ -\frac{1}{2} (a - \hat{a})^T H (a - \hat{a}) \right\}, \tag{5.5}$$

where H is the Hessian matrix of the log-likelihood function.

In this case, the covariance matrix, $V(\hat{a})$, of the estimated parameter vector is approximated as the inverted Hessian matrix taken at the maximum likelihood estimate $\hat{a}$:

$$V(\hat{a}) = \left[-\frac{\partial^2 l(x;a)}{\partial a^2}\bigg|_{a=\hat{a}} \right]^{-1} = H^{-1}. \tag{5.6}$$

This gives a symmetric uncertainty for each estimated parameter a_j as:

$$\hat{\sigma}_{a_j} = \sqrt{\hat{V}_{jj}(\hat{a})}. \tag{5.7}$$

This method is always an approximation of the true covariance matrix, because the likelihood function shape is approximated as a parabola.

A different approach is to profile the likelihood function in order to calculate a likelihood-based confidence interval. A likelihood ratio for a single parameter, a_k, and a set a (excluding a_k) is defined as:

$$\lambda_\mu(a_k) = -2 \cdot \ln\left(\frac{\mathcal{L}(x;a_k,\hat{\hat{a}})}{\mathcal{L}(x;\hat{a}_k,\hat{a})} \right), \tag{5.8}$$

where the numerator is the log-likelihood for a set of parameters $\hat{\hat{a}}$ estimated for some given value of a_k. The denominator with optimal values $\hat{a}$ and $\hat{a}_k$ is the likelihood evaluated at its global extremum. Given that $\lambda_\mu(a_k)$ is distributed according to a χ^2 distribution (Wilks' theorem) [131], one finds the minimum and maximum value of a_k that correspond to $\lambda_\mu(a_k) < \chi^2_{1,1-\alpha}$, where $1 - \alpha$ is the desired confidence interval. For example, for a 68% confidence interval (one standard deviation), $\lambda_\mu(a_k) \lesssim 1$.

In general, this leads to an asymmetric uncertainty interval with respect to the central extremum value $\hat{a}_k$. Note that in the asymptotic limit, both the profiling and Hessian matrix inversion methods provide the same results. However, reevaluating the likelihood for every given value of a_k requires minimisation of all other parameters, therefore the profiling method can become computationally intensive. It is often sufficient to apply the Hessian matrix inversion method, as long as additional evaluation is performed to ensure that the provided uncertainties are accurate.

In particle physics, a commonly used minimisation software is `Minuit` [132], [133]. It implements the Hessian matrix inversion as the `HESSE` method and the likelihood-based uncertainty estimation method as `MINOS`.

5.1.3 Extended maximum likelihood method

In particle physics, it is common to not only parametrise the shape of an observable's distribution but also to measure its absolute rate (normalisation). The standard setup for the maximum likelihood method does not allow determining the absolute normalisation. An additional term to the likelihood has to be introduced. In nature, if a measurement is performed repeatedly, its rate will fluctuate according to Poissonian statistics. Hence, the Equation (5.1), needs to have a Poisson term included:

$$\mathcal{L}(x;a) = \frac{\nu^N}{N!} \exp(-\nu) \prod_{i=1}^{N} f(x_i;a), \tag{5.9}$$

where N is the observed number of events and ν is the expected, or 'true', normalisation. Such a modified likelihood is called an *extended* likelihood. Taking the logarithm of Equation (5.9) gives:

$$l(x; a) = -\nu + N \ln \nu + \sum_{i=1}^{N} \ln(f(x; a)) + C,\qquad(5.10)$$

where C is independent of a and ν. The extended log-likelihood fitting otherwise follows the same procedure as Sections 5.1.1 and 5.1.2. As such, when optimising Equation (5.10) for a and ν, the constant parameter C can be ignored.

5.1.4   Unbinned maximum likelihood fitting

There are two ways that data can be arranged for a fit:

- Each event enters the likelihood function (Equation (5.3) or Equation (5.9)) independently,

- Events are first grouped in intervals of the observable x and the count of measurements falling into each interval are provided as inputs to the likelihood function.

The intervals are often referred to as *bins*. Consequentially, the techniques are referred to as *unbinned* and *binned* fits, respectively. In this thesis, unbinned maximum likelihood fits are used. They are computationally more intensive but are a more general than binned fits.

Throughout this thesis, fitting is implemented using the `zfit` framework [134]. It provides a Python-based library that is developed to fulfil particle physics fitting requirements. The `zfit` framework implements minimisation using the `iminuit` minimiser [135] which is a Python-friendly implementation of `Minuit`.

5.2   Classification

When collecting data in measurements, many independent observations (*events*) are usually performed in order to get a statistically significant sample. Resulting data sets generally contain a component of interest, often referred to as *signal*, and multiple other components which might show similar behaviour as the signal in certain distributions. The latter is commonly referred to as *background*. Disentangling the signal and background contributions in a given data set is an extremely important task in Big Data fields, where there may be thousands of subcomponents contributing to background that may be misclassified as signal.

In particle physics, the signal component usually refers to a single or a group of decay channels, whose properties are being measured. The most straightforward approach is to separate signal and background events by imposing requirements on observables that are typical or expected for signal. A requirement on some observable is often referred to as a *selection*. The downside to this is that there can be non-linear underlying correlations between different selections, which might make them less efficient at background process separation than they could be in conjunction with other observables. Therefore, a multidimensional observable space is often desired. However, as the number of observables used in the selection grows, the tuning of such a multidimensional selection becomes increasingly difficult.

This Section introduces some relevant techniques to combine the information of many observables of an event. Particularly, it overviews multivariate classification algorithm (MVA) concepts and boosted decision trees (BDTs). The material presented here is only a summary detailing techniques used in the analyses presented in the thesis. For a detailed overview of multivariate classification techniques, the reader is referred to Ref. [126], and for their underlying statistical framework to [136], [137].

5.2.1 Multivariate classification

In this Subsection, binary classification is implied, as this is the most relevant type for the work presented in this thesis. In general, a multi-class MVA classifier can be broken down into a series of binary MVA classifiers.

Let us assume an event is described by N observables, $X = \{x_1, x_2, ..., x_N\}$. This is also known as a *feature vector*. Formally, an MVA classifier is a mapping function, f, that maps the N-dimenstional feature vector, x, to a single real number, y:

$$\mathbb{R}^N \rightarrow \mathbb{R} : y = f(x). \tag{5.11}$$

This can then be interpreted as a hypersurface in an N-dimensional feature space.

The goal of multivariate classification is to build a classifier based on a set of pre-arranged examples $\{(\vec{x}_1, y_1), ..., (\vec{x}_J, y_J)\}$, where J is the number of events, and $y_j \in \{0, 1\}$ is an associated *target*. The building of such a classifier is called *training*, and the set of examples is called a *training sample*. In general, it is also possible to perform a training where no targets are supplied. These types of methods are called *unsupervised*. The methods used for data analysis in this thesis are *supervised* and presented in this Section.

Mathematically, a *loss function* is desired which returns the discrepancy between training targets, y and the prediction, $\hat{y}$, corresponding to the input data, $\vec{x}$:

$$\mathcal{L}(y, \hat{y}) = \mathcal{L}(y, f(\vec{x}; \vec{a})), \tag{5.12}$$

where $\vec{a}$ is a set of internal degrees of freedom describing the classification model. In classification problems, outputs of $f(\vec{x}; \vec{a})$ are conventionally mapped to a continuous output between 0 and 1 to represent corresponding categories. In this formalism, one can minimise the loss to extract a set of parameters that provide the smallest difference between y and $\hat{y}$. This problem is then equivalent to the one described in Section 5.1.1, with a different choice of a likelihood function.

Classifiers with a smaller number of degrees of freedom tend to be more robust against statistical fluctuations of a sample (smaller variance). However, this might make the model unable to learn the more intricate details in the training sample, and lead to bias when applied to a statistically-independent sample. The *bias-variance trade-off* is the balance between a model's ability to fit the training sample well and its ability to generalise to new unseen data. Usually, one uses a *validation* sample, which is equivalent but statistically independent of the training sample. An optimally-trained classifier has a similar performance on both of the samples. After the optimal training point of the classifier is reached, further training will only degrade the performance of the classifier on the validation sample. This regime is called *overtraining* and needs to be avoided. There are many ways to test for overtraining, but all

of them usually rely on comparing the classifier performance between the training sample and an additional, independent *testing sample*. In particle physics, it is common to use the validation sample also as a testing sample. This is only acceptable in the case of large training samples, where statistical fluctuations of data are negligible or where independent efficiency and background rejection studies are performed.

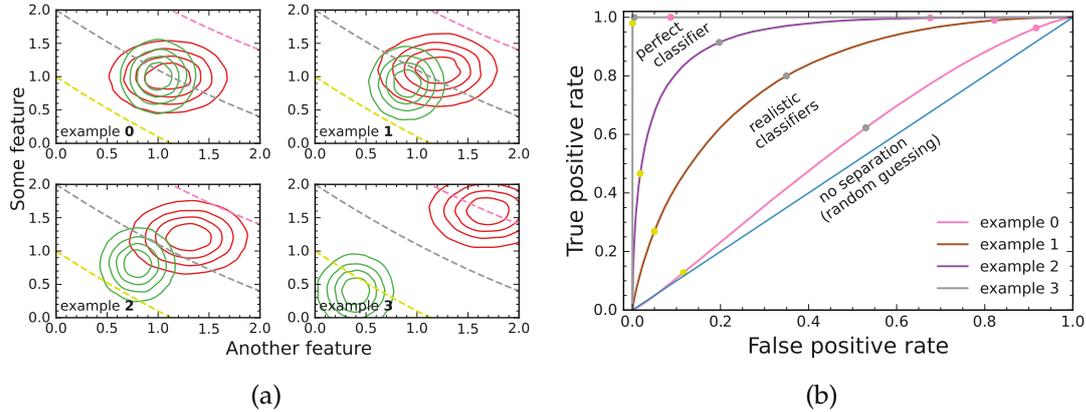(a)   (b)

Figure 5.1:   Two-dimensional distributions (red/green) in an example feature space are shown in (a) for four different example data sets (0-3). Each distribution is represented by a contour plot of 20% isoproportion lines. Dashed lines show the different selection thresholds for an example MVA model. In (b), the corresponding ROC curves are shown, calculated for examples 0-3. The working points that represent the dashed lines of (a) are shown in the same colour, for each of the four curves. A perfect classifier unambiguously separates green and red distributions, similar to example 3. On the other hand, a poor classifier would only be able to perform marginally better than random guessing, similar to example 0.

To assess the performance of a binary classifier, a receiver operating characteristic (ROC) is often used. A ROC curve represents the true positive rate as a function of the false positive rate and is calculated by iterating over different MVA classifier thresholds. A visual representation of the meaning of ROC curves is shown in Figure 5.1. Given an optimal classifier (no overtraining), one expects the ROC curve calculated on training and testing data to be similar. This can be qualitatively viewed by inspecting the corresponding curves, or, quantitatively, by defining an area-under-curve (AUC) score which is computed by integrating the ROC curve. An AUC = 1 would represent a false positive rate of 0, and a true positive rate of 1 (see Figure 5.1b). The least-optimal classifier would provide a random guess and have a false positive and true positive rate of 0.5, and AUC = 0.5.

In high-energy physics, alternative forms of false positive and true positive rates are popular. Namely, results are often expressed in terms of:

- *efficiency*: equivalent to the true positive rate;

- *background rejection*: equivalent to 1 − false positive rate;

- *purity*: fraction of signal with respect to the total sample size (signal+background); equivalent to positive predictive value.

5.2.2 (Boosted) decision trees

When combining many features, selecting an appropriately parametrised model (Equation (5.12)) can be complicated and suboptimal. Therefore, non-parametric models are often desired, and a simple way to implement them is by using *decision trees*.

A decision tree is a classification technique where the N-dimensional feature space is divided based on a series of binary selections. A simple model divides the feature space into two regions, where the division depends on the value of a single feature with respect to a certain threshold. The splitting process is then repeated in the two resulting regions. The exact value to split by is optimised in each region independently by evaluating a loss function. For classification problems cross-entropy is commonly used:

$$\mathcal{L}(y, \hat{y}) = -\sum_{j=1}^{J} y_j \log \hat{y}_j + (1 - y_j) \log(1 - \hat{y}_j), \tag{5.13}$$

where y is a target and $\hat{y}$ is a predicted label. The value of $\mathcal{L}(y, \hat{y}) \to 0$ as the fraction of correctly predicted values grows. Each final region which has not been subdivided is called a *leaf*. A set of selections leading to a leaf is known as a *branch*. Therefore, at the end of every branch, there is a node with two leaves. The maximum number of selections in a branch is called the *depth* of the classifier.

The tree could be grown indefinitely, and this will quickly lead to overtraining, highlighting the aforementioned bias-variance trade-off. In practice, a technique called *pruning* is employed which removes nodes from the tree based on their overall impact on the performance of the decision tree. This produces a smaller tree and the 'strength' of pruning is tunable with appropriate parameters.

Decision trees are easily interpretable as they are a series of binary decisions that are intuitive and closely resemble human reasoning. However, they also tend to be highly dependent on the statistical fluctuations of the training data set. With pruning, their performance on independent samples usually turns out to be weaker than many other MVAs, and the misclassification rate is only slightly better than random guesses. A single decision tree is, therefore, considered a *weak classifier*.

This and other problems surrounding decision trees are addressed by employing *boosting* and forming boosted decision trees (BDTs). Boosting enhances the performance of an ensemble of weak classifiers by combining their outputs. The performance of such a boosted ensemble can be much better even when all the input classifiers are weak.

During the first training iteration of a BDT, a regular decision tree would be trained, as explained before, with a shallow limit on the depth. Next, there would follow a second training iteration, where the events that were misclassified in the last training iteration are given a larger weight. The procedure is repeated M times to train a set of classifiers $g_m(x)$, where $m = 1, 2, .., M$. The final output is a weighted 'majority vote' of all these classifiers:

$$G(x, \alpha_m) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m g_m(x)\right), \tag{5.14}$$

with α_m being the weight of the m-th weak classifier. Conventionally, the output of the classifier is transformed, such that $G(x) \to G'(x) \in [0, 1]$.

The exact procedure to assign weights to misclassified events between training iterations and the computation of α_m depends on the boosting technique used. Consider an initial guess G_1, which corresponds to $M = 1$ case in Equation (5.14). One can then, given some learning rate (*shrinkage*) $0 < \nu \leq 1$, update the prediction with every training iteration:

$$G_m(x) = G_{m-1}(x) + \nu \alpha_m g_m(x), \tag{5.15}$$

by minimising the loss function for parameter α_m:

$$\alpha_m = \arg \min_{\alpha_m} \left(\sum_{i=1}^{N} \mathcal{L}(y_i, G_{m-1}(x_i) + \alpha_m g_m(x_i)) \right). \tag{5.16}$$

One way to solve Equation (5.16) in a numerically-optimal way is to compute the gradient of the loss function. Consequentially, a classifier built this way is called a *gradient boosted decision tree* [138].

To further increase the robustness of gradient-BDTs, one may only sample (without replacement) a part of the data set in each training iteration. In this case, any statistical fluctuations would average out over the sum of all trees. This technique is called stochastic gradient-BDTs [139]. The fraction of data used at every training iteration is called the *sampling rate*. BDTs are often considered as 'black box' models because their inner structure is not easily interpretable, making it difficult to understand how they arrive at their predictions. This can be seen as a disadvantage because the model's transparency is limited.

In the analyses present in this thesis, `FastBDT` [140] is used as the MVA for training stochastic gradient-BDTs. For the training of the classifier, `FastBDT` employs 4 parameters that were described in this Section: number of trees, maximum depth of each tree, shrinkage and sampling rate.

5.3   Unfolding

In physics, we are only able to probe nature through the experimental setup that we build. Any hypothetical detector is therefore subjected to random effects related to its finite resolution and acceptance. Consequentially, the measured results (distributions) will be convoluted with these effects (*smeared*). If two different experiments are comparing their measurements, it only makes sense if external factors are of minimal importance. The process of disentangling the true distribution from the measured distribution is called *unfolding*. In this way, it can be considered the *opposite* of measurement.

This Section provides a brief overview of the main concepts that feature in the results presenting this thesis. A thorough summary of the theory behind unfolding methods can be found in Refs. [126], [127] and a comparison of different methodologies common in particle physics in Refs. [141]–[143].

5.3.1   Mathematical basis

Consider a true distribution $f(x)$ of observable x being measured in some space $a < x < b$. The corresponding measured distribution $g(y)$ as a function of the measured value y is re-

lated to $f(x)$ as:

$$g(y) = \int_a^b A(y, x)f(x)dx, \tag{5.17}$$

which is the Fredholm integral equation of the first kind [127]. The function $A(y, x)$ is the Kernel (or response) function, and it holds the information about the physical nature of the measurement. In the process of unfolding, one seeks to find $f(x)$ given $g(y)$. Such a problem does not have a general solution and in the cases when there is a solution it can be highly dependent on small changes in $g(y)$ [144]. Figure 5.2 visually illustrates the smearing of a true distribution for an arbitrary observable given two smearing effects. These effects, which would contribute to the form of the $A(y, x)$, are not always accurately known in reality as they include resolution effects, interactions with detector material, impurities in experimental setup etc. The main method in high-energy physics is to extract an implicit form of this function from simulated samples.
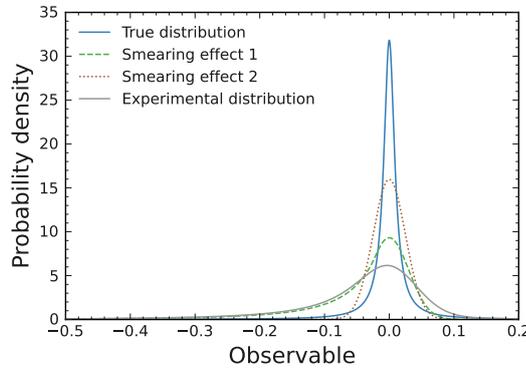

Figure 5.2: A visual illustration of the smearing of an arbitrary observable. Considering some 'true' distribution (blue, full line), depicted by a sharp Lorentzian, the experimentally observed result (grey, full line) will be affected by, usually, unknown detector response and resolution effects. In this case, the unknown experimental effects are illustrated by Gaussian (dotted) and Crystal Ball (dashed) functions.

In science, experimental data are often presented in the form of a histogram. The results are integrated over several shorter intervals (bins) and only the integral values are shown at the centre of each bin. As a result, a simpler matrix form of Equation (5.17) can be used which allows a numerical solution:

$$\hat{A}\vec{x} = \vec{y}, \tag{5.18}$$

where:

- $\vec{x}$ is an m-dimensional vector of a binned unknown true distribution;

- $\vec{y}$ is an n-dimensional vector of a binned measured distribution;

- $\hat{A}$ is an $n \times m$-dimensional response matrix.

In general, Equations (5.17) and (5.18) may contain additional terms dependent on y corresponding to background contributions and statistical fluctuations.

The elements A_{ij} of $\hat{A}$ correspond to a probability to measure an event in bin j that was produced in bin i. In particle physics, the response matrix is generally built using Monte Carlo techniques by using simulated particle collision samples and modelling the detector response. Consider the number of simulated events generated in a bin i, N_i^{gen}. In general, some of these events may be measured in a different bin j, and the number of such events is denoted as N_{ij}^{meas}. Then one builds the response matrix as:

$$A_{ij} = \frac{N_{ij}^{\text{meas}}}{N_j^{\text{gen}}}. \tag{5.19}$$

There are several methods to solve Equation (5.18), discussed in the following Sections.

5.3.2   Bin-by-bin unfolding

The simplest approach is to scale the measured distribution in each bin i by a factor determined from simulated samples:

$$x_i = y_i \cdot \frac{N_i^{\text{gen}}}{N_i^{\text{meas}}}, \tag{5.20}$$

where N_i^{meas} is $\sum_j N_{ij}^{\text{meas}}$. It follows that uncertainties propagate trivially:

$$\Delta x_i = \Delta y_i \cdot \frac{N_i^{\text{gen}}}{N_i^{\text{meas}}}, \tag{5.21}$$

where Δx_i and Δy_i are uncertainties associated with these quantities.

This method is only possible if the number of bins in the true distribution is equal to the number of bins in the measured distribution. Furthermore, it does not take any cross-bin effects into account, i.e. effectively treats each bin as uncorrelated. Therefore, it is usually used when the necessary corrections are small or when statistical uncertainty dominates the experimental result. The method is referred to as *bin-by-bin* unfolding, or bin-by-bin correction method. It has been the main method applied in the works of this thesis (see Section 6.13.6).

5.3.3   Matrix inversion method

The matrix inversion method is the simplest method which does not include assumptions about the measured data (e.g. uncorrelated bins). In this case the Equation (5.18) is solved by inverting the response matrix:

$$\vec{x} = \hat{A}^{-1}\vec{y}. \tag{5.22}$$

The number of bins of the true and the measured distribution must be the same, as only square matrices may be inverted. While the result is statistically fully correct, it tends to introduce large negative correlations between neighbouring bins in the unfolded distribution $\vec{x}$ [141]. The simple reasoning for this is the fact that statistical fluctuations in neighbouring bins will get amplified/suppressed by the matrix multiplication, creating an "oscillatory" pattern. This effect is undesirable and goes against the common sense that (generally) smooth distributions are expected in nature. The correction of such correlations is called *regularisation*[1].

---

[1]Note that regularisation in the context of machine learning usually refers to techniques that prevent overfitting. The term used here refers to the unfolding regularisation of inverse problems.

5.3.4 Other unfolding methods and regularisation

As discussed in the last Subsection, unfolding introduces large correlations, leading to 'non-physical' shapes of unfolded distributions. Regularisation is a delicate task because it arises from a conscious prejudice towards smooth measured distributions and the artificial introduction of a bias to smoothen them. Unsurprisingly, there exist many methods to perform regularisation. Although an in-depth discussion is irrelevant to the work presented in this thesis, some of the most common ones in particle physics are listed here:

- `TUnfold` method [145]: performs a least-squares minimisation of $\vec{y} - \hat{A}\vec{x}$ and includes a damping term for oscillations in x;

- D'Agostini method [146], [147]: inversion of response matrix using Bayes' theorem and iterative improvement of the unfolding result, stopping iterations early to not require regularisation;

- Iterative dynamically-stabilised method [148]: combines elements of d'Agostini's and bin-by-bin correction methods, preserving the normalisation of data in each bin;

- Singular value decomposition method [149]: solves Equation (5.18) by decomposing the response matrix into its singular values and adds a weighted prior condition to the solution.

In particle physics, unfolding is often implemented using the `RooUnfold` package [143], which can perform all the methods described in Section 5.3. In the analysis presented in this thesis, regularisation was not applied following the results of studies in Section 6.13.6.

5.4 Blinded analysis

In order to avoid an experimenter's bias, analyses are often performed in a *blinded* way [150]. Such a bias can occur when using the analysed data directly to perform optimisations or compare several techniques. An experimenter may enhance statistical or unknown systematic effects (even unintentionally) which bias the sample to give a desired or expected result.

Generally, one restricts access to the analysed sample until after the measurement setup is finalised. In particle physics, this is done by first performing the analysis on simulated data, and next using independent data samples to ensure that appropriate test metrics are matched. Normally, a certain range of an interesting observable is hidden, for example, in the case of $B \rightarrow X_s \gamma$ analysis presented here, it can be the photon energy, E_γ. This method is well-suited when performing measurements where the signal region is known in advance, as is the case in this thesis. The removal of a constraint and the application of the analysis technique on real data is called *unblinding*. It is usually performed as the last step, once other steps in the analysis have been scrutinised and validated. This is also the procedure that was applied in the inclusive $B \rightarrow X_s \gamma$ studies described in this thesis.

Chapter 6

Measurement of $B \to X_s \gamma$ with hadronic-tagging

So far the thesis laid out the theoretical foundation which motivates the study of $B \to X_s \gamma$ and experimental techniques necessary to perform this measurement. In this Chapter, the two will be connected, describing the main topic of the doctoral thesis: a measurement of the photon energy spectrum of $B \to X_s \gamma$ decays using the hadronic-tagging technique with the Belle II data. This measurement is the first of such kind in Belle II and, generally, since the previously discussed BaBar result [88]. It sets up the experimental procedure for the Belle II experiments following this analysis technique in the future.

    The measurement is performed in several (interconnected) steps, which are broken down in Section 6.1. The implication of the results and the outlook for the analysis will be discussed in the following Chapter 7.

6.1   Analysis strategy

This Section serves as an overview of the analysis and is intended to guide the reader through different steps in extracting the $B \to X_s \gamma$ photon energy spectrum. It also introduces key terminology that is used throughout the entire text. In this entire Chapter, superscripts are used to denote observables in a particular frame of reference. Given an observable $\mathcal{X}$ (laboratory frame), the same value in the frame of colliding electrons (centre-of-mass) is denoted as $\mathcal{X}^*$. Observables in the rest frame of the decaying B meson adopt the notation $\mathcal{X}^B$.

    The final goal of the analysis is to measure the partial branching fractions (see Equation (2.3)) as a function of E_γ^B (i.e. the photon energy spectrum in the signal B meson rest frame):

$$\frac{d\mathcal{B}(B \to X_s \gamma)}{dE_\gamma^B}\bigg|_i = \mathcal{U}_i \cdot \frac{N_i(B \to X_s \gamma)}{\varepsilon_i N_B}, \tag{6.1}$$

where i is a given E_γ^B interval, $N_i(B \to X_s \gamma)$ is the number of B mesons measured in the the interval i, ε_i is the average efficiency for selection and reconstruction of $B \to X_s \gamma$ decays in the interval i, N_B is the total number of B mesons in the analysed sample, and $\mathcal{U}_i$ are unfolding correction factors (bin-by-bin unfolding is implied here). The results of the integrated

branching fraction, $\mathcal{B}(B \rightarrow X_s\gamma)$, are evaluated by performing a sum over partial branching fractions computed in every interval i.

The analysed data sets are introduced in Section 6.2. A simulated data set, based on the expectations of the Standard Model, is used to prepare the analysis procedure. Only after the full analysis procedure is set up, the results on the Belle II data will be unblinded (see Section 5.4).

Section 6.3 introduces the reconstruction of Belle II experimental and simulated $B \rightarrow X_s\gamma$ samples. It also discusses the main processes which mimic $B \rightarrow X_s\gamma$ signal (backgrounds) and the strategies to suppress them. The background sources for $B \rightarrow X_s\gamma$ can be divided as:

- *Signal-side background*: where a photon candidate is originating in a non-$B \rightarrow X_s\gamma$ decay. In particular, either $e^+e^- \rightarrow q\bar{q}$ ($q \in \{u, d, s, c\}$), henceforth – *continuum*, or from a different B decay, e.g. where a $\pi^0 \rightarrow \gamma\gamma$ is created in the decay chain, henceforth – *generic B* background;

- *Tag-side background*: where the B, recoiling against the candidate $B \rightarrow X_s\gamma$, is reconstructed incorrectly, or $e^+e^- \rightarrow q\bar{q}$ collision products are combined to resemble a B decay. Such decays are referred to as *combinatorial B* or continuum background, respectively;

- $B \rightarrow X_d\gamma$ component: which is an irreducible background at the current analysis setup.

The multivariate optimisation strategies for signal and tag-B meson background suppression are described in Sections 6.4 to 6.7. They rely on selections to suppress the signal-side background and a dedicated BDT training which aims to suppress tag-side backgrounds.

After the full background suppression procedure, a thorough setup of the fitting model is described in Section 6.8. The fitting procedure is aimed at removing the combinatorial tag-B meson backgrounds, and hence the fit of the M_{bc} variable is performed (Equation (3.4)).

Lastly, an irreducible signal-side background component remains, particularly from generic signal-B meson decays. The E_γ^B spectrum is extracted by the simulation-dependant subtraction of remaining background processes as described in Section 6.9.

Section 6.10 explains the validation procedure of the analysis technique on Belle II simulated samples, whereas Sections 6.11 to 6.13 explore that on the Belle II recorded data samples. Section 6.14 condenses the observations from the validation and quantitively assigns systematic uncertainties. The unblinding and the final extraction of results and unfolding of Belle II data is presented in Section 6.15.

6.2 Data samples

6.2.1 Experimental data sets

This measurement uses data sets of e^+e^- collisions produced by the SuperKEKB accelerator and collected by the Belle II detector in 2019-2021. There are two data-collection modes in Belle II:

- *on-resonance data*: data sets collected at the collision energy $\sqrt{s} \approx 10.58$ GeV, corresponding to the mass of $\Upsilon(4S)$ meson;

- *off-resonance data*: data sets collected 60 MeV below the on-resonance threshold. Such data, by definition, does not contain $B\bar{B}$ events and is an excellent testing and validation sample to understand continuum processes.

The integrated luminosity, henceforth denoted as $\int\mathcal{L}$, corresponding to the on(off)-resonance sample is 189 fb^{-1} (18 fb^{-1}). The on-resonance data set contains approximately 198 million $B\bar{B}$ pairs.

6.2.2 Simulated data sets

In order to prepare the analysis procedure, calculate the signal-selection efficiencies and perform a validation adhering to the principles of a blinded analysis, large simulated samples are utilised. These samples are significantly larger than the data sets that were anticipated to be analysed by this analysis, to ensure that uncertainties due to limited simulation samples are minimal. The overview of the samples is discussed in this Subsection, but a quick overview is provided in Table 6.1.

Table 6.1: The overview of simulated samples used in the measurement described by this thesis. More in-depth discussion for each sample is present in the text.

Simulated sample	Size	Generators used
Generic-B^0 Generic-B^+	1+0.6 ab^{-1}	EvtGen [64]
continuum $u\bar{u}$ continuum $d\bar{d}$ continuum $c\bar{c}$ continuum $s\bar{s}$		KKMC [151] interfaced to Pythia 8 [152]
$B^+ \to X_s\gamma$ $B^0 \to X_s\gamma$	100 million	EvtGen, BTOXSGAMMA model [64]
$B^+ \to K^{*+}(892)\gamma$ $B^0 \to K^{*0}(892)\gamma$	10 million	EvtGen, SVP_HELAMP model [64]

All the simulated samples correspond to the official Belle II simulation production campaign and are based on Monte Carlo methods. Therefore, henceforth, the simulated samples are referred to as MC. In all cases, the detector response and readout are simulated using Geant 4 [153]. The following samples are used for the analysis:

- four $e^+e^- \to q\bar{q}$ ($q \in \{u, d, s, c\}$) simulated sample categories, referred to as continuum MC,

- two $\Upsilon(4S) \to B\bar{B}$ categories for charged and neutral B modes, referred to as generic-B MC.

Altogether, the above two categories are referred to as *generic* MC. Normally, $e^+e^- \to \tau^+\tau^-$ events are also considered as background, however, when using hadronic-tagging this type of background is observed to be suppressed to negligible levels.

The analysis is set up on 1.6 ab^{-1} of MC, which is more than 6 times larger than the on-resonance data set for this analysis. For the background subtraction step described in Section 6.9, which has the strongest dependence on limited-MC sample size, a significantly larger data set is crucial.

The generic-B MC includes $B \to X_s\gamma$ decays, however, the number of such events is expected to be small, further diminished by the efficiency of the hadronic-tagged analysis procedure, as discussed in Section 3.2.3. For this reason, additional samples are used:

- 100 million B^+B^- samples, where at least one B is guaranteed to decay as $B^+ \to X_s\gamma$ based on the Kagan-Neubert model [63] (see Section 2.5);

- 100 million $B^0\bar{B}^0$ samples, where at least one B is guaranteed to decay as $B^0 \to X_s\gamma$ based on the Kagan-Neubert model [63] (see Section 2.5);

- 10 million B^+B^- samples, where at least one B is guaranteed to decay as $B^\pm \to K^*(892)^\pm\gamma$ based on the `EvtGen SVP_HELAMP` model [64];

- 10 million $B^0\bar{B}^0$ samples, where at least one B is guaranteed to decay as $B^0 \to K^*(892)^\pm\gamma$ based on the `EvtGen HELAMP` model [64].

The first two are used for general analysis setup, reported in Sections 6.3 to 6.7, and related studies. The $B \to K^*\gamma$ samples is combined with $B \to X_s\gamma$ in a hybrid model approach [154] discussed in Section 6.2.3.

6.2.3   Signal model

As it was introduced in Section 2.5, inclusive decay models, by design, assume quark-hadron duality and therefore do not describe specific resonances but a smooth spectrum. For $B \to X_s\gamma$ this is an excellent approximation given the experimental resolution, as well as the fact that many different decay resonances combine and interfere.

A model based on all measured resonances (Table 3.1) that are included in the Belle II official MC for generic-B is shown in Figure 6.1. The generic-B model includes resonances based on their branching fractions given in Table 3.1 and scales down the Kagan-Neubert model globally to account for the inclusion of these. There are several resonances included (those that have been observed) in the generic-B model, but their sharp structure is largely smoothed-out, even without experimental smearing. On the other hand, the $B \to K^*(892)\gamma$ is isolated and sharply peaking. Although this model proves a slightly better explanation of the high-E_γ^B region, it fails at the tail region: there the inclusive model gets scaled down significantly lower than expected. This is because a global scaling is applied to $B \to X_s\gamma$, whereas most of the added resonances lie in the high-E_γ^B region (resonances at lower-E_γ^B have not yet been observed). Furthermore, the generic-B model uses older $B \to X_s\gamma$ photon energy spectrum parameters m_b and λ_1 (see Section 2.5), which makes the spectrum description suboptimal.

To ensure that the tail region is described correctly, while the resonant part is accounted for, a *hybrid model* approach is prepared for this analysis. It is implemented in a modified approach proposed by Ref. [154], where it was used for $B \to X_u\ell\nu$. The $B \to X_s\gamma$ photon
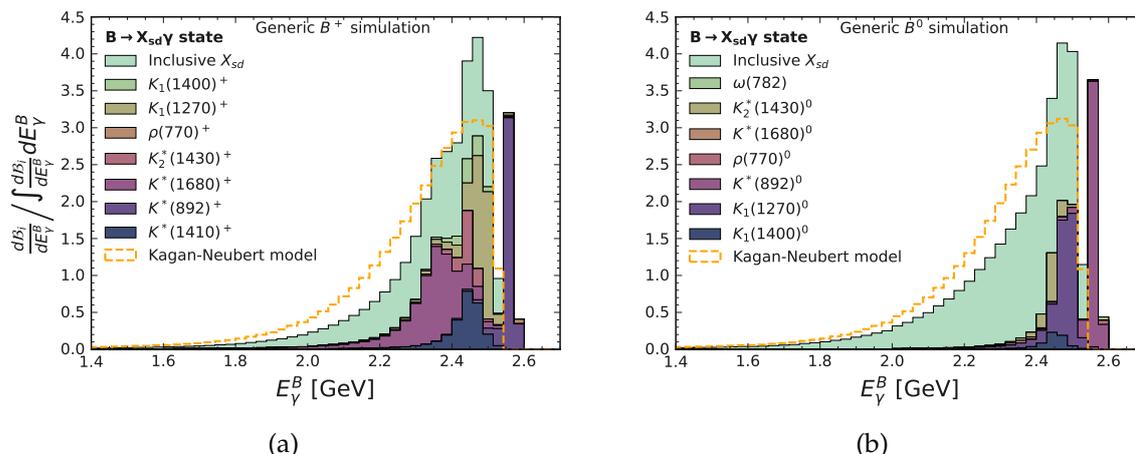
Figure 6.1: The model for $B \to X_s\gamma$ used in the Belle II official generic-B^+ (a) and generic-B^0 (b) MC. Although only several resonances are included, their structure is smoothed out. Unknown or unmeasured resonances, as well as non-resonant decays, are modelled with the inclusive Kagan-Neubert model, but the model is globally scaled down to account for the phase space covered by the exclusive decays. The branching fractions used here are taken from Table 3.1 or upper-limits from Refs. [16], [84]. The dashed line shows the Kagan-Neubert model. The Figures are produced using 500 000 event data sets for each mode produced by `EvtGen` equivalently to the Belle II simulation.

energy spectrum is subdivided into multiple intervals, referred to as *hybrid bins*. The scaling, called *hybrid weight*, instead of being global, is then calculated for each hybrid bin, such that:

$$H_i = I_i \cdot h_i + \Sigma R_i, \tag{6.2}$$

where h_i is a hybrid weight for hybrid bin i; I_i is the inclusive model prediction for the given hybrid bin; ΣR_i is the prediction of all resonances in the hybrid bin; H_i is the hybrid model prediction in the hybrid bin.

As mentioned before, for the I_i part, the Kagan-Neubert model is used (Section 2.5). In order to use more up-to-date parameters of the spectrum, the spectrum is reweighted to be compatible with parameters from Equation (2.34). In the kinetic scheme, compatible with `EvtGen` implementation of the Kagan-Neubert model, this amounts to:

$$m_b^{\text{kin}} = 4.624 \pm 0.045 \text{ GeV}/c^2; \quad \lambda_1^{\text{kin}} = -0.35 \pm 0.08 \text{ GeV}^2/c^4. \tag{6.3}$$

Taking into account the correlation of these parameters as described in Ref. [62], four additional up and down variations in the eigendirections are generated. The envelope of the variations is used as the inclusive model uncertainty in this analysis. The reweighted inclusive model is shown in Figure 6.2.

For the R_i part, it was decided to only include the $B \to K^*(892)\gamma$ resonance. The main reason to exclude other resonances is the fact that most are not known precisely, which would lead to a larger modelling uncertainty. Furthermore, the expected statistical precision (see discussion in Section 6.8.2) is not high enough to be sensitive to fine details in the spectrum.
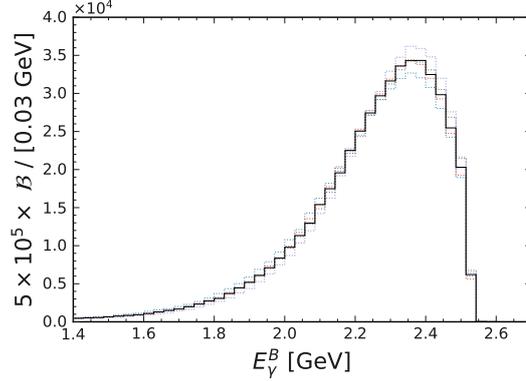
Figure 6.2: The inclusive X_s model based on Equation (6.3). The dashed lines show 4 up and down m_b and λ_1 variations based on their correlated uncertainties. The maximum deviations (envelope) are taken as the modelling uncertainty.

The set of hybrid bins was selected as displayed in Figure 6.3. An underflow and overflow bin is selected at 1.6 and 2.3 GeV, respectively. The overflow bin is chosen to include the majority of the $B \to K^*(892)\gamma$ resonance. The finer 0.1 GeV-wide bins in between contain small amounts of $B \to K^*(892)\gamma$ events. In each E_γ^B bin, an appropriate hybrid weight is calculated, such that the sum of the reweighted inclusive model (Figure 6.2) and $B \to K^*(892)\gamma$ contribution matches the partial branching fraction within that E_γ^B bin, based on Equation (6.2). This captures the desired description of the low-E_γ^B region with the inclusive model, and a K^* (892)-dominated behaviour at high-E_γ^B.
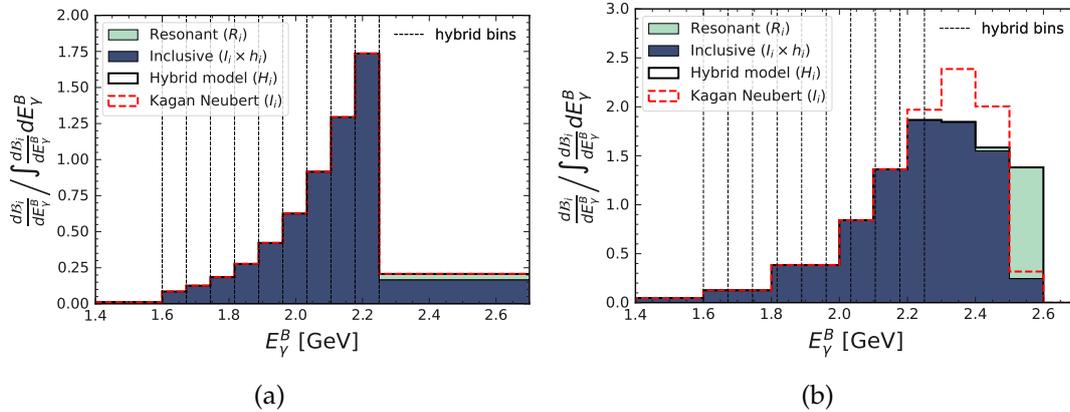

(a)

(b)

Figure 6.3: The hybrid model constructed for this analysis shown in two different binnings. In (a), the binning that matches the hybrid binning is chosen; in (b), the binning that will be chosen in this analysis (see Section 6.8.2 and Table 6.14). In both cases, the spectra are normalised such that the total area under them is 1. The hybrid model describes the tail adequately, while also taking into account the resonant contributions (compare with Figure 6.1). The different components corresponding to Equation (6.2) are labelled appropriately.

The hybrid bins need not match the binning used in the analysis or plotting. As seen in Figure 6.3b, the hybrid model can be used for different binnings which do not coincide with the hybrid bins. In the Figure, the model is shown in the binning that will later be used for the analysis, after optimising based on expected signal and background contributions.

The uncertainty of the hybrid model is evaluated by taking into account:

- Branching fraction uncertainty of the $B \rightarrow K^*(892)\gamma$, σ_{res};

- Branching fraction uncertainty of the inclusive $B \rightarrow X_s\gamma$ decay, σ_{incl};

- Inclusive model parameter variation envelope, as shown in Figure 6.2, σ_{var}.

These components are evaluated based on the values discussed in this Section and Table 3.1. They are visually shown for a selected $B \rightarrow X_s\gamma$ photon energy spectrum binning in Figure 6.4. The uncertainty is dominated by $B \rightarrow X_s\gamma$ model variation uncertainties, except in the interval where $B \rightarrow K^*(892)\gamma$ dominates. There the branching fraction uncertainty of the resonant decay is leading. The uncertainty related to the $B \rightarrow X_s\gamma$ inclusive branching fraction model is smaller than the other two.
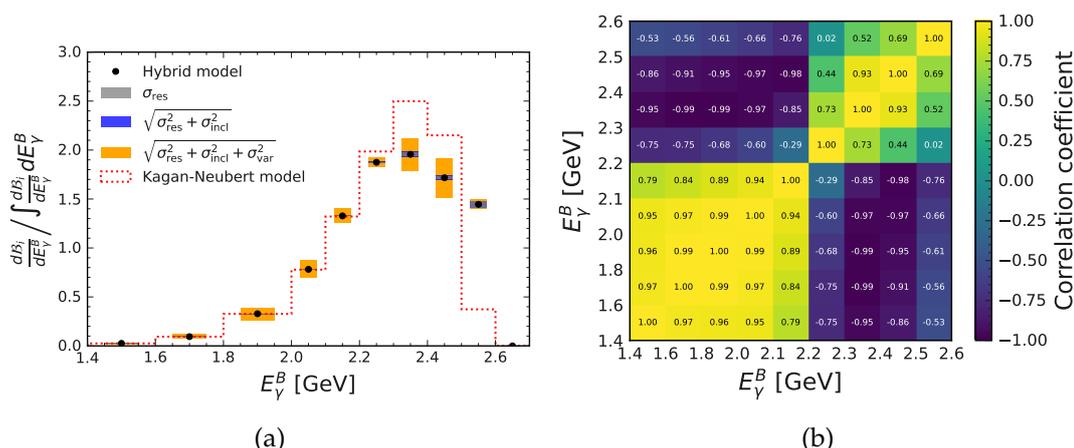

Figure 6.4: The inclusive X_s model based on Equation (6.3) is shown in (a). For comparison, the Kagan-Neubert model with the same parameters is also shown. The hybrid-signal model is overlaid with uncertainties from the resonant, inclusive branching fractions and the inclusive model parameters. The definitions of the uncertainties are given in the text. The correlation of the total uncertainty is given in (b).

6.3 Event reconstruction

The aim of the inclusive $B \rightarrow X_s\gamma$ analysis is to reconstruct an inclusive sample of all possible X_s states, as described before (e.g. Chapter 3). This means that explicit requirements on the momentum, number of tracks, angles etc. of the X_s system may introduce a direct bias on the 'inclusiveness' of the measurement. Assessing the impact of such selections on X_s in a model-independent way is difficult. Therefore, the X_s system is treated in a completely

'missing-momentum' approach, such that no direct requirements on it are imposed. The reconstruction requirements are only applied on the candidate tag-B meson and the candidate high energy photons from $B \rightarrow X_s\gamma$.

6.3.1   Tag-B meson candidate reconstruction

The analysis begins with the reconstruction of tag-B meson candidates in each event using the Belle II Full Event Intepretation (FEI) algorithm [155], [156], which is part of `basf2`. It is a hierarchical six-stage reconstruction chain, which begins with the identification of all tracks, displaced vertices (tracks that do not originate near the interaction point) and ECL clusters. The algorithm begins by combining track and ECL cluster information to reconstruct final-state candidate particles, such as $e^{\pm}$, $\mu^{\pm}$, γ, $\pi^{\pm}$, $K^{\pm}$ and K_L^0. In the next stage, the final-state particles are combined to form intermediate particles, such as $\pi^0, K_S^0, D^{(*)}$. In later stages, intermediate and final-state particles are combined into B mesons. At every stage of the procedure, the probability for the combined particle to be correct is evaluated by a BDT which maps input features related to the particle (four-momentum, vertex position, angles between daughter particles etc.) to a single classifier output score. The final output score related to the quality of reconstruction of the B meson is denoted as $\mathcal{P}_{\mathrm{FEI}}$. The schematic visualisation of the reconstruction process is shown in Figure 6.5.
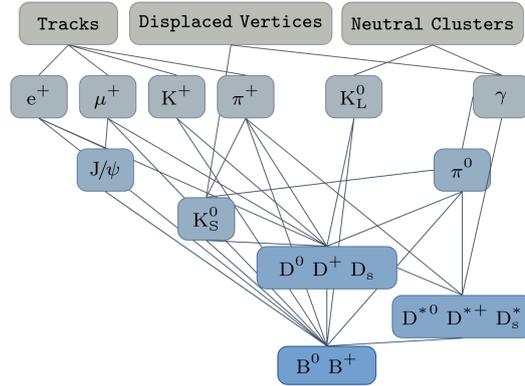

Figure 6.5: The schematic overview of FEI. The algorithm reconstructs B^+ (B^0) candidates in 36 (32) hadronic decay chains in six reconstruction stages that combine final-state and intermediate particles. Credit to Ref. [156].

In total, FEI reconstructs $\mathcal{O}(10000)$ distinct decay chains and provides B^+ meson candidates in 36 hadronic decay modes, and B^0 candidates in 32 hadronic decay modes. As a result, two FEI modes are differentiated: `feiB`$^+$, which combines $B^{\pm}$ meson candidates; and `feiB`^{0}, which combines B^0 meson candidates. Each event may have more than one candidate reconstructed in the same and/or different decay channels and/or FEI modes. The reconstructed decay channels for `feiB`$^+$ and `feiB`^{0} modes are shown in Table 6.2.

This thesis uses data and simulation samples following the standard Belle II approach, where sub-samples of data and MC with the FEI algorithm applied are produced centrally, referred to as *FEI skims*. In order to make the FEI algorithm more computationally efficient, event selections are made to reject events highly incompatible with one of the B mesons de-

Table 6.2: The *B* meson decay modes reconstructed by the FEI algorithm. FEI modes reconstructed as `feiB`$^+$ and `feiB`^{0} are listed separately.

	`feiB`$^+$ modes	`feiB`^{0} modes
1.	$B^+ \to \bar{D}^0 \pi^+$	$B^0 \to D^- \pi^+$
2.	$B^+ \to \bar{D}^0 \pi^+ \pi^0$	$B^0 \to D^- \pi^+ \pi^0$
3.	$B^+ \to \bar{D}^0 \pi^+ \pi^0 \pi^0$	$B^0 \to D^- \pi^+ \pi^0 \pi^0$
4.	$B^+ \to \bar{D}^0 \pi^+ \pi^+ \pi^-$	$B^0 \to D^- \pi^+ \pi^+ \pi^-$
5.	$B^+ \to \bar{D}^0 \pi^+ \pi^+ \pi^- \pi^0$	$B^0 \to D^- \pi^+ \pi^+ \pi^- \pi^0$
6.	$B^+ \to \bar{D}^0 D^+$	$B^0 \to \bar{D}^0 \pi^+ \pi^-$
7.	$B^+ \to \bar{D}^0 D^+ K_S^0$	$B^0 \to D^- D^0 K^+$
8.	$B^+ \to \bar{D}^{0*} D^+ K_S^0$	$B^0 \to D^- D^{0*} K^+$
9.	$B^+ \to \bar{D}^0 D^{+*} K_S^0$	$B^0 \to D^{-*} D^0 K^+$
10.	$B^+ \to \bar{D}^{0*} D^{+*} K_S^0$	$B^0 \to D^{-*} D^{0*} K^+$
11.	$B^+ \to \bar{D}^0 D^0 K^+$	$B^0 \to D^- D^+ K_S^0$
12.	$B^+ \to \bar{D}^{0*} D^0 K^+$	$B^0 \to D^{-*} D^+ K_S^0$
13.	$B^+ \to \bar{D}^0 D^{0*} K^+$	$B^0 \to D^- D^{+*} K_S^0$
14.	$B^+ \to \bar{D}^{0*} D^{0*} K^+$	$B^0 \to D^{-*} D^{+*} K_S^0$
15.	$B^+ \to D_s^+ \bar{D}^0$	$B^0 \to D_s^+ D^-$
16.	$B^+ \to \bar{D}^{0*} \pi^+$	$B^0 \to D^{-*} \pi^+$
17.	$B^+ \to \bar{D}^{0*} \pi^+ \pi^0$	$B^0 \to D^{-*} \pi^+ \pi^0$
18.	$B^+ \to \bar{D}^{0*} \pi^+ \pi^0 \pi^0$	$B^0 \to D^{-*} \pi^+ \pi^0 \pi^0$
19.	$B^+ \to \bar{D}^{0*} \pi^+ \pi^+ \pi^-$	$B^0 \to D^{-*} \pi^+ \pi^+ \pi^-$
20.	$B^+ \to \bar{D}^{0*} \pi^+ \pi^+ \pi^- \pi^0$	$B^0 \to D^{-*} \pi^+ \pi^+ \pi^- \pi^0$
21.	$B^+ \to D_s^{+*} \bar{D}^0$	$B^0 \to D_s^{+*} D^-$
22.	$B^+ \to D_s^+ \bar{D}^{0*}$	$B^0 \to D_s^+ D^{-*}$
23.	$B^+ \to \bar{D}^0 K^+$	$B^0 \to D_s^{+*} D^{-*}$
24.	$B^+ \to D^- \pi^+ \pi^+$	$B^0 \to J/\psi K_S^0$
25.	$B^+ \to D^- \pi^+ \pi^+ \pi^0$	$B^0 \to J/\psi K^+ \pi^-$
26.	$B^+ \to J/\psi K^+$	$B^0 \to J/\psi K_S^0 \pi^+ \pi^-$
27.	$B^+ \to J/\psi K^+ \pi^+ \pi^-$	$B^0 \to \Lambda_c^- p \pi^+ \pi^-$
28.	$B^+ \to J/\psi K^+ \pi^0$	$B^0 \to \bar{D}^0 p \bar{p}$
29.	$B^+ \to J/\psi K_S^0 \pi^+$	$B^0 \to D^- p \bar{p} \pi^+$
30.	$B^+ \to \Lambda_c^- p \pi^+ \pi^0$	$B^0 \to D^{-*} p \bar{p} \pi^+$
31.	$B^+ \to \Lambda_c^- p \pi^+ \pi^- \pi^+$	$B^0 \to \bar{D}^0 p \bar{p} \pi^+ \pi^-$
32.	$B^+ \to \bar{D}^0 p \bar{p} \pi^+$	$B^0 \to \bar{D}^{0*} p \bar{p} \pi^+ \pi^-$
33.	$B^+ \to \bar{D}^{0*} p \bar{p} \pi^+$	
34.	$B^+ \to D^+ p \bar{p} \pi^+ \pi^-$	
35.	$B^+ \to D^{+*} p \bar{p} \pi^+ \pi^-$	
36.	$B^+ \to \Lambda_c^- p \pi^+$	

caying hadronically. This decision is based on tracks and clusters as per standard Belle II reconstruction guidelines with additional selections, summarised in Table 6.3. Overall, they ensure that only energetic tracks originating from the interaction point are selected. They also minimise the impact of beam background clusters or clusters for which no track information can be associated (outside of CDC acceptance).

Table 6.3: Definitions for objects used in FEI selections.

Object name	Definition
Cleaned tracks	$\lvert d_0 \rvert < 0.5$ cm, $z_0 < 2$ cm, $p_T > 0.1$ GeV/c
Cleaned ECL clusters	$17\,^\circ < \theta < 150\,^\circ$, $E > 0.1$ GeV

Using the definitions of cleaned tracks and ECL clusters, reconstructed events are filtered and only the events that pass the requirements are analysed by the FEI algorithm. Such requirements are summarised in Table 6.4. The selection of events with at least 3 cleaned tracks in the event and at least 3 cleaned ECL clusters is based on the fact that $B\bar{B}$ events produce ~ 10 charged tracks and neutral particles [96]. Furthermore, the measured energy of the event is required to exceed 4 GeV in the e^+e^- collision centre-of-mass frame. This is a purely pragmatic requirement: because no neutrinos or missing-momentum are present in a hadronic decay, the energy cannot be much lower than 5.28 GeV. Finally, the total deposited energy registered by the ECL in the event is required to be between 2 and 7 GeV. Hadronic events are expected to deposit significantly more energy than 2 GeV. On the other hand, many lower energy particles should be stopped within PXD, SVD, CDC or TOP, meaning that their energy deposit in the ECL would be negligible. Therefore, a 7 GeV ECL energy upper limit ensures that low-multiplicity events, such as $e^+e^- \rightarrow e^+e^-$, are immediately removed to ensure a better-optimised workflow.

Table 6.4: Selections applied before running the FEI algorithm. Cleaned tracks and clusters are defined in Table 6.3.

Selection description	Selection
Number of cleaned tracks in event	≥ 3
Number of cleaned ECL clusters in event	≥ 3
Total measured centre-of-mass energy in the event	> 4 GeV
Total energy of cleaned ECL clusters and deposits associated with cleaned tracks	2 GeV $< E < 7$ GeV

The events that pass the requirements of Table 6.4 are analysed by the FEI algorithm. In each event, multiple FEI candidates can be reconstructed (see Figure 6.6a). To focus only on the candidates that are correctly reconstructed, selections on ΔE and M_{bc} are made, as well as a loose requirement on $\mathcal{P}_{\mathrm{FEI}}$. The selections shown in Table 6.5 are standard selections that are applied on the Belle II FEI skims.

The tag-side candidates that pass the nominal FEI requirements undergo a kinematic fit [157], where the particles used to reconstruct the tag-B candidate are fitted with a common vertex constraint. Candidates that fail the fit are rejected. This improves the resolution of the B meson momentum for correct candidates but may shift their momentum. Therefore, to

Table 6.5: Additional selections that reduce the data sets after applying FEI, focusing only on well-reconstructed tag-side candidates. These FEI skim selections are the nominal ones, which are applied on all FEI skimmed data sets in Belle II. In this analysis, only the selection on the tag-B meson is tightened in order to remove the edge effects. Such effects arise after applying a kinematic fit of the tag-side products.

Variable	FEI skim selections	Selections in this analysis
$M_{bc}(tag)$	> 5.24 GeV	> 5.245 GeV
$\Delta E(tag)$	−0.15 to 0.1 GeV	
$\mathcal{P}_{FEI}$ E	> 0.001	

avoid distribution-edge effects, a tighter M_{bc} selection is used in this analysis, as illustrated in Table 6.5. The tag-side selections used in this analysis do not affect $B \to X_s\gamma$, as correct tag-side candidates with lower M_{bc}, higher $|\Delta E|$ or lower $\mathcal{P}_{FEI}$ (compared to Table 6.5) are uncommon.

6.3.2 Candidate photon reconstruction

Only the photon from $B \to X_s\gamma$ can be reconstructed while ensuring a model-independent inclusive measurement. In order to reduce the quickly growing number of background photon candidates, only events where at least one photon satisfies $E_\gamma^* > 1.2$ GeV are considered. Photons must also be within the CDC acceptance ($17 - 150\,°$). These requirements are summarised in Table 6.6. Reconstructed photon energy is boosted to the signal B meson rest frame based on the Lorentz transformation in Appendix C.

Table 6.6: Requirements for photons in reconstructed events.

Selection description	Selection
Number of photons with E_γ^* >1.2 GeV	$N(E_\gamma^* > 1.2$ GeV$) \geq 1$
Polar angle of photon	$17\,° < \theta < 150\,°$

6.3.3 Overview of the selected sample

After the reconstruction, based on the MC samples, an event can have up to 20 tag-B candidates. The sample is broken down to show the relative fraction of the total number of tag-side B meson candidates in Figure 6.6a. About 62% (72%) of events for `feiB`$^+$ (`feiB`^{0}) modes have only one tag-side candidate. About 21% (18%) of events for `feiB`$^+$ (`feiB`^{0}) modes have two tag-side candidates, and 8% (5%) have three. The number of candidates per event reduces quickly, but faster for B^0 modes, with roughly 2% (1%) of events having more than five candidates for `feiB`$^+$ (`feiB`^{0}). The same event can have a B^+ and B^0 candidate reconstructed.

A similar distribution for the number of signal-side photon candidates with a threshold of $E_\gamma^B > 1.4$ GeV applied is shown in Figure 6.6b. The highest energy photon is the sole candidate in the event in 98% of the cases in generic MC. Similar studies on signal MC show that the highest energy photon is expected to come from $B \to X_s\gamma$ in 99.9% of the cases. The

$E_\gamma^B > 1.4$ GeV selection is chosen pragmatically to maintain a reasonable data set memory size without losing signal events. As the number of photon candidates grows swiftly with decreasing energy, this threshold still provides access to the majority of the $B \to X_s \gamma$ decay phase space.

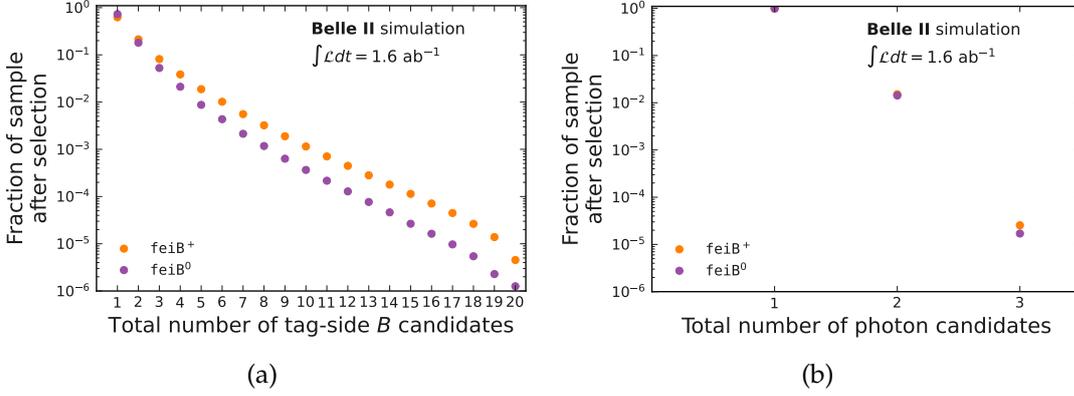(a)   (b)

Figure 6.6: Relative fractions of events for the number of reconstructed B meson candidates (a) and reconstructed photon candidates (b) in the generic MC sample. In (a), the overall volume of candidates is similar for feiB$^+$ and feiB^0 modes, with around one in two events only having a single candidate per event. Conversely, as depicted in (b), the vast majority of events contain only a single signal photon with $E_\gamma^B > 1.4$ GeV. Two or more photon candidates are present only $\mathcal{O}(1)$% of the time.

The reconstructed $B \to X_s \gamma$ spectrum in generic MC with the previously laid-out requirements are shown in Figure 6.7. Note that these events can contain multiple combinations of a photon and tag-side candidate per event. Overall, it may seem that the feiB^0 mode has a higher signal-to-background ratio compared to feiB$^+$. However, one has to take into account that feiB$^+$ and feiB^0 modes result from different reconstruction chains. Furthermore, feiB^0 has fewer modes than feiB$^+$ (see Section 6.3.1). Therefore, without additional studies that follow in Sections 6.7.1 and 6.7.2 such a conclusion cannot be unambiguously drawn. On the other hand, inspecting Figures 3.2a and 6.7, it is clear that a better signal-to-background ratio can already be observed even without any additional background treatment.

The tag-side probability distributions provided by the FEI classifier are shown in Figure 6.8. They further emphasise the differences between tag candidates reconstructed in feiB$^+$ and feiB^0 modes. A selection on the $\mathcal{P}_{\text{FEI}}$ variable is not trivial; even after disregarding the differences between feiB$^+$ and feiB^0 modes, the $\mathcal{P}_{\text{FEI}}$ values may be different within the same-charged B mode. This is shown and discussed in Appendix D. Tight direct selection may result in a selection of reconstruction channels but not necessarily the quality of reconstruction, as seen in Figures D.1 to D.2. To avoid such bias, further $\mathcal{P}_{\text{FEI}}$ thresholds are not considered in this analysis.
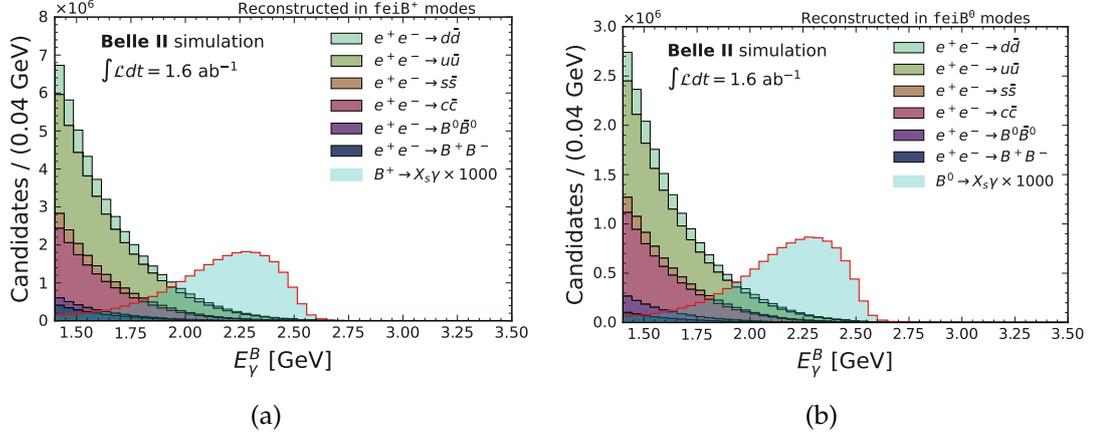
Figure 6.7: $B \to X_s \gamma$ spectrum in generic MC after event reconstruction in `feiB`$^+$ and `feiB`^{0} modes. Overlaid are events from signal MC, where the photon comes from $B \to X_s \gamma$, multiplied by a scaling factor. These Figures may include multiple tag-B and photon entries per event and can be compared with Figure 3.2a.
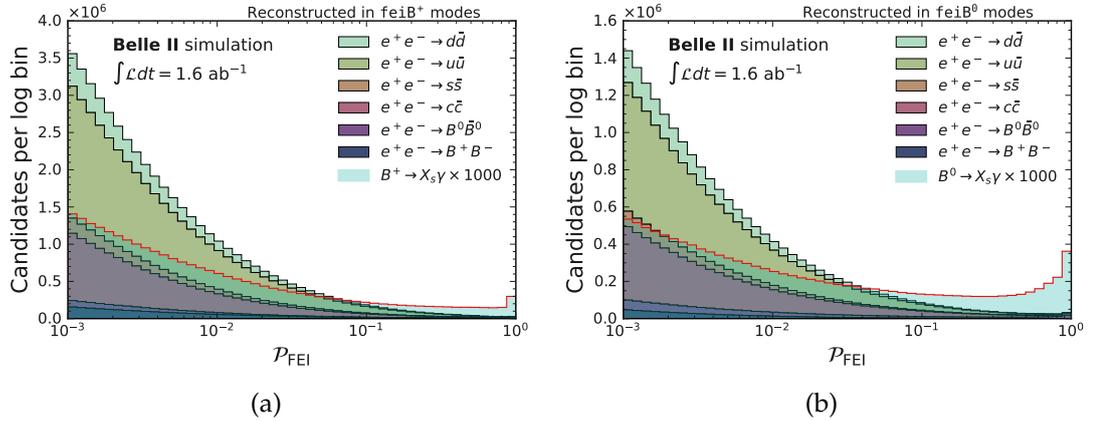

Figure 6.8: Tag-side $\mathcal{P}_{\text{FEI}}$ after reconstructing $B \to X_s \gamma$ events in generic MC in `feiB`$^+$ and `feiB`^{0} modes. Overlaid are events from signal MC, where the photon comes from $B \to X_s \gamma$, multiplied by a scaling factor. These Figures may include multiple tag-B and photon entries per event.

6.4   Photon candidate selection

The previous Section overviewed the samples that are reconstructed using basic requirements laid out in Sections 6.3.1 and 6.3.2. In this Section concrete selections are discussed that lead to background suppression, the best photon candidate and the best tag candidate selections.

6.4.1   Primary photon candidate selection

Contrary to the tag-side, a selection of the best photon candidate in the range $E_\gamma^B > 1.4$ GeV is effectively trivial based on the discussion in Section 6.3. For more than 99% of the signal MC sample, the highest E_γ^B photon is the correct photon originating from $B \rightarrow X_s\gamma$ decay. Therefore, it is chosen as the best photon candidate requirement with virtually no signal efficiency loss. Judging from Figure 6.6b, this provides an approximately 3% background suppression. For the rest of the thesis, this selection is always implied when referring to signal-photon candidates.

6.4.2   Main photon background sources

Based on Figure 6.7, the number of photon and tag candidates originating in non-$B\bar{B}$ events is significantly larger than that of B meson events. The proportion of $q\bar{q}$ to $B\bar{B}$ event candidates is 92.5% to 7.5% for feiB$^+$ mode; and 91.7% to 8.3% for feiB^0.

The majority of background photon candidates originate in $\pi^0 \rightarrow \gamma\gamma$ or $\eta \rightarrow \gamma\gamma$ decays. This in total accounts for roughly 85% of background photon candidates. Photon candidates, broken down by their mother particle, are shown in Figure 6.9. Other sources, such as (in decreasing order) initial-state radiation, neutron annihilation, parton shower final-state radiation, $\omega(782)$, η' decays each make up $0.5 - 3\%$ of the sample. All other sources individually make up less than 0.5% and include various hadron decays that are produced in continuum or B events. The backgrounds are similar for both feiB$^+$ and feiB^0 modes. Note that some photon candidates can be misidentified, in particular neutrons. This is further discussed in Section 6.4.3.

The reason why $\pi^0 \rightarrow \gamma\gamma$ and $\eta \rightarrow \gamma\gamma$ decays are such a prominent background is related to the fact that they can often be produced in hadronic decays. Because the π^0 and η are produced boosted, their diphoton decays can be asymmetric in the e^+e^- collision rest-frame, where one photon has a much larger energy than the other. The hadronic decay, overall, mimics the hadronised X_s system, whereas the more energetic photon is taken as the high energy photon candidate. However, for B decays, this background drops off rapidly with photon energy, and at high-E_γ^B becomes negligibly small because processes producing photons with $E_\gamma^B \approx m_b/2$ in B decays are rare. No such constraint exists for continuum events where light hadrons can be created in large numbers. Therefore, π^0 and η suppression, while important for B decays, also highly coincides with continuum event suppression.

At this stage, $B \rightarrow X_s\gamma$ events make up 0.05% of the feiB$^+$ sample and 0.07% of the feiB^0 sample. To reduce the discussed background components the following strategy is adopted:

- Suppress misidentified photons (different particle species);

- Suppress $\pi^0 \rightarrow \gamma\gamma$ and $\eta \rightarrow \gamma\gamma$ decays;
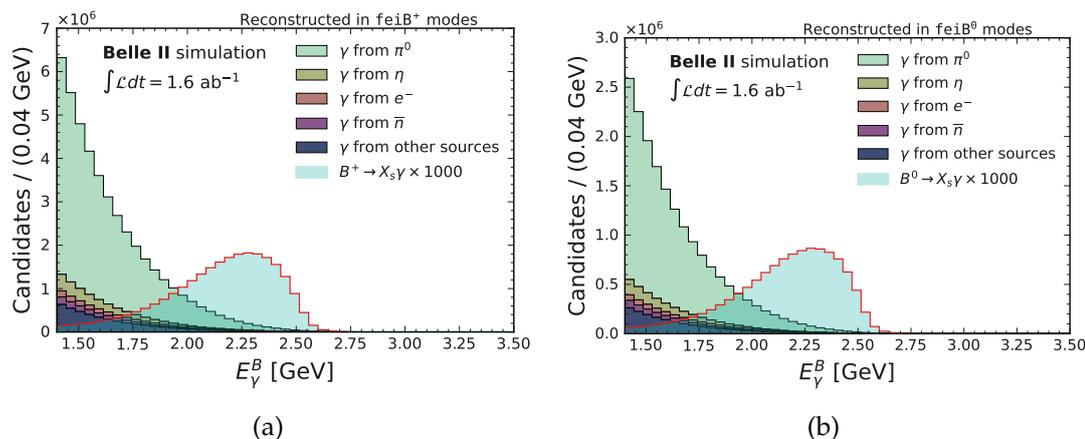
(a) (b)

Figure 6.9:   The background photon distribution after reconstruction in `feiB`$^+$ and `feiB`^{0} modes, stacked by the photon mother-particle species. A scaled $B \rightarrow X_s \gamma$ spectrum is also overlaid. Only one photon candidate per event is shown, but it may still be paired with multiple tag-side candidates. Roughly 85% of candidates originate in $\pi^0 \rightarrow \gamma\gamma$ or $\eta \rightarrow \gamma\gamma$ decays. Other important backgrounds are photons from initial-state radiation, bremsstrahlung, and neutron-annihilation processes. These account for approximately 3%, each. The leftover 10% originate in various other decays.

- Suppress $e^+ e^- \rightarrow q\bar{q}$ events;

- Reoptimise all selections simultaneously to adopt a final set of selections.

6.4.3   Misidentified photon suppression

Neutrons, K_L^0, protons, electrons and other charged hadrons for which tracking for the particle did not succeed may leave clusters in the ECL which are misidentified as photons. The photon misidentification rate is given in Table 6.7. The main misidentified photon candidates originate from neutrons with a small contribution from electron and K_L^0 showers.

Table 6.7:   Photon misidentification rates after reconstruction. The majority of photons are identified correctly. The largest component coming from misidentified neutron showers and K_L^0 deposits. The rates are similar for `feiB`$^+$ and `feiB`^{0} modes which is consistent with the fact that this property is independent of the decaying B charge.

Particle species	γ		n^0		e^-		K_L^0		p^-		Other
Candidate rate (FEI B^+ | B^0)	96.1%	96.0%	2.4%	2.5%	0.5%	0.5%	0.5%	0.5%	0.3%	0.3%	0.2%

Generally, the total energy deposit and distribution between ECL crystals, also known as the *shower shape*, is different depending on the particle species due to their different radiation lengths. This can be used to distinguish photon clusters using MVA methods. A technique achieving this, which uses the moments of Zernike polynomials, is documented in Ref. [158]. This approach is implemented in `basf2` and used in this analysis. Here, a condensed overview of the approach is provided.

A complete set of complex two-dimensional polynomials is defined as:

$$V_{nm}(\rho\cos\alpha, \rho\sin\alpha) = R_{nm}(\rho)e^{im\alpha}, \tag{6.4}$$

where $x = \rho\cos\alpha, y = \rho\sin\alpha$ are polar coordinates, m is an integer and $R_{nm}(\rho) = V(\rho, 0)$ is a polynomial of degree n. The expression for a Zernike polynomial is given as:

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}\rho^{n-2s}. \tag{6.5}$$

The moments of a function $f(\rho\cos\alpha, \rho\sin\alpha)$ are expressed in terms of V_{nm} as:

$$Z_{nm} = \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 V_{nm}^*(\rho\cos\alpha, \rho\sin\alpha)f(\rho\cos\alpha, \rho\sin\alpha)\rho d\rho d\alpha. \tag{6.6}$$

Z_{nm} are called Zernike moments. They have many useful properties that make them usable in image recognition, the field of optics, and more importantly, particle identification algorithms (see. Ref. [158] and references therein).

A Dirac comb is defined to parametrise a particle shower in the ECL as:

$$f(\vec{x}) = \sum_i \delta(\vec{x} - \vec{x}_i)\frac{w_i E_i}{\sum w_i E_i}, \tag{6.7}$$

where $\vec{x}$ is a dimensionless crystal position in the transverse plane, i is a crystal index, summing over all crystals in a given particle shower, and E_i is the energy of the i-th crystal. As showers can overlap, w_i is the fraction of energy in a crystal that is associated with the currently investigated shower. It can be shown, that Zernike moments for ECL showers can then be expressed as [158]:

$$|Z_{nm}| = \frac{n+1}{\pi}\frac{1}{\sum_i w_i E_i}\left|\sum_i R_{nm}(\rho_i)e^{-im\alpha_i}w_i E_i\right|. \tag{6.8}$$

The work in Ref. [158] selects the best combination of eleven $|Z_{nm}|$ which provides the strongest separation between hadronic showers and electromagnetic showers. The chosen combination of $|Z_{nm}|$ is combined using a BDT and produces a single output, hereafter referred to as `zernikeMVA` $\in (0, 1)$. The `zernikeMVA` distributions for $B \rightarrow X_s\gamma$ candidates in generic MC and signal MC events are shown in Figure 6.10.

Overall, for true photons this distribution is strongly peaking in the $0.8 - 1$ region. Misidentified hadrons peak close to 0. For non-$B \rightarrow X_s\gamma$ photon candidates, this distribution remains relatively uniform in the $0 - 0.8$ region. Moreover, `zernikeMVA` provides a good separation against real photon candidates that originate in neutron annihilation events. This is shown by a `zernikeMVA` distribution exclusively for true photon candidates in Figure 6.11.
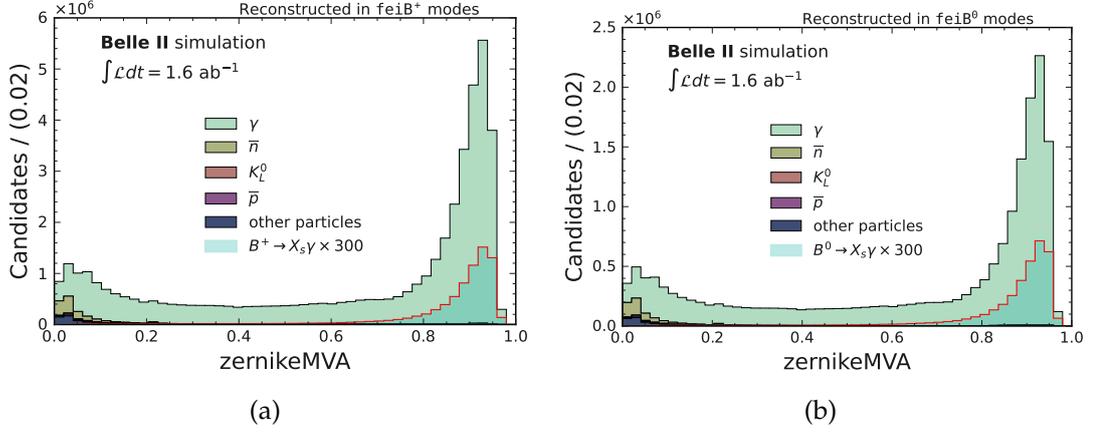
Figure 6.10: The distributions of the `zernikeMVA` for different particle species that are reconstructed as photon candidates in `feiB`$^+$ and `feiB`^{0} modes. The candidates presented in these Figures are the same as those in Figure 6.9. A scaled `zernikeMVA` distribution for $B \rightarrow X_s\gamma$ events is overlaid. A good separation is observed between real photons and hadronic showers misidentified as photons.
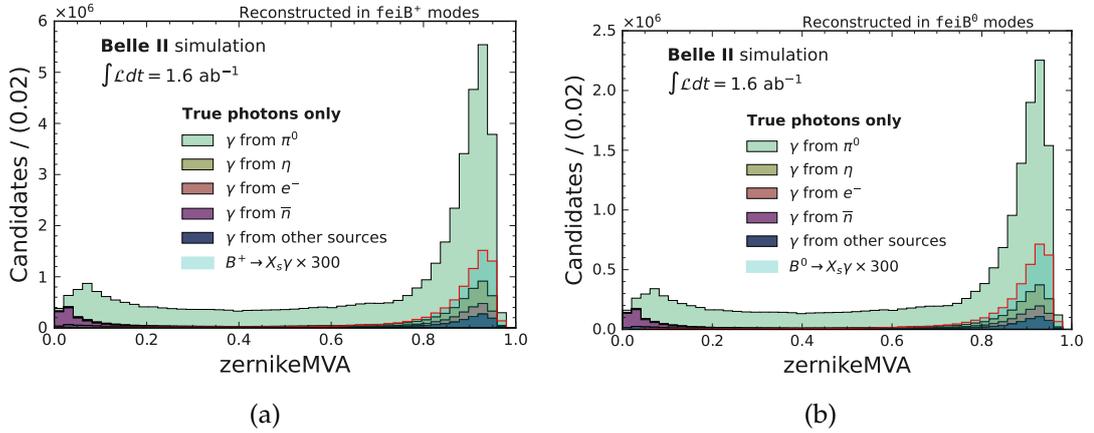

Figure 6.11: The distributions of the `zernikeMVA` for different photon sources in generic MC reconstructed in `feiB`$^+$ and `feiB`^{0} modes. The candidates presented here are only those which are true photons in Figure 6.10. A scaled `zernikeMVA` distribution for $B \rightarrow X_s\gamma$ events is overlaid. Photons associated with neutron annihilation events are separated.

6.4.4   Suppression of π^0 and η diphoton decays

Aboout 85% of background photons in this analysis originate from photons that are produced in $\pi^0 \to \gamma\gamma$ or $\eta \to \gamma\gamma$ decays. A lot of such light mesons originate in continuum events, but they are prominent even in $B\bar{B}$ events. Therefore, an efficient mechanism to suppress π^0- and η-related photon candidates is required.

In this analysis, a suppression tool, called π^0 *and η veto* is utilised. It is implemented as part of basf2 and is a standard Belle II approach for suppression of radiative backgrounds from light mesons. Here, an overview of the training and the validation process is provided which is performed by an independent analysis.

The general idea of the π^0 and η veto is to pair the high energy photon signal candidate (*hard* photon) with lower energy photons (*soft* photons) in the event. The compatibility of the combination with a $\pi^0 \to \gamma\gamma$ or $\eta \to \gamma\gamma$ decay is evaluated and a probability-like quantity is calculated to quantify it.

The soft photon candidate is selected with an energy of 30 MeV (40 MeV in the backward ECL endcap) for π^0 or 60 MeV for η. The photon is also required to have deposited the energy in two or more crystals. Furthermore, photon candidates are required to have an associated cluster time less than one standard deviation away from zero. These selections ensure that beam background photons, neutral hadrons misidentified as photons and misreconstructed charged particles are not included in the soft photon sample.

The soft photons that pass these selections are combined with the hard photon candidate. The following observables are then calculated and used to train a MVA classifier:

- Invariant mass of the soft photon and hard photon combination;

- Soft photon energy in the laboratory frame;

- Soft photon ECL cluster polar angle;

- Distance between the soft photon ECL cluster and the nearest track extrapolated to the ECL;

- Helicity angle of the combination.

The classifier for $\eta \to \gamma\gamma$ includes additional observables to increase the separation power:

- zernikeMVA of the soft photon;

- Number of crystals where the soft photon has deposited energy;

- Ratio of soft photon energy in 3-by-3 crystals around the central crystal to soft photon energy in the 5-by-5 crystals with the corner crystals removed.

For every combination of a soft and hard photon, the MVA produces an output between 0 and 1. The same hard photon is paired with all soft photons in a given event, and the largest MVA output is assigned to it as the π^0 or η probability. This MVA output is denoted as $\mathcal{P}(\pi^0 \to \gamma\gamma)$ or $\mathcal{P}(\eta \to \gamma\gamma)$, respectively. Note that despite the nomenclature, this variable is only probability-like (i.e. $\mathcal{P} \in (0,1)$) but does not truly represent a probability. The distributions for $\mathcal{P}(\pi^0 \to \gamma\gamma)$ and $\mathcal{P}(\eta \to \gamma\gamma)$ are shown in Figure 6.12. In all cases, $B \to X_s\gamma$ can be
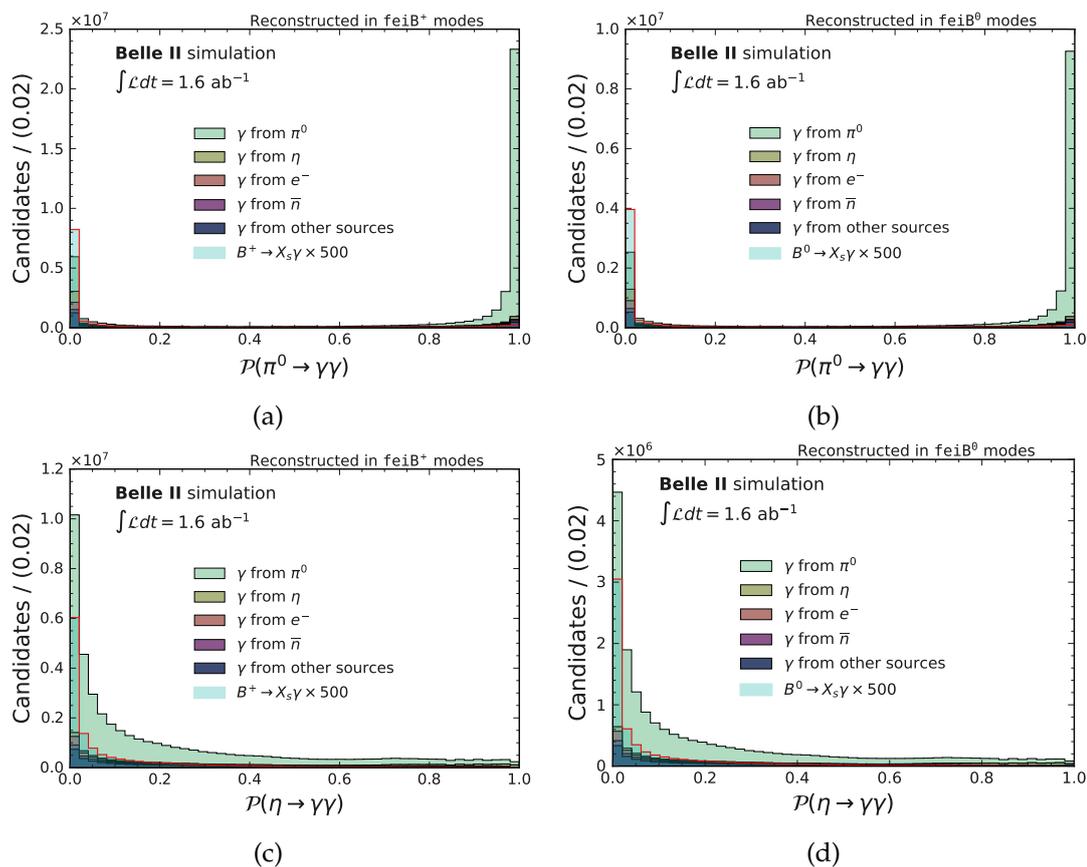
Figure 6.12: The distributions of $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$ ((a) and (b)) and $\mathcal{P}(\eta \rightarrow \gamma\gamma)$ ((c) and (d)) for different photon sources in generic MC reconstructed in `feiB`$^+$ and `feiB`^{0} modes. This is shown for all photon candidates included in Figure 6.9. Scaled respective veto probability distributions for $B \rightarrow X_s\gamma$ events are overlaid. The separation power of $\mathcal{P}(\eta \rightarrow \gamma\gamma)$ is hidden by a large number of $\pi^0 \rightarrow \gamma\gamma$ events in (c) and (d). It is apparent again when π^0 decays events are removed from the sample as shown in Figure 6.13.

seen to be strongly peaking near 0, consistent with photons that do not originate from light unflavoured meson decays.

For the case of π^0 veto, shown in Figures 6.12a and 6.12b, an excellent separation is observed between photons originating in π^0 decays and other photons. $B \to X_s\gamma$ and other non-π^0 photon candidates also show a small peak at high-$\mathcal{P}(\pi^0 \to \gamma\gamma)$ values, which alludes to a small inefficiency of the algorithm. However, compared to the separation power provided, this is an acceptable trade-off. For the $\mathcal{P}(\eta \to \gamma\gamma)$, the separation is less clear. The reason for this is the fact that the generic MC sample is dominated by $\pi^0 \to \gamma\gamma$ decays which are not targeted by the $\mathcal{P}(\eta \to \gamma\gamma)$ classifier. Removing $\pi^0 \to \gamma\gamma$ decays from the sample, a clear separation of photon candidates originating in η decays from other types of decays becomes apparent (see Figure 6.13).
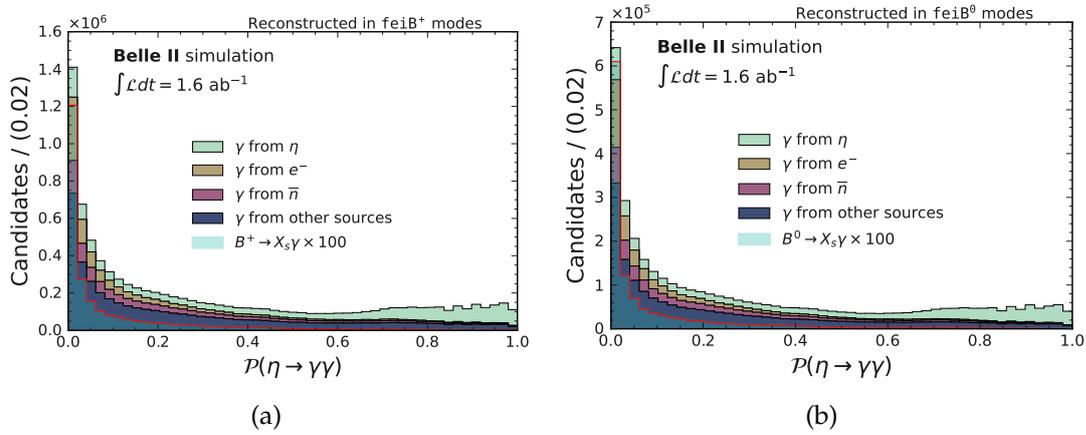

Figure 6.13:   The distributions of $\mathcal{P}(\eta \to \gamma\gamma)$ for different photon sources in generic MC reconstructed in `feiB+` and `feiB0`, but with photons that are associated with $\pi^0 \to \gamma\gamma$ removed. This is equivalent to Figures 6.12c and 6.12d with the π^0 component not included. A scaled $\mathcal{P}(\eta \to \gamma\gamma)$ distribution for $B \to X_s\gamma$ events is overlaid. Although the separation power is not as strong as in the case of $\mathcal{P}(\pi^0 \to \gamma\gamma)$ (Figures 6.12a and 6.12b), a clear peak at low-$\mathcal{P}(\eta \to \gamma\gamma)$ can be seen for $B \to X_s\gamma$.

6.4.5   Signal-photon background suppression correlation

Even though no direct selection is applied on the X_s system, through direct or higher-order correlations with E_γ^B, a bias may be introduced to the photon energy. To ensure that no such effect is introduced, a correlation study is performed for $\mathcal{P}(\pi^0 \to \gamma\gamma)$, $\mathcal{P}(\eta \to \gamma\gamma)$ and `zernikeMVA` observables. In principle, it is acceptable if the selection introduces a bias to the background, as long as this bias is well reproduced in simulation. The latter will be validated in Sections 6.11 and 6.13. Therefore, the study is performed exclusively focusing on $B \to X_s\gamma$ events, as it is aimed to ensure that the photon energy spectrum itself is minimally biased.

A two-dimensional map of $\mathcal{P}(\pi^0 \to \gamma\gamma)$, $\mathcal{P}(\eta \to \gamma\gamma)$, `zernikeMVA` versus E_γ^B is given in Figure 6.14. Because the distributions of the three variables used for background suppression
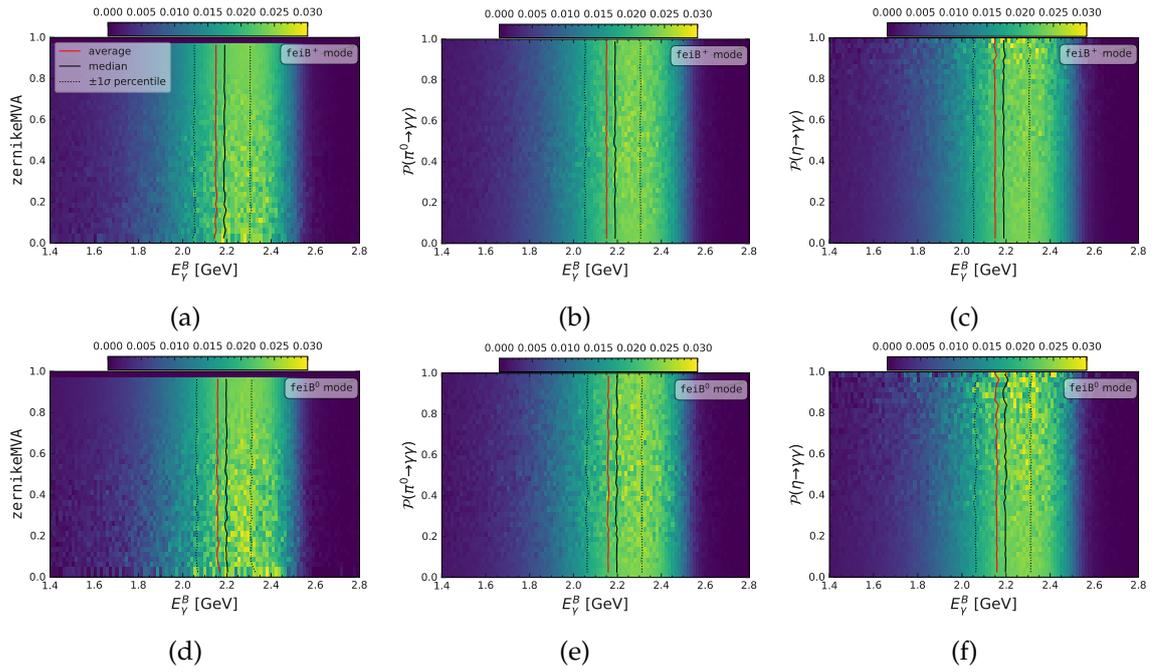
Figure 6.14: Correlation tests for background suppression observables described in Section 6.4, depicted as a 2D histogram. Each row is normalised such that all bins within that row add up to 1. For signal $B^+ \rightarrow X_s\gamma$ events, the tests are shown in the upper panels (a) – (c), and for $B^0 \rightarrow X_s\gamma$ in the lower ones (d) – (f). In the red line, the average photon energy, $\langle E_\gamma^B \rangle$, is shown as a function of the tested observable. In black and black-dotted lines: the median and $\pm 1\sigma$ percentile values of E_γ^B, respectively. No strong dependence can be observed in any of the quantities or the 2D maps.

are not uniform, each row is normalised such that the sum of each row is equal to unity. This makes the comparison between differently populated bins simpler. The Figure also denotes the average, the median and $\pm 1\sigma$ percentiles of E_γ^B. No strong bias is introduced by any of the observables to any of these quantities. Furthermore, the structure itself remains constant across all bins and no clear dependence on E_γ^B can be seen. No significant differences between different FEI modes are observed. It is therefore concluded that the selections are unbiasing and suitable for signal-side photon background suppression. The exact selections on these observables are optimised simultaneously with continuum event suppression in Section 6.6.

6.5 $e^+e^- \to q\bar{q}$ event suppression

Section 6.4 introduced the strategy for selecting the best photon candidate, as well as selection to suppress events where the photon is misidentified or originates from sources different than $B \to X_s\gamma$. As was seen before in, for example, Figure 6.7, $e^+e^- \to q\bar{q}$ events provide the vast majority of photon candidates. Therefore, a dedicated event selection for this type of background is devised. It takes advantage of different event topologies expected for $\Upsilon(4S) \to q\bar{q}$ and $e^+e^- \to q\bar{q}$ events. Events, where a $\Upsilon(4S) \to B\bar{B}$ decay is present, tend to be more 'spherical' when compared with $e^+e^- \to q\bar{q}$ events exhibiting a 'jet-like' distribution. This is related to the fact that e^+e^- collisions at B-factory experiments have just enough energy to produce a B pair almost at rest, which means that its decay products, on average, tend to be distributed uniformly in polar and azimuthal angles. On the other hand, light quark pairs, produced in e^+e^- collision events, also gain a substantial amount of back-to-back momentum and, as a result, the decay products are collimated along a symmetry axis. The schematic idea of this is shown in Figure 6.15. This Section will provide an in-depth discussion on how the discrimination between $B \to X_s\gamma$ and continuum is achieved using a BDT.
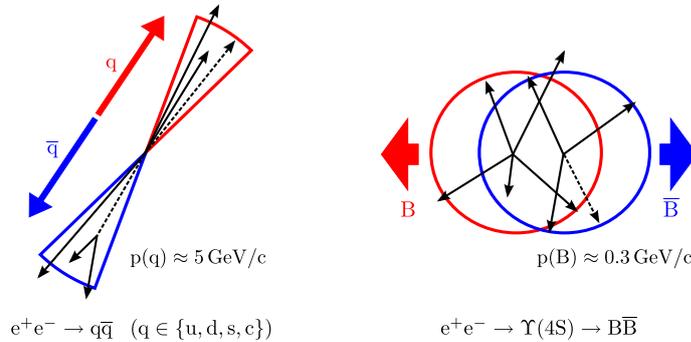

Figure 6.15: Schematic illustration of continuum and $B\bar{B}$ events created in an e^+e^- collision in B-factories. Events, where a B meson is produced, are generally more spherical since the $\Upsilon(4S)$ is produced at rest and its decay products tend to not have a preferred direction. Typical momenta of light-quark and $B\bar{B}$ mesons are shown. The specific directions shown are illustrative only.

6.5.1 Training sample pre-selection

Before a BDT is trained, it is generally desirable to prepare the data sets such that the classifier learns based on relevant data. Such data preprocessing is performed based on the variables described in Section 6.4. The continuum classifier is then trained on the reduced data set.

In order to find the optimal selections, a figure-of-merit study is performed for each observable. Two figure-of-merit options were considered for this analysis: a more standard figure-of-merit FOM_1:

$$\mathrm{FOM}_1 = \frac{\mathsf{S}}{\sqrt{\mathsf{S} + \mathsf{B}}}, \tag{6.9}$$

and FOM_2 defined in Ref. [159] (often referred to as 'Punzi' figure-of-merit):

$$\mathrm{FOM}_2 = \frac{\mathsf{S}}{\mathsf{S}_0} \frac{1}{\frac{3}{2} + \sqrt{\mathsf{B}}}. \tag{6.10}$$

In both equations, S is the number of signal events after selection, B is the number of background events after selection, and S_0 is the number of signal events before selection. Although Equation (6.10) was derived with search-like analyses in mind, it is used in this analysis to minimise signal model dependency: the ratio S/S_0 reduces many model-dependent effects.

For each figure-of-merit calculation, background events (B) are counted based on generic MC, whereas signal events, S, are counted based on signal MC to ensure a significant sample size. In the case of Equation (6.9), an appropriate luminosity scaling for S is also used. Each data set has duplicate tag-B candidates randomly removed by picking a random tag-side candidate. Each figure-of-merit is then calculated for 200 equally spaced selections in the target observable. The maximum figure-of-merit point is taken as the optimal pre-selection for each of the variables. This procedure is shown for FOM_2 in Figure 6.16. Results for FOM_1 are used as a cross-check for FOM_2 and turn out to be consistent. This is shown in Appendix E. The results are consistent between the two figure-of-merits as well as B^+ and B^0 modes. Therefore, due to the model independence of FOM_2, this figure-of-merit is the only one discussed henceforth.

At this stage, it is unnecessary to choose the 'best' selection, as another simultaneous optimisation will be performed in Section 6.6, together with continuum suppression BDT output. The main goal is to reduce the sample size to include only relevant data such that the trained BDT can make decisions for difficult cases that are not easily distinguishable using a simple selections. The pre-selections are chosen to suppress background but retain most of the signal, based on Figure 6.16. They are set to be considerably looser than their optimal selection. Specifically, the requirements for a loose selection are tailored such that roughly 75% of $B \rightarrow X_s \gamma$ candidates are retained. They are shown in Table 6.8.

The pre-selections improve the signal-to-background ratio by roughly an order of magnitude. This is seen by comparing Figure 6.17 with Figure 6.7. The $B \rightarrow X_s \gamma$ signal MC scale differs by about a factor of 10, highlighting the background suppression efficiency.

Finally, many combinatorial tag-side candidates in $B\overline{B}$ events may still contribute to the analysis at this stage. A more detailed definition for a 'well-reconstructed' tag is explored in Section 6.7.3. At this stage, it is sufficient to acknowledge that the vast majority of correctly reconstructed tag-side candidates are expected to have $M_{\mathrm{bc}} > 5.27$ GeV/c^2. Therefore, this requirement is also adopted for optimisation studies and training in Sections 6.5.2 to 6.5.4.
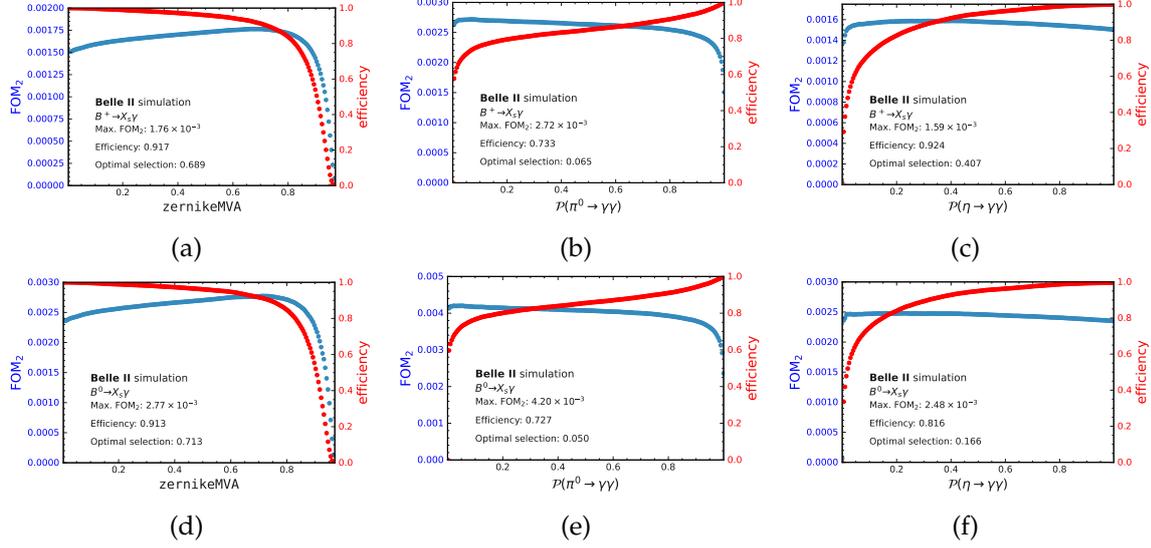
Figure 6.16: Optimal selection calculation for observables described in Section 6.4 based on FOM$_2$ (see Equation (6.10)). For $B^+ \rightarrow X_s\gamma$ events the tests are shown in (a) – (c), and for $B^0 \rightarrow X_s\gamma$ in (d) – (f). The Figures show the efficiency and FOM$_2$ score calculated by scanning 200 thresholds of $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$, $\mathcal{P}(\eta \rightarrow \gamma\gamma)$ and `zernikeMVA`. The maximum value of FOM$_2$, the corresponding threshold and efficiency are shown as well.

Table 6.8: Selections that remove background and misreconstructed candidates, preparing the reconstructed data sets (Section 6.3.3) for continuum BDT training (Section 6.5.3). A later optimisation will be used for a final candidate selection in Section 6.6.

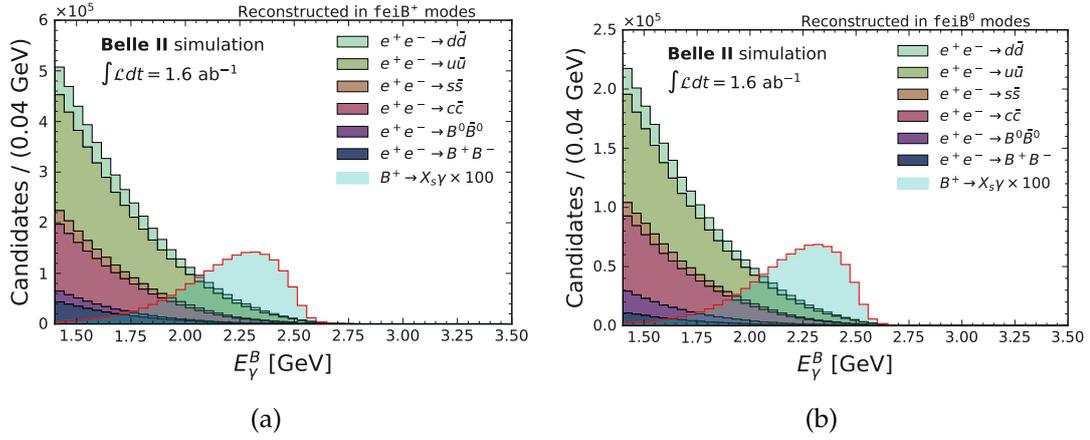| Variable	Loose selections
`zernikeMVA`	> 0.5
$\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$	< 0.4
$\mathcal{P}(\eta \rightarrow \gamma\gamma)$	< 0.4
B^+ mode: γ candidate retention efficiency	75.4%
B^0 mode: γ candidate retention efficiency	76.7%

Figure 6.17: $B \rightarrow X_s\gamma$ spectrum in generic MC reconstructed in `feiB⁺` and `feiB⁰` modes after pre-selection for training of the continuum BDT classifier. Overlaid are events from signal MC where the photon comes from $B \rightarrow X_s\gamma$, multiplied by a scaling factor. Compared to Figure 6.7, the effectiveness of background suppression so far is apparent. These Figures may include multiple tag-B entries per event.

6.5.2 Continuum suppression feature selection

Belle II has a large selection of observables that can be used for continuum suppression that are suitable to be used as input features to a BDT. These observables describe the event topology and other collective particle decay properties. They are optimised to provide optimal separation between $B\bar{B}$ and $q\bar{q}$ events. Two caveats have to be kept in mind for the $B \rightarrow X_s\gamma$ analysis:

- Generally, the $B \rightarrow X_s\gamma$ event topology may be different compared to $B\bar{B}$ events. $B \rightarrow X_s\gamma$ decays have a single jet-like X_s system, while the other B meson decays hadronically. This leads to a somewhat middle-ground between a generic-$B\bar{B}$ event and an $e^+e^- \rightarrow q\bar{q}$ event, as it was illustrated in Figure 6.15.

- Many of these observables contain momenta, angles or other parameters of some (or even all) particles in the event – including the X_s system and the photon. This may lead to a bias in the E_γ^B spectrum. Furthermore, even relatively small biases over many training features may be learnt by the BDT and introduced to the spectrum.

- Some features may perform differently in real data compared to simulation due to unexpected differences in alignment, calibration or background distributions. As simulated data sets are used to train a BDT in this analysis, such a comparison is crucial.

Given the mentioned points, it is important to test that the observables used for the training provide adequate separation between $B \rightarrow X_s\gamma$ and $q\bar{q}$, while no bias is introduced to the photon energy spectrum. Furthermore, this has to be well-modelled in Belle II data.

In this analysis, the following observable categories are considered for separation between $e^+e^- \rightarrow q\bar{q}$ and $B \rightarrow X_s\gamma$:

- Thrust-based observables (Appendix F.1);

- Sphericity and aplanarity (Appendix F.2);

- Harmonic moments (Appendix F.3);

- Fox-Wolfram moments (Appendix F.4);

- Modified Fox-Wolfram moments (Appendix F.5);

- CLEO cones (Appendix F.6);

- Tag-B meson vertex observables (Appendix F.7);

- Flavour tagger output for the tag-B meson (Appendix F.8).

In total, this provides 75 potential training features that are tested to be uncorrelated to the photon energy spectrum and adequately described in simulation. The tests use a metric of *total divergence to the average* (often called *Jensen-Shannon distance*) [160], which is used to quantitatively evaluate the similarity between two distributions. The Jensen-Shannon distance is bounded by 1 for two given probability distributions, $\mathbb{X}_1$ and $\mathbb{X}_2$:

$$0 \leq \text{JSD}(\mathbb{X}_1 || \mathbb{X}_2) \leq 1, \tag{6.11}$$

where exactly similar distributions have a score of 0. The score tends towards 1 when the distributions are highly different.

Two tests are performed:

- **Test 1**: $\mathbf{E_\gamma^B}$, $\mathbf{E_\gamma^*}$ **and tag-side** $\mathbf{M_{bc}}$ **bias test**: to ensure that the classifier does not indirectly select particular X_s or tag-side B channels, each potential training feature is separated into 5 equally populated regions (*slices*) of $B \to X_s \gamma$ events in signal MC. For this test, $B^+ \to X_s \gamma$ and $B^0 \to X_s \gamma$ are merged. In each of these regions, the slices of E_γ^B, E_γ^* and M_{bc} are compared. The Jensen-Shannon distance is required to not be larger than 0.06 between any two given slices of a training feature. The requirement to pass **Test 1** has been chosen by observing the typical values of the agreement shown by the tested unbiased distributions. If this requirement is not passed by at least one of the distributions (E_γ^B, E_γ^* or M_{bc}), the feature is excluded from the list of final BDT training features.

- **Test 2**: $\mathbf{e^+e^-} \to \mathbf{q\overline{q}}$ **data-simulation similarity test**: to ensure that MC adequately describes the Belle II data sets. This test is only performed if **Test 1** is passed. As this is a blinded analysis, off-resonance data samples are used, which only contain $e^+e^- \to q\overline{q}$ events. In this case, the Jensen-Shannon distance is calculated between the area-normalised distribution of a training feature in the off-resonance data set and the area-normalised distribution of a training feature in the continuum MC. The metric is required to be no larger than 0.1. A looser requirement is adopted here because some difference is expected between the distributions: the collision energy in the off-resonance data is different to the one in the on-resonance simulation. Furthermore, an overall smaller off-resonance data set (~ 19 fb^{-1}, see Section 6.2) may have certain differences due to statistical fluctuations.

The tested distributions include the selections from Section 6.5.1, except for the case of the off-resonance data, where the $M_{\mathrm{bc}} > 5.27$ GeV/c^2 requirement is lifted. For every event, when more than one tag-side candidate B candidate exists, a random one is picked. The procedure of **Test 1** with exact definitions for the 75 observables is given in Appendix F.

Out of 75 potential training features, 29 pass the requirements of **Test 1**. These are passed to **Test 2**. This requirement only removes a single feature which attests to the good representation of data by the Belle II MC. The results for the 26 final observables that pass both test requirements and are used as features in the BDT training are shown in Table 6.9.

6.5.3 Continuum suppression training

As it was argued in Section 6.5.2, events containing $B \to X_s\gamma$ decays may have slightly different kinematic properties compared to a generic-$B\bar{B}$. Although these differences are not large, as seen in Appendix G, training a classifier to separate generic $\Upsilon(4S) \to B\bar{B}$ and $e^+e^- \to q\bar{q}$ events may lead to a suboptimal separation of $B \to X_s\gamma$.

A more effective setup is to remove $B\bar{B}$ events from generic MC and supplement the leftover events with $B \to X_s\gamma$ events from signal MC. In such a scenario, the classifier learns to distinguish between the signal decays and continuum events without the additional ambiguity of including generic $B\bar{B}$ decays.

The training samples are prepared by creating a mixture of 100000 $e^+e^- \to q\bar{q}$ events and 100000 $B \to X_s\gamma$ events from the signal MC sample. All selections from Table 6.8 are employed for the training data sets. In each event, one γ and tag-B candidate combination is randomly chosen. This requirement ensures that the same event cannot contribute multiple training entries. The target label for the training is defined as a `flag` which follows

$$\texttt{flag} = \begin{cases} 0, & \text{for } e^+e^- \to q\bar{q} \text{ events,} \\ 1, & \text{for } B \to X_s\gamma \text{ events.} \end{cases} \tag{6.12}$$

Half of the $e^+e^- \to q\bar{q}$ training sample is taken from the `feiB`$^+$ mode, and the other half is from `feiB`^{0} modes. For signal, half is taken from the $B^+ \to X_s\gamma$ signal mode, the other half from the $B^0 \to X_s\gamma$, irrespective of the FEI mode. An equivalent sample is prepared as the testing sample for the training.

The training is performed using a `FastBDT` algorithm, introduced in Section 5.2.2. Four hyperparameters within the `FastBDT` framework have to be chosen. Hyperparameter optimisation is performed in a grid-like search, based on two quantities:

$$\begin{aligned} &\mathrm{AUC}_{\text{test}}; \\ \Delta\mathrm{AUC} &\equiv |\mathrm{AUC}_{\text{train}} - \mathrm{AUC}_{\text{test}}|, \end{aligned} \tag{6.13}$$

where $\mathrm{AUC}_{\text{train/test}}$ is the AUC score for the training or testing samples. A set of hyperparameters is sought, such that $\mathrm{AUC}_{\text{test}}$ is maximised while $\Delta\mathrm{AUC}$ is minimised. The results of hyperparameter optimisation are summarised in Table 6.10 and shown in Figure 6.18. Larger depth or number of trees are not explored to avoid non-linear biases which may be learnt by the classifier and would require additional studies to pinpoint. Large shrinkage values are undesired to ensure an adequate learning rate of the classifier.

Table 6.9: The training features for the $e^+e^- \to q\bar{q}$ suppression that pass the requirements of **Test 1** (see Appendix F) and **Test 2** (see Appendix G). The Table also shows the value of the Jensen-Shannon distances for each observable for the different requirements of both tests. Exact definitions of these quantities are provided in Appendix F. Observable groups follow those introduced in the text.

Feature name	Jensen Shannon Distances [$\sqrt{\text{bit}}$]			
	Test 1			Test 2
	E_γ^B	E_γ^*	M_{bc}	Data–Sim.
Thrust related				
$\cos\theta_{\text{TB}\wedge\text{TO}}$	0.03	0.01	0.04	0.02
$\cos\theta_{\text{TB}\wedge z}$	0.01	0.01	0.02	0.01
T_{B}	0.04	0.03	0.04	0.06
$\cos\theta_{\text{T}}$	0.02	0.02	0.01	0.02
Harmonic moments				
B_1^T	0.05	0.05	0.02	0.01
B_3^T	0.04	0.03	0.01	0.03
CLEO cones				
CC_0^B	0.04	0.03	0.03	0.03
CC_1^B	0.03	0.03	0.02	0.03
CC_2^B	0.04	0.03	0.02	0.01
CC_3^B	0.05	0.05	0.02	0.02
CC_0	0.05	0.05	0.02	0.02
CC_3	0.06	0.06	0.02	0.02
Modified Fox-Wolfram moments				
H_{c4}^{so}	0.05	0.04	0.02	0.02
H_{m2}^{so}	0.03	0.03	0.02	0.02
H_{m4}^{so}	0.03	0.03	0.01	0.01
H_0^{oo}	0.03	0.02	0.02	0.04
Tag-B meson vertex observables				
z of tag-B	0.01	0.01	0.01	0.02
Δx of tag-B	0.02	0.02	0.03	0.03
Δy of tag-B	0.01	0.01	0.03	0.04
Δz of tag-B	0.02	0.02	0.04	0.02
$\Delta\tau$	0.04	0.03	0.02	0.02
Δz	0.04	0.03	0.02	0.02
Δz_B	0.04	0.03	0.02	0.02
$\chi^2_{B_{\text{ROE}};\text{IP}}$	0.05	0.05	0.01	0.00
$x_{B_{\text{ROE}}}$	0.03	0.03	0.01	0.02
$z_{B_{\text{ROE}}}$	0.03	0.03	0.01	0.05

Table 6.10: Hyperparameter optimisation based on a grid-search method. The four hyperparameters for the FastBDT algorithm are defined in Section 5.2.2. The optimal values are chosen based on criteria defined in Equation (6.13). They are shown in the rightmost column and taken as the parameters for the training. The corresponding AUC$_{\text{test}}$ and ΔAUC are shown in Figure 6.18 and Figure 6.19b.

Hyperparameter	Tested grid values	Chosen optimal parameter
depth	$\{1, 2, 3\}$	2
number of trees	$\{100, 200, 400, 600, 1000\}$	400
shrinkage	$\{0.01, 0.05, 0.1, 0.3, 0.5\}$	0.1
training subsample	$\{0.2, 0.4, 0.5, 0.6, 0.8\}$	0.8
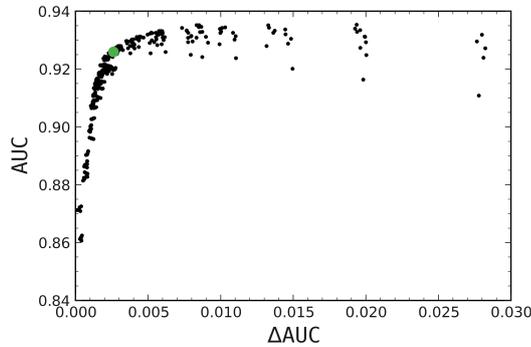
Figure 6.18: Hyperparameter optimisation as a function of quantities defined in Equation (6.13). In total 375 working points are tested. The choice of parameters shown in Table 6.10 is represented by the enlarged green point. It is chosen in the threshold area, where AUC$_{\text{test}}$ gain saturates and ΔAUC gain accelerates.

The training is performed with features from Table 6.9 and hyperparameters from Table 6.10. The normalised classifier output for test and train samples is shown in Figure 6.19a. It is seen that the classifier shows almost no bias, as the train and test samples agree very well. This is further alluded to by inspecting the ROC curve in Figure 6.19b, where no significant differences are observed. The difference in the AUC scores is smaller than 1%.
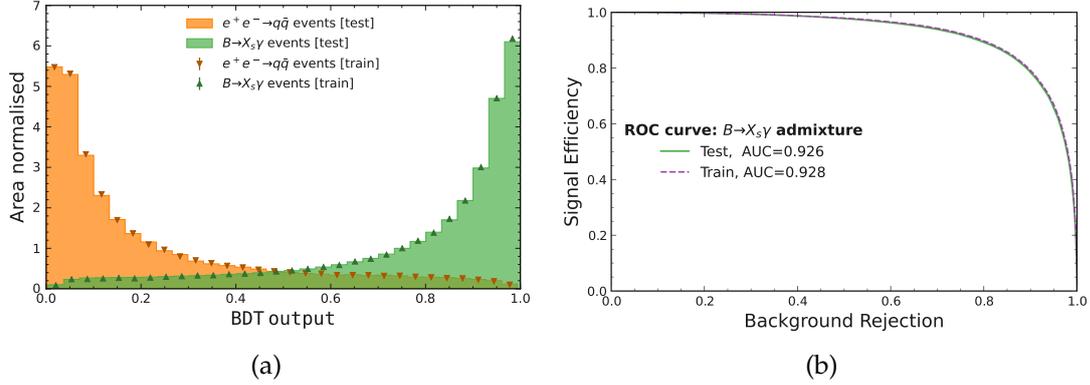

(a)   (b)

Figure 6.19:   The training evaluation for this analysis.   Excellent separation between $e^+e^- \to q\bar{q}$ and $B \to X_s\gamma$ samples and good agreement between corresponding test and train samples is seen in (a). The ROC curve of the training is shown in (b). The test and train sample AUC scores are above 0.9 and differ little, alluding to a high-separation power that is observed and no evidence of overtraining.

Using the tools provided by the `FastBDT` algorithm, a *relative feature importance* is computed. Particularly for `FastBDT`, it is computed by evaluating the decrease of the AUC score if the feature is not included in the training data set (for more details see Ref. [140]). Therefore, it can be considered a quantitative measure of the impact of a feature on the final classifier output. The relative training observable importance is shown in Figure 6.20. It is seen that $\cos\theta_{\mathrm{TB \wedge TO}}$ (the angle between the thrust axis of the tag-B candidate and everything else in the event) has by far the largest impact on the classifier. This is not surprising after inspecting the individual separation power for the current problem in Appendix G. Other important separation features come from Δz_B & Δz_B, related to the longitudinal distance between the decay vertices, thrust of the tag-B meson, T_B, and CLEO cone variables $\mathtt{CC}_i$. For consistency, the events used in the training are removed from further analysis.

6.5.4   Continuum suppression validation

The resulting `BDT output` is tested to not bias the photon energy spectrum, further ensuring the validity of the training. Following the tests for bias of the average, median and 1σ percentiles in Section 6.4.5, the same study is performed for the training classifier output. The result is shown in Figure 6.21a for the $B \to X_s\gamma$ admixture sample with the same requirements as the training sample. No significant bias to any of the photon energy spectrum quantities is observed in any of the intervals. In Figure 6.21b, the trained classifier output is also applied to the off-resonance data. Comparing the shapes of off-resonance data and simulation an adequate agreement is observed, particularly in the high-`BDT output` region where
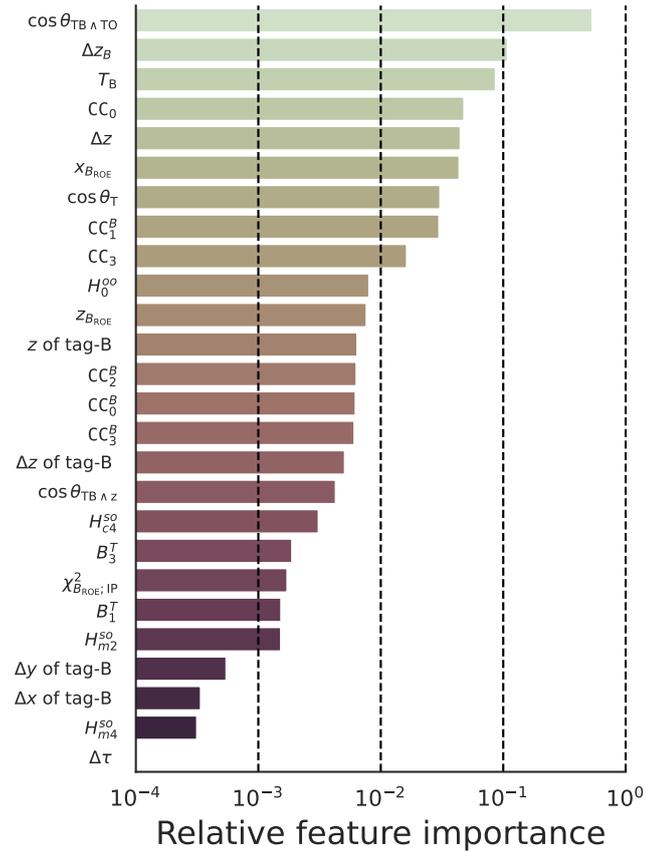
Figure 6.20: The relative feature importances of different observables used in the training. The definitions of these observables are provided in Appendix F. The feature importance highlights a relative change in AUC score when the observable is not included in the training.

most signal lies. The results seen in Figure 6.21 validate the fact that the BDT is prepared in an unbiased way and strongly suppresses the continuum events.
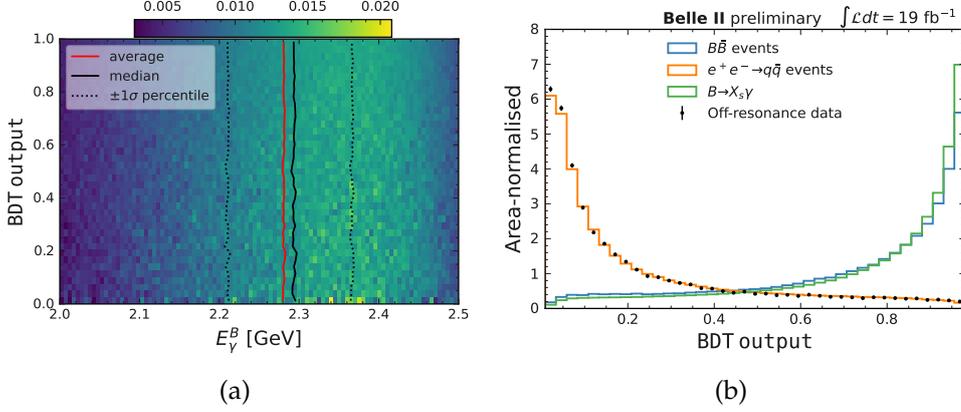

(a)   (b)

Figure 6.21: Validation of the training described in Section 6.5.3, for the mixture of `feiB`$^+$ and `feiB`^{0} modes. The correlation test as a 2D histogram for the trained classifier output is shown in (a). Each row is normalised such that all bins within that row add up to 1. The red line shows the average photon energy, $\langle E_\gamma^B \rangle$, as a function of the tested observable. In black and black-dotted lines: median and $\pm 1\sigma$ percentile values of E_γ^B, respectively. No strong dependence can be observed in any of the quantities or the 2D maps. The shape of $e^+e^- \to q\bar{q}$ and off-resonance data is compared in (b). It is seen that both simulated and off-resonance data sets are well separated from $B\bar{B}$ and $B \to X_s\gamma$ simulation.

6.6   Final selection optimisation

The last two Sections introduced the selection to suppress photons originating in non-$B \to X_s\gamma$ decays, particularly from π^0 and η decays (Section 6.4), and the strategy to suppress $e^+e^- \to q\bar{q}$ events that were the dominant component in the selected data set (Section 6.5). Although a preselection was already developed to prepare an adequate training sample for the BDT, a more optimal ('tighter') selection is desired to ensure the optimal efficiency and purity of the selected sample. This Section describes the approach taken to find such optimal selection and calculate the efficiency loss for all applied selections.

6.6.1   Simultaneous selection optimisation

After the pre-selection that prepared the data for training a BDT in Table 6.8, a more robust strategy for tighter selections is developed. In particular, the BDT output, $\mathcal{P}(\pi^0 \to \gamma\gamma)$, $\mathcal{P}(\eta \to \gamma\gamma)$ and `zernikeMVA` may be interconnected in the sense that applying the selection on one of them influences a selection on the others. To find an optimal selection point, each threshold is optimised in an iteration-based approach. At each step, one variable threshold is optimised to a value that gives the best figure-of-merit score, while keeping the other requirements unvaried. Then, this is repeated for other variables. Each individual optimi-

sation is equivalent to that shown in Figure 6.16 and uses figure-of-merit FOM$_2$, defined in Equation (6.10).

In order to maximise the efficiency of the optimisation on correctly reconstructed events without adhering to a more strict definition at this stage, this procedure is performed on one randomly selected peaking tag-B candidate per event ($M_{\rm bc} > 5.27$ GeV/c^2) combined with the highest energy photon. The starting point for each selection corresponds to the values in Table 6.8. The starting BDT output selection is chosen at 0.5. The $B \to X_s \gamma$ admixture of charged and neutral modes is used. The $e^+ e^- \to q\overline{q}$ and generic $B\overline{B}$ background events from feiB$^+$ and feiB^0 modes are merged. This aims to reproduce 'realistic' data conditions, where different background efficiencies may be observed due to different behaviours of feiB$^+$ and feiB^0 modes.

After performing the optimisation for each selection, the optimisaton steps are repeated 9 more times. The selections converge and do not vary after round 3 of optimisation. The converged values are shown in Table 6.11.

Table 6.11: Optimal selections chosen for this analysis, based on the iterative approach described in Section 6.6.1. The values for BDT output and zernikeMVA are chosen near those that are found optimal. For $\mathcal{P}(\pi^0 \to \gamma\gamma)$ and $\mathcal{P}(\eta \to \gamma\gamma)$ the choice is made based on the availability of data-simulation agreement studies performed at Belle II. At the time of preparing the analysis, only studies with $\mathcal{P}(\pi^0 \to \gamma\gamma)$ and $\mathcal{P}(\eta \to \gamma\gamma)$ thresholds up to 0.4 were performed (see Section 6.11.2).

Variable	Figure-of-merit maximised at	Final chosen
zernikeMVA	> 0.629	0.6
$\mathcal{P}(\pi^0 \to \gamma\gamma)$	< 0.258	0.4
$\mathcal{P}(\eta \to \gamma\gamma)$	< 0.036	0.4
BDT output	> 0.798	0.8

For $\mathcal{P}(\pi^0 \to \gamma\gamma)$ and $\mathcal{P}(\eta \to \gamma\gamma)$, the found selection is relatively tight, if inspecting Figures 6.12a to 6.12d. Furthermore, at the time of preparation of the analysis described, studies regarding the $\mathcal{P}(\pi^0 \to \gamma\gamma)$ and $\mathcal{P}(\eta \to \gamma\gamma)$ applicability to such tight selections were not available. Therefore, it was decided to not tighten this selection further than the pre-selection value obtained in Table 6.8. Repeating the study while keeping $\mathcal{P}(\pi^0 \to \gamma\gamma)$ and $\mathcal{P}(\eta \to \gamma\gamma)$ selection at 0.4 yields compatible results to those shown in Table 6.11. Other selections are retained based on the optimal value from the initial 10 iterations.

6.6.2 Summary and efficiency of all analysis selections

Table 6.12 summarises all the selections and BDT training results from Sections 6.4 to 6.6, and lists the final $B \to X_s \gamma$ candidate retention. The retention, in this case, is defined as:

$$r_{\rm cand} = \frac{N_{B \to X_s \gamma} \text{ candidates after cut}}{N_{B \to X_s \gamma} \text{ no cut}},$$
(6.14)

which is an approximation as it may include multiple tag-B candidates. In the table, the $M_{\rm bc} > 5.27$ GeV/c^2 requirement is no longer applied and all tag-B meson candidates are kept (i.e. the same high energy photon candidates may contribute more than once).

Table 6.12: The summary of all selections and their retentions, based on Equation (6.14). The selections listed here are applied on official Belle II `feiB+` and `feiB0` samples, described in Section 6.3.3. The columns show efficiency for $B \to X_s\gamma$ events, calculated on signal MC, continuum and $B\bar{B}$ events, both of which are calculated on generic MC. It can be seen that continuum events are suppressed by roughly two orders of magnitude, whereas generic-$B\bar{B}$ decays by more than an order of magnitude.

`feiB+` mode reconstruction

Selection	$B \to X_s\gamma$	Continuum	$B\bar{B}$ events
		Retention	
none	1.000	1.000	1.000
E_γ rank = 1	0.998	0.966	0.976
`zernikeMVA > 0.6`	0.944	0.654	0.696
$\mathcal{P}(\pi^0 \to \gamma\gamma) < 0.4$	0.831	0.215	0.314
$\mathcal{P}(\eta \to \gamma\gamma) < 0.4$	0.921	0.764	0.768
`BDT output > 0.8`	0.562	0.025	0.485
tag-$M_{bc} > 5.245$ GeV/c^2	0.949	0.886	0.929
all	0.421	0.005	0.073

`feiB0` mode reconstruction

Selection	$B \to X_s\gamma$	Continuum	$B\bar{B}$ events
		Retention	
none	1.000	1.000	1.000
E_γ rank = 1	0.998	0.968	0.979
`zernikeMVA > 0.6`	0.945	0.657	0.690
$\mathcal{P}(\pi^0 \to \gamma\gamma) < 0.4$	0.841	0.222	0.324
$\mathcal{P}(\eta \to \gamma\gamma) < 0.4$	0.927	0.782	0.774
`BDT output > 0.8`	0.554	0.025	0.479
tag-$M_{bc} > 5.245$ GeV/c^2	0.946	0.884	0.923
all	0.421	0.005	0.073

The background suppression procedure, shown in Table 6.12, roughly halves the number of available $B \to X_s\gamma$ events in the sample. However, the background candidates from $e^+e^- \to q\bar{q}$ processes are reduced 200 times. Furthermore, generic-$B\bar{B}$ event contribution is estimated at 7% of the original, which means more that an order of magnitude suppression is achieved. The photon energy spectrum, after these selections, is shown in Figure 6.22. Compared with the previous versions of this Figure, e.g. Figure 6.7, a much better signal-to-background ratio is obtained.
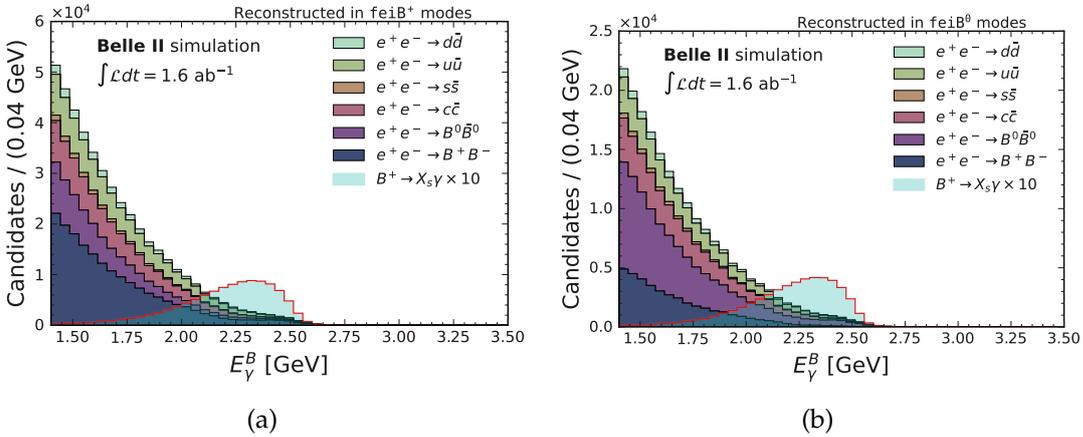

(a)

(b)

Figure 6.22: $B \to X_s\gamma$ spectrum in generic MC after event reconstruction in `feiB+` and `feiB0` modes with optimal background suppression selections listed in Table 6.12. Overlaid are events from signal MC where the photon comes from $B \to X_s\gamma$, multiplied by a scaling factor, with the same selections applied. These Figures may include a high energy photon combined with multiple tag-B entries per event and can be compared directly with Figure 6.7 where it is seen that the signal-to-background ratio for $B \to X_s\gamma$ is 100 times higher.

6.7 Tag-side B candidate selection

In Section 6.3.3 it was discussed that in about half of all reconstructed events there exists more than one tag-side candidate. That does not take into account the overlap between `feiB`$^+$ and `feiB`^{0} mode, which further enhances this effect. Performing the best tag-B candidate selection is important, as multiple entries per event should not be included in the final sample. However, the interest in this analysis lies in the signal side which decays as $B \rightarrow X_s \gamma$, which means that a standard Belle II 'truth-matching' procedure is too strict. In principle, the requirement is to only reconstruct a sample of tag-B mesons that *provide good kinematic constraints to the signal side*. In this Section, the best tag-side candidate selection and a concrete definition for tag-B mesons with correctly reconstructed kinematic properties is introduced.

6.7.1 Selection within the same FEI mode

The number of tag-side candidates for `feiB`$^+$ and `feiB`^{0} modes, after the optimised selections in Table 6.12, is shown in Figure 6.23. Overall, comparing to Figure 6.6a, the candidate fractions are similar. This attests to the fact that background (and particularly continuum) suppression was done without introducing a bias in preferentially selecting events with large $\mathcal{P}_{\text{FEI}}$ in Sections 6.4 and 6.5. About 67% (74%) of events for `feiB`$^+$ (`feiB`^{0}) modes have only one tag-side candidate. About 19% (17%) of events for `feiB`$^+$ (`feiB`^{0}) modes have two tag-side candidates and 7% (5%) have three. The number of candidates per event reduces quickly, but faster for B^0 modes, with roughly 2% (1%) of events having more than 5 candidates for B^+ and B^0. Note that the same event can have a B^+ and B^0 event reconstructed.
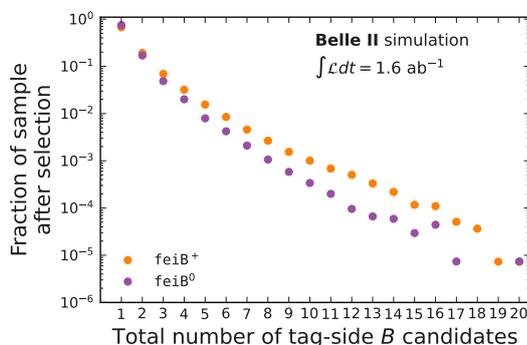

Figure 6.23: The relative fractions of events for the number of B meson candidates in the generic MC data set after the background suppression selections in Table 6.12. This Figure can be directly compared with Figure 6.6a. The overall fractions are similar, confirming a valid background suppression procedure which does not introduce a preference towards specific tag-side modes.

However, even after all selections there often exists more than one B meson + photon combination. The first step of the selection of best tag-B is choosing one candidate each in `feiB`$^+$ and `feiB`^{0} modes. While a general approach could be developed, it was observed that at this stage a particular choice of the tag does not influence the resolution or the average value of the spectrum strongly. This is visualised in Figure 6.24. For both neutral and charged

$B \to X_s \gamma$ modes the distributions look similar whether the highest-$\mathcal{P}_{\text{FEI}}$ candidate is selected in each event or a random tag-B meson is chosen as the main candidate. On the other hand, the M_{bc} distribution, as expected, has a more distinct peak for the case where the highest $\mathcal{P}_{\text{FEI}}$ candidate is picked in each event. The latter result for $B \to X_s \gamma$ is shown in Figure 6.25. The Figure also includes a similar M_{bc} test for the continuum events.

(a)   (b)
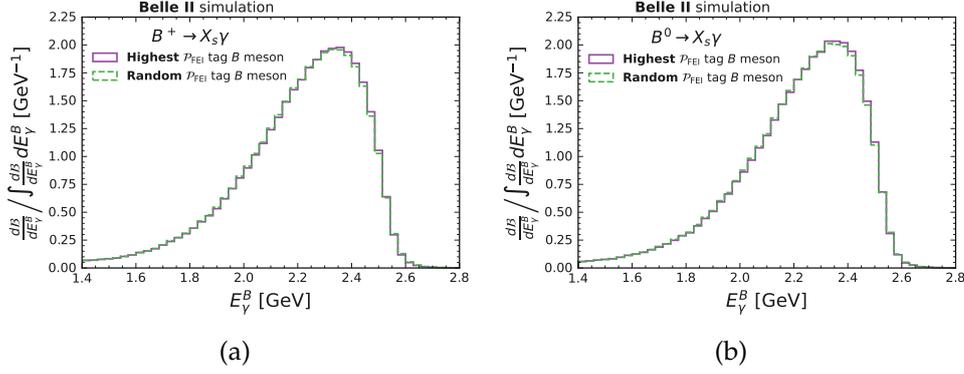
Figure 6.24: The photon energy spectrum after selecting a single tag-B meson candidate per event either randomly or by requiring the largest $\mathcal{P}_{\text{FEI}}$. This is shown for $B^+ \to X_s \gamma$ events in (a) and for $B^0 \to X_s \gamma$ in (b). The Figures are normalised to their total integral value such that a shape comparison can be performed. The difference between the distributions is negligible.

As it is desirable to emphasise the contrast between continuum and B events for the fitting step that will follow (see Section 6.8) the highest $\mathcal{P}_{\text{FEI}}$ candidate in each event is chosen as the B candidate with virtually no bias to the resolution. However, the study here, as of yet, does not address the cases when a candidate in the same event is reconstructed in the feiB$^+$ and feiB^0 modes. Therefore, for now, both candidates are kept in such events and the study is continued in Section 6.7.2.

6.7.2   Selection between feiB$^+$ and feiB^0 modes

Section 6.7.1 showed that one can select the highest $\mathcal{P}_{\text{FEI}}$ candidate from feiB$^+$ and feiB^0 without a significant effect on the E_γ^B resolution and with an enhancement of the M_{bc} distribution peak. It reduced each event to a single tag-B and photon combination in most events. However, it did not address the case when there is a candidate reconstructed in both feiB$^+$ and feiB^0 modes: implying that events may still have up to two combinations. Such cases are evaluated to happen roughly 10.5% of the time. For the sample where two B candidates exist, two quantities are calculated

$$\mathcal{A}_{\text{tag}} = \frac{\mathcal{P}_{\text{tag}}(\texttt{feiB}^+) - \mathcal{P}_{\text{tag}}(\texttt{feiB}^0)}{\mathcal{P}_{\text{tag}}(\texttt{feiB}^+) + \mathcal{P}_{\text{tag}}(\texttt{feiB}^0)}, \tag{6.15}$$

which is called the asymmetry of $\mathcal{P}_{\text{FEI}}$ between a feiB$^+$ and feiB^0 candidate in the same event, and

$$\Delta(M_{\text{bc}}) = M_{\text{bc}}(B^+) - M_{\text{bc}}(B^0), \tag{6.16}$$
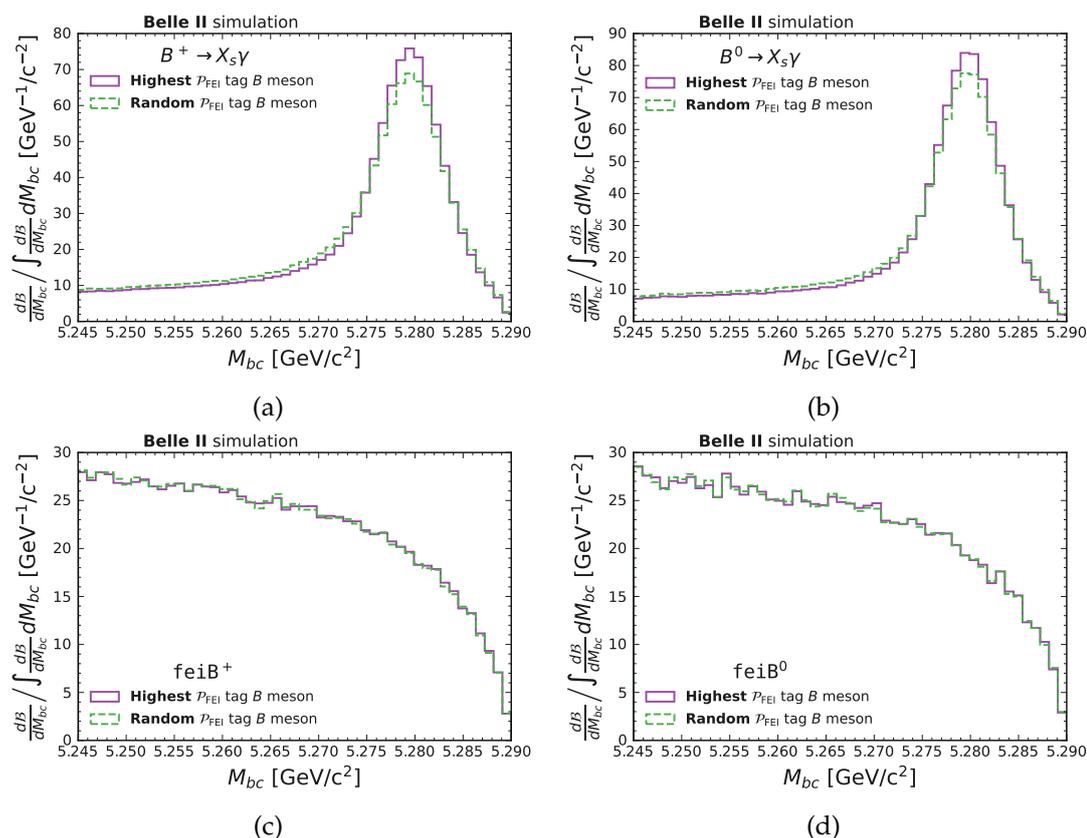
Figure 6.25: The M_{bc} shapes for $B \to X_s\gamma$ signal MC ((a) and (b)) and $e^+e^- \to q\bar{q}$ events from generic MC ((c) and (d)) after selecting a single tag-B meson per event either randomly or by requiring the largest $\mathcal{P}_{\text{FEI}}$. As seen in (a) and (b), the difference in the M_{bc} distribution for $B^+ \to X_s\gamma$ and $B^0 \to X_s\gamma$ mostly pronounced in the peak region. On the other hand, (c) and (d) show no strong dependence in shape irrespective of the way the tag-side candidate is chosen. The Figures are normalised to their total integral value such that a shape comparison can be performed. This observation motivates the selection of the highest-$\mathcal{P}_{\text{FEI}}$ candidate.

which is the difference in M_{bc} value of the two candidates. The $\mathcal{A}_{\mathrm{tag}}$ tends to zero if they both have a similar $\mathcal{P}_{\mathrm{FEI}}$ and to $\pm$ unity if one of the candidates has a much larger $\mathcal{P}_{\mathrm{FEI}}$. The $\Delta(M_{\mathrm{bc}})$ is a difference in M_{bc} between both of the candidates. These quantities are visualised in a two-dimensional grid in Figure 6.26. The sample is split into two subsamples, where a real B^+ (Figure 6.26a) or B^0 (Figure 6.26b) hadronic decay is present on the tag-side. If a B^+ candidate is present, $\mathcal{A}_{\mathrm{tag}} \approx 1$ and $\Delta(M_{\mathrm{bc}}) \gtrsim 0$ for the majority of the candidates. For B^0 candidates the opposite is true: $\mathcal{A}_{\mathrm{tag}} \approx -1$ and $\Delta(M_{\mathrm{bc}}) \lesssim 0$ for the majority of the candidates. This result implies that if the true candidate is a B^+ (B^0), then the value of $\mathcal{P}_{\mathrm{FEI}}$ is higher for feiB$^+$ (feiB^0) candidates. The $\Delta(M_{\mathrm{bc}})$ distribution is interpreted by acknowledging that the incorrect candidate is more likely to be present in the M_{bc} tail, rather than the peak region.
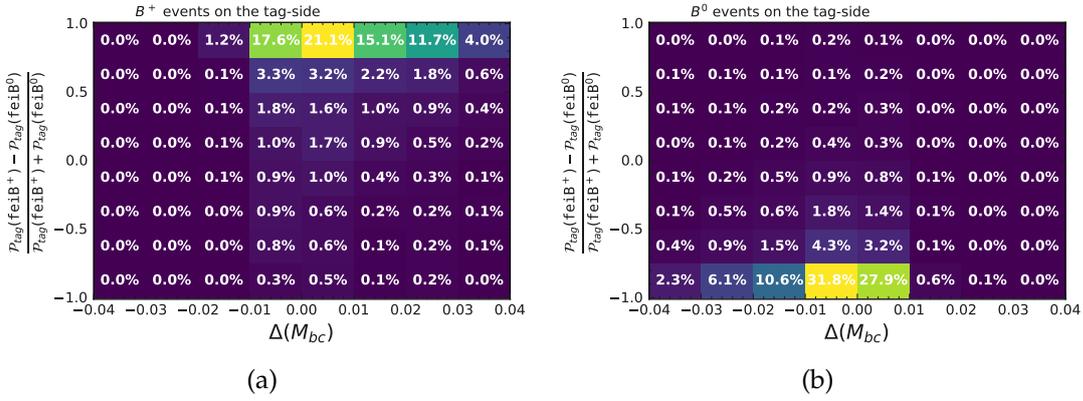

Figure 6.26: A two-dimensional grid of $\mathcal{A}_{\mathrm{tag}}$ (Equation (6.15)) and $\Delta(M_{\mathrm{bc}})$ (Equation (6.16)) for events that have two B candidates from feiB$^+$ and feiB^0 modes. Candidates with a real tag-side B^+ decay (a) and a real tag-side B^0 decay (b) are shown. Most candidates lie near $\mathcal{A}_{\mathrm{tag}} \approx +1(-1)$ and $\Delta(M_{\mathrm{bc}})$ value tends to be positive (negative) for $B^+(B^0)$ candidates. The percentages show the fraction of candidates falling in a two-dimension bin.

Based on the observations discussed in this Section, it can be concluded that it is appropriate to select the FEI candidate with the highest signal probability even when selecting between different feiB$^+$ and feiB^0 modes. This result finalises the best-candidate selection in this analysis. It is now ensured that every photon candidate corresponds to a unique tag-side B meson candidate.

6.7.3   Truth-level tag-B mesons

In particle physics simulation it is possible to rely on the information provided by the MC generators to trace back the measured and reconstructed particles to the generated particles. This procedure is called *truth-matching* and is the usual way to associate 'measured objects' and 'generated objects'. In particular, the truth-matching requirements ensure that all the particles have the correct particle species hypotheses, momenta and energies determined. Furthermore, if unstable, the same should be true for their entire decay chain. In signal MC, this allows to study, for example, exclusively the the signal events with successful reconstruction. However, in the case of tag-B meson reconstruction, it is not particularly important to know that the B meson and its entire decay chain are reconstructed fully correctly. The most crucial

requirement for a tag-*B* is the consistent kinematic constraint that it can provide to infer the information about the signal-side. In this analysis, a good kinematic constraint manifests as resonant behaviour in M_{bc} for the tag-*B* meson distribution.

Using `basf2` truth-level information, the reconstructed tag-*B* mesons are subdivided into 11 categories based on the differences from the generated decay chains. The categories are as follows:

1. The tag-*B* and all its daughters are correctly identified and associated;

2. A final-state radiation photon is not reconstructed;

3. The tag-*B* contains more non-final state particles (i.e. resonances) that were not reconstructed;

4. The tag-*B* was reconstructed from the secondary decay product which implies that a wrong species hypothesis was used;

5. The tag-*B* has a missing neutrino;

6. The tag-*B* has a photon missing;

7. The tag-*B* has a massive final-state particle missing;

8. The tag-*B* has a K_L^0 missing (special category compared to 6);

9. The tag-*B* has a final state particle associated that has a wrong-signed charge;

10. The tag-*B* has a (n-th) daughter non-final-state particle which belongs to a different particle;

11. Different errors in the truth-matching procedure.

Furthermore, all possible combinations between these are created (e.g. the tag-*B* has a photon *and* a massive final-state particle missing). In total, 108 combined categories are observed in the $B \rightarrow X_s \gamma$ signal MC samples for *B* meson candidates. Most categories, particularly higher-order combinations, do not have any entries in simulated data samples due to their rareness or kinematic inconsistency.

For all tag-*B* meson candidates in signal MC within each category, an M_{bc} distribution is produced and the Jensen-Shannon distance is calculated with respect to category 0 (attributed to perfect reconstruction). As category 0 (by definition) provides the *best possible kinematic constraint*, the Jensen-Shannon distance is sought to be as close to zero as possible. Such a result would imply that the difference of the reconstructed tag-*B* meson is inconsequential to the quality of the constraint. As an additional metric, for each category, the fraction of *B* meson candidates with $M_{\text{bc}} < 5.26$ GeV/c^2 is evaluated. In a perfect reconstruction case, this number is practically zero.

The two-dimensional distribution of Jensen-Shannon distances and the fraction of *B* meson candidates with $M_{\text{bc}} < 5.26$ GeV/c^2 for all 108 combined categories is shown in Figure 6.27. For charged, neutral and mixture of the two samples, the fraction of *B* meson candidates with $M_{\text{bc}} < 5.26$ GeV/c^2 begins to swiftly grow at Jensen-Shannon distance ≈ 0.3.

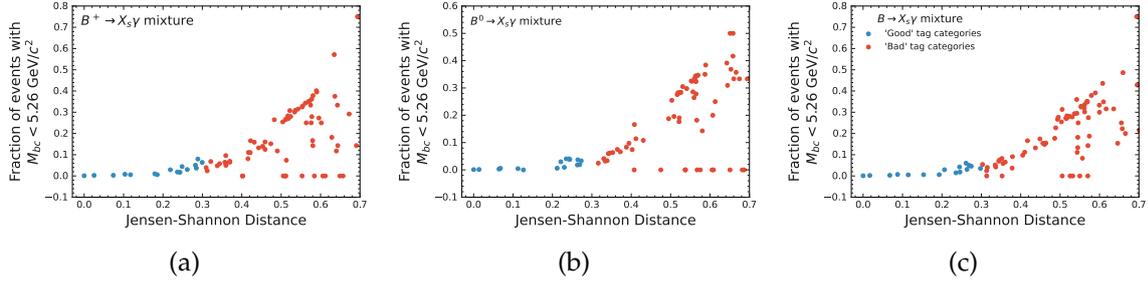Figure 6.27: The two-dimensional distribution of Jensen-Shannon distances and the fraction of B meson candidates with $M_{bc} < 5.26$ GeV/c^2 shown for $B^+ \to X_s\gamma$ (a), $B^0 \to X_s\gamma$ (b), and the mixture of the two (c). 108 categories in total have their M_{bc} distributions evaluated. The legend is shared between the Figures. The blue dots are chosen as the definition of tags providing a good kinematic constraint (good tags). The choice is based on a threshold of Jensen-Shannon distance ≈ 0.3, where it can be seen that the fraction of B meson candidates with $M_{bc} < 5.26$ GeV/c^2 begins to swiftly grow.

The requirement of Jensen-Shannon distance < 0.3 is adopted as the threshold to consider a B meson reconstruction category as providing a good kinematic constraint. The results between different categories are consistent: the same modes are extracted for B^+, B^0 and the mixture of two. The categories associated with a good kinematic constraint are summarised in Table 6.13.

Table 6.13: Categories of tag-B reconstruction that provide a good kinematic constraint. These categories correspond to the blue points in Figure 6.27c. The definitions of each category are provided in the text of Section 6.7.3.

Category number	Jensen-Shannon distance
1	0
9	0.02
8	0.07
9, 8	0.09
3	0.11
9, 3	0.15
5	0.19
8, 3	0.21
5, 3	0.24
6	0.24
8, 5, 3	0.25
9, 6	0.26
9, 5	0.26
8, 6	0.27
6, 5	0.27
9, 5, 3	0.30

The M_{bc} distributions for $B \to X_s\gamma$ with categories providing a good kinematic constraint and a bad one are shown in Figure 6.28a. The same Figure also highlights the difference if no procedure such as the one described in this Subsection would be introduced. One can see a resonant behaviour of imperfectly reconstructed tags in Figure 6.28b, which is mostly

absorbed into the definition of a good tag provided. This highlights the importance of the study presented in this Section – it would be difficult to separate the imperfectly reconstructed tag-B meson distribution from the perfectly reconstructed one in an M_{bc} fit that will follow (see Section 6.8), due to a large correlation between their shapes. For the rest of this thesis, a tag-B meson that provides a good kinematic constraint will simply be referred to as a *peaking* tag, referencing their behaviour in M_{bc}.

(a)   (b)
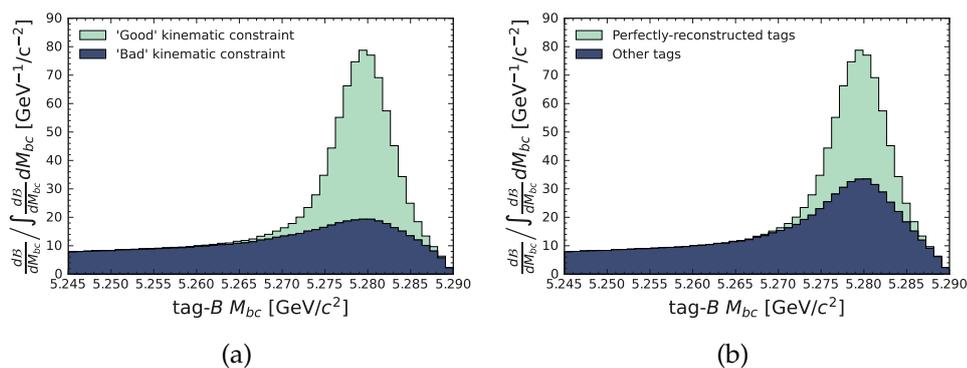
Figure 6.28: M_{bc} distributions of $B \to X_s \gamma$ split by good/bad kinematic constraint criterion (a) and the conventional perfect/imperfect reconstruction criterion (b). The definition of the tag-B meson described in Section 6.7.3 accurately represents the resonant M_{bc} structure.

6.8 M_{bc} **fitting setup**

By this point in the analysis, all the selections have been finalised and discussed. Furthermore, clear definitions for tag-B mesons that properly kinematically constrain $B \to X_s \gamma$ are presented, such that E_γ^B is evaluated accurately. However, as can be seen in Figure 6.22, even if continuum and $B\bar{B}$ backgrounds are suppressed heavily compared to the initial sample (Figure 6.7), there is still a significantly more background events than $B \to X_s \gamma$ signal events. Many of these, particularly continuum background, originate from incorrect tag-B meson reconstruction (see Figure 6.28) and can therefore be estimated in data using an M_{bc} fitting procedure. In this Section, a thorough overview of the M_{bc} fit is presented which extracts the counts of good tag-B in different E_γ^B intervals. All fits shown in this Section are unbinned extended negative log-likelihood fits, as presented in Section 5.1.

6.8.1 Components in the selected data set

There are three types of events in the e^+e^- collision data set following all the selections described so far:

- Generic $B\bar{B}$ (including $B \to X_s \gamma$) that are tagged with a good tag-B;

- Generic $B\bar{B}$ (including $B \to X_s \gamma$) that are tagged with a misreconstructed tag-B;

- Photon candidates originating in $e^+e^- \to q\bar{q}$.

These three components are referred to as 'peaking', 'combinatorial $B\bar{B}$' and 'continuum' throughout Section 6.8. Extracted from generic MC, they are visualised Figure 6.29.
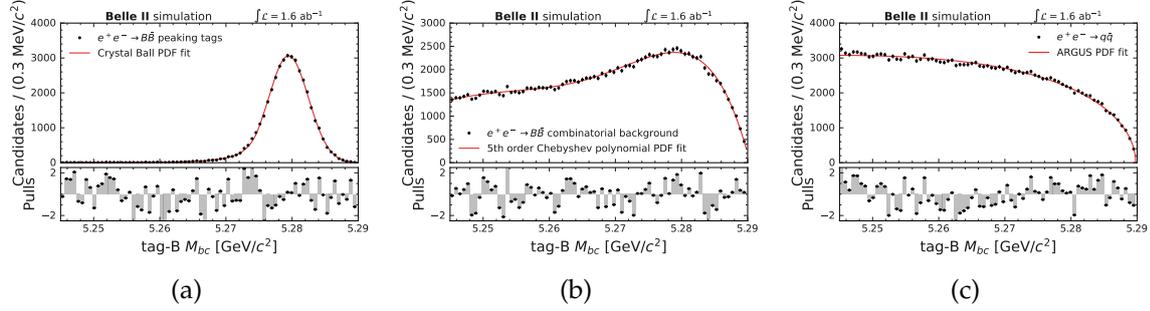

Figure 6.29: Separate components that are present in generic MC after selections that suppress background (Table 6.12). The individual components are defined in the text of Section 6.8.1. Each distribution contains an unbinned fit to the data points, and the subpanels show the pull (Equation (6.17)) in each case. The fitting function for peaking tag-B mesons (a) is chosen as the Crystal Ball function; for combinatorial tag-B mesons (b) it is chosen as the 5th order Chebyshev; for continuum $e^+e^- \to q\bar{q}$ events (c) it is chosen as the ARGUS function.

The fitting model is prepared to describe the three components and, particularly, to extract the number of tag-B candidates that correspond to the 'peaking' component. The strategy to describe each component is as follows:

- Peaking M_{bc} distributions are often fitted using a Crystal Ball function. It is defined in Appendix H.1 and can be understood as a Gaussian distribution with a polynomial tail. Figure 6.29a illustrates the suitability to describe the peaking M_{bc} distribution.

- Continuum M_{bc} distributions are conventionally described by the ARGUS function designed specifically for continuum events, defined in Appendix H.2. This function is used to describe $e^+e^- \to q\bar{q}$ simulated backgrounds in this analysis and shown in Figure 6.29c.

- The particular shape of the combinatorial $B\bar{B}$ background is generally dependent on the signal mode and does not have a conventional method of description. In this analysis, the usage of FEI and the fact that background events are conservatively suppressed to avoid signal-side biases lead to a wide but slightly peaking shape. Several options were assessed, but it was found to be suitably described by a Chebyshev polynomial, which can be adapted to a necessary functional shape. The definition of Chebyshev polynomials is given in Appendix H.3. A 5th-order polynomial describes the combinatorial $B\bar{B}$ background distribution in Figure 6.29b,

In Figure 6.29, the subpanels show the pull distribution, defined as:

$$\text{pull}(x) = \frac{x - \mu}{\sigma}. \tag{6.17}$$

Throughout this thesis it is used to evaluate the quality of the fit, as repeated measurements of a random variable x should fluctuate around a mean value μ with a statistical width σ.

Therefore, any dependencies or structures observed in the pull distribution would be indicators of poor fit quality.

6.8.2 Photon energy intervals for the fit

It is clear from Figure 6.22 that the signal-to-background ratio changes across all E_γ^B bins. Furthermore, even continuum-to-$B\bar{B}$ event fractions are not constant. This results from photons related to $e^+e^- \to q\bar{q}$ backgrounds which can extend to high-E_γ^B values because they do not originate from a B meson which is always produced with an energy $\approx \sqrt{s}/2$ at Belle II. The goal of the fit, as discussed in the introduction of this Section, is to remove combinatorial $B\bar{B}$ and continuum events from further analysis. While a total M_{bc} fit could be performed, this approach loses event-level information, such as the energies of individual photons, and only provides the event counts in the fitted E_γ^B region. Moreover, the background composition is expected to vary with E_γ^B, hence such an overall fit may generally be suboptimal.

Instead, in this analysis, the photon spectrum is divided into multiple E_γ^B intervals, and the M_{bc} distributions belonging to each interval are fitted using the functions described in Section 6.8.1. This approach reduces the existing data set to multiple E_γ^B intervals with known good tag-B counts, completely removing continuum and combinatorial $B\bar{B}$ events from the data set. Such an approach means that the final photon energy spectrum is provided with the same binning as used for fitting. It is therefore important to optimise the chosen intervals for the fit with respect to the expected $B \to X_s\gamma$ signal-to-background ratio, even if the primary goal of the fit is not signal extraction.

Three scenarios are tested for $E_\gamma^B \in (1.4, 2.8)$ GeV: 50 MeV, 100 MeV and 200 MeV wide bins. The test is performed by evaluating the statistical significance, with a definition equivalent to Equation (6.9), on a data set scaled to 189 fb^{-1}. In this case, the background is considered as non-$B \to X_s\gamma$ events that remain after the M_{bc} fit: only correctly tagged $B\bar{B}$ events (no combinatorial or continuum background). The result of the study of the statistical significance for the hybrid-signal model (Section 6.2.3) is shown in Figure 6.30.

In general, the highest statistical significance is observed with the widest bins, but this method, by definition, contains the least amount of information about the spectrum. Irrespective of the bin width, $E_\gamma^B \in (1.4, 1.8)$ and $E_\gamma^B > 2.7$ GeV show a statistical significance lower than unity, therefore the selected binning should attempt to maximise it. In the case of $E_\gamma^B \in (1.9, 2.0)$ GeV, it is observed that a similar statistical significance can be achieved in $E_\gamma^B \in (1.8, 2.0)$ GeV. This motivates choosing the wider bin to achieve a lower threshold than the one in the analysis by BaBar [88]. On the other hand, the studies performed in Section 6.13.3 show that the resolution of E_γ^B is around 40 MeV. Choosing a finely binned E_γ^B, compared to the resolution, may lead to complications in the unfolding procedure. As bin-by-bin unfolding is employed in this analysis, the 50 MeV bins were not considered. These considerations lead to choosing the following eleven E_γ^B bins for the analysis:

- Three 200 MeV bins for $E_\gamma^B \in (1.4, 2.0)$ GeV;

- Seven 100 MeV bins for $E_\gamma^B \in (2.0, 2.7)$ GeV;

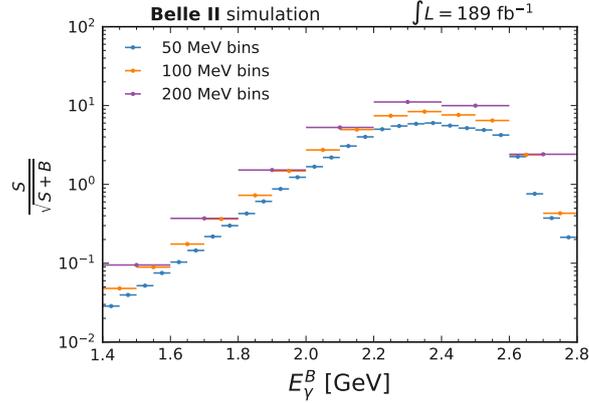- A single, inclusive overflow bin $E_\gamma^B > 2.7$ GeV.

Figure 6.30: The statistical significance based on Equation (6.9). Here, S is taken as the number of $B \to X_s\gamma$ events after fitting and B is taken as non-$B \to X_s\gamma$ events expected after fitting in each E_γ^B interval. Three scenarios are tested: 50 MeV, 100 MeV and 200 MeV wide bins. The final binning chosen is a hybrid scenario described in Section 6.8.2. The hybrid model described in Section 6.2.3 is used for this study.

This choice provides a compromise between statistical significance, E_γ^B spectrum resolution and complications in unfolding. The expected continuum, combinatorial, correctly-tagged non-$B \to X_s\gamma$ and correctly-tagged $B \to X_s\gamma$ event count expectations for the chosen binning are provided in Table 6.14. The Table entry containing combinatorial tag-B backgrounds may include events where $B \to X_s\gamma$ are present. However, they are rejected as background in the $M_{\rm bc}$ fits. On the other hand, the peaking-$B\bar{B}$ background, which is not rejected by the $M_{\rm bc}$ fit and treated in Section 6.9, is separated from $B \to X_s\gamma$. The Table also highlights the importance of a valid binning; several bins can achieve a comparable statistical significance to what could be achieved if just a single bin was used.

Table 6.14: The expected number of events as a fraction of the data set after selections in Table 6.12, for the binning chosen in Section 6.8.2. The Table also shows corresponding statistical significance for a 189 fb^{-1} sized data set.

E_γ^B bins [GeV]	Continuum frac.	Combinatorial $B\bar{B}$ (incl. $B \to X_s\gamma$)	Peaking $B\bar{B}$ frac (excl. $B \to X_s\gamma$)	Peaking $B \to X_s\gamma$	$\frac{S}{\sqrt{S+B}}$ at 189 fb^{-1}
1.4 − 1.6	0.22	0.20	0.047	0.00010	0.1
1.6 − 1.8	0.14	0.094	0.028	0.00031	0.38
1.8 − 2.0	0.088	0.046	0.016	0.00097	1.56
2.0 − 2.1	0.028	0.013	0.0048	0.0010	2.80
2.1 − 2.2	0.020	0.0079	0.0026	0.0016	5.05
2.2 − 2.3	0.013	0.0048	0.00093	0.0019	7.5
2.3 − 2.4	0.0075	0.0033	0.00031	0.0019	8.4
2.4 − 2.5	0.0042	0.0019	0.00024	0.0015	7.6
2.5 − 2.6	0.0019	0.00062	0.00013	0.0011	6.5
2.6 − 2.7	0.00059	0.000098	0.000016	0.00014	2.38
2.7 − 5.0	0.00021	0.000011	< 0.000001	0.000005	0.46
All	0.52	0.37	0.10	0.011	6.61

Because of the low expected statistical significance, $E_\gamma^B \in (1.4, 1.8)$ and $E_\gamma^B > 2.7$ GeV are chosen as validation regions. They are referred to as *sidebands*. The *signal region* of the analysis is defined as $E_\gamma^B \in (1.8, 2.7)$ GeV.

6.8.3 M_{bc} **fit model building**

The functions introduced in Section 6.8.1 and Appendix H are now considered in terms of the defined binning in Section 6.8.2. They are used to build an M_{bc} fit model for the total data set. The implementation of an N-th order Chebyshev polynomial, a Crystal Ball function and an ARGUS function results in $N + 6$ model parameters and 3 normalisation parameters for each E_γ^B bin. For fit stability purposes, it is desirable to reduce this number. Firstly, a primary fit model is prepared, where some of the parameters can be pre-determined or shared amongst the bins. The key idea is that all the functions are fitted once, separately, on the subsets of the simulated data that they aim to describe. This also provides the initialisation values of the parameters that are used on the final fit of the total entire data set. The primary fits for every component are discussed in the following Subsections.

Crystal Ball primary fit

The Crystal Ball function describes the distribution of peaking tag-B mesons in M_{bc} in terms of shape parameters μ, σ, α and n and the normalisation of the PDF, $\mathcal{N}$ (see Equation (H.1)). In principle, the reconstruction of the tag-B meson should not be strongly influenced by the signal-B meson, which means that the shape parameters should not depend on E_γ^B. This hypothesis is tested in Figure 6.31. It can be seen that strong correlations between E_γ^B and parameters μ and σ are absent. This test is not performed on parameters α and n, as they tend to be less stable than μ and σ, depending more on the fluctuations of the data set. It is therefore concluded that a single Crystal Ball shape is sufficient, where all E_γ^B bins share the same shape parameters. The discussed primary fits are performed simultaneously on the peaking tag-B meson M_{bc} distributions in MC for all E_γ^B bins. They are shown in Figure I.1. The evaluated shape parameters are fixed for the final M_{bc} fit model. The parameter values and uncertainties from the primary fit are shown in Table 6.15.

Chebyshev polynomial primary fit

The Chebyshev PDF takes N parameters, k_i ($i = \{1, ..N\}$), and a normalisation, $\mathcal{N}$ (Equation (H.5)). N is therefore referred to as the *order* of the Chebyshev PDF. Third, fourth and fifth order Chebyshev polynomials are tested for suitability to describe the combinatorial $B\bar{B}$ background distribution. The fifth-order result was already introduced in Figure 6.29b. The results of the best fit for lower-order polynomials are presented in Figure 6.32. From the pull distributions in the subpanels and the general inspection of the fit, it is clear that using a polynomial of order lower than five is insufficient for an adequate description of combinatorial $B\bar{B}$ events.

The coefficients k_{1-5} of the Chebyshev polynomial cannot be easily connected to physical observables, and therefore it is hard to evaluate their dependence on E_γ^B. As such, three different Chebyshev PDFs are evaluated for the following intervals: $E_\gamma^B \in (1.4, 1.6)$, $E_\gamma^B \in (1.6, 1.8)$, $E_\gamma^B > 1.8$ GeV. The last interval is not subdivided further because of the data set size (see Table 6.14), each interval is optimised to contain an approximately equal expected number of events. The discussed primary fits are performed simultaneously on the M_{bc} distributions of the combinatorial $B\bar{B}$ in all E_γ^B bins. They are shown in Figure I.3. The determined coefficients
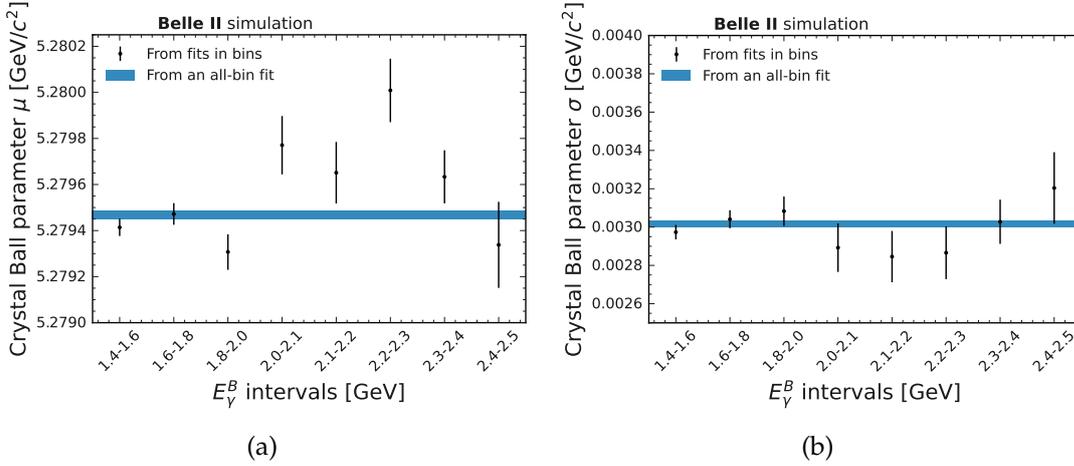
Figure 6.31: The parameters from M_{bc} fits of peaking tag-B mesons in generic $B\bar{B}$ simulation, using a Crystal Ball function. The datapoints showcase the estimated parameters μ (a) and σ (b) for different E_γ^B intervals. This can be compared with the overall shape (blue band), which is determined if the entire $E_\gamma^B \in (1.4, 2.8)$ GeV region is fitted. The parameters of the blue band correspond to the fit in Figure 6.29a. No strong dependence on E_γ^B is observed.
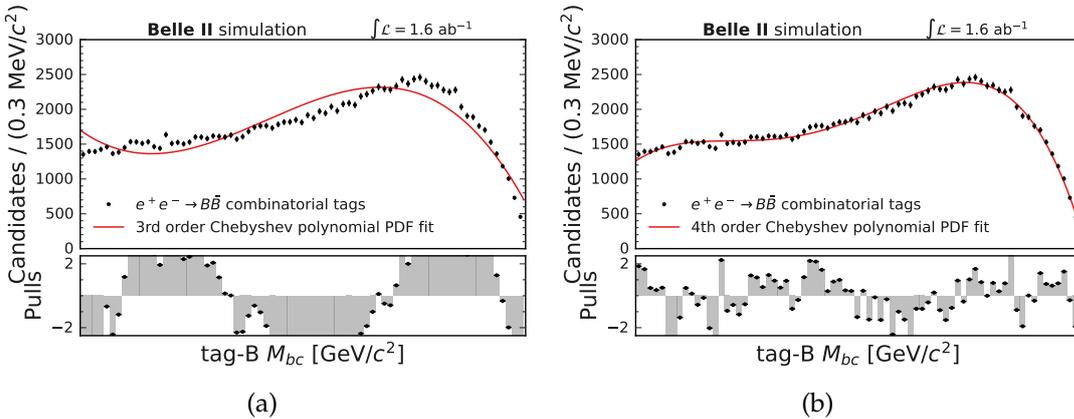

Figure 6.32: The M_{bc} fits of combinatorial $B\bar{B}$ events with a third-order (a) or fourth-order (b) Chebyshev PDF. These Figures can be compared to Figure 6.29b. Lower-order Chebyshev polynomials are unable to accurately describe the combinatorial $B\bar{B}$ data.

k_{1-5} are fixed in the final M_{bc} fit model. The primary fit results are presented in Table 6.15.

ARGUS primary fit

Similarly to the previous cases, the ARGUS function has 2 parameters, c and m_0, and a normalisation $\mathcal{N}$ (Equation (H.3)). Unlike the coefficients of the Chebyshev, c and m_0 are easier to understand intuitively: with m_0 corresponding to the maximum allowed M_{bc} values in the distribution and c to the shape of continuum background. More generally, the variation of c and m_0 accounts for possible background shape differences between MC and data. It offers an additional layer of flexibility through a variation of relative ARGUS and Chebyshev normalisation values. This means that allowing a degree of variation in the ARGUS shape is beneficial to account for possible background differences in data.

Following these considerations, a setup is adopted with independent shapes for $E_\gamma^B \in (1.4, 1.6)$ GeV, $E_\gamma^B \in (1.6, 1.8)$ GeV, $E_\gamma^B > 1.8$ GeV. In each respective range, one single shape parameter $c^{(1.4-1.6)}$, $c^{(1.6-1.8)}$, $c^{(>1.8)}$ is determined from the total fit. To account for possible differences between simulated and data $\sqrt{s}$, the m_0 parameter is shared between all bins and is also determined from the fit of the combined data set. The discussed primary fits are performed simultaneously on continuum M_{bc} distributions in all E_γ^B bins. They are shown in Figure I.2. The initial values for the final M_{bc} fit model are presented in Table 6.15.

Final fit model

Based on the primary fits of the subcomponents of the total data set, the number of fitting parameters is reduced: certain fit parameters are shared between different E_γ^B intervals, whereas other parameters are kept at their initialised value and not varied. The final M_{bc} fitting model for the analysis is summarised in Table 6.15. It shows the initial values for each parameter and highlights which parameters are kept fixed in the final M_{bc} fit. It can be seen, that the previously introduced $11 \cdot (5+2+4+3) = 154$ parameters are reduced to 37 fitting parameters. The fitting model is then applied to the total data set in Section 6.8.4 in order to estimate the main parameter of interest: normalisations of the Crystal Ball in every E_γ^B bin (i.e. the number of good tag-B in each E_γ^B bin). The primary M_{bc} fits determining the parameters in Table 6.15 on corresponding data sets are given in Appendix I.

6.8.4 M_{bc} **fit of the total simulated data set**

The prepared fit, discussed in Section 6.8.3, is applied to the total generic MC data set. The initial values of the fit are given in Table 6.15 and all bins are fitted simultaneously. The fit results are shown in Figure 6.33. The overall M_{bc} distributions are accurately described for all E_γ^B bins, irrespective of the overall number of data points that are being fitted. This is already an important achievement: a single fitting setup is adaptive enough to cover different signal-to-background ratios, statistical sample sizes and overall shapes of the distributions.

As was already mentioned, the main goal of the M_{bc} fit introduced in this Section is to extract the good tag-B counts in each E_γ^B bin. This is equivalent to the normalisation parameter of the Crystal Ball PDF, $\mathcal{N}_{\mathrm{CB}}$. These normalisations are shown in Figure 6.34a. The exact
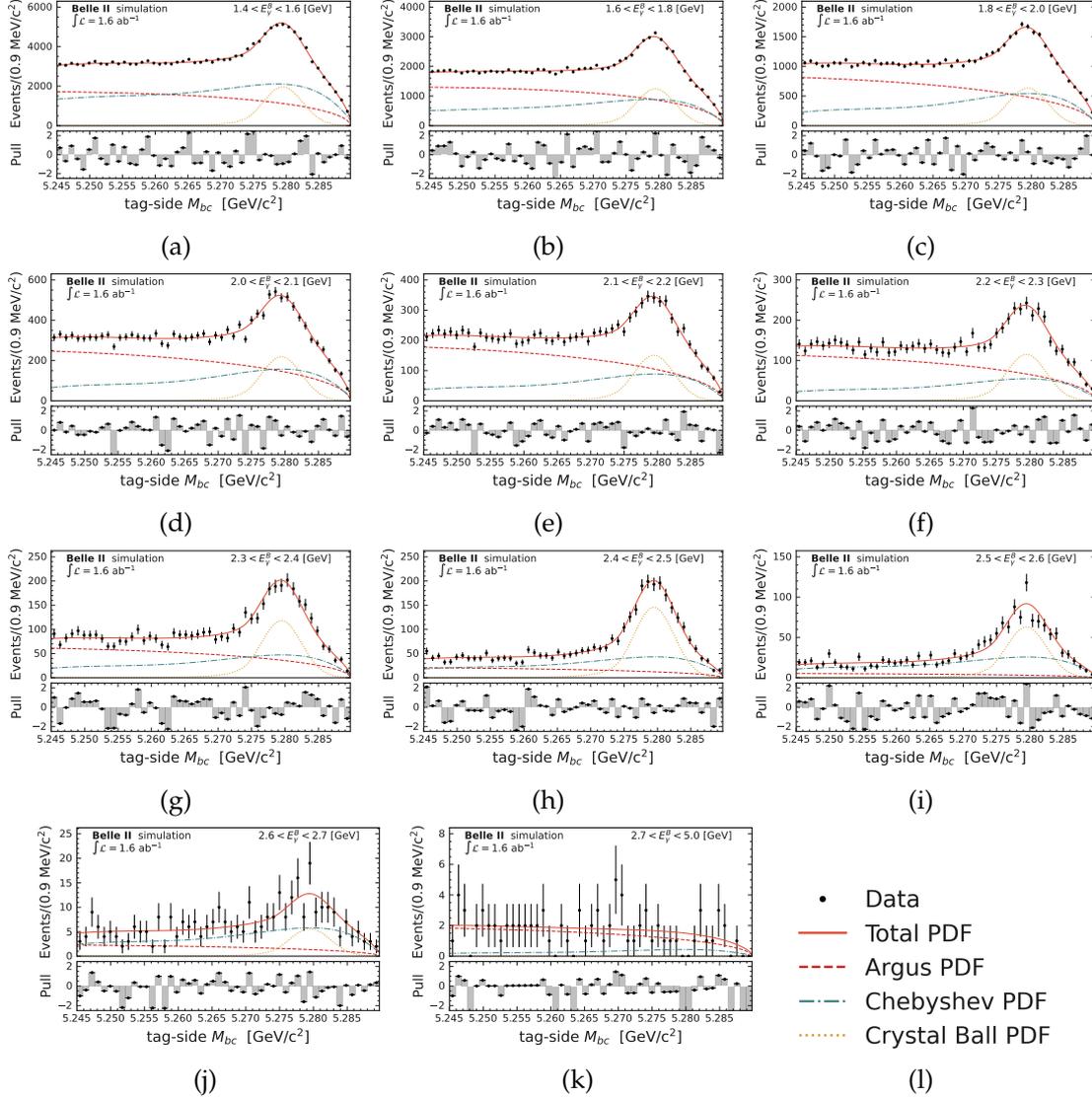
Figure 6.33: The fits of the 1.6 ab^{-1} data set of generic MC, using the fitting model from Table 6.15. A good description of the M_{bc} distributions can be seen throughout the E_γ^B bins. This highlights that the fit setup accurately describes large simulated data samples. The legend is shown in (l). For the rest of the thesis, this legend is always implied for M_{bc} fits.

Table 6.15: The summary of the fitting model used in this analysis for the M_{bc} fit. For the final M_{bc} fit, the parameters are initialised at the values that are listed, corresponding to the ones determined in the primary fitting steps, explained in Section 6.8.3. The values that are bolded in the Table are not estimated from the final M_{bc} fit but are kept at their initialised values. On the other hand, all non-bolded values can vary in the final fit. Uncertainties are evaluated using the `HESSE` method in the primary fitting steps. In the Table, they are omitted if the relative uncertainty is lower than 0.1%.

E_γ^B bin	Crystal Ball					Chebyshev						Argus		
	$\mathcal{N}_{CB}$	μ	σ	α	n	$\mathcal{N}_{cheb}$	k_1	k_2	k_3	k_4	k_5	$\mathcal{N}_{ARGUS}$	c	m_0
1.4 − 1.6	17294 ± 131					71636 ± 267	−0.0751 ± 0.0071	−0.3125 ± 0.0073	−0.2516 ± 0.0064	−0.1381 ± 0.0064	−0.0296 ± 0.0062	78613 ± 280	−33.11 ± 0.79	
1.6 − 1.8	10218 ± 101					34172 ± 185	−0.01 ± 0.01	−0.33 ± 0.01	−0.283 ± 0.009	−0.143 ± 0.009	−0.033 ± 0.009	51809 ± 228	−28.03 ± 1.00	
1.8 − 2.0	5947 ± 77					16389 ± 128						32021 ± 179		
2.0 − 2.1	1938 ± 44					4325 ± 66						10220 ± 101		
2.1 − 2.2	1246 ± 35					2622 ± 51						7119 ± 84		
2.2 − 2.3	909 ± 30	**5.279**	**0.003**	**1.573 ± 0.035**	**3.561 ± 0.22**	1598 ± 40	**0.119 ± 0.011**	**−0.362 ± 0.012**	**−0.348 ± 0.010**	**−0.193 ± 0.010**	**−0.029 ± 0.010**	4575 ± 67		**5.2897**
2.3 − 2.4	985 ± 31					1211 ± 35						2724 ± 52	−25.07 ± 0.92	
2.4 − 2.5	1213 ± 35					785 ± 28						1520 ± 39		
2.5 − 2.6	629 ± 25					311 ± 52						682 ± 26		
2.6 − 2.7	62 ± 8					52 ± 7						216 ± 15		
2.7 − 5.0	1 ± 1					7 ± 2						75 ± 9		

values of the rest of the parameters are not needed for further analysis, as long as an unbiased and accurate description of good tag-B counts is achieved through $\mathcal{N}_{CB}$. The main goal of keeping them as free parameters in the fit is to ensure that the M_{bc} background shape and yield estimations are sufficiently flexible to account for potential data-simulation discrepancies. Therefore, for the rest of the thesis, other parameters will not be explicitly mentioned, unless their values are relevant to the discussion.

To ensure that the fit is receptive to changes of $B \rightarrow X_s\gamma$ signal, two simple tests are devised, where the fit is performed on generic MC but

- all $B \rightarrow X_s\gamma$ signal events are removed;

- $B \rightarrow X_s\gamma$ signal shape in generic MC is reweighted using the hybrid model weights.

The results are shown in Figures 6.34b and 6.34c. It is clear that if the $B \rightarrow X_s\gamma$ events are completely removed or reweighted, the M_{bc} fit presented in this Subsection responds appropriately. Therefore, although a particular simulation model was used when preparing the fit, no particular bias towards any $B \rightarrow X_s\gamma$ model is introduced.

6.9 Remaining $B\bar{B}$ background subtraction

After the fitting procedure in Section 6.8 the good tag-B counts are estimated, and in particular:

- all $e^+e^- \rightarrow q\bar{q}$ contributions have now been removed;

- events where the tag-side B mesons are not properly reconstructed have been removed.

However, it may seem odd that after performing the M_{bc} fit (see Figure 6.34), the result is still not quite comparable to an E_γ^B spectrum, resembling that of the background seen in, e.g.,
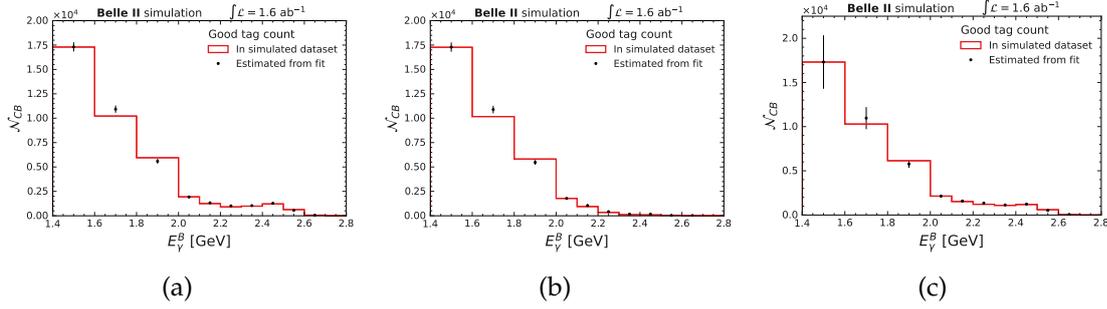
(a)   (b)   (c)

Figure 6.34: The comparisons of estimated peaking tag-B counts in each E_γ^B bin, $\mathcal{N}_{\text{CB}}$, for the total 1.6 ab^{-1} generic MC data set (a). Additionally, results are shown for the fits of the generic MC data set but with all $B \rightarrow X_s\gamma$ events removed (b), and for generic MC data set but with the hybrid-signal model reweighting introduced (c). In each case, the uncertainties are HESSE uncertainties estimated by the fit. It can be seen that the fit is receptive to the expected count of peaking tag-B mesons. Larger uncertainties in (c) are due to known issues in weighted error estimation. They do not affect this analysis as weighted fits are not performed later.

Figure 6.22. Since in the inclusive treatment the X_s is not constrained, a component of *irreducible* background will always be present. In the case of this analysis, the good tag-B counts after the optimal selection in Table 6.12 and the fitting procedure presented here contain:

- correctly tagged $B \rightarrow X_s\gamma$ events,

- correctly tagged-$B\bar{B}$ events other than $B \rightarrow X_s\gamma$.

Although in the future versions of this analysis it may be possible to further diminish the second component through improved selection and fitting procedures (see discussion in Section 7.2), some irreducible background component will always remain.

The final step before the full $B \rightarrow X_s\gamma$ spectrum extraction in MC is evaluating the size of the remaining $B\bar{B}$ background. Two strategies are considered:

- Count good tag-B mesons in each E_γ^B interval that correspond to non-$B \rightarrow X_s\gamma$ in MC and use these values to subtract remaining $B\bar{B}$ events from the fitted values. This is a simple method which is independent of the fitting procedure.

- Perform the fit again on the simulated sample with $B \rightarrow X_s\gamma$ events removed. This way, good tag-B meson counts are estimated with fitting effects included. This effectively corresponds to the data points in Figure 6.34b.

Although both choices are valid and have certain advantages, in this analysis, the second method is chosen. The main reason for this choice is the fact that biases or systematic effects in the fit result, if such exist, could be suppressed when subtracting. It also requires a similar validation procedure (e.g. to be an unbiased estimator) as the total fit. Therefore, after the M_{bc} fit, the number of correctly-tagged $B \rightarrow X_s\gamma$ events in data will be measured as follows:

$$N_{\text{DATA}}^{B \rightarrow X_s\gamma} = \mathcal{N}_{\text{DATA}}^{\text{good tags}} - \frac{189 \text{ fb}^{-1}}{1600 \text{ fb}^{-1}} \cdot \mathcal{N}_{\text{MC}}^{\text{good tags with B} \rightarrow X_s\gamma \text{ removed}} . \tag{6.18}$$

6.10 Analysis strategy validation in simulation

Sections 6.8 and 6.9 introduced the fitting procedure and $B\bar{B}$ background subtraction. Together with the optimal selections from Section 6.6, this fully defines the analysis strategy from the Belle II simulated data sets to the $B \to X_s\gamma$ spectrum. However, the defined fit has to be validated in simulation to give an unbiased estimation of $B \to X_s\gamma$ events, with a good resolution and signal efficiency. The studies in this Section prove such results.

6.10.1 Validation of M_{bc} fit of reduced sample size

The results in Figure 6.34 are obtained by a fit of 1.6 ab^{-1} of MC, which is about an order of magnitude larger than is expected in the case of this analysis. Therefore, the generic MC data set is pseudorandomly split into 10 smaller subsets, corresponding to 160 fb^{-1}, and each of them is fitted independently. The choice of 160 fb^{-1}, rather than 189 fb^{-1} which is the sample size of the Belle II data used in the analysis, is due to anticipated data-simulation differences (see discussion of data-simulation differences in Section 6.11, particularly FEI calibration in Section 6.11.1). Indeed, a 190 fb^{-1} data set corresponds to approximately 160 fb^{-1} in simulation due to differences in tag-B reconstruction efficiency.

The resulting 10 fits and the estimated $\mathcal{N}_{CB}$ corresponding to each E_γ^B bin are shown in Figure 6.35. The expected number of events in each bin is always equal to one-tenth of that in the generic MC data set. It can be observed that all data points, and their average, are statistically compatible with the expectation. These results indicate that despite using a 10 times larger data set to define the PDFs, this M_{bc} fit model produces reliable and stable results. Further tests, particularly a test ensuring that the fit is unbiased, are performed in Section 6.10.

6.10.2 Validation of subtraction of remaining-$B\bar{B}$ background

The strategy to extract the $B \to X_s\gamma$ photon energy spectrum and suppress the remaining $B\bar{B}$ background was laid out in Section 6.9. In particular, the full generic MC data set is modified, such that $B \to X_s\gamma$ events are absent from it. Then, the M_{bc} fit discussed in Section 6.8.3 is performed. To test that this procedure is viable, the subtraction is performed for the results shown in Figure 6.35. Although the 10 fits are performed on 160 fb^{-1} data sets, the background subtraction is done with a 1.6 ab^{-1} data set. Therefore, the statistical uncertainty from the fit of the smaller data set is dominating. The subtracted result is shown in Figure 6.36.

The 1.6 ab^{-1} data set and the 160 fb^{-1} subsets are largely correlated, therefore the result has a smaller spread than one might expect from a unit Gaussian, based on the statistical uncertainty provided by the fit. However, the purpose of this test is to showcase that the setup extracts values that are statistically compatible with the scaled-down original data set. The results showcased in Figures 6.35 and 6.36 clearly show that the central values of the E_γ^B spectrum extracted follow the number of $B \to X_s\gamma$ events in the data set. In this Section so far, no particular modelling of $B \to X_s\gamma$ spectrum has been assumed. On the other hand, following the setup that is taken to remove $B\bar{B}$ backgrounds after the M_{bc} fit, the analysis is heavily dependent on the background model. For this reason, special emphasis will be put on testing the *background* description validation in Sections 6.11 and 6.12.
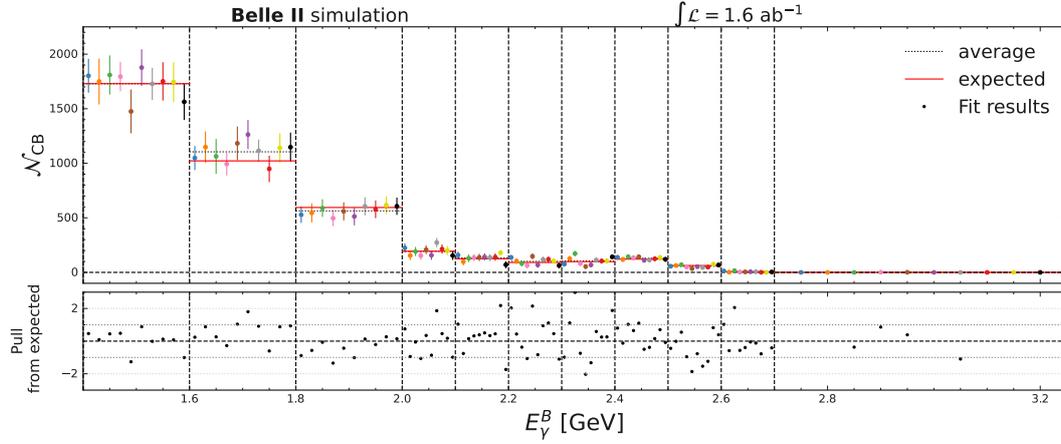
Figure 6.35: The estimated $\mathcal{N}_{CB}$ values from fits of one-tenth of generic MC, corresponding to 160 fb^{-1} of simulation. The dashed lines represent different E_γ^B bins, each bin showing one data point corresponding to a simultaneous fit of all E_γ^B bins. The dotted lines show the average of all 10 points in each bin, whereas the full lines show the number of good tag-B events in the original 1.6 ab^{-1} data set, scaled down 10 times ('expected'). The subpanels show the pull of each data point from the expected number of events. These results show that the fit extracts a valid result on a data set that is an order of magnitude smaller.
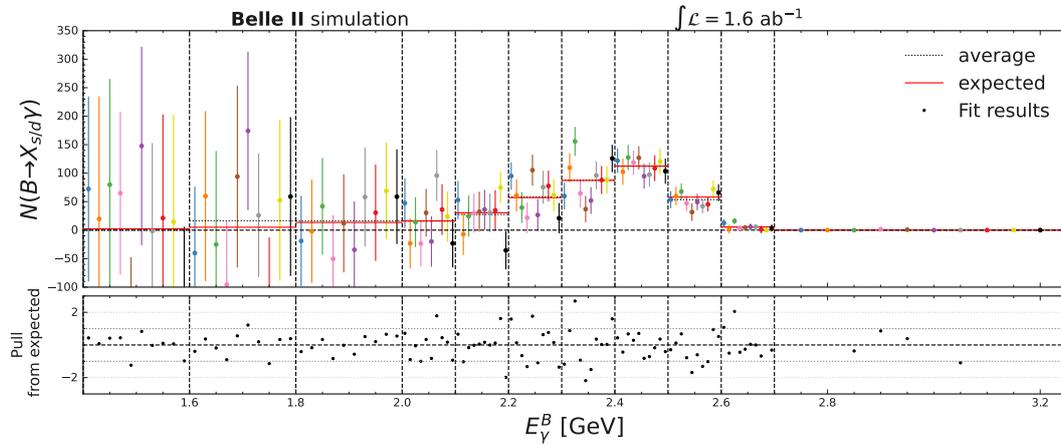
Figure 6.36: The estimated $\mathcal{N}_{CB}$ with subtracted background, based on the Equation (6.18). The values before $B\bar{B}$-background subtraction are shown in a corresponding Figure 6.35. The remaining events originate from $B \to X_s\gamma$ in fitted MC data sets, and can be seen to agree well with the generic MC expectations. To ensure that minimal statistical background subtraction uncertainty is introduced, the full 1.6 ab^{-1} data set is chosen to subtract the background. The uncertainties of each data point are those of the M_{bc} fit of the background data set and on the tested subset, added in quadrature.

6.10.3 Closure test of the M_{bc} fit

If the uncertainties estimated by a fit are correct, then fitting pseudodata sets generated from PDFs fitted on test data must yield statistically compatible results. This verifies two important aspects of the fit:

- the estimated fit parameter central values are reproduced when fitting a statistically equivalent data set;

- the estimated fit parameter uncertainties appropriately describe the statistical fluctuations of the central values.

More concisely, the pull distribution of an unbiased fit, in this case, calculated as:

$$\mathrm{pull} = \frac{\mathcal{N}_{\mathrm{CB}} \cdot \mathrm{scale} - \mathcal{N}_{\mathrm{CB}}^{\mathrm{pseudo}}}{\mathrm{fit\ error}}, \tag{6.19}$$

must be described by a unit Gaussian (assuming the central limit theorem is applicable for the pseudodata sample size). In the case of this analysis, $\mathcal{N}_{\mathrm{CB}}$ is the estimated number of good tag-B mesons in the generic MC sample. On the other hand, $\mathcal{N}_{\mathrm{CB}}^{\mathrm{pseudo}}$ are the normalisations estimated in a a randomly sampled data set that follows the PDFs fitted on the generic MC data set. The fit error is the corresponding uncertainty, in this analysis estimated by the HESSE method. The scale is used to equate the sample size between the sampled and total simulated data set. Tests of this type are known as *closure tests* and allow verifying that the central values reproduced by the fit fluctuate as indicated by the PDF uncertainties.

The closure test in this analysis is done on pseudodata sets of equivalent size as the Belle II collected data. First, 1000 pseudodata sets equivalent to 160 fb^{-1} are sampled from the PDF that was fitted in Figure 6.33. Since in all of the cases $\mathcal{N}_{\mathrm{CB}}$ and fit error are known, the M_{bc} fit is used on the pseudodata set, and a pull is calculated based on Equation (6.19). The pull distributions for every E_γ^B bin are shown in Figure 6.37.

To test the statistical validity of M_{bc} fits in each bin, a Gaussian PDF is fitted on the distributions, with parameters μ and σ being estimated. The parameters correspond to the mean value and the width of the Gaussian distribution, respectively. The parameter estimation is performed as an unbinned negative log-likelihood fit. The corresponding Gaussian fit results and the parameters are also included in Figure 6.37 and also summarised in Figure 6.38.

In all cases the results are compatible with a unit Gaussian within to 2 standard deviations. As the pulls distributions are Gaussian-like, the central limit theorem regime is reached. The only exception is the parameter μ in $E_\gamma^B > 2.7$ GeV. Here, the central value of the pulls appears to be biased towards lower values. This is attributed to statistical effects, as that bin has only a handful of entries (even at 1.6 ab^{-1}) and is therefore strongly affected by statistical background fluctuations. As it is not expected to observe any statistically significant signal events in that region, it is chosen to not define any correction based on the observed bias.

These results allow concluding that the M_{bc} fit is an unbiased estimator. In other words, the central values are unbiased and the uncertainties adequately cover the statistical fluctuations of the results.
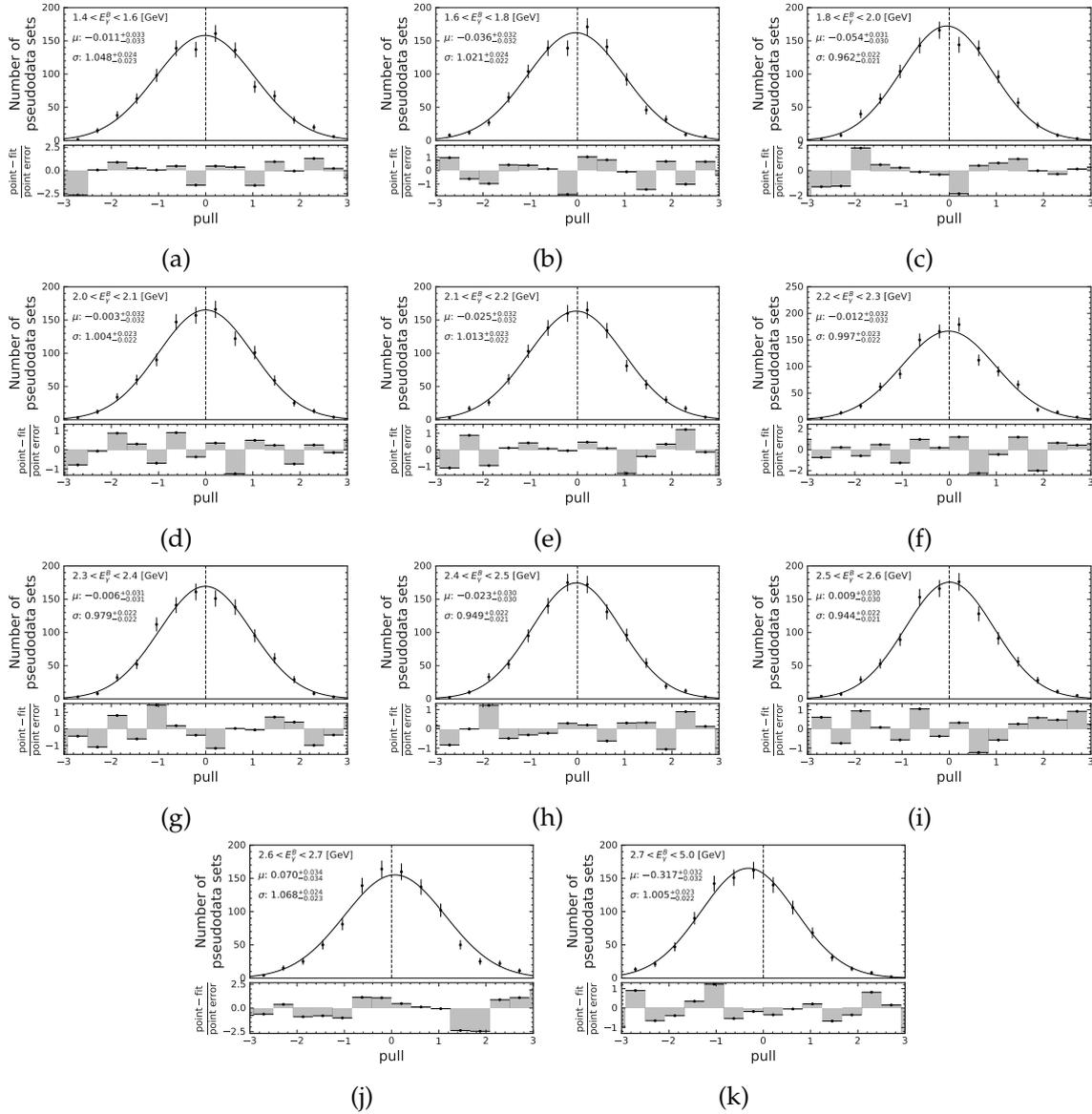
Figure 6.37: The pull distributions used in the M_{bc} fit closure test, corresponding to results of 1000 pseudodata sets generated based on the PDF fitted on the total generic MC data set. The definition of the pull for this test is given in Equation (6.19). The data points show the counts of values in the given pull intervals, and the statistical uncertainty. The pulls are also fitted with a Gaussian distribution and the mean value, μ, as well as the width, σ, are extracted. The Gaussian fit is shown as a solid line and is an unbinned fit (i.e. not the fit to the shown data points).

(a)   (b)
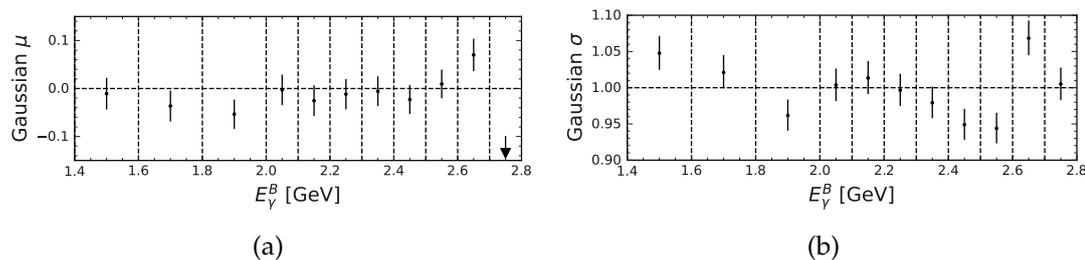
Figure 6.38: Summarised means and widths (μ and σ, respectively) of the Gaussian fits of the pull distributions in each E_γ^B bin. The fits for evaluation of μ and σ are shown in Figure 6.37. The results are compatible with a unit Gaussian except in the case of parameter μ for $E_\gamma^B > 2.7$ GeV, where statistical effects play a large role.

6.10.4   Linearity test of the M_{bc} fit

The fit is further validated using the so-called *linearity* test. Any valid extended fit model ought to provide a behaviour such that the estimated normalisation, $\mathcal{N}$, grows linearly with the increase of the corresponding component in the fitted data set. To perform such a test in this analysis, up to 25000 $B \rightarrow X_s\gamma$ signal-MC events where a good tag-B meson has been identified are combined with the generic MC data set. For a valid fitting setup, N events 'injected' in the fitted data set should yield approximately an increase of N in the estimated $\mathcal{N}_{CB}$. Such a test of the expected linear behaviour is summarised in Figure 6.39.
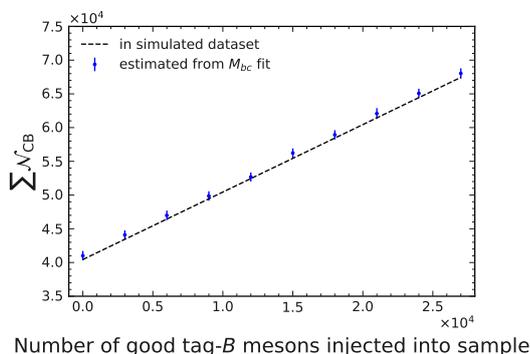
Figure 6.39: The summary of the linearity test for the M_{bc} fit of this analysis. As the number of events corresponding to the peaking tag-B mesons increases, the extracted sum of normalisations across all bins, $\Sigma\mathcal{N}_{CB}$, also increases. The increase is compatible with a linear increase with a slope of unity.

As the number of 'injected' good tag-B events grows, the sum of the normalisations for all E_γ^B bins should increase linearly. This behaviour is reproduced, as $\sum \mathcal{N}_{CB}$ grows linearly with the increase of the data set. It can therefore be concluded that the behaviour of the M_{bc} fit is linear.

6.10.5   Correlation matrix of the parameters of the M_{bc} fit

Using the pseudodata sets generated in Section 6.10.3, relationships between all parameters estimated by the M_{bc} fit can be calculated. In particular, a Pearson-R correlation coefficient for some collection of paired values $(x_1, y_1), (x_2, y_2,)...,(x_n, y_n)$:

$$R_{x,y} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}. \tag{6.20}$$

For every pair of parameters estimated by the M_{bc} fit, the Pearson-R value is evaluated. This results in a 36-by-36 matrix, corresponding to all parameters estimated by the M_{bc} fit in this analysis, and shown in Figure 6.40. The 37th parameter, m_0, is not included, because no significant correlations with that parameter are observed in the fit.

Several important insights into the fit can be understood by observing the correlation matrix:

- Firstly, consider the correlation of $\mathcal{N}_{\mathrm{CB}i}$ with $\mathcal{N}_{\mathrm{CB}j}$. The correlations are low, implying that increases in one bin do not induce strong differences in other bins, as is to be expected. Small correlations that can be observed are likely a combination of correlations through other parameters (see later) and statistical fluctuations. In general, most of the values are correlated by less than 10%.

- Secondly, consider the correlation of $\mathcal{N}_{\mathrm{CB}i}$ with $\mathcal{N}_{\mathrm{CHEB}i}$. It can be seen that the diagonal elements are strongly anti-correlated. This is understood as a consequence of the fact that both Chebyshev and Crystal Ball PDFs contain a degree of peaking behaviour in M_{bc}. Therefore, larger $\mathcal{N}_{\mathrm{CHEB}}$ values lead to lower $\mathcal{N}_{\mathrm{CB}}$ simply because the peaking behaviour of one parameter diminishes the other.

- Thirdly, consider the correlation of $\mathcal{N}_{\mathrm{CB}i}$ with $\mathcal{N}_{\mathrm{ARGUS}i}$. This correlation is positive and understood as a direct result of the second point. If $\mathcal{N}_{\mathrm{CB}}$ is evaluated as larger, the Chebyshev polynomial, which is suppressed as a result, can also less adequately describe the low-end of M_{bc}. That region must therefore be described by $\mathcal{N}_{\mathrm{ARGUS}}$, inducing a chain of correlation: $\mathcal{N}_{\mathrm{CB}} \uparrow \rightarrow \mathcal{N}_{\mathrm{CHEB}} \downarrow \rightarrow \mathcal{N}_{\mathrm{ARGUS}} \uparrow$.

- Fourthly, correlations of $\mathcal{N}_{\mathrm{CB}i}$ with c_j. This can be understood in a similar way as the correlation with $\mathcal{N}_{\mathrm{CHEB}i}$. The parameter c controls the shape of the ARGUS PDFs. Notably, larger and positive values of c tend to produce a PDF that has relatively more area at high-M_{bc} than at low-M_{bc}, producing peaking behaviour. Therefore, largely positive values of c introduce smaller values of $\mathcal{N}_{\mathrm{CB}}$.

- Finally, $\mathcal{N}_{\mathrm{CB}}$ correlations with off-diagonal elements of other parameters. In these cases, the correlations are very small, similar like the $\mathcal{N}_{\mathrm{CB}i,j}$ correlations, but they follow the trends of the diagonal elements. This behaviour is interpreted through the effects of the shared parameter c. As the parameter c modifies the shape of the ARGUS PDF, the changes are reflected in $\mathcal{N}_{\mathrm{ARGUS}}$ which propagate as differences to the normalisation parameters of other PDFs, through discussed relations. Particularly, because the variations of c cause an increase to several bins simultaneously, correlations between off-diagonal elements $\mathcal{N}_{\mathrm{CB}i,j}$ are also observed.
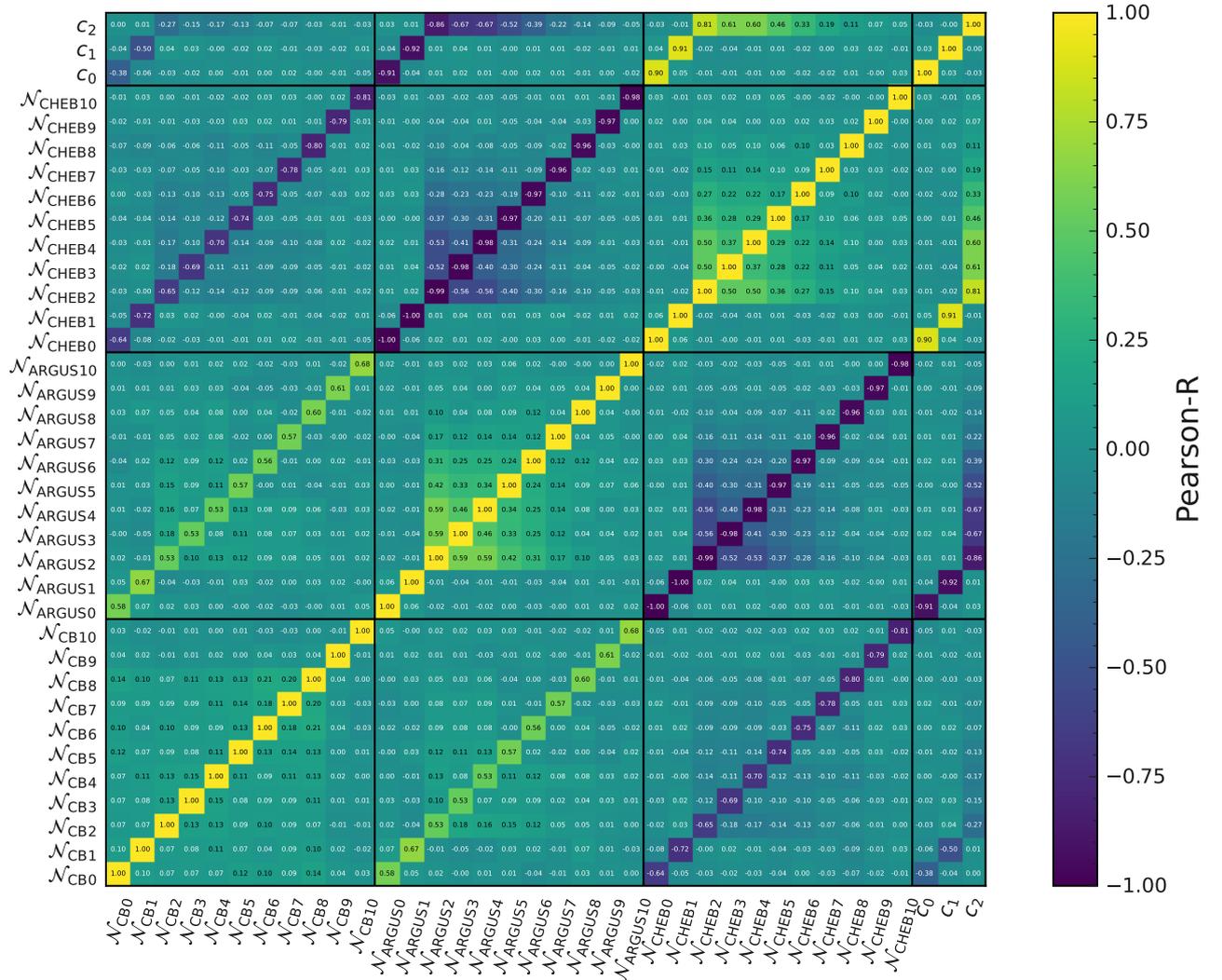
Figure 6.40: The correlation matrix of the M_{bc} fit, generated using a 1000 pseudodata set. Every pair of parameters has their correlation evaluated as the Pearson-R coefficient (Equation (6.20)). Parameter names correspond to those in Table 6.15 and the numbering $0 - 10$ correspond to the bin number, starting from $1.4 - 1.6$ GeV.

A study with more than 1000 pseudodata sets may be necessary in the future to better understand the indirect off-diagonal correlations present in this analysis. As correlations between $\mathcal{N}_{\mathrm{CB}i,j}$ are observed as small, they are not considered to significantly affect the result in this analysis. Therefore, it is concluded that no unexpected behaviour of the fit is observed, with correlations between different parameters explainable through the differences in the PDF shape they describe.

6.11 Simulation-to-data corrections

The fitter and background subtraction procedure, introduced in Sections 6.8.3 and 6.9, have been thoroughly validated in Section 6.10 in MC. The real challenge, as usual, is ensuring that the conclusions and results observed in MC generalise correctly to real Belle II data. The key concept of a blinded analysis dictates that one must validate the analysis procedure in control samples or regions: collections of data that are abundant, well-understood and provide insight into the behaviour of signal in the detector while being signal free. In this Section, FEI validation, π^0 and η veto validation, photon detection efficiency and background modelling is presented. The combined results from Sections 6.11.1 to 6.11.4 are shown in Table 6.16.

Table 6.16: The corrections for background (and signal in Section 6.13.2) efficiency in the hadronic-tagged $B \to X_s\gamma$ photon energy spectrum measurement. FEI calibration calculations are discussed in Section 6.11.1. Derivation of correction for the π^0 and η veto are presented in Section 6.11.2. The photon detection efficiency study is described in Section 6.11.3. Background modelling corrections are calculated in Section 6.11.4. The FEI, π^0 and γ corrections are averaged values corresponding to the respective E_γ^B bin, as the candidate-level information is lost after the M_{bc} fit. The signal region is highlighted.

E_γ^B [GeV]	FEI calibration	π^0 and η veto correction	γ efficiency correction	Leftover-$B\bar{B}$ background
1.4 − 1.6		1.090 ± 0.050	0.991 ± 0.023	1.021 ± 0.022
1.6 − 1.8		1.074 ± 0.048	0.995 ± 0.022	1.039 ± 0.032
1.8 − 2.0		1.064 ± 0.046	0.996 ± 0.021	1.05 ± 0.04
2.0 − 2.1		1.055 ± 0.046	0.996 ± 0.021	1.03 ± 0.05
2.1 − 2.2		1.050 ± 0.047	0.997 ± 0.021	1.02 ± 0.05
2.2 − 2.3	0.6630 ± 0.0229	1.046 ± 0.047	0.997 ± 0.021	1.00 ± 0.04
2.3 − 2.4		1.045 ± 0.047	1.000 ± 0.020	1.036 ± 0.024
2.4 − 2.5		1.047 ± 0.047	1.001 ± 0.019	1.25 ± 0.04
2.5 − 2.6		1.050 ± 0.047	1.001 ± 0.019	0.994 ± 0.033
2.6 − 2.7		1.050 ± 0.046	0.998 ± 0.019	1.000 ± 0.032
> 2.7		1.053 ± 0.046	0.998 ± 0.018	1.0 ± 0

6.11.1 Calibration of the FEI algorithm

The working principle of FEI has already been discussed in Section 6.3.1. It combines many classifiers which perform reconstructions of the hadronic decays of B mesons in various decay chains. Furthermore, the training of the algorithm happens in MC. To ensure that the algorithm appropriately acts on Belle II data, its performance must be studied or *calibrated*.

The calibration study is performed on data collected by Belle II, for every simulation campaign, and the work is not part of the original work presented in this thesis. Full details of the calibration method are presented in Ref. [161], but the main details that are relevant to the work of the thesis are summarised here.

The calibration study uses $B \to X_{u/c}\ell\nu$ decays due to the branching fraction of almost 20 % and a clean experimental signature, where $X_{u/c}$ denotes an inclusive state originating from the c or u quark, similarly to the $X_{s/d}$ notation. Firstly, in each event, only the highest FEI probability tag-B candidate is selected with loose requirements on Fox-Wolfram moments (see Appendix F.4) and ΔE to ensure adequate $e^+e^- \to q\bar{q}$ suppression. Next, a high energy lepton $p_\ell^B > 1$ GeV/c is required in each event. This candidate is required to originate near the interaction point and its identification information from all sub-detectors is required to be consistent with a lepton.

After the selection, a binned likelihood fit for M_{bc} is set up, which contains three binned PDFs: signal $B \to X_{u/c}\ell\nu$ decays, secondary or misidentified leptons, and $e^+e^- \to q\bar{q}$ events. Here, secondary leptons identify leptons coming from the B mesons other than $B \to X_{u,c}\ell\nu$ decay. Misidentified leptons are used as a broad term for hadrons whose identification information is consistent with that of either an electron or a muon. The signal $B \to X_{u/c}\ell\nu$ PDF is composed of four sub-PDFs, particularly: $B \to D\ell\nu$, $B \to D^*\ell\nu$, $B \to X_u\ell\nu$ and the rest of $B \to X_c\ell\nu$ modes. The fit is performed separately for the following combinations of tag-B mesons and lepton:

- B^+ and e^-,

- B^+ and μ^-,

- B^0 and e^-,

- B^0 and μ^-.

This is shown in Figure 6.41.

The branching fractions of $B \to X_{u/c}\ell\nu$ are evaluated from the fitted distributions. These values are then directly compared with the world average values. A correction factor, C_{FEI}, is derived such that the two values become compatible. The leading evaluated systematic uncertainties come from the imperfect experimental knowledge of the $B \to X_u\ell\nu$ branching fractions and their form factors, the fit model composition, tracking and particle identification uncertainties. For the Belle II simulation campaign used in this analysis, and averaged for both lepton modes, the result is as follows:

$$C_{\mathrm{FEI}}(B^+) = 0.6599 \pm 0.0225; \quad C_{\mathrm{FEI}}(B^0) = 0.6695 \pm 0.0237, \tag{6.21}$$

where two different calibration factors are presented for $\mathtt{feiB}^+$ and $\mathtt{feiB}^0$ modes, respectively. Therefore, for an adequate comparison with Belle II data, any Belle II MC involving the use of FEI is henceforth scaled appropriately.

6.11.2 Calibration of π^0 and η suppression tools

It was seen in Section 6.4.4 that one of the strongest tools for background suppression in this analysis is the π^0 and η suppression tool. Consequentially, any data-simulation discrepancies
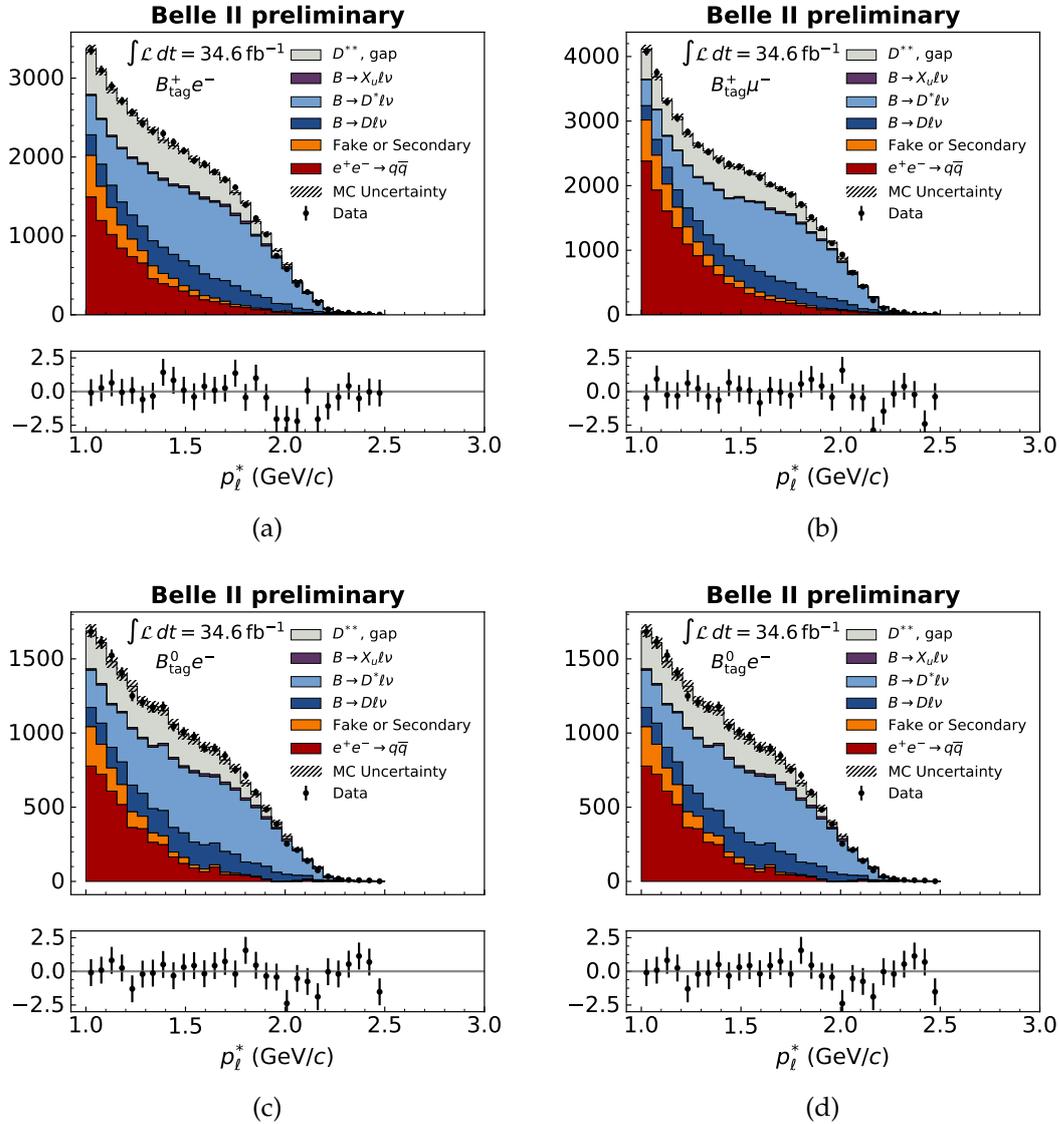
Figure 6.41: Illustration of the fits to $B \rightarrow X_{u/c} \ell \nu$ decays in the FEI calibration study. Results are shown for the combinations of charged and neutral tag-B modes with e^- ((a) and (c)) and μ^- ((b) and (d)). Different fit components are shown in the legend and the subpanels contain the pulls of the fit. Figures are taken from Ref. [161].

have a high impact on the final result. The calibration of the π^0 and η veto is performed in an independent study and is not part of the original work presented in this thesis but the relevant calibration study is discussed in this Section. Although the calibration analysis only studies $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$, it is assumed that the corrections are also valid for $\mathcal{P}(\eta \rightarrow \gamma\gamma)$ selections. The main concern for the $B \rightarrow X_s\gamma$ analysis is the primary (signal) photon efficiency: the number of photons that do not originate in light meson decays and get rejected given a certain $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$ selection.

The calibration study uses $B^+ \rightarrow \bar{D}^0[\rightarrow K^+\pi^-]\pi^+$ and $B^0 \rightarrow D^-[\rightarrow K^+\pi^-\pi^-]\pi^+$ decays, where the square brackets denote a subdecay of the D meson. The π^+, originating in the primary B decay, is combined with all other photon candidates in the event in a strategy described in Section 6.4.4 assuming a null-mass hypothesis. This produces many π^0-like combinations (*pseudo*$-\pi^0$) which yield a $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$ score with minimal background from real π^0 decays.

The reconstruction requires good-quality tracks that originate near the interaction point. The identification information from Belle II subdetectors is used to distinguish between pions and kaons. Because the π^+ from the primary B decay is combined with other photons, a massless hypothesis is used for calculations of the invariant mass and the helicity angles for the MVA. After constructing the pseudo-π^0, selections on $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$ are performed accordingly to those chosen in this analysis. Therefore, two distributions are created: one with no $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$ requirement, and a subset distribution with $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma) < 0.4$. In both cases, the charged and neutral B channels are combined.

An unbinned M_{bc} fit is performed on distributions with and without the $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$ selections. The M_{bc} is modelled by a Crystal Ball function for signal decays and an ARGUS function for continuum background. All parameters of Crystal Ball and continuum are unconstrained. An additional PDF to model the peaking non-signal $B\bar{B}$ components is used. This PDF is initialised in MC as a sum of a Gaussian and an Argus. The shape parameters and normalisation of the $B\bar{B}$ background PDF are not estimated but kept at the initialised values. The fits of the data in the case of no $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$ selection, and a $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma) < 0.95$ selection are given in Figure 6.42.

The fit extracts the counts of $B \rightarrow D\pi^+$ events, $N_{B \rightarrow D\pi^+}$, as the normalisation parameter of the Crystal Ball. An efficiency, $\varepsilon \equiv N_{B \rightarrow D\pi^+}/N_{B \rightarrow D\pi^+}^{\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)<0.4}$ is defined, which corresponds to the primary photon efficiency. If the fit is performed in MC and data, an efficiency ratio can be used as a correction factor:

$$R_{\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)} = \frac{N_{B \rightarrow D\pi^+}/N_{B \rightarrow D\pi^+}^{\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)<0.4}|_{\text{data}}}{N_{B \rightarrow D\pi^+}/N_{B \rightarrow D\pi^+}^{\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)<0.4}|_{\text{MC}}}. \tag{6.22}$$

The corrections are calculated in 200 MeV intervals of the laboratory frame energy of the primary π^+. Results for corresponding to selection chosen in this analysis are given in Figure 6.43. The internal Belle II study providing these corrections was only performed in the range of 1.8 to 3.0 GeV in the laboratory frame energy of the primary π^+. A linear extrapolation is performed to estimate the values outside the range. It is observed that the linear extrapolation is consistent, within errors, with the corrections in $1.8 - 3.0$ GeV. Therefore, a correction factor of 1.10 ± 0.05 is chosen for events outside the range covered by the calibration study. This value is consistent with the correction factors in other E_γ bins.
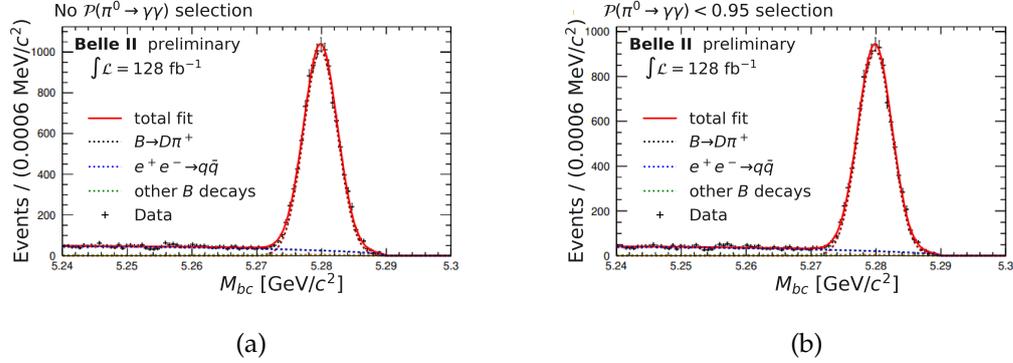
(a)   (b)

Figure 6.42: The fit estimating the number of $B \to D\pi^+$ events in Belle II data. Different PDFs used in the fit are shown in the legend and explained in the text. The fit is performed on a sample without (a) and with (b) the $\mathcal{P}(\pi^0 \to \gamma\gamma)$ selection applied. The extracted values from MC and data are then combined to calculate $\mathcal{P}(\pi^0 \to \gamma\gamma)$ correction factors (see Equation (6.22)). These Figures are produced by a Belle II internal study of the π^0 veto and are not part of the original work in this thesis. Only the labels and legends have been adapted.
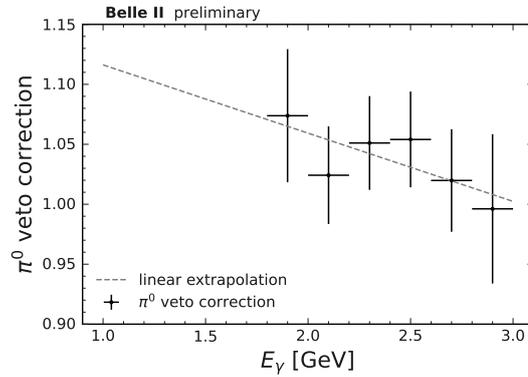
Figure 6.43: The corrections $R_{\mathcal{P}(\pi^0 \to \gamma\gamma)}$ for the $\mathcal{P}(\pi^0 \to \gamma\gamma) < 0.4$ selection used in this analysis. The results cover $1.8 - 3.0$ GeV energies in the laboratory frame, E_γ. Because the laboratory frame energies cannot be trivially transformed to the B meson rest frame energies, a linear extrapolation to lower energies is performed as identified by the dashed line.

6.11.3   Belle II calorimeter photon detection efficiency

One of the main necessities of a measurement involving photons in the final state, is, of course, a precise and accurate simulation of the ECL. Although it is designed with precision and resolution suitable for flavour physics studies, exact data-simulation differences have to be evaluated. A Belle II calorimeter photon detection efficiency study has been performed. The initial setup of the analysis has been prepared by Dr. Natalia Kovalchuk and Prof. Dr. Torben Ferber. However, as part of the original work presented in this thesis, the analysis was updated and reworked for later versions of Belle II data. It also supplemented the initial studies with a full systematic uncertainty evaluation. While it is a critical study for the $B \rightarrow X_s \gamma$ analysis, the results are also routinely used in other analyses that utilise photons in their final states. The study is summarised in a Belle II public note [162]. Here, the main measurement concepts and the results relevant to the measurement of $B \rightarrow X_s \gamma$ are presented.

To measure the photon detection efficiency, one must first have the knowledge that a photon has been created in an event and then search for it within the calorimeter. In this efficiency study, $e^+e^- \rightarrow \mu^+\mu^-$ scattering events are employed. In particular, collision events where a high energy photon is radiated in the initial state are sought. The concept of the efficiency measurement is sketched in Figure 6.44.
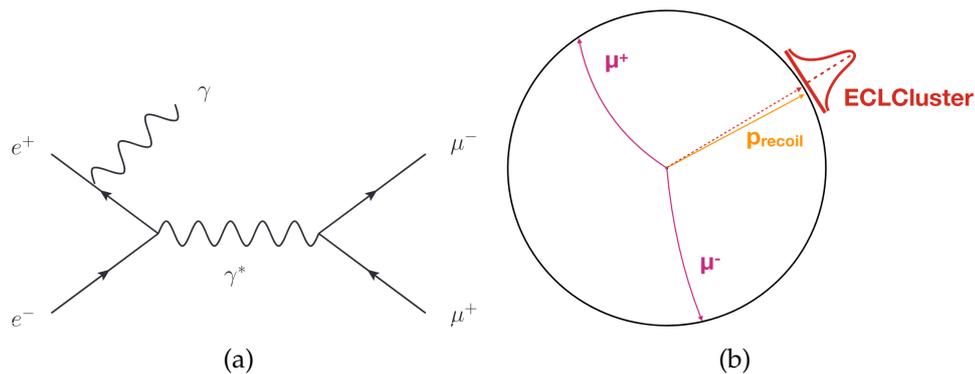
Figure 6.44:   The main concepts of the measurement of the Belle II photon detection efficiency. The Feynman diagram of the $e^+e^- \rightarrow \mu^+\mu^-$ events, where a photon is radiated in the initial collision state, is shown in (a). Due to the radiated photon, the resulting dimuon system will have a missing momentum with respect to the usual collision energy $\sqrt{s} \approx 10.58$ GeV. The direction of the missing momentum can be extrapolated to search for photon clusters in the calorimeter, as sketched in (b).

The main goal is to reconstruct two muon tracks in each event and evaluate their total momentum and energy. If a high energy photon (further called initial-state radiation or ISR) is created before the collision, the energy of the dimuon system has a certain degree of missing-momentum whose direction coincides with that of the emitted photon. This missing momentum direction is called *recoil* and is characterised by the magnitude (equivalent to the photon energy), polar angle and azimuthal angle. Therefore, by selecting events where such recoil is present, one looks for photon clusters corresponding to the angle and the energy within the calorimeter. This gives a photon detection efficiency estimate through a simple event count-

ing relation:

$$\varepsilon_\gamma(|\vec{p}_{\text{recoil}}|, \theta_{\text{recoil}}, \phi_{\text{recoil}}) = \frac{N(\text{photon found} \cup \text{recoil found})}{N(\text{recoil found})}, \tag{6.23}$$

which can be evaluated as a function of missing momentum with a magnitude $|\vec{p}_{\text{recoil}}|$ and corresponding angles θ and ϕ.

Many background processes degrade the efficiency by producing events with sufficient recoil momentum. A notable example is the $e^+e^- \to \tau^+\tau^-$ events which may produce two muons and four neutrinos through subsequent τ decays. The presence of neutrinos creates a missing momentum which mimics the recoil that originates from an ISR photon. Furthermore, tracking inefficiencies can lead to an incorrect measurement of the direction or magnitude of the recoil vector. Finally, two or more ISR photons per event further complicate the photon finding procedure, resulting in an overall drop in efficiency. Therefore, Equation (6.23) is more correctly referred to as *photon finding efficiency*, rather than photon detection efficiency.

The photon finding inefficiency effects are suppressed to a certain degree through a double ratio measurement:

$$R_\gamma = \frac{\varepsilon_\gamma^{\text{DATA}}}{\varepsilon_\gamma^{\text{MC}}}, \tag{6.24}$$

where ε_γ are respective values of photon finding efficiency calculated in data or simulation, based on Equation (6.23). The double ratio, R_γ, is considered the photon detection efficiency data-to-simulation ratio in this analysis.

First, tracks consistent with $e^+e^- \to \mu^+\mu^-$ processes are selected by requiring each event to have exactly two charged tracks that:

- have a high-momentum requirement $p > 1$ GeV/c;

- are consistent to have originated from the interaction point;

- act as a minimum ionising particle in the ECL: leave energy deposits smaller than 300 MeV and less than 80% of the total momentum.

The two muon tracks passing these requirements are used to evaluate the missing energy and momentum of the event, requiring the recoil magnitude $p_{\text{recoil}} > 0.2$ GeV/c. Backgrounds from various $e^+e^- \to$ hadrons or $e^+e^- \to \tau^+\tau^-$ are strongly suppressed by the mass associated with the missing momentum requirement, $m_{\text{Recoil}}^2 < 2$ GeV/c^4. This requirement ensures that the particle associated with the missing momentum is consistent with a photon. Generally, this does not have to be true for non $e^+e^- \to \mu^+\mu^-(\gamma)$ events where more than two tracks are present. Additional checks, such as sufficient isolation of the muon tracks and the recoil are required to ensure adequate separation between their energy deposits in the ECL. If all the aforementioned requirements are passed, an event is considered to have a *recoil found*. The distributions for events with a recoil found are shown in Figure 6.45a.

The photon candidates are selected by requiring them to have a centre-of-mass energy of at least 75 MeV and a timing of the associated cluster at ±200 ns. These requirements were optimised to reduce the impact of beam background photons. No tighter selections on

photons or their reconstruction quality are made to ensure that no bias is introduced in the detection efficiency.

The recoil candidates are matched to photon clusters in the ECL via two matching requirements:

- Photons have to be within 0.3 rad cone around the recoil direction;

- The cluster energy to recoil momentum ratio must satisfy $1.2 > E_\gamma/p_{\text{Recoil}} > 0.5$.

If both requirements are fulfilled the event is considered to have a *photon found*. The distribution of events where the recoil is successfully matched to a photon is given in Figure 6.45b.
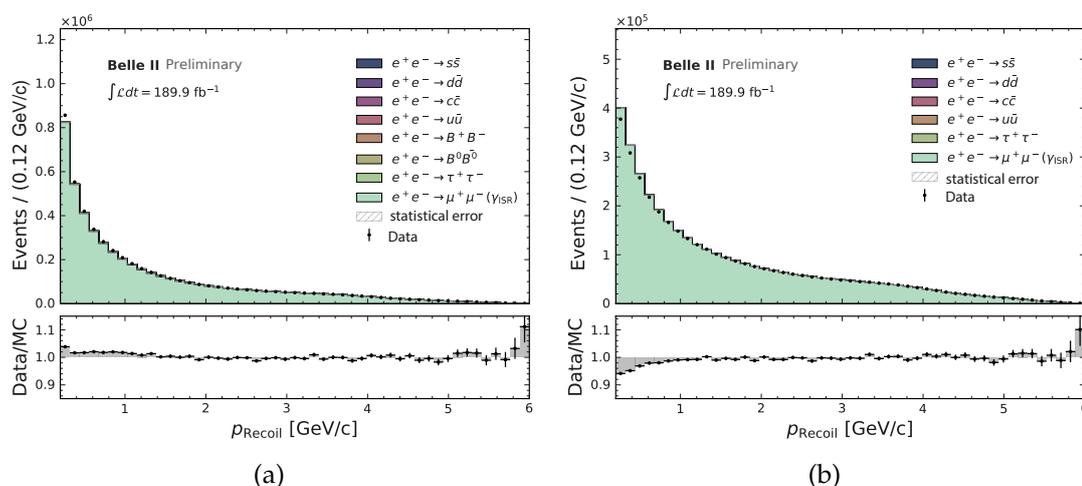
Figure 6.45: Distribution of $e^+e^- \rightarrow \mu^+\mu^-$ events with a photon radiated in the initial interaction state, as a function of the magnitude of the missing momentum of the dimuon system. Events where a missing momentum has been found are shown in (a). Events where a missing momentum has been found and it was consistent with a photon in the calorimeter, as discussed in the text, are shown in (b). Various sources of background events are also included, and they can be seen to be at a low level. Overall, the signal and background events describe data accurately. The subpanels show the data-to-simulation ratio. These Figures contain only statistical uncertainties. They show the results for the data sets of a size equivalent to the ones used in this analysis.

The events represented in Figure 6.45a reflect the content of the numerator of Equation (6.23), whereas Figure 6.45b – that of the denominator. The finding efficiencies and their ratio in data and MC are calculated. The expected backgrounds, evaluated in MC, are subtracted from the data distributions. The results are shown in Figure 6.46. The data-to-simulation ratio is generally high and approximately equal to unity for photons above 1 GeV. A drop-off for low energy photons (low p_{Recoil}) is attributed to the presence of soft ISR photons in the event which affect the direction of the recoil. This effect becomes larger with the lower photon energy, where the impact of a second ISR photon grows.

The systematic uncertainties are also calculated. Most of the selections are tightened and loosened to evaluate the dependence of efficiency on selection requirements. The maximum
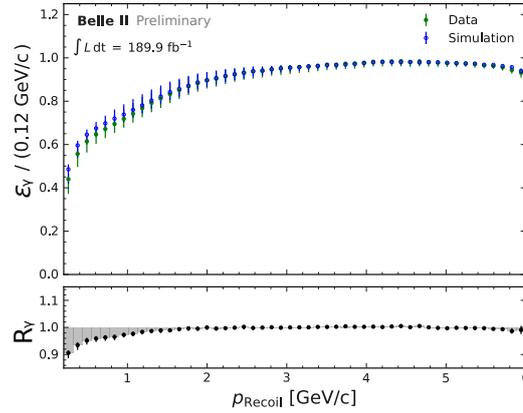
Figure 6.46:   The photon finding efficiency as a function of the missing momentum of the dimuon system, p_{Recoil}. The subpanel shows the ratio of data and simulation finding efficiencies, as per Equation (6.24). The Figures include systematic uncertainties, which are correlated between data and MC. They show the results for the data sets of a size equivalent to the ones used in this analysis.

shifts in the central value are assigned as systematic uncertainties. A full shift to the central value without the remaining background subtraction in data is also added as a systematic uncertainty. The largest systematic uncertainties arise from the leftover background modelling, m^2_{Recoil} selection variations and muon calorimeter energy deposit variation requirements. In total they are at $O(1\%)$ level.

Equivalent distributions to Figure 6.45 as 3-D functions of p_{Recoil}, ϕ_{Recoil} and θ_{Recoil} are produced and 3-D efficiency maps are calculated. Based on the detected photon direction and energy, appropriate corrections are applied. In the $B \to X_s\gamma$ analysis, the event-level information is lost after performing the M_{bc} fit and subtracting the remaining $B\bar{B}$ background. Therefore, the average corrections are evaluated with the $B \to X_s\gamma$ hybrid-signal model based on the E^B_γ spectrum binning. These values are provided in Table 6.16.

6.11.4   Modelling of remaining-$B\bar{B}$ background processes

While the results of previous Sections correct for the procedures used in the removal of photon and tag-side backgrounds, they do not account for any discrepancies that may be introduced when generating the MC. Although a full study of all possible background modes that may contribute to the leftover $B\bar{B}$ background and their description in the Belle II simulation is outside of the scope of this work, the adequacy of background simulation is studied for the analysis presented here. In general, the Belle II MC is validated and known to produce an accurate and precise simulation of most processes that are common to B-factories. However, our knowledge of the nature and the Standard Model is constantly improving, therefore, the branching fractions or other parameters used in the generation of the generic $B\bar{B}$ simulation may not be updated in time as the simulation campaigns happen roughly yearly. The goal of this study is to check that the generated branching fractions of the main backgrounds that contribute to $B \to X_s\gamma$ match those reported by Ref. [16].

First, only events contributing to good tag-B mesons based on studies in Section 6.7.3 are selected. In each E_γ^B interval, all (non-$B \to X_s\gamma$) B decay modes that produce high-energy photon candidates are selected. The modes are ranked by their relative abundance within that E_γ^B interval. Particularly in the low-E_γ^B region, there are hundreds of different B meson decay channels that can contribute to the background. Pragmatically, only background B decays that contribute at least 1% to the background in at least one E_γ^B interval are further considered. These requirements encompass 53 B^+ modes and 39 B^0 decay modes. They are listed in Table J.2 and Table J.1, respectively. These Tables also contain their relative abundances in every given bin. As hundreds of B decay modes contribute at $O(< 1\%)$ relative abundance, these requirements may not cover all background modes. However, it is assumed that the corrections for non-dominant background modes should on average be unity, as large branching fraction differences in Belle II simulation are not expected.

The main sources of background photons come from various $B \to D$ transitions, where photons originate from either subsequent D decays or the accompanying particle, e.g. $B \to D\rho$. Semileptonic B decays are also a major source of background. In general, $B \to D$ transitions are largely dominating up to $2.2 - 2.3$ GeV. At higher E_γ^B, other types of decays become prominent, but more sporadically and without clearly dominating modes. In particular, various rare B decays, such as $B \to K$, $B \to \pi$ and $B \to \eta$ transitions become more prominent.

The selected background B decay modes have their branching fraction in the Belle II simulation files compared with that reported by Ref. [16]. Based on the findings, the following correction, for every mode in every E_γ^B bin is derived:

$$R_{\mathrm{BB}}^{B \to X}(E_\gamma^B) = f(E_\gamma^B) \times \frac{\mathcal{B}_{\mathrm{PDG}}}{\mathcal{B}_{\mathrm{Belle\ II}}}, \tag{6.25}$$

where $f(E_\gamma^B)$ is the relative fraction of a background mode $B \to X$ in a given E_γ^B interval. $\mathcal{B}_{\mathrm{PDG}}$, $\mathcal{B}_{\mathrm{Belle\ II}}$ denote the branching fractions found in the Particle Data Group summary [16] and Belle II simulation, respectively. If available, the $\mathcal{B}_{\mathrm{PDG}}$ is varied according to the provided uncertainty, otherwise, a 100% variation is taken and the central value shifts calculated with appropriate variations are assigned as uncertainties for $R_{\mathrm{BB}}^{B \to X}(E_\gamma^B)$. The results are summed together in each E_γ^B bin and their uncertainties are propagated, taking the correlations, resulting from the fact that the same decay mode may contribute in multiple E_γ^B bins. The correction factors from B^+ and B^0 background modes are averaged, assuming no correlation between different B charges. The final correction factors and their uncertainties are shown in Table 6.16. The values are (nearly) all consistent with unity, as expected considering the high quality of the Belle II simulation.

The correlation matrix of the correction factors arising from the fact that similar decay modes occur across many bins is shown in Figure 6.47. It can be seen that mostly low-E_γ^B bins are correlated, whereas high-E_γ^B bins show smaller correlations. Indeed, abundant background processes fall off quickly with increasing E_γ^B. In the high-E_γ^B region, background photons are rarer and often originate as outliers from a variety of rare decays, decorrelating the bins.
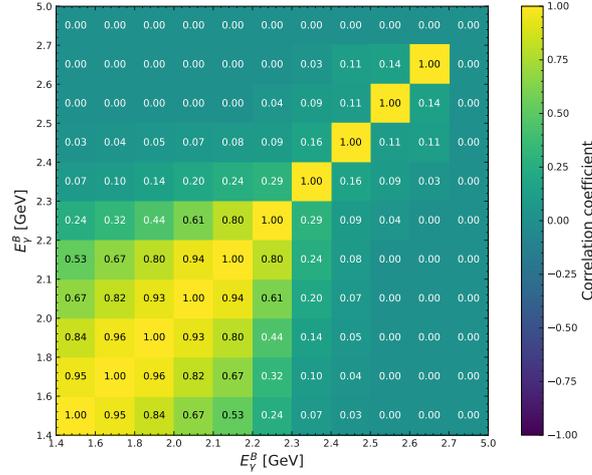
Figure 6.47:   The correlation matrix of $B\bar{B}$ corrections.   Particularly low-E_γ^B bins are largely correlated because similar background modes contribute in these regions. On the other hand, high-E_γ^B background photons originate more sporadically and from fewer sources, thereby reducing the correlation.

6.11.5   Out-of-time photon suppression

Although no special corrections are calculated, an additional selection is added to ensure that photons produced by the beam background[1], and products of previous collision events are not included in the candidate photon selection. Such machine-related backgrounds vary based on the experimental conditions and these effects are not well captured by the Belle II run-period independent simulation that is used in preparation for this analysis. Therefore, a timing selection requires the photon to be registered in a time window $|\tau_\gamma| < 200$ ns around the collision time. Furthermore, to reject photons associated with low-quality cluster reconstruction, a $|\tau_\gamma|/\Delta\tau_\gamma < 2$ requirement is added, where $\Delta\tau_\gamma$ is the uncertainty of the photon timing measurement. Using a technique analogous to $e^+e^- \rightarrow \mu^+\mu^-$ recoil study, it is observed that it degrades the photon finding efficiency by less than 2%. Therefore, these selections are employed with no additional data-to-simulation correction.

6.12   Background validation studies

Up until now the discussion of the analysis revolved around MC studies. The full analysis procedure using simulation is defined and produces an unbiased, stable result in Section 6.10. Then, Section 6.11 looked at the corrections required to accurately correct differences expected in MC to better represent data. At this stage, all appropriate measures have been taken so that the analysis can be fully applied to Belle II data. The concepts of a blinded analysis, however, dictate that to ensure no biases are present, the full analysis procedure must be performed in validation samples. For the $B \rightarrow X_s\gamma$ analysis, four validation samples are defined:

---

[1]Particles that are not from the primary collision but are produced by the interactions of the accelerator beam and residual gas in the beam pipe or the material of the detector.

- $e^+e^- \to q\bar{q}$ sample that contains collision data collected 60 MeV below the $\Upsilon(4S)$ centre-of-mass energy.

- Sample with an enhanced $e^+e^- \to q\bar{q}$ component, where the `BDT output` score requirement is inverted: `BDT output < 0.4`.

- Sample with an enhanced $B\bar{B}$ background component, where the π^0 and η veto requirements are inverted.

- Sideband $E_\gamma^B \in (1.4, 1.8)$ GeV, where the signal-to-background ratio is small, and $2.7 < E_\gamma^B$ GeV, where the signal is kinematically forbidden, respectively. Note that some signal events may still be present in the latter due to resolution effects, but they are not expected to be statistically significant.

In this Section, the analysis selections, $M_{\rm bc}$ fitting and leftover-$B\bar{B}$ background subtraction are investigated thoroughly using these samples.

6.12.1 Validation on the $e^+e^- \to q\bar{q}$ off-resonance sample

The validation on $e^+e^- \to q\bar{q}$ events is performed using only $e^+e^- \to q\bar{q}$ simulation. The goal of this validation is to ensure that continuum backgrounds are described by the simulated samples correctly. Although Section 6.5.4 partially ensures this, only the distribution shape requirements are tested there. Furthermore, the best candidate selection, which was developed on $B\bar{B}$ samples, may change the conclusions that were found earlier.

All the corrections for π^0 and η veto, in addition to the FEI calibration, are applied as discussed in Section 6.11. The full analysis selection procedure involving the most optimal selections and the best tag-side candidate selection is applied, as presented in Sections 6.6.2 and 6.7.2. The resulting continuum-only E_γ^B distributions are shown in Figure 6.48.

       (a)   (b)   (c)
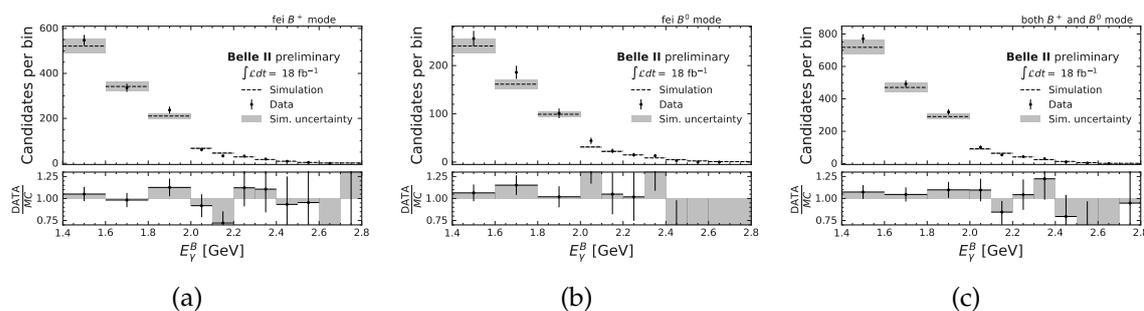
Figure 6.48: Validation of the E_γ^B distribution of $e^+e^- \to q\bar{q}$ events. Good agreement is observed in `feiB`$^+$ (a), `feiB`^{0} (b) and the combined sample (c). The uncertainty for data contains only the statistical component. The MC also contains systematic uncertainties corresponding to Table 6.16.

Overall, the agreement between continuum data and MC is good. Due to the rather low size of the off-resonance data sample and the strong continuum suppression in this analysis,

the statistical uncertainty is relatively large. It is concluded that the E_γ^B distribution of $e^+e^- \rightarrow q\bar{q}$ events is well-modelled in simulation and follows the expectations seen in earlier Sections.

Due to known issues with beam energy values in the off-resonance data, which affect the M_{bc} calculation (but not the overall validity of other values), the $e^+e^- \rightarrow q\bar{q}$ off-resonance M_{bc} distributions do not accurately represent the M_{bc} values of continuum events. Therefore, these samples are not used for M_{bc} distribution and M_{bc} fitting validation.

6.12.2   Validation on the $e^+e^- \rightarrow q\bar{q}$ enhanced sample

As it was mentioned in Section 6.12.1, the continuum data samples were not used to validate the M_{bc} fitting. However, an alternative validation sample is prepared, where the continuum component is enhanced. This is achieved by inverting the `BDT output` selection (see Section 6.5) thereby suppressing $B\bar{B}$ events. To ensure a minimal amount of $B \rightarrow X_s\gamma$ events in the sample, $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$ and $\mathcal{P}(\eta \rightarrow \gamma\gamma)$ are also inverted. This creates a sample with mostly $e^+e^- \rightarrow q\bar{q}$ events and small components of $B\bar{B}$ events. The inverted values are chosen as `BDT output` < 0.2 and $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma) > 0.4$ and $\mathcal{P}(\eta \rightarrow \gamma\gamma) > 0.4$. The resulting E_γ^B spectra for both FEI modes are shown in Figure 6.49.

| (a) | (b) | (c) |
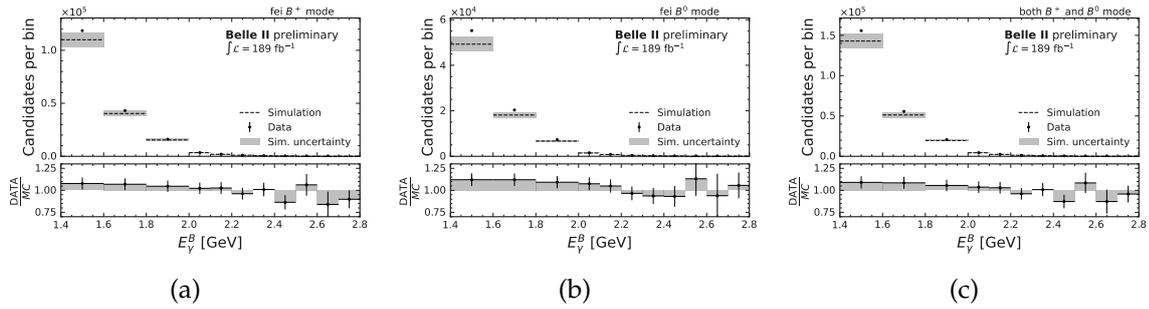
Figure 6.49: The E_γ^B distribution of $q\bar{q}$ enhanced samples (see Section 6.12.2). Adequate agreement is observed in `feiB`$^+$ (a), `feiB`^{0} (b) and the combined sample (c). The uncertainty for data contains only the statistical component. The MC uncertainty also contains systematic uncertainties corresponding to Table 6.16.

Although the agreement is generally adequate, particularly in the signal region, a small excess of events is observed in the low-E_γ^B region. Similarly, the resulting M_{bc} distributions are shown in Figure 6.50. A striking difference from the generally good agreement observed so far can be seen for `feiB`$^+$, `feiB`^{0} and the combined sample. A larger amount of continuum events, particularly at low-M_{bc}, is present. Moreover, the upper threshold of M_{bc} is shifted. The overall data-to-simulation discrepancy is as high as 20%.

These differences are understood as a result of data-taking period independent simulation used in this analysis. In normal data collection conditions, $\sqrt{s}$ is not perfectly stable: minute variations or drifts can occur over time. The data-taking period independent simulation does not account for these changes, where the collision energy is simply set to a predetermined average value. The M_{bc} endpoint is directly affected by $\sqrt{s}$, as seen in Equation (3.4), with lower values of $\sqrt{s}$ pushing the M_{bc} threshold lower. On the other hand, a larger overall amount of $q\bar{q}$ events is understood as a consequence of the fact that collecting the data at
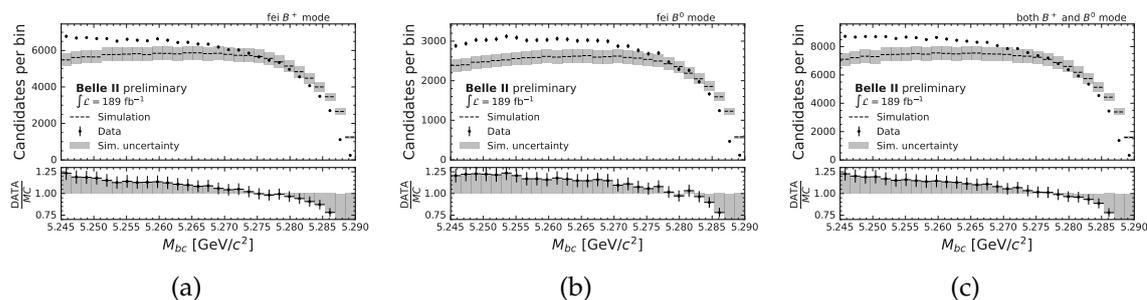
Figure 6.50: The M_{bc} distribution of $q\bar{q}$ enhanced samples (see Section 6.12.2). Some clear differences in `feiB`$^+$ (a), `feiB`^{0} (b), and the combined sample (c) can be observed. Particularly, it is evident that there are more continuum events at low M_{bc}. The M_{bc} endpoint is shifted to lower values. These results motivate the modification of the M_{bc} fitting procedure, summarised in Table 6.17.

lower collision energies (but not lower than $\Upsilon(4S)$ energy) enhances the $e^+e^- \to q\bar{q}$ process cross-section. Altogether, this leads to more continuum events present in the data sample than predicted by the data-taking period independent simulation.

For good tag-B mesons, the shift would not occur. This is a result of the fact that p_B^* (as seen in Equation (3.4)) is directly related to the total energy of the collision. The resonant-like behaviour in M_{bc} is driven by the B meson mass and, therefore, for correctly reconstructed objects is valid irrespective of $\sqrt{s}$. Such constraints are not present for misreconstructed events, therefore, the shifts are expected to happen there. Although the most robust solution is the usage of data-taking period dependent simulation, at the time of preparation of this analysis such MC was not yet available at Belle II. While future studies will be able to rely on it, in this analysis additional steps were taken to account for for the lack of it.

While M_{bc} is strongly affected, the E_γ^B spectrum is still well-described. This is a consequence of the fact that E_γ^B does not depend as strongly on $\sqrt{s}$ as M_{bc}. Therefore, a correction is only necessary for the M_{bc} distribution *and* only for combinatorial $B\bar{B}$ and continuum events. An *ad hoc* approach is developed, where the M_{bc} distribution for MC events is shifted manually. The procedure is as follows:

- Count the frequencies of each $\sqrt{s}$ value occurring in the Belle II on-resonance data set;

- Randomly remove half of the beam energies in the simulated Belle II data set of *events where no good tag-B mesons are present*;

- Replace the removed beam energies with the values of the first step, based on the frequencies they occur at in the Belle II on-resonance data set.

Only 50% of energies are replaced to minimise any potential conscious bias that such a procedure could introduce. Note that the replacement of $\sqrt{s}$ only affects the M_{bc} calculation and not other observables which may rely on $\sqrt{s}$ in their definition. The result of the correction on the M_{bc} distribution of the sample of both FEI modes combined is shown in Figure 6.51. Although a perfect agreement is not achieved via the *ad hoc* correction, the $M_{bc} > 5.27$ GeV/c^2
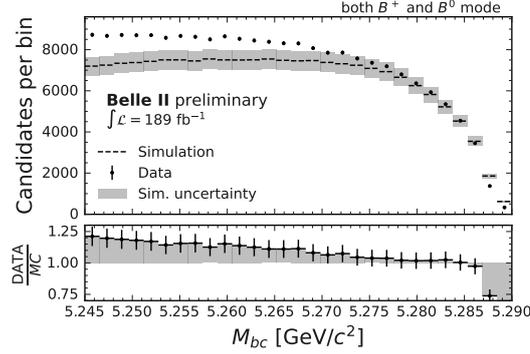
Figure 6.51: The M_{bc} distribution of $q\bar{q}$ enhanced samples (see Section 6.12.2), where an M_{bc} correction has been applied to the simulated distribution. Although the correction does not perfectly align the distributions, the signal region ($M_{bc} \approx 5.28$ GeV/c^2) is described correctly.

region description improves. This is seen when comparing Figure 6.51 and Figure 6.50.

A key point to discuss here is the effect the different M_{bc} shape may have on the M_{bc} fitter. The differences in the tail and the end-point are expected to not strongly affect the result because the M_{bc} fitter is prepared with shape differences in mind (see Section 6.8.3). However, the initial parameters of the M_{bc} fitter in Table 6.15 have to be updated to emphasise that a different corrected M_{bc} is the new fitting observable. The M_{bc} fitter is therefore updated, following the same procedures as Section 6.8.3 and the new values are given in Table 6.17.

Table 6.17: The summary of the M_{bc} fitting model used in this analysis after updating the fitting variable to the corrected M_{bc}. The parameters are initialised at the values that are listed, corresponding to the ones determined in the primary fitting steps, explained in Section 6.8.3, with M_{bc} replaced by a corrected M_{bc} value. The values that are bolded in the Table are not estimated from the final corrected M_{bc} fit but are kept at their initialised values. On the other hand, all non-bolded values can vary in the final fitter. Uncertainties are evaluated using the HESSE method in the primary fitting steps. In the Table, they are omitted if the relative uncertainty is lower than 0.1%.

E_γ^B bin	Crystal Ball					Chebyshev						Argus		
	N_{CB}	μ	σ	α	n	N_{cheb}	k_1	k_2	k_3	k_4	k_5	N_{ARGUS}	c	m_0
1.4 − 1.6	17294 ± 131					70507 ± 266	−0.150 ± 0.007	−0.382 ± 0.007	−0.272 ± 0.006	−0.132 ± 0.006	−0.003 ± 0.006	76798 ± 277	−26.35 ± 0.81	
1.6 − 1.8	10218 ± 101					33666 ± 183	−0.084 ± 0.010	−0.411 ± 0.010	−0.300 ± 0.009	−0.140 ± 0.009	−0.003 ± 0.009	50658 ± 225	−21.08 ± 0.99	
1.8 − 2.0	5947 ± 77					16192 ± 127						31228 ± 176		
2.0 − 2.1	1938 ± 44					4279 ± 65						9983 ± 100		
2.1 − 2.2	1246 ± 35					2589 ± 51						6951 ± 83		
2.2 − 2.3	909 ± 30	**5.279**	**0.003**	1.573 ± 0.035	3.561 ± 0.22	1581 ± 40	0.030 ± 0.011	−0.438 ± 0.011	−0.377 ± 0.009	−0.175 ± 0.010	−0.007 ± 0.010	4470 ± 67	−18.64 ± 0.93	**5.2897**
2.3 − 2.4	985 ± 31					1197 ± 35						2663 ± 52		
2.4 − 2.5	1213 ± 35					779 ± 28						1482 ± 39		
2.5 − 2.6	626 ± 25					310 ± 18						662 ± 26		
2.6 − 2.7	62 ± 8					52 ± 7						211 ± 15		
2.7 − 5.0	1 ± 1					6 ± 2						73 ± 9		

This fitter is applied to the continuum-enhanced samples of Belle II data used in this analysis, and the fitted M_{bc} distributions are shown in Figure 6.52. The Figures also contain the number of good tag-B mesons that are estimated by the fitter. It can be seen that no statisti-

cally significant peak in M_{bc} is extracted. Therefore it can be concluded that the validation of the M_{bc} fitter is successful: enough flexibility is seen to describe the slight difference in M_{bc} shapes between MC and real data, without introducing a bias to the extracted number of good tag-B mesons.
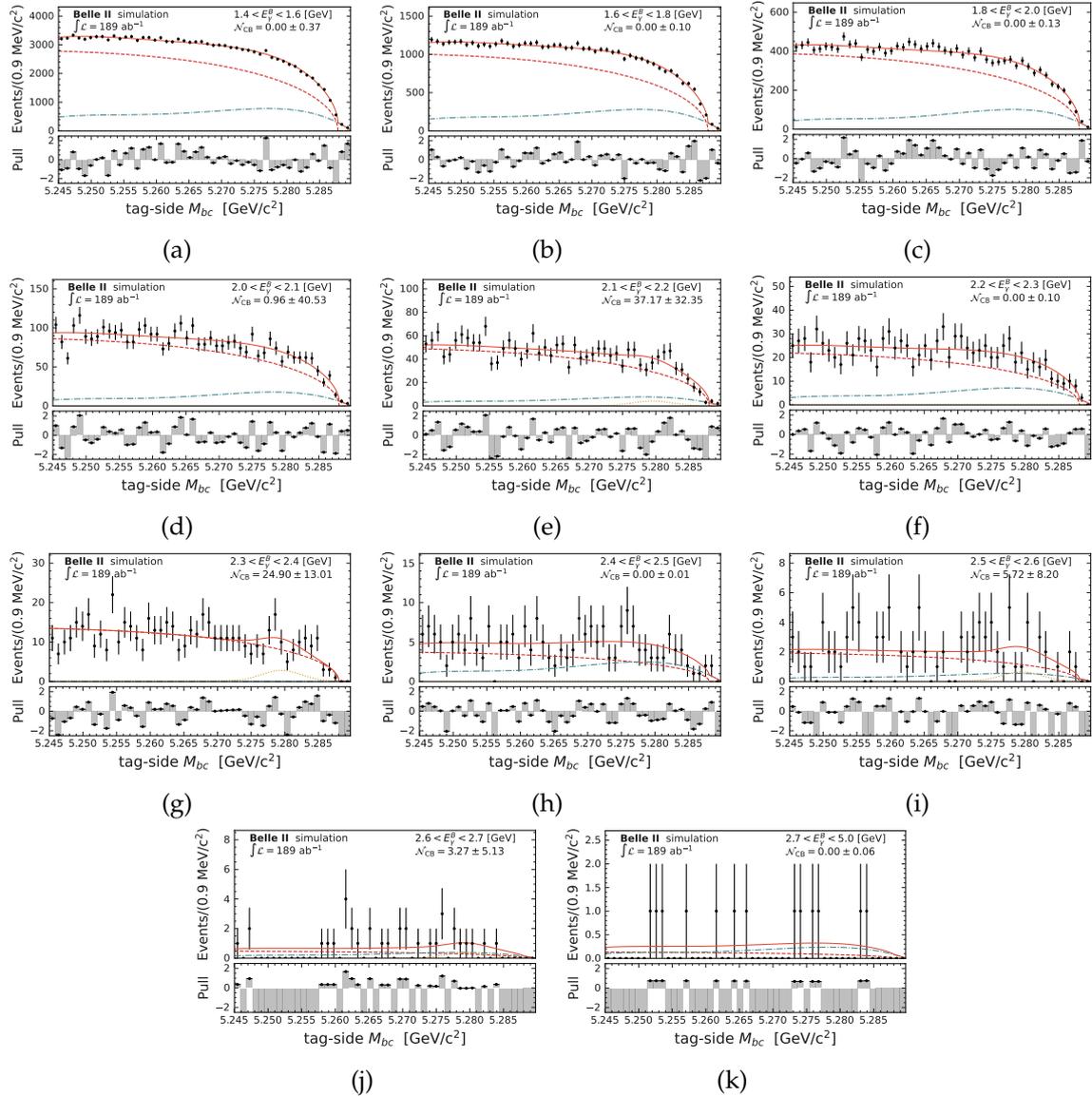
Figure 6.52: The fits of Belle II data set corresponding to 189 fb^{-1} with selections that enhance $e^+e^- \to q\bar{q}$ events, as discussed in Section 6.12.2. The fitting model from Table 6.17 is used, defined on the corrected M_{bc} to account for variations of $\sqrt{s}$ in Belle II data. A good description of the M_{bc} distributions can be seen throughout the E_γ^B bins. As the continuum component in each E_γ^B bin is enhanced, no good tag-B mesons are expected.

6.12.3   Validation on the $B\bar{B}$-background enhanced sample

In the last Section, the validation on a sample with $e^+e^- \to q\bar{q}$ events was performed. Another important validation, given the background subtraction step (Section 6.9), is the $B\bar{B}$ background description in MC.

Firstly, a $B\bar{B}$-background enhanced sample is prepared. This is done in almost the same manner as Section 6.12.2, except with the continuum suppression requirement unchanged from the optimal selection. In this case, only the $\mathcal{P}(\pi^0 \to \gamma\gamma)$ and $\mathcal{P}(\eta \to \gamma\gamma)$ selections are inverted. The selections are chosen as $\mathcal{P}(\pi^0 \to \gamma\gamma) > 0.6$ and $\mathcal{P}(\eta \to \gamma\gamma) > 0.6$, as these selections ensure that the signal-to-background ratio is less than 0.1%. As the selections are inverted, so must the corrections that account for them be modified (Table 6.16). In particular, the corresponding corrections for the $\mathcal{P}(\pi^0 \to \gamma\gamma)$ and $\mathcal{P}(\eta \to \gamma\gamma)$ are propagated as follows:

$$\text{Corr}_{>0.x} = \frac{1}{\text{Corr}_{<0.x}}; \quad \sigma(\text{Corr}_{>0.x}) = \frac{1}{\text{Corr}_{<0.x}^2} \times \sigma(\text{Corr}_{<0.x}). \tag{6.26}$$

The E_γ^B distribution of the $B\bar{B}$-enhanced sample is shown in Figure 6.53. Overall, the distributions show excellent agreement between data and MC. The previously seen discrepancy in normalisation (see Figure 6.49) is no longer apparent. This can be interpreted based on the fact that $e^+e^- \to q\bar{q}$ events are strongly suppressed, namely by the requirements of `BDT output`.
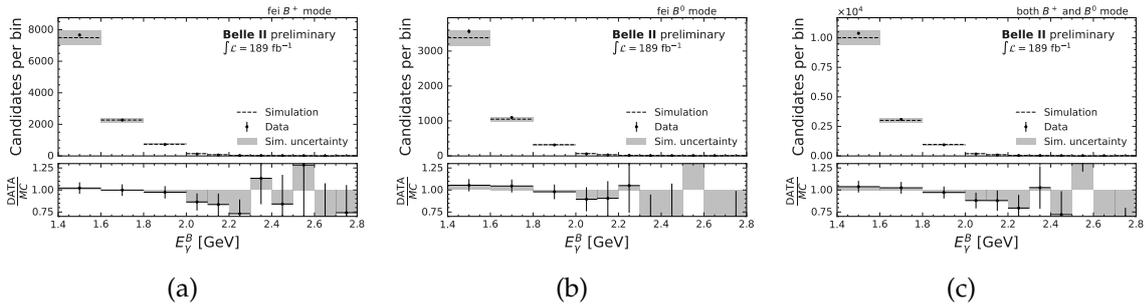

Figure 6.53:   The E_γ^B distribution of $B\bar{B}$-background enhanced samples (see Section 6.12.3). Compared to Figure 6.48c, it is clear that the $B\bar{B}$ background drops off faster with increasing E_γ^B than $e^+e^- \to q\bar{q}$. Overall, the data-simulation agreement is excellent and this is attributed to the fact that continuum events, which were accredited to causing a discrepancy in Section 6.12.2, are highly suppressed in the $B\bar{B}$-background enhanced sample. The uncertainties in data are statistical, whereas MC includes statistical and systematic uncertainties from the corrections in Table 6.16.

Following the observations in Section 6.12.2, similar modifications to M_{bc} are necessary here, too. The corrected M_{bc} distribution is shown in Figure 6.54. The results are shown for the combined `feiB`$^+$ and `feiB`^{0} sample only, although the individual FEI modes also show similar results. In particular, one can see that the agreement between data and MC is closer than what was observed in Section 6.12.2. This goes in line with the previous explanation related to the discrepancy of $e^+e^- \to q\bar{q}$ modelling which is not apparent here due to their suppression. Observing the Data/MC ratio distribution in the subpanel of Figure 6.54 it is

clear that, generally, fewer events seem to be present in the peak region and more in the tail region. Again, this is in line with the previous explanation: the variations of $\sqrt{s}$ in real data modify the $\Upsilon(4S)$ and $e^+e^- \to q\bar{q}$ production cross-sections.
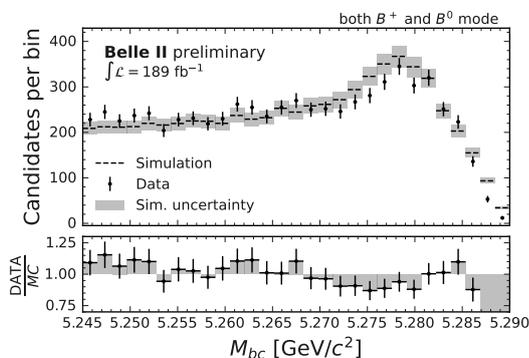
Figure 6.54: The M_{bc} distribution of $B\bar{B}$-background enhanced samples (see Section 6.12.3), where an M_{bc} correction has been applied to the simulated distribution. Only the `feiB`$^+$ and `feiB`^{0} combined sample is shown, but the individual ones show similar results. Due to a lower number of $e^+e^- \to q\bar{q}$ events, the low-M_{bc} region disagreement is less pronounced than in Figure 6.51. The uncertainties in data are statistical, whereas MC also includes systematic uncertainties from the corrections in Table 6.16.

Finally, a fit of the $B\bar{B}$-enhanced M_{bc} distributions is performed, according to the corrected M_{bc} fitting model in Table 6.17. Unlike in Figure 6.52, the number of expected $B\bar{B}$ events is no longer negligible. Therefore, this validation serves as a test of the fit bias and the background subtraction procedure. The M_{bc} fits of Belle II MC and data are shown in Figures 6.55 and 6.56, respectively. The Figures also show the extracted good tag-B meson counts as normalisations of the Crystal Ball PDF, $\mathcal{N}_{\text{CB}}$.

Despite a vastly varying number of events and shapes of the total distribution throughout different E_γ^B intervals, the fitter performs well. The extracted $\mathcal{N}_{\text{CB}}^{\text{DATA}}$ and $\mathcal{N}_{\text{CB}}^{\text{MC}}$ (corresponding to good tag-B meson yields in data and simulation, respectively) are directly compared. Correcting the simulation based on Table 6.16 with inverted $\mathcal{P}(\pi^0 \to \gamma\gamma)$ and $\mathcal{P}(\eta \to \gamma\gamma)$ correction, as reported in Equation (6.26), and calculating $\mathcal{N}_{\text{CB}}^{\text{DATA}} - \mathcal{N}_{\text{CB}}^{\text{MC}}$ is expected to yield a value consistent with zero. The resulting difference, with appropriate statistical and systematic uncertainties applied, is shown in Figure 6.57. The hypothesised result is observed, confirming the adequacy of the fitter, the validity of the background subtraction procedure and the applied corrections.

6.12.4 Validation outside of the E_γ^B signal region

In Sections 6.12.1 to 6.12.3 it was seen that the background simulation of $q\bar{q}$ and $B\bar{B}$ events, although not perfect, is described adequately by the M_{bc} fitter yielding a correct and valid estimation of good tag-B mesons in data and MC. The last validation performed for background simulation is done outside of the E_γ^B signal region. As discussed in Section 6.8.2, the $E_\gamma^B \in (1.4 - 1.8)$ GeV and $E_\gamma^B > 2.7$ GeV intervals are selected as sideband regions, due to a
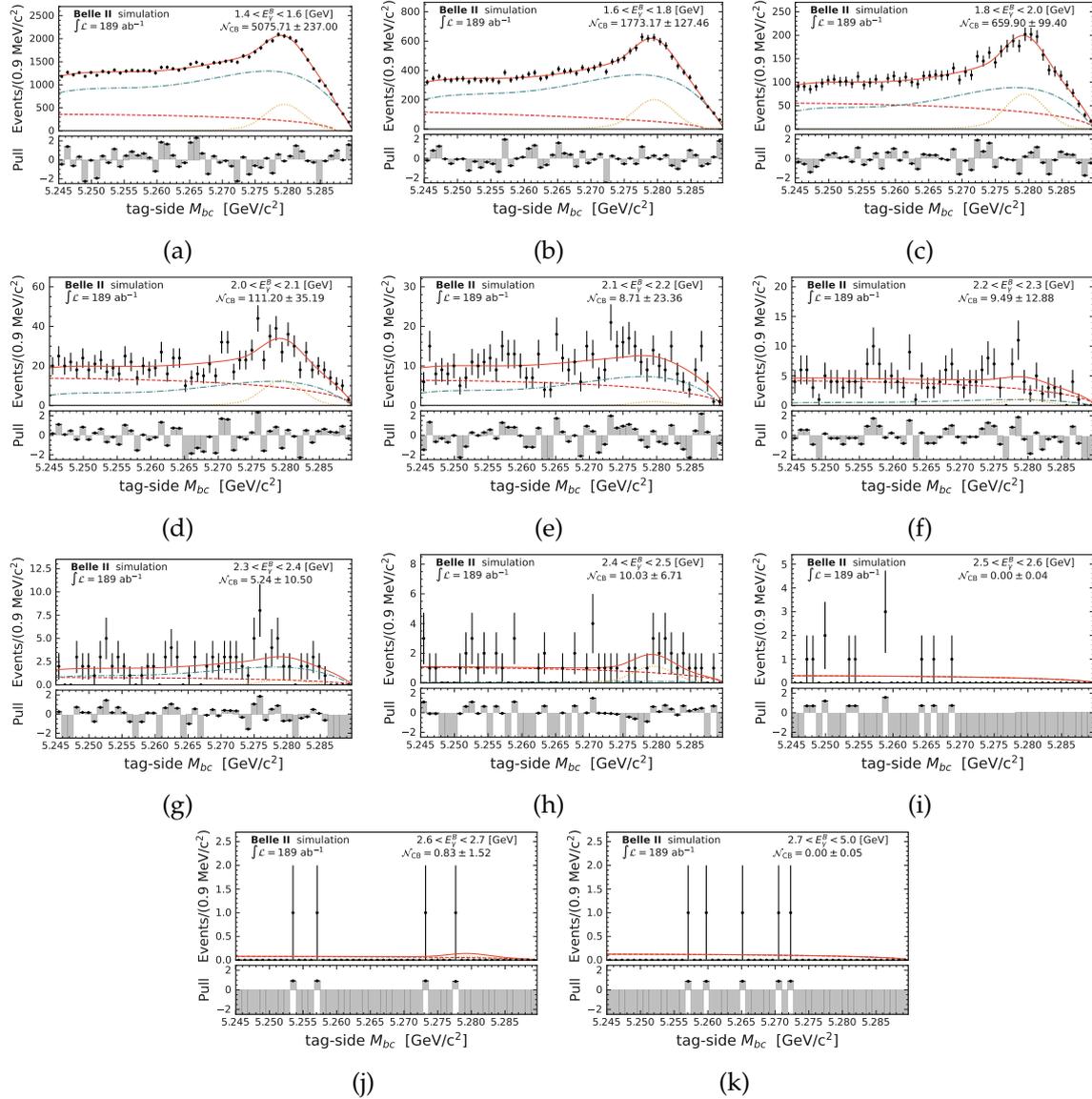
Figure 6.55: The fits of Belle II MC corresponding to 1.6 ab^{-1} with the selection that enhances non-$B \to X_s\gamma$ events, as discussed in Section 6.12.3. The fitting model from Table 6.17 is used, which is defined on the corrected M_{bc} to account for variations of $\sqrt{s}$ in Belle II data. A good description of the M_{bc} distributions can be seen throughout the E_γ^B bins.
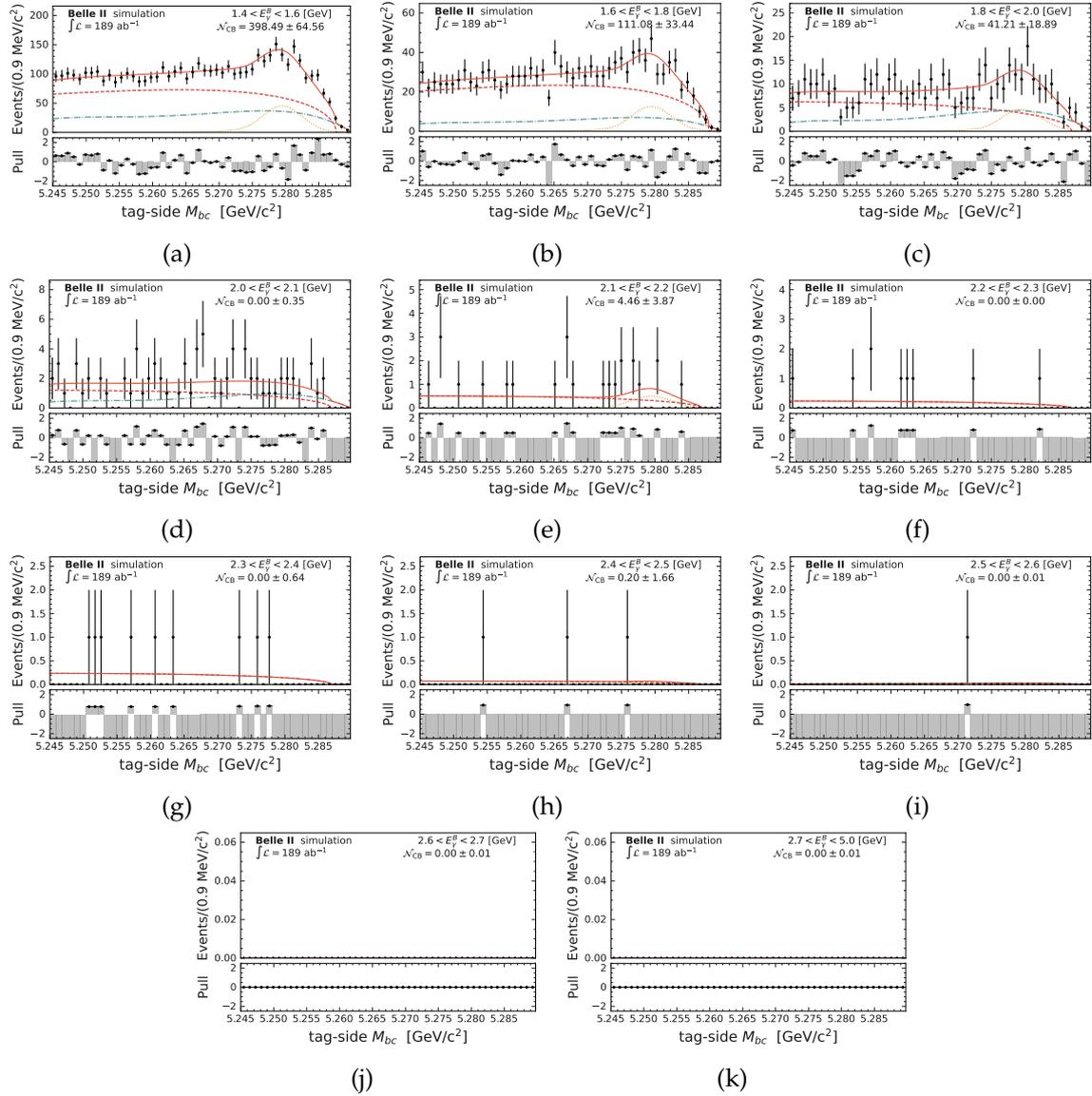
Figure 6.56: The fits of Belle II data corresponding to 189 fb^{-1} with the selection that enhances non-$B \to X_s \gamma$ events, as discussed in Section 6.12.3. The fitting model from Table 6.17 is used, which is defined on the corrected $M_{\rm bc}$ to account for variations of $\sqrt{s}$ in Belle II data. A good description of the $M_{\rm bc}$ distributions can be seen throughout the E_γ^B bins.
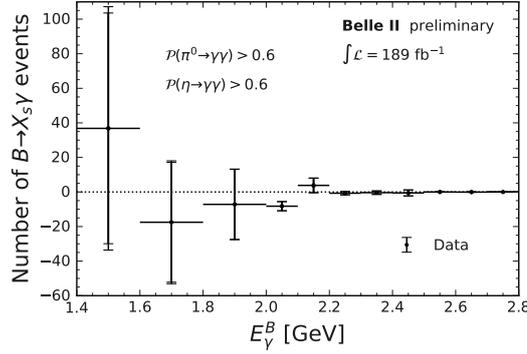
Figure 6.57: The difference between the number of good tag-B meson counts extracted from fits in Belle II MC (Figure 6.55) and that in Belle II data (Figure 6.56) in the $B\bar{B}$-background enhanced samples (see Section 6.12.3). The simulated values are corrected for luminosity and to better represent data based on studies in Section 6.11. The background subtraction procedure is further detailed in Sections 6.9 and 6.10.2. Here, an agreement with zero is expected in all E_γ^B bins. The inner (outer) error bars correspond to the statistical (total) uncertainty.

small number of $B \to X_s\gamma$ events and a low signal-to-background ratio expected there. The same arguments make the regions excellent for background validation.

The E_γ^B distribution for the three E_γ^B sideband intervals is shown in Figure 6.58. A striking difference in normalisation of nearly 20% is observed, seemingly similar to that observed in Figure 6.49. Interestingly, here, the $e^+e^- \to q\bar{q}$ component is thought to be strongly suppressed by BDT output. To better understand this discrepancy, the corrected M_{bc} distributions in each E_γ^B sideband bin are inspected. This is shown in Figure 6.59.

| (a) | (b) | (c) |
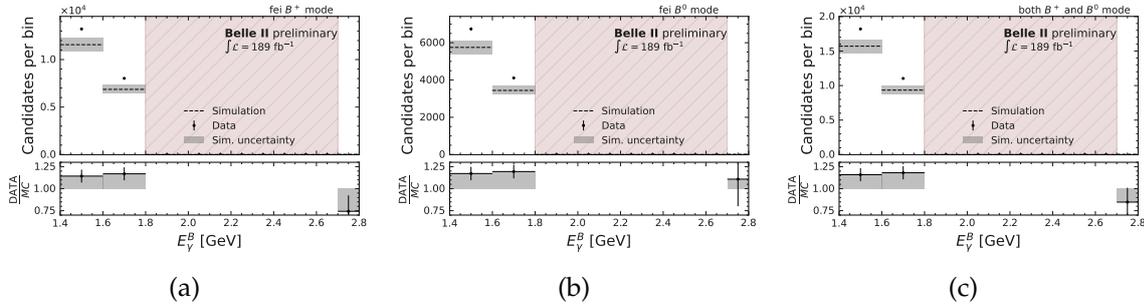
Figure 6.58: The E_γ^B distribution of the E_γ^B sideband regions (see Section 6.12.4). The low-E_γ^B region side sees a roughly 20% discrepancy. The shaded area represents the signal region which is blinded: it was not analysed during the validation step. The uncertainties in data are statistical, whereas MC includes systematic uncertainties from the corrections in Table 6.16.

The results of Figures 6.58 and 6.59 indicate that Belle II data, especially the low-M_{bc} region, has a clear excess compared to MC. While some discrepancy is expected considering the results of Figure 6.51, the larger scale of the discrepancy is confusing, given the fact that this was not observed in Figure 6.54. Although more studies on this subject are necessary

(a) (b) (c)
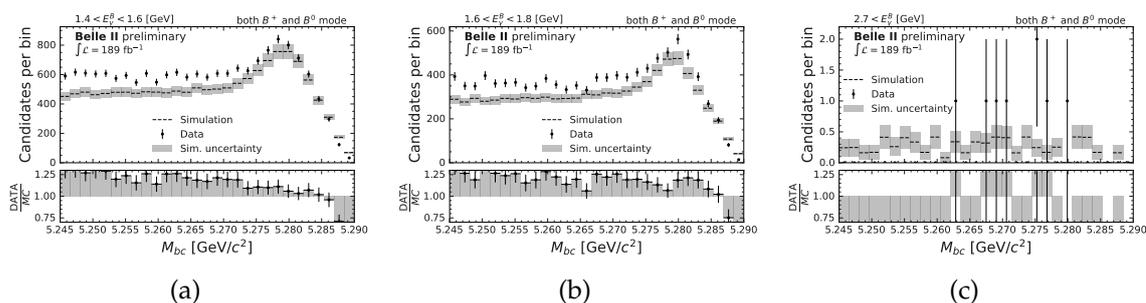
Figure 6.59: The M_{bc} distribution of the E_γ^B sideband regions (see Section 6.12.4). The Figures showcase different E_γ^B ranges, as indicated in the left corner of each Figure. Interestingly, a similar low-M_{bc} discrepancy is observed with the enhanced-continuum sample, shown in Figure 6.50. The uncertainties in data are statistical, whereas MC also includes systematic uncertainties from the corrections in Table 6.16.

to fully understand the discrepancy, it is attributed to a $B\bar{B}$-background component which is not captured in the Belle II MC. The potential origin is shortly discussed here.

In particular, consider the removal of the `zernikeMVA` selection. The resulting E_γ^B sideband distribution and the M_{bc} distribution are shown in Figure 6.60. In this case, the agreement between data and simulation in the E_γ^B-sideband spectrum appears to be near-perfect, as seen in Figure 6.60a. Indeed, even considering the M_{bc} distribution for $E_\gamma^B \in (1.4, 1.6)$ GeV in Figure 6.60b, one observes a better overall agreement. These observations strongly support the hypothesis that hadronic clusters targeted by `zernikeMVA` are differently affected in MC and data. The component cannot be common to $e^+e^- \to q\bar{q}$ events, because the effect was not seen in off-resonance data in Section 6.12.1. As a result, this component is suppressed in MC by a selection on the `zernikeMVA` observable, but this does not happen in data.

(a) (b) (c)
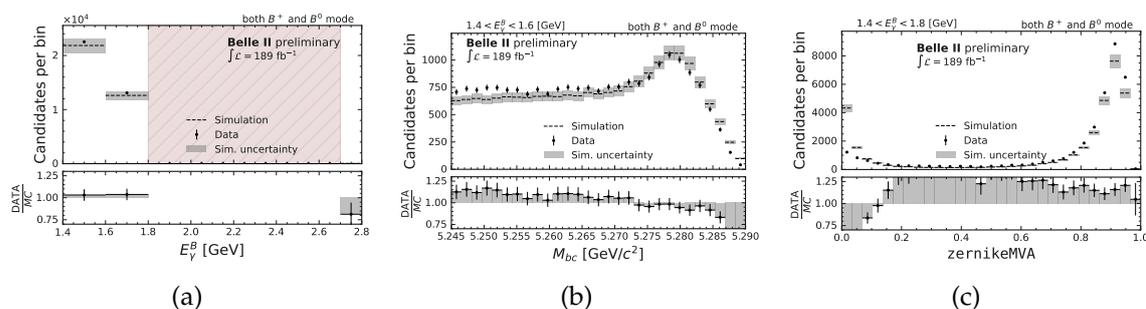
Figure 6.60: The E_γ^B distributions in the E_γ^B sideband region (a) and the M_{bc} distribution in $1.4 < E_\gamma^B < 1.6$ GeV regions (b). In both cases, the requirement on the `zernikeMVA` is removed to compare with Figures 6.58c and 6.59c, respectively. The full range of the `zernikeMVA` distribution is shown in (c). Without the `zernikeMVA`, the agreement between data and simulation appears to be improved. This indicates that a component in data is not simulated adequately. The presence of a mismodelled component in the Belle II simulation is apparent when inspecting the `zernikeMVA` distribution.

These considerations are supported by inspecting the `zernikeMVA` distribution in the $E_\gamma^B \in (1.4, 1.8)$ GeV region, as seen in Figure 6.60c. While MC contains a sharp peak near 0, this is not evident in Belle II data, which has an excess at larger values of `zernikeMVA`. Hence, the background suppression efficiency is not well-represented in MC. These results are also conflated with the differences in the M_{bc} endpoint, making an exact evaluation of the effect difficult at this stage. Independent studies of the `zernikeMVA` distributions, performed similarly to the photon detection efficiency study described in Section 6.11.3, did not indicate the presence of such a peak. This may imply that this type of selection is specific for photon candidates from $B\bar{B}$ events misidentified as high-energy photons. Therefore, it was concluded that additional studies of `zernikeMVA` in the context of radiative and inclusive analyses will be necessary for future versions of this analysis.

While these observations are alarming, so far all the results have shown that the M_{bc} fitter and background subtraction procedure are robust against the continuum M_{bc} distribution shape differences, as seen in Sections 6.12.2 and 6.12.3. Therefore, this analysis does not replace or remove the `zernikeMVA` requirement.

The validity of such an assumption is tested by performing the M_{bc} fit of the full sample of data. The results from the signal region, $E_\gamma^B \in (1.8, 2.7)$ remain blinded at this stage. The individual M_{bc} fits of the E_γ^B sideband regions are shown for data in Figure 6.61, and for MC in Figure 6.62.

(a)   (b)   (c)
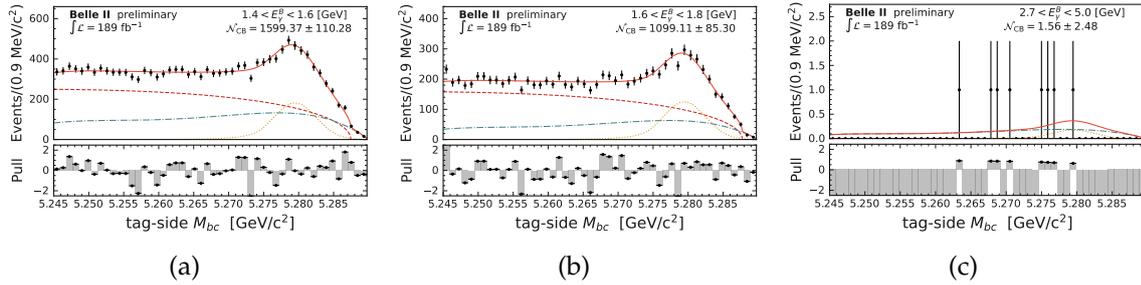
Figure 6.61: The M_{bc} fits of Belle II data corresponding to 189 fb^{-1} of data in E_γ^B sideband regions, as discussed in Section 6.12.4. The fitting model from Table 6.17 is used, which is defined on the corrected M_{bc} to account for variations of $\sqrt{s}$ in Belle II data. A good description of the M_{bc} distributions can be seen throughout the E_γ^B bins.

The summarised results of the good tag-B meson yields estimated in the M_{bc} fits are shown in Figure 6.63. In the high-E_γ^B sideband, no peaking tag-B mesons are observed in data or MC, which is exactly consistent with the expectations. In the low-E_γ^B region, a large number of events is observed. The data points are compatible with the background expectation, although the estimates in both intervals are higher than the expected background. Subtracting the background expectation from the good tag-B meson count in data results in Figure 6.64a. A similar observation follows: although both values are (nearly) compatible with zero, the central values tend to positive values.

The total number of events in the low-E_γ^B sideband is 2698 ± 139 (expected to be predominantly background, see Section 6.8.2). The background expectation from MC in the same region is 2483 ± 130. The ratio of these values evaluates 1.087 ± 0.080. As this value is not
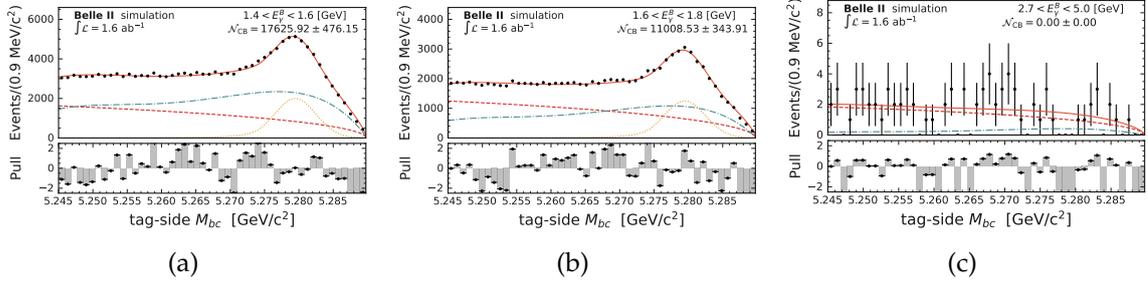
Figure 6.62: The M_{bc} fits of Belle II MC corresponding to 1.6 ab^{-1} of data in the E_γ^B sideband regions, as discussed in Section 6.12.4. Any $B \to X_s\gamma$ events in these data sets are removed. The fitting model from Table 6.17 is used, which is defined on the 'corrected'-M_{bc} to account for variations of $\sqrt{s}$ in Belle II data. A good description of the M_{bc} distributions can be seen throughout the E_γ^B bins.
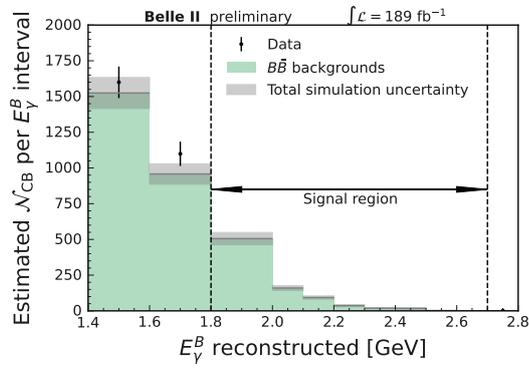
Figure 6.63: The results of fitting the M_{bc} on the sideband region in data (see Section 6.12.4). The values corresponding to data fits are estimated through an M_{bc} fit shown in Figure 6.61. The remaining $B\bar{B}$ background expectations are estimated through M_{bc} fits in Figure 6.62 and Figure 6.71. The signal region in this Figure is blinded. The extracted results from the data fit and simulated background expectations are compatible within their full uncertainty, but both points are higher than the background estimation.
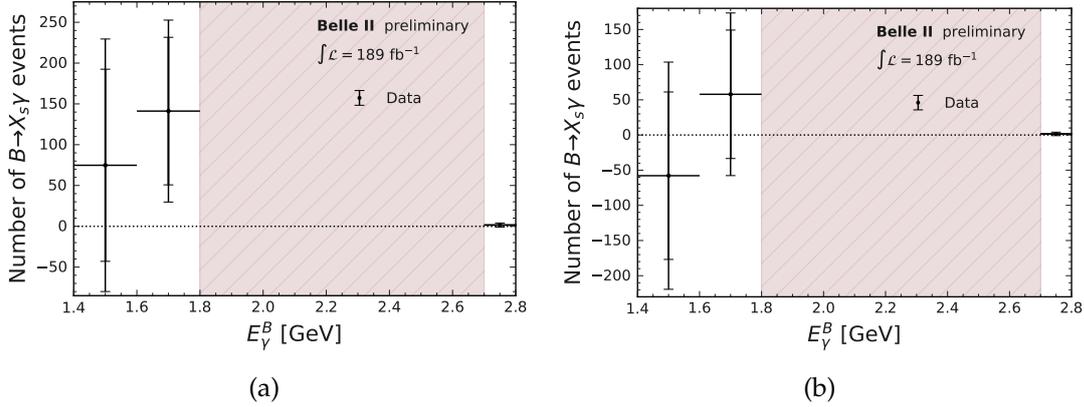
Figure 6.64: The results of subtraction of the remaining $B\bar{B}$ background after the fit of Belle II data. The results with no correction factor applied are seen in (a). An 8.7% scaling factor for the simulated background values is included in (b). The signal region, denoted by the shaded area, is blinded at this stage. In this analysis, it is chosen to scale the background (i.e. the scenario shown in (b)), as that region is expected to contain no $B \rightarrow X_s\gamma$ events. This Figure only includes systematic uncertainties related to corrections described in Section 6.11.

compatible with unity within 1 standard deviation, a background scaling of 8.7% is adopted. The 100% of the scaling is also adopted as a systematic uncertainty later. The scaled and background-subtracted data fit yields are shown in Figure 6.64b. By construction, they are fully compatible with zero.

6.13 Signal modelling and efficiency studies

Section 6.12 shows that the background distributions are adequately represented in simulation, which proves that the analysis setup on MC is valid for data. It was also seen that the M_{bc} fitter extracts values consistent with zero, where no $B \rightarrow X_s\gamma$ signal was expected. The analysis strategy in Sections 6.6, 6.8 and 6.9 does not strongly depend on the signal model. No strong assumptions are made about the signal shape at any point in the analysis so far. Therefore, following the fitting and background subtraction procedures, the number of $B \rightarrow X_s\gamma$ events as a function of E_γ^B in the analysed Belle II data sample can be evaluated.

In order to transform the measured numbers of $B \rightarrow X_s\gamma$ events to partial branching fractions of $B \rightarrow X_s\gamma$ decays, efficiency corrections and unfolding is necessary. In this Section, the expected signal efficiency and E_γ^B resolution of $B \rightarrow X_s\gamma$ events is investigated. The hybrid-signal model is then used to derive unfolding correction factors.

6.13.1 Efficiency of $B \rightarrow X_s\gamma$ decays

The $B \rightarrow X_s\gamma$ selection efficiency is evaluated using the Belle II MC and corrected based on the studies that have been discussed in Section 6.11. The signal efficiency is assumed to be factorisable:

$$\varepsilon_{B \rightarrow X_s\gamma} = \varepsilon_{\mathrm{FEI}} \cdot \varepsilon_{\mathrm{selection}}, \tag{6.27}$$

where ε_{FEI} is the FEI tagging efficiency and $\varepsilon_{\text{selection}}$ is the selection efficiency related to requirements shown in Section 6.6.2. The factorisation assumption is a valid one as ε_{FEI} is related to the reconstruction of the tag-B meson, whereas $\varepsilon_{\text{selection}}$ is fully a signal-B meson quantity.

The FEI tagging efficiency is evaluated as:

$$\varepsilon_{\text{FEI}} = \frac{N(B \rightarrow X_s\gamma)_{\text{good tags}}}{N(B \rightarrow X_s\gamma)_{\text{untagged}}} \tag{6.28}$$

The numerator, $N(B \rightarrow X_s\gamma)_{\text{good tags}}$, is equal to the number of $B \rightarrow X_s\gamma$ events associated with good tag-B mesons after running FEI. It is evaluated using the good-tag definition in Section 6.7.3. The denominator, $N(B \rightarrow X_s\gamma)_{\text{untagged}}$, is equal to the number of $B \rightarrow X_s\gamma$ events on an equivalent sample, where FEI is not run. In both cases, the hybrid-signal model is used.

The evaluated tagging efficiency is shown in Figure 6.65a. The efficiency is evaluated as a function of $\tilde{E}_\gamma^B$, which denotes that the *true* photon energy is used, as opposed to the reconstructed value. This is done, as the untagged inclusive sample cannot have a meaningful comparison in terms of reconstructed E_γ^B. It can be seen that the efficiency increases with $\tilde{E}_\gamma^B$, but the overall increase is around 10% throughout the considered range. As a direct connection between $\tilde{E}_\gamma^B$ and E_γ^B is difficult to evaluate, the average efficiency value is chosen as the tagging efficiency:

$$\varepsilon_{\text{FEI}} = 0.006659 \pm 0.000006, \tag{6.29}$$

where the uncertainty is only statistical. The correction and the systematic uncertainty related to FEI calibration is evaluated in Section 6.14.3. Conversely, the signal modelling uncertainty is expected to be small because any deviations would be suppressed in the ratio in Equation (6.28).

The $B \rightarrow X_s\gamma$ selection efficiency is evaluated using three different signal models as:

$$\varepsilon_{\text{selection}} = \frac{N(B \rightarrow X_s\gamma)_{\text{after selection}}}{N(B \rightarrow X_s\gamma)_{\text{before selection}}}, \tag{6.30}$$

here $N(B \rightarrow X_s\gamma)_{\text{after(before) selection}}$ is the count of $B \rightarrow X_s\gamma$ events in the FEI tagged sample with(without) the background suppression selections, given in Table 6.11. This is evaluated on three models: the Kagan-Neubert model, the Belle II generic MC signal model and the hybrid-signal model. The results are shown in Figure 6.65b. All three models show compatible results. The $\varepsilon_{\text{selection}}$ grows approximately linearly from 30% at 1.4 GeV to 60% at 2.6 GeV and begins to drop. The values of the hybrid-signal model are chosen as central values of efficiency. Finally, the results of Figure 6.65 are combined to evaluated the total simulated efficiency based on Equation (6.27).

6.13.2 Validation of $B \rightarrow X_s\gamma$ efficiency

The results of Section 6.13.1 are based on Belle II simulation only, therefore, have to be validated on data. Some selections, such as π^0 and η suppression tools, photon detection efficiency and FEI are validated in external and independent studies. However, the selection on

(a)   (b)
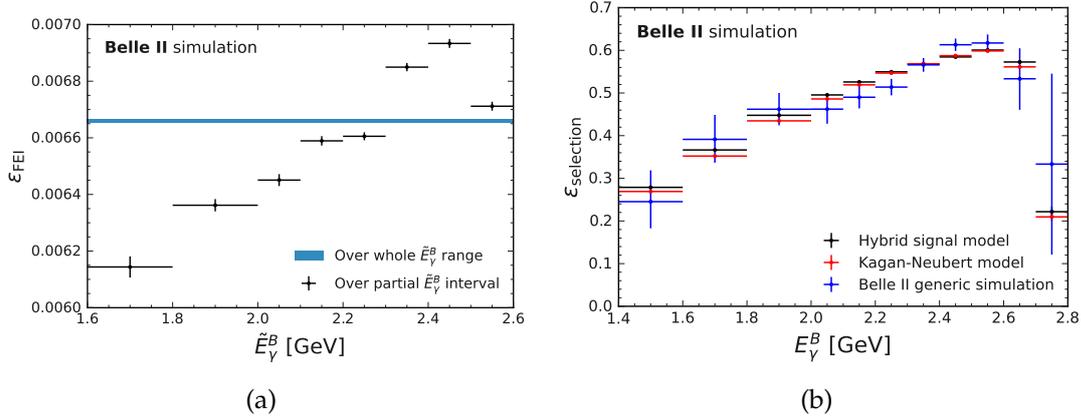
Figure 6.65: The efficiency evaluation of $B \to X_s\gamma$ events in simulated samples based on the two factorised components in Equation (6.27). ε_{FEI}, shown in (a), is seen to vary lightly, no more than 10% accross the $\tilde{E}_\gamma^B$ range. $\varepsilon_{\text{selection}}$, shown in (b) for three different models, grows with E_γ^B approximately linearly and starts to drop at $E_\gamma^B \approx 2.6$ GeV. The three models show consistent results, strengthening the argument of a signal model-independent analysis.

`BDT output` and `zernikeMVA` do not have such dedicated studies (and with the latter some hints of performance differences were discussed). To perform the validation for these quantities, yet maintain the analysis blinded, ratio-based efficiency tests are used on data:

$$\varepsilon' = \frac{N_{\text{with selection}}(M_{\text{bc}} > 5.27)}{N_{\text{with looser selection}}(M_{\text{bc}} > 5.27)},\tag{6.31}$$

where $N_{\text{with selection}}$ is the number of events in Belle II data, given some selection, and $N_{\text{with looser selection}}$ is the number of events recomputed with a looser selection. In both cases, this is evaluated with an $M_{\text{bc}} > 5.27$ GeV/c^2 requirement to ensure that the focus is mainly on good tag-B mesons. Finally, in order to maximise the number of $B \to X_s\gamma$ events, the $2.5 < E_\gamma^B < 2.6$ GeV interval is used, as it contains primarily $B \to X_s\gamma$ events and very low background contamination (see Table 6.14)

Four selection configurations are tested:

- regular: where `zernikeMVA > 0.6` or `BDT output > 0.8` are maintained at their optimal slection values.

- looser: where `zernikeMVA > 0.4` or `BDT output > 0.6` are loosened to include more background.

- tighter: where `zernikeMVA > 0.8` or `BDT output > 0.9` are tightened to suppress more background.

- none: no selection on `zernikeMVA` or `BDT output`.

These selection configurations are then used to compute efficiencies and the binomial uncertainties based on Equation (6.31). The results are shown in Table 6.18.

Table 6.18: Resulting efficiencies in simulation, ε_{MC}, and data, $\varepsilon_{\text{DATA}}$, after applying selection variations based on Equation (6.31). The selection configurations are defined in Section 6.13.2. The uncertainties are calculated as Clopper-Pearson intervals for a binomial ratio.

Configuration	zernikeMVA		BDT output	
	ε_{MC}	$\varepsilon_{\text{DATA}}$	ε_{MC}	$\varepsilon_{\text{DATA}}$
tighter/regular	$0.936\pm^{0.009}_{0.011}$	$0.867\pm^{0.034}_{0.042}$	$0.659\pm^{0.018}_{0.019}$	$0.657\pm^{0.049}_{0.053}$
regular/looser	$0.965\pm^{0.007}_{0.008}$	$0.963\pm^{0.017}_{0.028}$	$0.667\pm^{0.015}_{0.015}$	$0.691\pm^{0.039}_{0.042}$
tighter/looser	$0.903\pm^{0.011}_{0.012}$	$0.835\pm^{0.037}_{0.044}$	$0.440\pm^{0.016}_{0.016}$	$0.454\pm^{0.044}_{0.043}$
regular/none	$0.895\pm^{0.011}_{0.012}$	$0.882\pm^{0.030}_{0.037}$	$0.182\pm^{0.006}_{0.006}$	$0.136\pm^{0.014}_{0.013}$

Overall, the patterns of variations in efficiencies in simulation, ε_{MC}, and data, $\varepsilon_{\text{DATA}}$, show similar behaviour. The central values are generally compatible within $1\sigma - 2\sigma$. The largest difference between ε_{SIM} and $\varepsilon_{\text{DATA}}$ for zernikeMVA is observed between tighter and regular configurations, which is consistent with observations of Figure 6.60c. Tightening the selection suppresses additional background in data, but not in MC, where such component is not present, causing different ϵ' behaviour. Based on these observations, a variation $\Delta\varepsilon \approx 0.069 \pm 0.039$ is observed (the asymmetric binomial uncertainty has been symmetrised here). For a conservative approach, this analysis, therefore, adopts a 10% efficiency uncertainty based on zernikeMVA modelling. The BDT output variations are smaller, with the largest variation observed in the regular/looser configuration. The variation is evaluated at $\Delta\varepsilon \approx 0.024 \pm 0.043$. For a conservative approach, a 3% efficiency uncertainty is adopted, associated with the BDT output modelling.

The results of this Section, Section 6.13.1 and corrections from Section 6.11 are combined in Equation (6.30) to calculate the selection efficiency of $B \rightarrow X_s \gamma$ as a function of E^B_γ. The results are visualised in Figure 6.66. Using the factorised relation in Equation (6.27) to combine the result with the average FEI tagging efficiency (given in Equation (6.29)), these values form the final efficiency.

6.13.3 Resolution studies of $B \rightarrow X_s \gamma$ events

The resolution of E^B_γ is related to both photon detection resolution and tag-side B meson reconstruction. Therefore, it can depend on the choice of a good tag-B meson definition.

The resolution is modelled as the width of the distribution related to

$$\tilde{E}^B_\gamma - E^B_\gamma \tag{6.32}$$

where $\tilde{E}^B_\gamma$ is the true energy of a photon in the signal-B meson rest frame, and E^B_γ is the measured photon energy in the signal-B meson rest frame. To extract the width of the distribution, a double-sided Crystal Ball function is fitted. The double-sided Crystal Ball function follows the same definition as previously discussed in Appendix H.1, however, it includes two additional parameters α_2 and n_2, which introduce a polynomial behaviour to both sides of the central Gaussian. The resolution is assumed to be represented by the Gaussian width parameter σ.
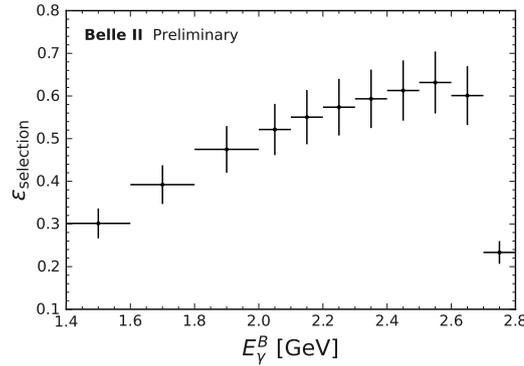
Figure 6.66: $B \to X_s\gamma$ signal selection efficiency, evaluated using Equation (6.30). The central values represent the efficiency of the hybrid-signal mode and include corrections and a full systematic uncertainty, based on independent studies in Section 6.11 and blinded data efficiency studies in Section 6.13.2. Multiplied by the FEI tagging efficiency in Equation (6.29), this yields the full efficiency of this analysis for $B \to X_s\gamma$.

The fitting strategy of the hybrid-signal model sample with a full selection used in this analysis is presented in Appendix K. Each fit is performed in $\tilde{E}_\gamma^B$ intervals and estimates the parameter σ and its uncertainty. The summary of the results is given in Figure 6.67. The fits are done with two choices of tag-B mesons: firstly, with the good tag-B mesons as defined in Section 6.7.3, and secondly, for comparison, with tag-B mesons that have been reconstructed perfectly. The photon energy resolution grows with E_γ^B from 25 MeV to 40 MeV but the ratio $\sigma/\tilde{E}_\gamma^B$ stays approximately constant. The use of good tag-B mesons degrades the efficiency by approximately 10%. As the evaluated resolution is $\mathcal{O}(10$ MeV$)$, this (retroactively) justifies the selection of 100 MeV wide bins in Section 6.8.2.
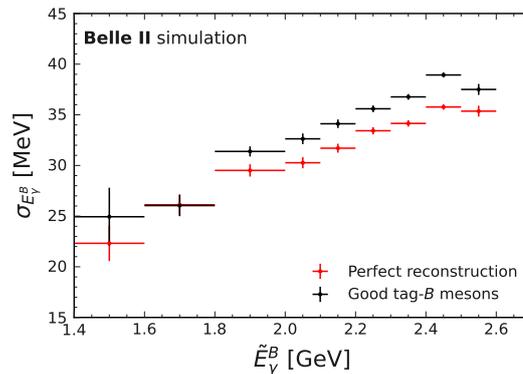
Figure 6.67: The resolution of $\tilde{E}_\gamma^B$, as estimated from the fits in Figure K.1 and Figure K.2. Comparing the resolution of E_γ^B using the good tag-B meson definition, with that using only perfectly reconstructed tag-B mesons a $\sim 10\%$ difference is seen. The resolution grows approximately linearly with $\tilde{E}_\gamma^B$.

6.13.4 Tag-side and signal-side correlation study

The purpose of this analysis is an unbiased inclusive measurement of the $B \rightarrow X_s \gamma$ decays. Although careful validation for any potential biases to the photon energy spectrum is performed at all stages, the correlation of the tag-B and signal-B has so far not been investigated in depth. This is tested by counting $B \rightarrow X_s \gamma$ events in an untagged sample and that with FEI tagging applied, similar to Section 6.13.2. The number of daughter particles that the X_s system hadronises to is then evaluated in the inclusive $B \rightarrow X_s \gamma$ signal model. The results are shown in Table 6.19.

Table 6.19: Fraction of events with a given X_s multiplicity in FEI tagged and untagged sample. Although FEI tagging slightly prefers lower multiplicities, the overall inclusivity is retained with no strong biases.

X_s multiplicity	Fraction of sample	
	FEI tagged	Untagged
2	0.621	0.594
3	0.236	0.242
4	0.090	0.099
5 or more	0.053	0.065

It can be seen that although $B \rightarrow X_s \gamma$ decay channels where X_s hadronises to two particles are slightly preferred over other cases, the overall composition of the sample is not strongly affected. Indeed, such behaviour seems expected: the more neutral and charged final-state particles that are produced in the detector, the more (incorrect) combinations to reconstruct a tag-B meson become available for the FEI chain of classifiers.

Although this study is performed in simulated samples only, the results generalise well to Belle II data: the number-of-track dependence will affect the classifiers in the same way, as it only relates to the combination of the decay products to good tag-B mesons. The other differences that are related to tag-side efficiency mismodelling would be captured by the FEI correction factors in Equation (6.21).

6.13.5 Modelling of $B \rightarrow X_d \gamma$ component

As part of the inclusive analysis, a separation between X_s and X_d is challenging. Up until now, the X_d component has been neglected and not included in the discussion. However, when extracting the photon energy spectrum in Belle II data, the presence of the X_d component is unavoidable. Without additional consideration, the measured result could only be interpreted as $B \rightarrow X_{s/d} \gamma$.

The measured branching fraction of $B \rightarrow X_d \gamma$, as seen in Equation (3.1) and Equation (2.22), is at least an order of magnitude smaller than $B \rightarrow X_s \gamma$, although the uncertainties for the measurement are large. Based on Equation (2.12), neglecting corrections and additional terms in the Lagrangian, the $B \rightarrow X_d \gamma$ branching fraction is suppressed by [16]:

$$\left| \frac{V_{td}}{V_{ts}} \right|^2 \approx 0.042. \tag{6.33}$$

Two assumptions are made in this analysis:

- $B \to X_d\gamma$ photon energy spectrum shape is the same as the $B \to X_s\gamma$ shape,

- $B \to X_d\gamma$ event selection efficiency is the same as $B \to X_s\gamma$ event selection efficiency.

These assumptions are a valid approximation because the same underlying processes (electroweak radiative transitions) govern the decay. While the endpoint region for the spectra $E_\gamma^B \gtrsim 2.6$ would be different (X_d is dominated by $\rho(770)$ which is wider than $K^*(892)$), at the experimental precision anticipated, the difference is not expected to be significant.

Therefore, the measured $B \to X_{s/d}\gamma$ branching fractions will be lowered by an amount equivalent to Equation (6.33):

$$
\begin{aligned}
N_{B \to X_s\gamma} &= N_{B \to X_{s/d}\gamma} - N_{B \to X_d\gamma} \\
&= \frac{1}{1.042} N_{B \to X_{s/d}\gamma}.
\end{aligned}
\tag{6.34}
$$

The full value of the correction, $(1 - 1/1.042) \cdot N_{B \to X_{s/d}\gamma}$, is assigned as a systematic uncertainty related to the modelling of $B \to X_d\gamma$.

6.13.6   Unfolding of the measured photon energy spectrum

The measured E_γ^B spectrum is smeared due to resolution effects. This can be seen in Figure 6.68a, where the true and measured photon energies are compared. In both cases, the hybrid-signal model is used. The overall peak of the spectrum is shifted towards lower-E_γ^B after a measurement. Therefore, the measured result of the Belle II data has to be unfolded. Unfolding was already introduced in Section 5.3. The unfolding strategy of the E_γ^B spectrum to the true energy $\tilde{E}_\gamma^B$ is presented in this Section.

Firstly, a response matrix (see Equation (5.19)) is calculated. It shows the fraction of events that are generated in a given $\tilde{E}_\gamma^B$ interval, but are measured in a given E_γ^B interval. The response matrix is shown in Figure 6.68b.

As was discussed in Section 5.3, the bin-by-bin correction method is used for unfolding. It was chosen after testing several unfolding techniques, including those that involve a regularisation of the unfolded result. An example comparison of the singular value decomposition, matrix inversion and bin-by-bin unfolding techniques is shown in Figure 6.69. The singular value decomposition method also includes a regularisation strength parameter, $k = 7$. The Figure shows the expected $B \to X_s\gamma$ photon energy spectrum (based on the hybrid-signal model) before and after measuring. The unfolded points follow the true distribution perfectly in this case, as the hybrid-signal model is also what is used for the calculation of the response matrix.

The statistical and total uncertainties are evaluated as the average uncertainty from pseudodata fits in Section 6.10.2, corrected for background and signal modelling based on discussions in Sections 6.11 and 6.13. They are then propagated through the full unfolding procedure. For bin-by-bin unfolding the propagation is governed by Equations (5.20) and (5.21). The resulting uncertainties due to the matrix inversion method are larger by more than a factor of two compared to the bin-by-bin unfolding method in many photon energy intervals.
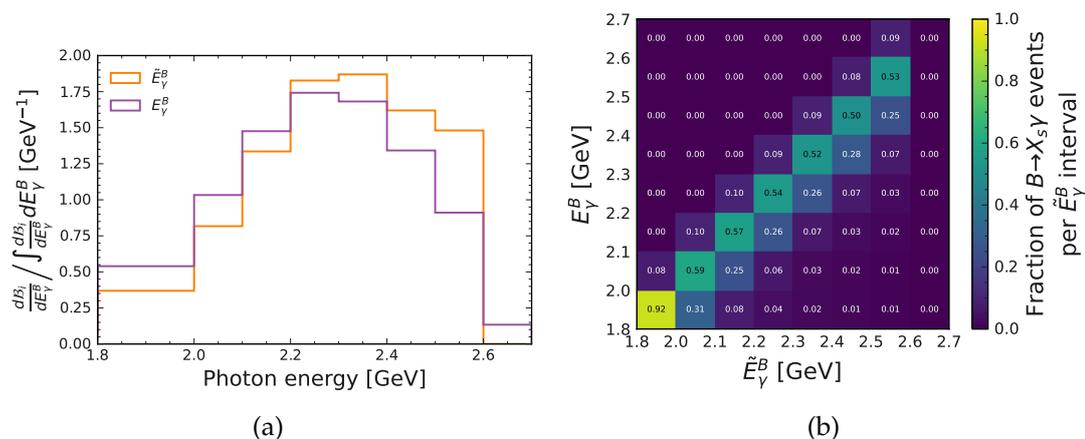
Figure 6.68: The comparison of true photon energy ($\tilde{E}_\gamma^B$), and the measured photon energy (E_γ^B) is seen in (a). The spectrum is slightly shifted to lower energies after the measurement. A corresponding response matrix is built for these distributions and shown in (b). The largest $B \to X_s\gamma$ event fractions reside on the diagonal, meaning that resolution effects are not larger than the photon energy interval size.

At the same time, this increase in uncertainties is reduced by introducing a weak regularisation requirement using the singular value decomposition method. As the analysis results are statistically limited, the observed correlation of the expected number of events in different E_γ^B intervals is small. Therefore, it is concluded that for this analysis bin-by-bin unfolding method is sufficient, as it does not inflate the uncertainties and does not introduce additional correlations between bins via regularisation.

6.14 Systematic uncertainty overview

Most systematic uncertainties or the grounds for introducing them, have already been presented in this thesis. They mainly arise from selection, background modelling and efficiency corrections. Additional subleading uncertainties from unfolding and $B \to X_d\gamma$ component subtraction are included. All of them remain subdominant compared to the much larger statistical component, except for the low-E_γ^B region. Many systematic uncertainties are set to their conservative estimates, which highlights that the analysis can be further optimised for future versions, as the Belle II data set increases. In this Section, the already presented information about systematic uncertainties is condensed and finalised.

6.14.1 Background modelling uncertainties

Background modelling encompasses all uncertainties that are related to the background evaluation in the Belle II MC. That involves two uncertainties that have been already discussed:

- background suppression efficiency differences in data and MC;

- background normalisation differences in data and MC.
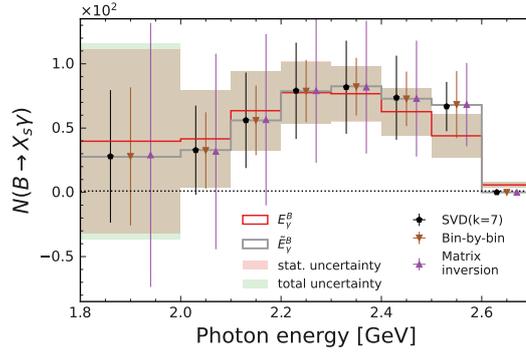
Figure 6.69: Comparison of the singular value decomposition, bin-by-bin correction and matrix inversion method unfolding strategies for the photon energy spectrum. The solid lines represent the true and measured expected photon energy spectra, based on the hybrid-signal model. As this model is also used to build the response matrix (see Figure 6.68b), the data points, corresponding to different unfolding methods line up exactly with the true result. The shaded area represents the expected measurement uncertainties, based on the pseudo-data study in Section 6.10.2. The systematic uncertainty involves corrections for background and signal simulation discussed in Sections 6.11 and 6.13.

Their values are provided in Table 6.20 and the strategy to evaluate them is explained further in this Subsection.

Table 6.20: The remaining $B\bar{B}$ background estimates, their statistical and systematic uncertainties. They are calculated in MC, based on the strategy laid out in Section 6.9, and the values shown here are scaled to those expected for 189 fb^{-1}. The corrections and their uncertainties related to Section 6.11 are applied. The uncertainty sources are discussed in detail in Section 6.14.1. The signal region is separated by horizontal lines.

E_γ^B interval [GeV]	$\mathcal{N}_{\mathrm{CB}}^{non-B\to X_s\gamma}$ (scaled to 189 fb^{-1})	Systematic uncertainties				
		FEI calibration	$\pi^0 \to \gamma\,\gamma$ and $\eta \to \gamma\,\gamma$ suppression	Photon detection efficiency	Background branching fraction modelling	Normalisation
1.4 − 1.6	1657.2 ± 44.8	±57.2	±76.6	±38.5	±35.9	±132.6
1.6 − 1.8	1041.3 ± 32.5	±35.9	±47.0	±22.8	±31.7	±83.3
1.8 − 2.0	549.8 ± 23.8	±19.0	±24.0	±11.7	±23.0	±44.0
2.0 − 2.1	173.9 ± 12.1	±6.0	±7.6	±3.7	±8.3	±13.9
2.1 − 2.2	101.6 ± 9.4	±3.5	±4.5	±2.1	±4.9	±8.1
2.2 − 2.3	39.8 ± 6.9	±1.4	±1.8	±0.8	±1.5	±3.2
2.3 − 2.4	20.4 ± 5.7	±0.7	±0.9	±0.4	±0.5	±1.6
2.4 − 2.5	18.6 ± 5.6	±0.6	±0.8	±0.4	±0.6	±1.5
2.5 − 2.6	1.7 ± 2.9	±0.1	±0.1	±0.0	±0.1	±0.1
2.6 − 2.7	0.0 ± 1.3	±0.0	±0.0	±0.0	±0.0	±0.0
> 2.7	0.0 ± 0.0	±0.0	±0.0	±0.0	±0.0	±0.0

Uncertainties due to background suppression modelling

The final result is affected by background modelling when background expectations in MC are subtracted from the M_{bc} fit of the Belle II data. The available modelling corrections based on results from independent studies are summarised in Section 6.11. The background yields

attain uncertainties from FEI calibration factors, $\mathcal{P}(\pi^0 \rightarrow \gamma\gamma)$ and $\mathcal{P}(\eta \rightarrow \gamma\gamma)$ modelling, photon detection efficiency and background modelling.

All the different uncertainties are considered as correlated in different E_γ^B intervals. The background mode branching fraction uncertainties follow the correlations depicted in Figure 6.47. The uncertainties related to other corrections are considered fully correlated across different E_γ^B intervals.

Uncertainties due to background normalisation

Section 6.12.4 concludes that the average number of good tag-B mesons in $E_\gamma^B \in (1.4, 1.8)$ GeV in data and MC differs by 8.7%. This difference is illustrated in Figure 6.64. It was discussed that the difference of 8.7% may be partially correlated to the differences in the M_{bc} endpoint (discussed broadly in Section 6.12.3). It could also partially be related to zernikeMVA discrepancy as shown in Figure 6.60c. As these effects are difficult to disentangle without additional independent studies, the full 8.7% correction is assigned as a systematic uncertainty. Therefore the background expectations are varied by ±8.7%, and the full extent of the variation is assigned as a systematic uncertainty to the measured $B \rightarrow X_s\gamma$ yields. The uncertainties are not considered correlated to account for the fact that the background shape can vary from that of simulation as a function of E_γ^B.

6.14.2 M_{bc} fitting model uncertainties

The M_{bc} fitting model uncertainties are related to the choice of the specific PDFs and the initialised parameter values of the fit model. They apply to the evaluated $\mathcal{N}_{\text{CB}}^{\text{DATA}}$ in the data fits. As a result, they are evaluated directly on data and are calculated as a last step before the full unblinding of the central values of the M_{bc} fit of data. Two uncertainties are evaluated:

- uncertainties due to the M_{bc} endpoint variation;

- uncertainties due to the fixed parameters in the M_{bc} fit.

Their values are provided in Table 6.21 and the strategy to evaluate them is explained further in this Subsection.

Uncertainties due to M_{bc} endpoint variations

As discussed in Section 6.12.3, the fit model is modified to account for $\sqrt{s}$ fluctuations in Belle II data manifesting as a shift in M_{bc}. To account for an imperfect correction to MC, a systematic uncertainty is assigned. The fit of data is performed with and without the M_{bc} correction introduced, with the two fit models in Tables 6.15 and 6.17, respectively. The full variation $\Delta\mathcal{N}_{\text{CB}}$ between the estimated $\mathcal{N}_{\text{CB}}$ in the two scenarios is assigned as a systematic uncertainty. The uncertainties are considered uncorrelated to account for possible changes in the model variations as a function of E_γ^B. The evaluation of these uncertainties is performed in a blinded way, i.e., only the $\Delta\mathcal{N}_{\text{CB}}$ is evaluated without the absolute counts. This also serves as a test ensuring that the fit converges on the full data set.

Table 6.21:   The uncertainties of $\mathcal{N}_{\mathrm{CB}}^{\mathrm{DATA}}$, relating to the M_{bc} fit model used in this analysis. They are evaluated directly on data, without unblinding the central values of evaluated $\mathcal{N}_{\mathrm{CB}}$. The uncertainty sources are discussed in detail in Section 6.14.2. The signal region is highlighted by the horizontal lines.

E_γ^B interval [GeV]	Fixed parameter uncertainty	M_{bc} endpoint variation uncertainty
1.4 − 1.6	±10.9	±97.2
1.6 − 1.8	±35.9	±71.3
1.8 − 2.0	±26.6	±56.9
2.0 − 2.1	±20.0	±3.6
2.1 − 2.2	±12.9	±5.9
2.2 − 2.3	±6.3	±3.1
2.3 − 2.4	±4.1	±4.8
2.4 − 2.5	±3.5	±0.8
2.5 − 2.6	±0.6	±1.0
2.6 − 2.7	±0.6	±1.9
> 2.7	±0.1	±0.1

Uncertainties due to fixed parameters in the M_{bc} fitting model

After initialising the Chebyshev PDF, described in Section 6.8.3, the estimated parameters are fixed in further fitting steps. Due to potential shape differences in data and the fact that the Chebyshev polynomial is introduced on a finite sample size, the parametrisation is only an approximation. The M_{bc} fitter provides an uncertainty estimation for each of the Chebyshev PDF parameters which are given in Table 6.17. Variations of the central value of the parameters $k_{1,..,5}$ based on their primary fit uncertainties are performed, recomputing $\mathcal{N}_{\mathrm{CB}}$ for every variation. This is performed directly on the Belle II data when unblinding the analysis.

Because the uncertainties for Chebyshev parameters k_i are correlated, they can be varied in their eigenspace simultaneously. For this, a principal component analysis is performed and simultaneous variations of all parameters k are evaluated, based on the covariance matrix produced by the `zfit` interface. The M_{bc} fits are repeated with correlated variations of k_i. The maximal shifts of $\mathcal{N}_{\mathrm{CB}}$ are assigned as a systematic uncertainty. The evaluation of the shifts is performed in a blinded way, i.e., they are evaluated without directly observing the $\mathcal{N}_{\mathrm{CB}}$ values estimated by the fitter. Although only a variation of k_i is performed, due to a high correlation between signal and background PDF shapes in the fit (see Figure 6.40), the results can be interpreted as a general fitting model uncertainty.

6.14.3 $B → X_s γ$ **efficiency uncertainties**

The $B → X_s γ$ efficiency factorises into two components shown in Equation (6.27). Consequentially, two uncertainties are evaluated:

- $B → X_s γ$ selection modelling uncertainty;

- $B → X_s γ$ tagging efficiency uncertainty.

The procedure and the results to evaluate the uncertainties are presented further in this Section.

$B \rightarrow X_s \gamma$ **selection modelling**

The $B \rightarrow X_s \gamma$ selection modelling affects the final result through signal efficiency calculations. The efficiency attains uncertainties from the same corrections as those listed in Section 6.14.1 except for background branching fraction uncertainties, which are not applicable. As before, all the different uncertainties are considered as fully correlated between different E_γ^B intervals. Furthermore, uncertainties due to the BDT output and zernikeMVA modelling are included, based on the findings of Section 6.13.2. These uncertainties are treated as uncorrelated to allow for variations in efficiency as a function of E_γ^B.

Table 6.22: The $B \rightarrow X_s \gamma$ selection modelling uncertainties. The central values and uncertainties are also visualised in Figure 6.66, where corrections from Table 6.16 and Section 6.13.2 are included. The uncertainty sources are discussed in Section 6.14.3. The signal region is separated by the horizontal lines.

E_γ^B interval [GeV]	Signal selection efficiency	Systematic uncertainties			
		zernikeMVA selection	BDT output selection	$\pi^0 \rightarrow \gamma\gamma$ and $\eta \rightarrow \gamma\gamma$ suppression	Photon detection efficiency
$1.4 - 1.6$	0.279			1.090 ± 0.050	0.991 ± 0.023
$1.6 - 1.8$	0.367			1.074 ± 0.048	0.995 ± 0.022
$1.8 - 2.0$	0.448			1.064 ± 0.046	0.996 ± 0.021
$2.0 - 2.1$	0.496			1.055 ± 0.046	0.996 ± 0.021
$2.1 - 2.2$	0.526			1.050 ± 0.047	0.997 ± 0.021
$2.2 - 2.3$	0.550	$\pm 10\%$	$\pm 3\%$	1.046 ± 0.047	0.997 ± 0.021
$2.3 - 2.4$	0.568			1.045 ± 0.047	1.000 ± 0.020
$2.4 - 2.5$	0.585			1.047 ± 0.047	1.001 ± 0.019
$2.5 - 2.6$	0.601			1.050 ± 0.047	1.001 ± 0.019
$2.6 - 2.7$	0.573			1.050 ± 0.046	0.998 ± 0.019
> 2.7	0.222			1.053 ± 0.046	0.998 ± 0.018

$B \rightarrow X_s \gamma$ **tagging**

The tagging efficiency is calculated in MC samples and given in Equation (6.28). The FEI calibration correction, evaluated in Section 6.11.1, is applied to the calculated efficiency to, which results in

$$\varepsilon_{\text{FEI}} = 0.00444 \pm 0.00015 \tag{6.35}$$

in data. The resulting uncertainty in Equation (6.35) is treated as a fully correlated systematic error across all E_γ^B intervals.

6.14.4 Other uncertainties

Other uncertainties, that are not included in the previous categories are summarised in this Subsection. Although individually these uncertainties are not related, their importance is sub-leading in most E_γ^B bins.

Uncertainties due to $B \to X_d\gamma$

The uncertainty due to $B \to X_d\gamma$ component modelling strategy is discussed in Section 6.13.5. The measured number of $B \to X_{s/d}\gamma$ events is corrected based on the theoretical $B \to X_d\gamma$ expectation, amounting to roughly 4% of the value. The full difference $N(B \to X_d\gamma) \equiv N(B \to X_{s/d}\gamma) - N(B \to X_s\gamma)$ is assigned as a systematic uncertainty. The systematic uncertainty is considered uncorrelated between different E_γ^B bins to account for possible spectrum shape variations.

E_γ^B spectrum unfolding uncertainties

The unfolding uncertainties are calculated for every unfolding factor calculated in Section 6.13.6. Each correction factor obtains an uncertainty based on the hybrid-signal model, shown in Figure 6.4a. The resulting unfolding correction factors and their uncertainties are summarised in Table 6.23. They are considered correlated based on the correlation matrix evaluated when building the hybrid-signal model as seen in Figure 6.4.

Table 6.23: Bin-by-bin correction factors for unfolding based on Figure 6.68. They are calculated on a large simulated sample and therefore have negligible statistical uncertainties. The systematic uncertainty calculation approach includes E_γ^B spectrum shape parameter uncertainties, $B \to X_s\gamma$ and $B \to K^*(892)\gamma$ branching fraction uncertainties as discussed in Section 6.2.3.

E_γ^B interval [GeV]	Bin-by-bin unfolding factor	Systematic uncertainty	Statistical uncertainty
$1.8 - 2.0$	0.6840	±0.1297	
$2.0 - 2.1$	0.7913	±0.0906	
$2.1 - 2.2$	0.9053	±0.0476	
$2.2 - 2.3$	1.0492	±0.0294	$< O(10^{-3})$
$2.3 - 2.4$	1.1121	±0.0934	
$2.4 - 2.5$	1.2073	±0.1411	
$2.5 - 2.6$	1.626	±0.0566	
$2.6 - 2.7$	0.0	±0.0	

Uncertainty on the number of B mesons in the sample

The number of B mesons in the analysed Belle II data sample is estimated by an independent study (not part of the work presented in this thesis) with a data-driven method in which off-resonance data are used to subtract the non-$B\bar{B}$ contribution from the on-resonance data. It is found to be:

$$N_B = 2 \cdot (198 \pm 3) \cdot 10^6. \tag{6.36}$$

6.15 Results of the analysis

All the discussion thus far involved either MC samples, Belle II data samples with a negligible amount of $B \to X_s\gamma$ events, or with the signal region ($E_\gamma^B \in (1.8, 2.7)$ GeV) hidden. After

designing the analysis with simulation, performing extensive validation and evaluating systematic uncertainties, the signal region is ready to be unblinded, as it was shown that no significant biases are expected. This Section presents the main results of the unblinded analysis. The discussion related to them, their interpretation and future prospects for $B \rightarrow X_s \gamma$ are discussed in Chapter 7.

6.15.1 M_{bc} **fit results**

Following the M_{bc} fitting strategy described in Section 6.8.3 and the model in Table 6.17, the fits of the signal region in Belle II data are performed. They are shown in Figure 6.70. Together with fits in Figure 6.61, this gives all the data fits for the E_γ^B intervals defined in Section 6.8.2. The extracted $\mathcal{N}_{\mathrm{CB}}$ are shown in the top right corner of each Figure.
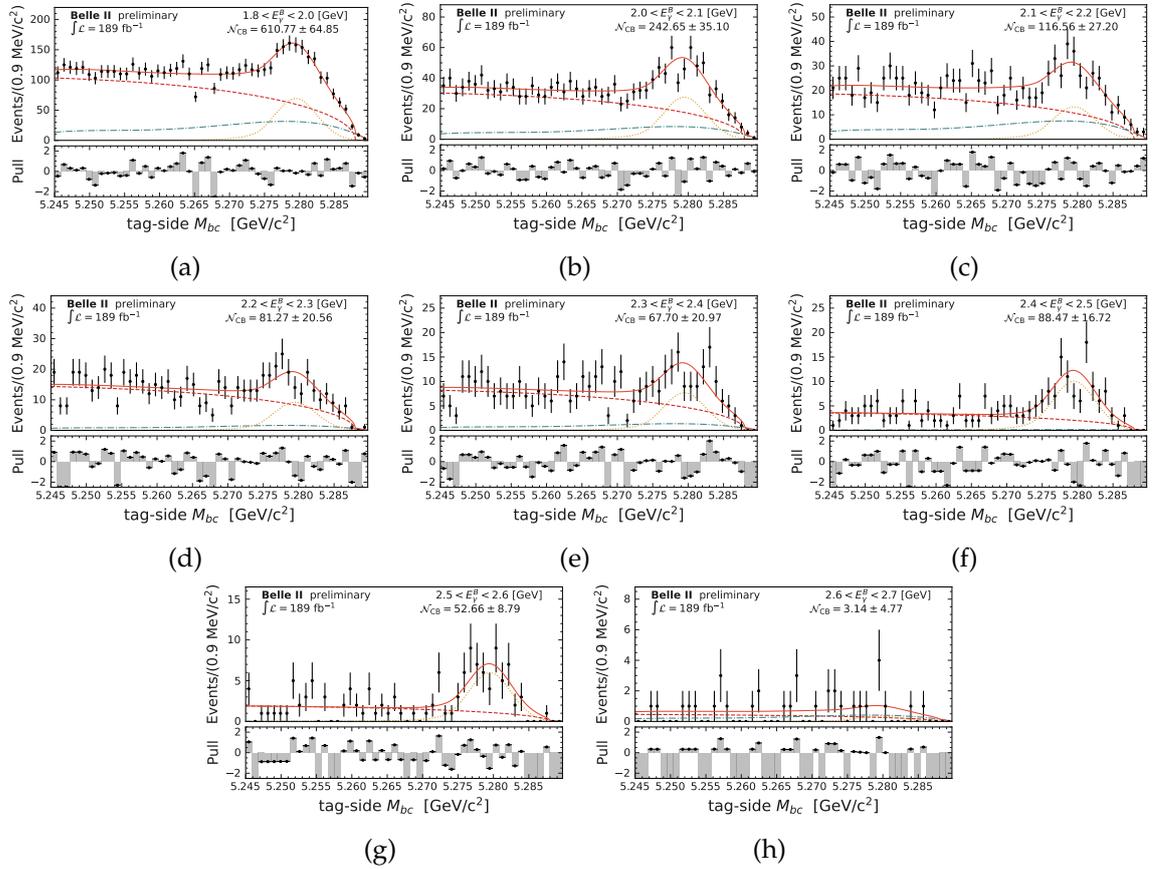
Figure 6.70: The E_γ^B signal region fits of the Belle II data, based on the fitting model in Table 6.17. The fits are performed as unbinned negative log-likelihood fits. The different E_γ^B intervals are shown in the top right corner of each Figure, together with the extracted good tag-B meson yield, $\mathcal{N}_{\mathrm{CB}}$, which in this case corresponds to $B \rightarrow X_{s/d} \gamma$ and other $B\bar{B}$ decay channels. The fits outside of the signal region are provided in Figure 6.61.

The fit results on the Belle II MC with all $B \rightarrow X_s \gamma$ events removed are shown in Fig-

ure 6.71. The extracted number of $B\bar{B}$ background events is shown in the top right corner of each Figure. These values are corrected, scaled, and are equal to the ones listed in Table 6.20. Together with fits in Figure 6.62, it gives all the MC fits for the defined E_γ^B intervals in Section 6.8.2.

(a)   (b)   (c)

(d)   (e)   (f)

(g)   (h)
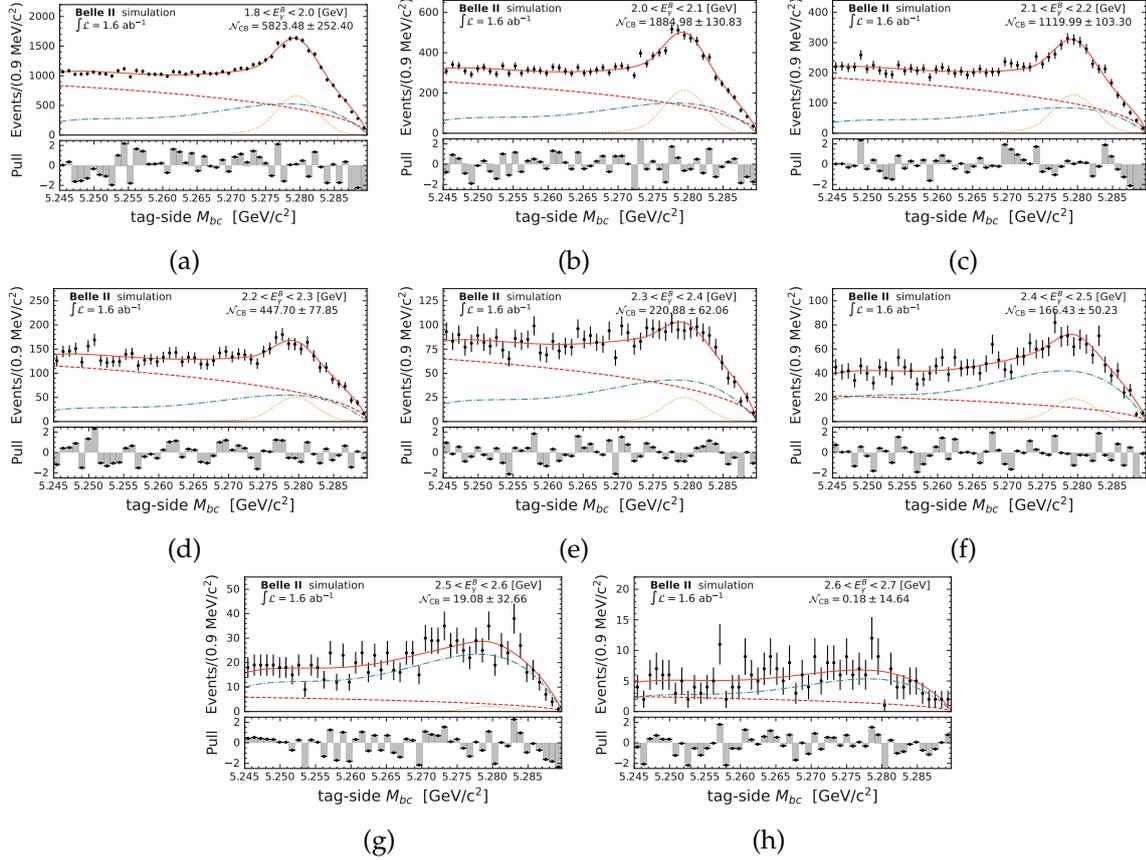
Figure 6.71: The E_γ^B signal region fits of the Belle II MC, where the $B \to X_{s/d}\gamma$ events have been removed. The fitting model is summarised in Table 6.17. The fits are performed as unbinned negative log-likelihood fits. The different E_γ^B intervals are shown in the top right corner of each Figure, together with the extracted good tag-B meson yield, $\mathcal{N}_{CB}$, which corresponds only to non-$B \to X_{s/d}\gamma$ events. The fits outside of the signal region are provided in Figure 6.61.

The number of good tag-B events, evaluated from fits in Belle II data (Figures 6.61 and 6.70), are summarised in Figure 6.72. The Figure also includes results from fits of Belle II MC with all $B \to X_s\gamma$ events removed (Figures 6.62 and 6.71). The latter provides the expectations of $B\bar{B}$ backgrounds that remain in Belle II data after the fit. The MC is corrected as discussed in Section 6.11, with appropriate uncertainties from Section 6.14 included. The Figure 6.72a shows the $\mathcal{N}_{CB}$ in linear scale, which makes the comparison of low-E_γ^B region easier. In this region, the number of signal events is expected to be much larger than any contribution from $B \to X_s\gamma$ events. On the other hand, Figure 6.72b shows the results in a

logarithmic axis, which makes it evident that an excess over the remaining $B\bar{B}$ background is present. This excess in data is evidence of $B \to X_{s/d}\gamma$ events.
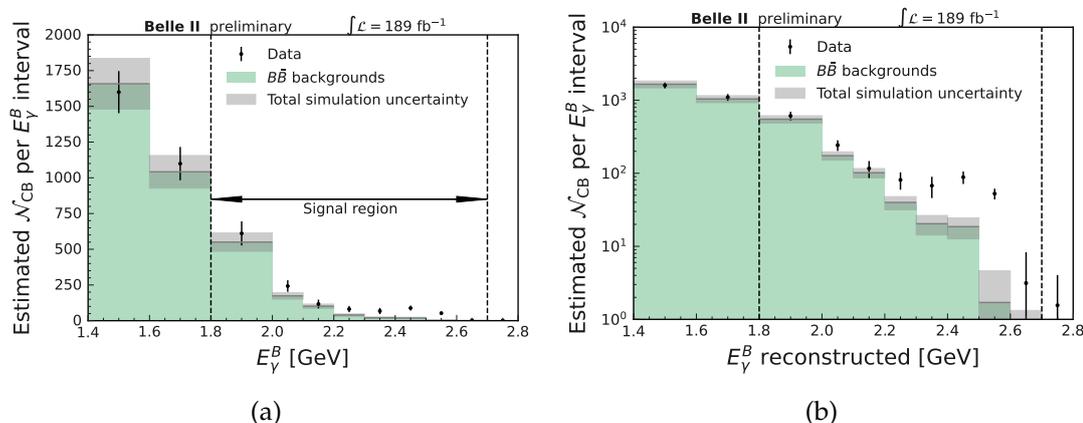
Figure 6.72: The evaluated $\mathcal{N}_{\text{CB}}$ in fits shown in Figures 6.61, 6.62, 6.70 and 6.71. All corrections discussed in Table 6.16 are applied, and relevant uncertainties from Tables 6.20 and 6.21 are included. The results with a linear, and with a logarithmic scale are shown in (a) and (b), respectively. The excess seen in Belle II data is evidence of the presence of $B \to X_{s/d}\gamma$ events.

6.15.2 Remaining-$B\bar{B}$ background subtraction results

The excess observed in Belle II data fits over background-only Belle II MC fits is attributed to the presence of $B \to X_{s/d}\gamma$ events. Following the background subtraction strategy shown in Sections 6.9 and 6.10.2, the remaining good tag-B meson backgrounds are subtracted. This corresponds to the subtraction of the filled (green) histogram from data points in Figure 6.72. Within uncertainties, the resulting difference is attributed to photons that originate in $B \to X_{s/d}\gamma$ decays. The background-subtracted photon energy spectrum is shown in Figure 6.73.

The background-subtracted photon energy spectrum includes all systematic uncertainties related to background subtraction (Section 6.14.1) and fitting uncertainties (Section 6.14.2). The results agree very well with the simulated expectation from the hybrid-signal model. The pulls (see Equation (6.17)) are calculated with respect to the hybrid-signal model expectation are evaluated and show that all the measured data points agree with the expectation within 1 to 2σ.

The measured number of events for the signal region in each E_γ^B interval, as well as a cumulative count are given in Table 6.24. The statistical and total uncertainties are depicted separately. The cumulative count is performed starting at the high-E_γ^B intervals and cumulatively adding the contributions from lower E_γ^B intervals (with systematic uncertainties summed appropriately to their correlation). This is done because the low end of the E_γ^B spectrum has the highest uncertainties and the smallest precision. Using Equation (6.34), the $B \to X_{s/d}\gamma$ counts are transformed to $B \to X_s\gamma$ event counts. The cumulative and binned results show excellent agreement with the expectation of the hybrid-signal model.

(a)　　　　　　　　　　　　　　　　(b)
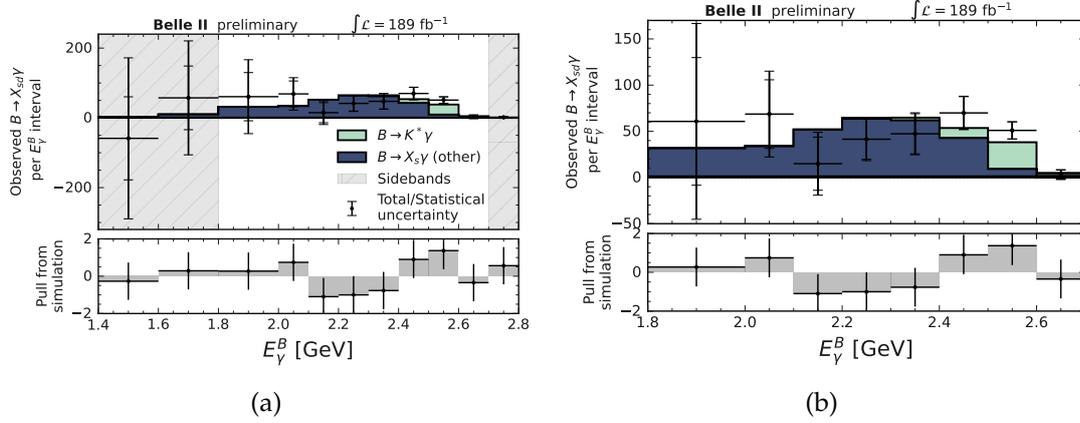
Figure 6.73: The counts of events consistent with radiative decays extracted from the fitted Belle II data after subtracting the remaining non-$B \to X_{s/d}\gamma$ background, based on results summarised in Figure 6.72. The total uncertainty includes background subtraction and fitting uncertainties, without efficiency corrections or unfolding. The data points are overlaid with the simulated expectation which shows the hybrid-signal model, defined in Section 6.2.3. All data points, including the sideband region, are shown in (a). On the other hand, (b) focuses only on the signal region. The legend is shared for both Figures.

Table 6.24: The observed number of events (before unfolding) that are consistent with $B \to X_{s/d}\gamma$ events in 189 fb^{-1} of Belle II data. The first half of the Table shows the central value, statistical uncertainty and total uncertainty (in brackets) for each E_γ^B interval. The second half of the Table shows the cumulative observed number of events (the summation is done from high-E_γ^B, where uncertainties are lower). The transformation between $B \to X_{s/d}\gamma$ and $B \to X_s\gamma$ is performed using the relation in Equation (6.34). The expected number of $B \to X_s\gamma$ events is provided based on the central values of the hybrid-signal model. All results are consistent with the expectations.

| E_γ^B interval [GeV] | Central value ± Statistical (Total) uncertainty | | | | | |
	Observed $B \to X_{s/d}\gamma$ events	Observed $B \to X_s\gamma$ events	Expected $B \to X_s\gamma$ value	Cumulative $B \to X_{s/d}\gamma$ events	Cumulative $B \to X_s\gamma$ events	Expected $B \to X_s\gamma$ value
1.4 − 1.6	−59 ± 119 (231)	−56 ± 114 (221)	3	357 ± 176 (310)	342 ± 168 (297)	357
1.6 − 1.8	57 ± 91 (163)	55 ± 87 (157)	10	416 ± 129 (207)	398 ± 124 (199)	354
1.8 − 2.0	61 ± 69 (106)	58 ± 66 (102)	32	358 ± 91 (127)	343 ± 88 (122)	344
2.0 − 2.1	69 ± 37 (46)	66 ± 36 (45)	34	298 ± 60 (69)	285 ± 57 (68)	312
2.1 − 2.2	15 ± 29 (34)	14 ± 28 (33)	52	229 ± 47 (52)	220 ± 45 (50)	278
2.2 − 2.3	41 ± 22 (23)	40 ± 21 (22)	65	214 ± 37 (39)	205 ± 35 (38)	226
2.3 − 2.4	47 ± 22 (23)	45 ± 21 (22)	65	173 ± 30 (31)	166 ± 29 (31)	162
2.4 − 2.5	70 ± 18 (18)	67 ± 17 (18)	54	126 ± 21 (21)	120 ± 20 (21)	97
2.5 − 2.6	51 ± 9 (9)	49 ± 9 (9)	38	56 ± 11 (11)	53 ± 10 (11)	43
2.6 − 2.7	3 ± 5 (5)	3 ± 5 (5)	5	5 ± 6 (6)	4 ± 5 (6)	5
> 2.7	2 ± 2 (2)	1 ± 2 (2)	0	2 ± 2 (2)	1 ± 2 (2)	0

6.15.3 Partial branching fraction measurement results

The measurement results, provided in Table 6.24 and visualised in Figure 6.73a, are used to calculate the partial branching fractions of $B \rightarrow X_s \gamma$ decays. To this end, observed $B \rightarrow X_s \gamma$ events are corrected for efficiency and unfolded. The efficiency used for the calculations is provided in Section 6.14.3, whereas the unfolding factors and strategy are described in Section 6.14.4. Note that up until now E_γ^B and $\tilde{E}_\gamma^B$ were used to explicitly differentiate the measured and the true (unfolded) photon energy, respectively. *In the following Sections only* the E_γ^B notation is used to denote the unfolded photon energy (i.e. the tilde notation is omitted).

Combining all the results presented in this thesis leads to the following form of Equation (6.1) for the partial branching fraction:

$$\frac{\Delta \mathcal{B}(B \rightarrow X_s \gamma)_i}{\Delta E_{\gamma,i}^B} = \frac{\mathcal{U}_i \cdot \left(\mathcal{N}_{\text{CB},i}^{\text{DATA}} - \mathcal{N}_{\text{CB},i}^{\text{non}-B \rightarrow X_{s/d}\gamma} - N_i^{B \rightarrow X_d \gamma} \right)}{\varepsilon_i \cdot N_B}, \tag{6.37}$$

where (left to right, top to bottom):

- i is a given E_γ^B interval;

- $\mathcal{U}_i$ is an unfolding factor for interval i, based on Table 6.23;

- $\mathcal{N}_{\text{CB},i}^{\text{DATA}}$ is the number of good tag-B mesons in Belle II data for interval i. These values are given in Table 6.24;

- $\mathcal{N}_{\text{CB},i}^{\text{non}-B \rightarrow X_s \gamma}$ is the number of good tag-B meson candidates where the signal B meson does not decay as $B \rightarrow X_{s/d}\gamma$, evaluated in Belle II MC (corrected for differences in luminosity and modelling) for an interval i. These values are given in Table 6.20;

- $N_i^{B \rightarrow X_d \gamma}$ is the number of $B \rightarrow X_d \gamma$ events contributing to an interval i;

- ε_i is the factorised signal efficiency for interval i, defined in Equation (6.27); with values taken from Table 6.22 and Equation (6.35);

- N_B is the number of B mesons in the analysed Belle II data sample, as given in Equation (6.36).

The results of the calculations based on Equation (6.37) with all results discussed and derived in this thesis are shown in Table 6.25. The statistical and systematic uncertainties are included, where the latter is broken down into four categories presented in Section 6.14. The calculations are only performed for the signal region $E_\gamma^B \in (1.8, 2.7)$ GeV where the unfolding factors are available. Furthermore, the $E_\gamma^B \in (2.6, 2.7)$ GeV interval is not given in Table 6.25, because the corresponding unfolding factor is zero (see Table 6.23).

Comparing the systematic and statistical uncertainties, the results in Table 6.25 are largely statistically dominated in high E_γ^B intervals. The systematic uncertainty is more comparable in the low-E_γ^B intervals, with background and fit modelling uncertainties being the dominant systematic errors in the low-E_γ^B region. On the other hand, signal selection modelling and unfolding uncertainties become the leading systematic errors at high-E_γ^B. These observations

Table 6.25: Results of the partial branching fraction measurement presented in this thesis, based on Equation (6.37). The first part of the Table shows the partial branching fractions for each E_γ^B interval, their statistical and systematic uncertainty components. The second part of the Table shows the breakdown of the systematic uncertainty into groups that are defined in Section 6.14. Statistical uncertainties remain the dominant component in the analysis. Note that signal efficiency and background modelling uncertainties are correlated due to the same correction factors used (see Section 6.11).

E_γ^B interval [GeV]	$\frac{\Delta \mathcal{B}(B \to X_s \gamma)_i}{\Delta E_{\gamma,i}^B}(10^{-4})$	Statistical uncertainty (10^{-4})	Systematic uncertainty (10^{-4})	Systematic uncertainty group			
				Background modelling	M_{bc} fit model	$B \to X_s\gamma$ efficiency	Other uncertainties
1.8 − 2.0	0.48	±0.54	±0.64	0.49	0.42	0.03	0.09
2.0 − 2.1	0.57	±0.31	±0.25	0.17	0.17	0.06	0.07
2.1 − 2.2	0.13	±0.26	±0.16	0.11	0.13	0.01	0.01
2.2 − 2.3	0.41	±0.22	±0.10	0.04	0.07	0.05	0.02
2.3 − 2.4	0.48	±0.22	±0.10	0.02	0.06	0.06	0.05
2.4 − 2.5	0.75	±0.19	±0.14	0.02	0.04	0.09	0.09
2.5 − 2.6	0.71	±0.13	±0.10	0.00	0.02	0.09	0.04

are consistent with the fact that the signal-to-background ratio grows with increasing E_γ^B (as seen in e.g. Table 6.14), and the fact that the number of $B \to X_s\gamma$ events is larger at higher E_γ^B values.

The systematic uncertainties in Table 6.25 are correlated: both between different systematic uncertainty groups (i.e. signal selection efficiency and background modelling) and between different E_γ^B bins. These correlations, as discussed in Section 6.14, are combined and evaluated. The correlation matrix is shown in Figure 6.74b. The statistical and systematic component is merged here. As the results have a large statistical uncertainty component, the correlation of uncertainties is found to be low.

6.15.4   Total branching fraction measurement results

As the measurement employs a selection of $E_\gamma^B > 1.4$ GeV, the full $\mathcal{B}(B \to X_s\gamma)$ cannot be measured. Furthermore, the $E_\gamma^B \in (1.4, 1.8)$ GeV range has a large systematic and statistical uncertainty component. Therefore, only an evaluation in the $E_\gamma^B \in (1.8, 2.6)$ GeV interval from experimental data is feasible. The results for thresholds of 1.8 GeV, 2.0 GeV and 2.1 GeV are given in Table 6.26. The combined value of the correlated systematic uncertainties becomes dominant at the low-E_γ^B threshold but remains comparable to the statistical uncertainty.

To compare with the results of other experiments and theoretical values, the theoretical extrapolation factors are used, as provided in Ref. [86], to evaluate the branching fraction at the threshold of $E_\gamma^B > 1.6$ GeV. As it was discussed in Section 2.4, for $B \to X_s\gamma$ total rate evaluation, $E_\gamma^B > 1.6$ GeV is a conventionally chosen threshold. The extrapolated results are also provided in the last column of Table 6.26. Note that such extrapolation is model-dependent and cannot replace accurate measurements with a lower threshold.

(a) (b)
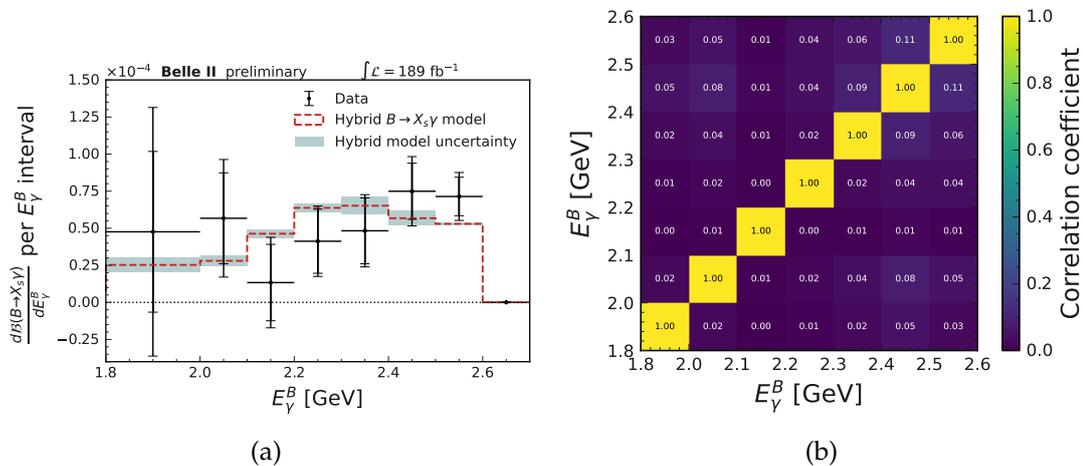
Figure 6.74: The visualisation of the results in Table 6.25. The partial branching fractions of $B \to X_s\gamma$ as a function of E_γ^B, measured in 189 fb^{-1} Belle II data are shown in (a). The inner error bars correspond to the statistical uncertainty, whereas the outer ones to the total. The measurement results are overlaid with the expectations for the hybrid-signal model, and the associated uncertainty to it, evaluated in Section 6.2.3. The data results show excellent agreement with the model. The correlation matrix of the partial branching fraction total uncertainties is shown in (b).

Table 6.26: The integrated $B \to X_s\gamma$ branching fractions for different low-E_γ^B thresholds measured with 189 fb^{-1} of Belle II data. They are evaluated by summing the partial branching fractions in Table 6.25. The systematic and statistical uncertainties are denoted in the brackets. The extrapolation factors are taken from [86]. They are not provided for the $E_\gamma^B = 2.1$ GeV threshold, therefore the result is evaluated by extrapolating linearly and assuming a monotonic increase in uncertainty. This results in an extrapolation factor of 0.870 ± 0.024. The number should be interpreted with the aforementioned caveats in mind only.

E_γ^B lower threshold [GeV]	$\mathcal{B}(B \to X_s\gamma)$ [10^{-4}]	$\mathcal{B}(B \to X_s\gamma)$ [10^{-4}] (extrapolated to 1.6 GeV) [86]
1.8	3.54 ± 0.78 (stat.) ± 0.83 (syst.)	3.65 ± 0.80 (stat.) ±0.86 (syst.) ±0.02 (extrap.)
2.0	3.06 ± 0.56 (stat.) ± 0.47 (syst.)	3.42 ± 0.62 (stat.) ±0.52 (syst.) ±0.06 (extrap.)
2.1	2.49 ± 0.46 (stat.) ± 0.35 (syst.)	2.86 ± 0.53 (stat.) ±0.40 (syst.) ±0.08 (extrap.)

6.15.5   Moments of the $B \to X_s\gamma$ photon energy spectrum

As discussed in Sections 2.4 and 2.5 the moments of the $B \to X_s\gamma$ spectrum are important for the understanding of $B \to X_s\gamma$ decay properties. The first and second moments of the $B \to X_s\gamma$ spectrum are calculated based on the results in Table 6.25. They are approximated as a weighted sum of the centre value of E_γ^B intervals, ΔE_γ^B, with weights corresponding to the partial branching fractions in that E_γ^B interval:

$$\left\langle E_\gamma^B \right\rangle = \frac{\sum_i f_c(\Delta E_{\gamma,i}^B) \cdot \frac{\Delta \mathcal{B}(B \to X_s\gamma)_i}{\Delta E_{\gamma,i}^B}}{\sum_i \frac{\Delta \mathcal{B}(B \to X_s\gamma)_i}{\Delta E_{\gamma,i}^B}}, \tag{6.38}$$

where f_c is used as a loose notation for a function that returns the central point of an interval. The calculated values of the first and second moments of E_γ^B for different lower thresholds are given in Table 6.27. The systematic and statistical uncertainty sources match those in earlier Subsections, and their correlation is accounted for as discussed before.

The measured values of the first moment in Table 6.27 (average of the photon energy spectrum) slightly decrease with the E_γ^B threshold. The total uncertainty changes from 2% to 4% as the E_γ^B lower threshold decreases from 2.1 GeV to 1.8 GeV. On the other hand, the second moment (the variance of the spectrum) highly depends on the threshold chosen. The uncertainties grow swiftly with a decreasing E_γ^B threshold. While the uncertainty with a $E_\gamma^B = 2.0$ GeV threshold is at 25%, this increases to 46% at 1.8 GeV. A larger uncertainty is attributed to the fact that the dispersion of the possible E_γ^B energies highly depends on a precise measurement of the tail. Conversely, in the case of the first moment, the peak region carries the highest importance.

Table 6.27: The moments of the integrated $B \to X_s\gamma$ photon energy spectrum for different low-E_γ^B thresholds measured on 189 fb^{-1} of Belle II data. These moments are evaluated by a weighted sum of the partial branching fractions in Table 6.25 according to Equation (6.38). The systematic and statistical uncertainties are denoted in the brackets.

E_γ^B lower threshold [GeV]	$\langle E_\gamma^B \rangle$ [GeV]	$\langle E_\gamma^{B2} \rangle - \langle E_\gamma^B \rangle^2$ [GeV^2]
1.8	2.284 ± 0.065 (stat.) ± 0.071 (syst.)	0.0502 ± 0.0157 (stat.) ± 0.0176 (syst.)
2.0	2.343 ± 0.036 (stat.) ± 0.026 (syst.)	0.0315 ± 0.0063 (stat.) ± 0.0045 (syst.)
2.1	2.410 ± 0.032 (stat.) ± 0.019 (syst.)	0.0147 ± 0.0057 (stat.) ± 0.0036 (syst.)

Chapter 7

Overview of the results and future prospects

The full analysis setup and the branching fraction measurement results in Sections 6.15.3 and 6.15.4 are the ultimate goal of this analysis. They represent the first implementation of an inclusive radiative analysis at Belle II. Moreover, it is the first implementation of a hadronic-tagged $B \to X_s\gamma$ measurement since 2007, when the result was reported by the BaBar collaboration (see Table 3.2). The results of this Belle II analysis are available in Ref. [163].

This Chapter is dedicated to the discussion and overview of the results, as well as their significance. The prospects of $B \to X_s\gamma$, focusing on the hadronic-tagged measurements, are also discussed.

7.1 Discussion of the Belle II hadronic-tagged $B \to X_s\gamma$ results

The resulting values in Table 6.26 are in perfect agreement with the SM predictions (see Section 2.4). They also agree with the past measurements of $B \to X_s\gamma$, which used both inclusive and sum-of-exclusive methods. The visual comparison of all inclusive measurements is provided in Figure 7.1.

The hadronic-tagged $B \to X_s\gamma$ analysis uncertainty is dominated by its statistical component as seen in Table 6.26. The systematic uncertainty is larger than the statistical uncertainty only if the lowest-E_γ^B signal-region interval is included in the integrated branching fraction evaluation, as the systematic uncertainty is driven by the number of background events in the post-fit sample. Judging from the partial branching fractions of $B \to X_s\gamma$ in Section 6.15.3, the largest contribution to the systematic uncertainty originates from background modelling. This contribution drops off quickly as the number of background events decreases. The M_{bc} fitting model uncertainties also drop quickly with E_γ^B, as some of the main fitting challenges reduce with the decreasing amount of combinatorial $B\bar{B}$ and continuum events. On the other hand, the uncertainties due to signal modelling and $B \to X_d\gamma$ contamination increase with E_γ^B, as the number of $B \to X_{s/d}\gamma$ events grows.

As discussed in Section 3.2, different analysis procedures are complementary to each other. Therefore, although individually the result of this thesis is not competitive with the most precise measurements of BaBar and Belle, which are performed with up to 3 times larger
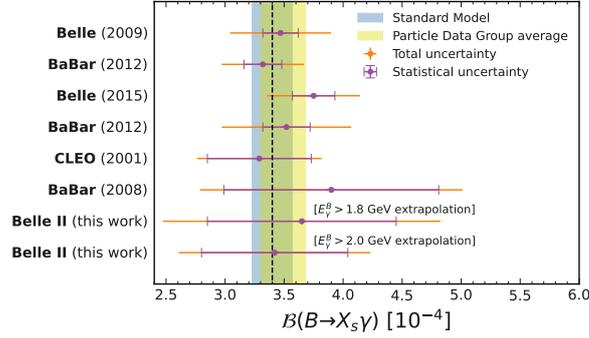
Figure 7.1: The results of the measurement of the $B \to X_s\gamma$ branching fractions compared with past results from other experiments. The results from other experiments correspond to those in Table 3.2. The results from this analysis are taken from Table 6.26. The values are extrapolated to $E_\gamma^B > 1.6$ GeV threshold, as discussed in Section 6.15.4. The Standard Model expectation corresponds to the Equation (2.22), whereas the Particle Data Group average [16] to Equation (3.1).

data sets, it demonstrates important consistency between different measurement techniques. Furthermore, it is the second-ever measurement of the hadronic-tagged $B \to X_s\gamma$: therefore it serves as a proof-of-concept of the measurement technique and its applicability across different experimental setups.

It is interesting to compare the uncertainties obtained in this analysis with those seen in the past. Consider the hadronic-tagged measurement of BaBar [88] (extrapolated to 1.6 GeV), which obtains:

$$\mathcal{B}(B \to X_s\gamma) = (3.90 \pm 0.91(\text{syst.}) \pm 0.64(\text{stat.})) \cdot 10^{-4}. \tag{7.1}$$

It is possible to compare this to the results of the analysis presented in this thesis to see that the statistical uncertainty of the Belle II result is $0.91/0.80 \approx 1.14$ times lower. The improvement of the statistical uncertainty is a combination of several reasons that include:

- A higher tagging efficiency at Belle II offered by the FEI algorithm compared to the one used at BaBar (compare $\varepsilon_{\text{FEI}} = (0.44 \pm 0.02)\%$ in Equation (6.35) with the 0.3% reported in Ref. [88]);

- Different continuum suppression strategy (this analysis uses a BDT, whereas BaBar used a Fischer discriminant);

- Different fitting setup (this analysis uses three PDFs, whereas BaBar perform the fit using a Crystal Ball and ARGUS PDF combination);

The systematic uncertainty highly depends on the lower-E_γ^B threshold. Therefore, a meaningful comparison is only possible where the E_γ^B threshold is the same. The BaBar measurement in Equation (7.1) is evaluated with a threshold of $E_\gamma^B > 1.9$ GeV. The systematic uncertainty of this analysis, evaluated at 1.8 and 2.0 GeV thresholds, can be approximated by linearly interpolating: $0.5 \cdot (0.52 + 0.86) \approx 0.69$. Therefore, the systematic uncertainty is

slightly higher but comparable, which is consistent with the lower-E_γ^B threshold setting. Furthermore, the evaluated uncertainties leave room for future improvement as will be discussed in Section 7.2.

Finally, the measurement of E_γ^B spectrum moments in Table 6.27 agrees well with the world average values [16]:

$$\langle E_\gamma^B \rangle = 2.314 \pm 0.011 \text{ GeV}; \quad \langle E_\gamma^{B2} \rangle - \langle E_\gamma^B \rangle^2 = 0.0303 \pm 0.0025 \text{ GeV}^2. \tag{7.2}$$

The uncertainties of the measured moments in Table 6.27 are several times larger than the world average values. They contain a comparable significant statistical and systematic uncertainty component.

Interestingly, at higher-E_γ^B thresholds, e.g. 2.1 GeV, while still larger, the systematic uncertainty is comparable to the total uncertainty of the world average. Interpreting the uncertainties from the partial branching fraction measurement in Table 6.25 indicates that background and fit modelling uncertainties are large contributors to the systematic uncertainty at lower-E_γ^B thresholds. Therefore, for improved accuracies of the hadronic tagged measurement, the upcoming larger Belle II data set will not be sufficient alone: parameter estimation, background subtraction and modelling uncertainties have to be further studied. The next Section discusess the anticipated prospects for this goal.

7.2 Future prospects for hadronic-tagged $B \to X_s\gamma$ analysis at Belle II

Belle II is an ongoing experiment, which means that more and more e^+e^- collision data will be recorded in the next decade. No other ongoing experiment can contribute to the inclusive radiative measurements. As it was clear from Section 6.15 and Section 7.1, at the moment the analysis is limited by the statistical uncertainty. With the larger Belle II data set, the importance of systematic effects will grow. Although in this analysis several systematic uncertainties are set at their conservative estimates, additional studies will allow reducing them. This was studied, as part of the original work for this thesis, and the results are available in Ref. [98]. The uncertainty projections for the hadronic-tagged $B \to X_s\gamma$ are summarised in Table 7.1.

The statistical uncertainties for the hadronic tagged $B \to X_s\gamma$ are expected to reach the 5% level with 5 ab^{-1} of Belle II data. The systematic uncertainty expectations are evaluated assuming that the main contributor to the systematic uncertainty is the remaining-$B\bar{B}$ background subtraction and FEI tagging. If the knowledge of remaining-$B\bar{B}$ background stays at the 10% level (8.7% evaluated in this analysis) and FEI calibration uncertainty is not improved, it is expected that a 6.5% total systematic uncertainty on the branching fraction of $B \to X_s\gamma$ can be achieved. On the other hand, if the remaining after-fit $B\bar{B}$ background modelling is understood to a 5% level, it is plausible to half the expected systematic uncertainty. The uncertainties on signal selection efficiency will further reduce as the understanding of background suppression tools (e.g. the π^0 veto, `zernikeMVA`) improves. The $B \to X_d\gamma$ component will be accurately subtracted when precise $B \to X_d\gamma$ measurements with Belle II are performed.

Table 7.1: The projected uncertainties for the hadronic-tagged $B \to X_s\gamma$ with the increased Belle II data set size. These projections are evaluated assuming the principal contributions in systematic uncertainty arise from background modelling and suppression uncertainties. The baseline case is presented for a scenario where the remaining good tag-B meson background is known to 10%, whereas the improved scenario corresponds to where it is known to a 5% accuracy.

Lower E_γ^B threshold	Statistical uncertainty				Baseline (improved) systematic uncertainty
	1 ab^{-1}	5 ab^{-1}	10 ab^{-1}	50 ab^{-1}	
1.4 GeV	10.7%	6.4%	4.7%	2.2%	10.3 % (5.2%)
1.6 GeV	9.9 %	6.1%	4.5%	2.1%	8.5 % (4.2%)
1.8 GeV	9.3 %	5.7%	4.2%	2.0%	6.5 % (3.2%)
2.0 GeV	8.3 %	5.1%	3.8%	1.7%	3.7 % (1.8%)

Summarising, world-leading hadronic-tagged $B \to X_s\gamma$ measurements with Belle II data sets of $1-5$ ab^{-1} are possible if remaining background contributions are understood to a 5% or higher precision. This is an important observation given the fact that other types of inclusive $B \to X_s\gamma$ analysis techniques with the Belle and BaBar data sets are already limited by systematic uncertainties (see Table 3.2). Therefore, a different approach, one that will be provided by the hadronic-tagged analyses, is necessary for further insights into the radiative $B \to X_s\gamma$ transitions. In the shorter term, as Belle has not reported a hadronic-tagged $B \to X_s\gamma$ analysis, a joint Belle and Belle II analysis may provide a total data set of approximately 1 ab^{-1}, enabling such results in the next couple of years.

7.3   Input of the results on the SIMBA global fit

In a collaborative effort with the SIMBA collaboration, the results of Table 6.25 have been used to evaluate the m_b and λ_1 and the C_7^{incl} parameters (see Section 2.5). Originally in Ref. [62], a simultaneous parameter estimation fit is performed using all available experimental results of the photon energy spectrum that include BaBar and Belle results of hadronic, sum-of-exclusive, lepton-tagged and untagged measurement strategies. The same procedure in terms of the fitting strategy is repeated with the additional Belle II result described in this thesis.

The fit results and the comparison of the effect on the estimated value and uncertainties of the fit parameters are shown in Figure 7.2. Note that the $E_\gamma^B \in (1.8, 2.0)$ GeV interval of the Belle II result is excluded from the fit due to the large uncertainty and differences in the interval widths compared to earlier $B \to X_s\gamma$ measurements. The results slightly shift the central values of the parameters but the overall result retains a similar uncertainty and is consistent with the earlier results.

With the new inputs from Belle II, the determined SIMBA values are:

$$m_b^{1S} = 4.748 \pm 0.043 \text{ GeV}/c^2; \quad \lambda_1^{inv} = -0.219 \pm 0.082 \text{ GeV}^2/c^4, \tag{7.3}$$

where the results can be directly compared with the previous values from SIMBA given in Equation (2.34). The uncertainties for both values, as before, combine fitting, theoretical and parametric components. The latter two are assumed to not have changed compared with the earlier SIMBA fit excluding the Belle II result. While the current impact of the Belle II results

(a)

(b)
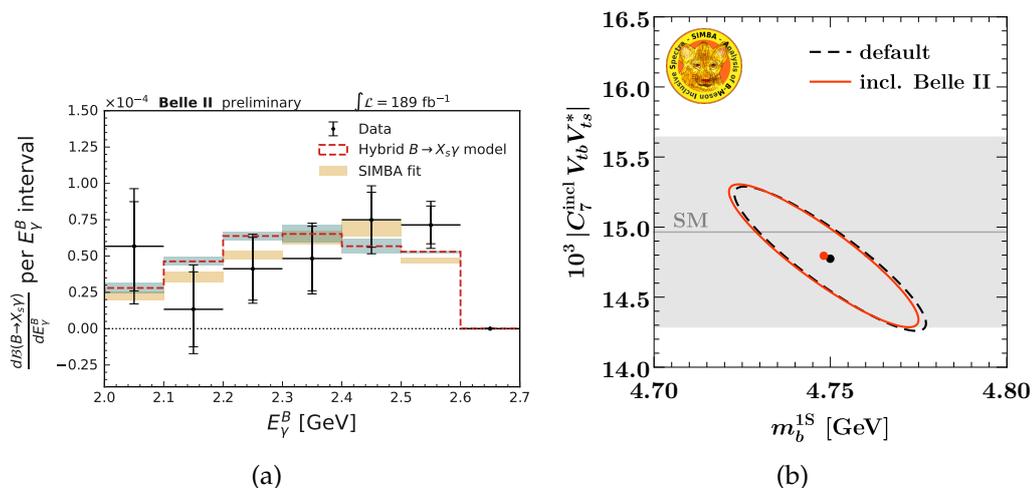
Figure 7.2: The results of the $B \to X_s \gamma$ spectrum parameter determination by the SIMBA collaboration which includes the results of the work presented in this thesis [62]. The results of the SIMBA fit superimposed on Figure 6.74a are given in (a). The corresponding $|C_7^{\text{incl}} V_{tb} V_{ts}^*|$ and m_b^{1S} values are visualised in (b). The dark dashed line (default) corresponds to the result without the Belle II results, whereas the red curve includes the Belle II result. As expected due to the current low statistical precision of the added $B \to X_s \gamma$ measurement of Belle II, the impact is small. Credit to the SIMBA collaboration for the fit and (b).

is small, the results of the hadronic-tagged Belle II analyses still have room for improvement, as discussed in Section 7.2. Moreover, they provide better sensitivity to the details of the E_γ^B spectrum, due to the direct access to the B meson rest frame. Therefore, future versions of the Belle II analysis presented in this thesis will be key inputs to the SIMBA results and other global fits, such as Ref. [164].

Chapter 8

Summary and outlook

The work presented in this thesis, radiative $B \to X_s\gamma$ decays have been studied using 189 fb^{-1} of e^+e^- collision data collected at the $\Upsilon(4S)$ resonance energy by the Belle II experiment. These decays are important probes of the Standard Model and beyond, providing relevant information to deepen the understanding of nature. Such information can be obtained through the measurements of $B \to X_s\gamma$ photon energy spectrum shape, its moments and the total branching fraction.

Chapters 2 to 5 provided an overview of the theoretical foundation and experimental status of $B \to X_s\gamma$ decays, introduced the state-of-the-art B-factory experiment Belle II and summarised the most relevant analysis techniques used in high energy physics. Chapter 6 described the analysis of the $B \to X_s\gamma$ decays using a hadronic-tagging approach, where the partnering B meson from the $\Upsilon(4S)$ decay is fully reconstructed. In an inclusive treatment, only the high energy photon is reconstructed, ensuring that all X_s states are selected. Hadronic-tagged $B \to X_s\gamma$ decays were measured for the first time with the Belle II experiment, and for the second time, in general. It is an important measurement technique which enables direct access to observables in the decaying B meson rest frame. Chapter 7 discussed the analysis results in the context of past measurements, future outlook and current impact.

To achieve the best result, this $B \to X_s\gamma$ analysis used the Full Event Interpretation algorithm, which is a series of boosted decision trees, to reconstruct the partner B meson. As a result, this analysis described the steps to suppress two types of background: those related to the signal-side and those related to the tag-side. Signal-side backgrounds have been suppressed by employing multivariate algorithms and selections. The tag-side backgrounds were suppressed by extracting the counts of correctly-reconstructed tag-B mesons in different photon energy bins. The remaining photon backgrounds were then removed by relying on simulation. The analysis selections and the overall procedure were tested and validated with numerous independent processes, including other B decays, $e^+e^- \to q\bar{q}$, and $e^+e^- \to \mu^+\mu^-$. The results obtained in this thesis, also available as preprint in Ref. [163], show excellent agreement with the world averages and the Standard Model predictions. They are compiled in Table 8.1.

The current version of the analysis contains a large statistical uncertainty component. This will be reduced in the future, as more and more data are collected by the Belle II experiment. The systematic uncertainty strongly depends on the lower photon energy threshold which

Table 8.1: Compilation of several highlight results of presented the thesis.

Observable [$E_\gamma^B > 1.8$ GeV]	Results of the analysis	World average value [16]
$\mathcal{B}(B \to X_s\gamma)$ [10^{-4}]	3.54 ± 0.78 (stat.) ± 0.83 (syst.)	3.49 ± 0.19
$\langle E_\gamma^B \rangle$ [GeV]	2.284 ± 0.065 (stat.) ± 0.071 (syst.)	2.314 ± 0.011
$\langle E_\gamma^{B2} \rangle - \langle E_\gamma^B \rangle^2$ [GeV^2]	0.0502 ± 0.0157 (stat.) ± 0.0176 (syst.)	0.0303 ± 0.0025

is employed to suppress background. Indeed, with a threshold of 1.8 GeV that was used in this analysis, background modelling uncertainties were seen to be some of the largest. In the future versions of this analysis, improved understanding of the background can reduce the systematic uncertainty down to $5 - 10\%$. Such precision is highly anticipated in many global fits, such as the one performed by the SIMBA collaboration [62]. While the current results may not provide a significant impact on the theoretical averages, they serve as a stepping stone for future Belle II radiative hadronic-tagged analyses that will be leading contenders in the field of flavour physics.

Appendices

Appendix A

Local operators in the effective Lagrangian

The effective Lagrangian given in Equation (2.12) includes the local operators O_i. In the equation, they denote the current-current O_{1-2}, four-quark penguin operators (O_{3-6}) and the photonic dipole (O_7) and gluonic dipole (O_8) operators [28] and have their full expressions given by [27], [29]:

$$O_1^u = (\bar{s}_L \gamma_\mu T^a u_L)(\bar{u}_L \gamma^\mu T^a b_L),$$
$$O_2^u = (\bar{s}_L \gamma_\mu u_L)(\bar{u}_L \gamma^\mu b_L),$$
$$O_1 = (\bar{s}_L \gamma_\mu T^a c_L)(\bar{c}_L \gamma^\mu T^a b_L),$$
$$O_2 = (\bar{s}_L \gamma_\mu c_L)(\bar{c}_L \gamma^\mu b_L),$$
$$O_3 = (\bar{s}_L \gamma_\mu b_L) \sum_q (\bar{q} \gamma^\mu q),$$
$$O_4 = (\bar{s}_L \gamma_\mu T^a b_L) \sum_q (\bar{q} \gamma^\mu T^a q),$$
$$O_5 = (\bar{s}_L \gamma_{\mu 1} \gamma_{\mu 2} \gamma_{\mu 3} b_L) \sum_q (\bar{q} \gamma^{\mu 1} \gamma^{\mu 2} \gamma^{\mu 3} q),$$
$$O_6 = (\bar{s}_L \gamma_{\mu 1} \gamma_{\mu 2} \gamma_{\mu 3} T^a b_L) \sum_q (\bar{q} \gamma^{\mu 1} \gamma^{\mu 2} \gamma^{\mu 3} T^a q),$$
$$O_7 = \frac{e}{16\pi^2} m_b (\bar{s}_L \sigma^{\mu\nu} b_R) F_{\mu\nu},$$
$$O_8 = \frac{g}{16\pi^2} m_b (\bar{s}_L \sigma^{\mu\nu} T^a b_R) G_{\mu\nu}^a,$$

where $q = u, d, c, s, t$. The $q_{L(R)}$ correspond to left-handed (right-handed) quark spinors. $F_{\mu\nu}$ and $G_{\mu\nu}^a$ are the field strength tensors of the electromagnetic and strong fields, respectively. T^a are the generators of the $SU(3)$ group and $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$ [12].

Appendix B

Major production cross-sections at B-factories

Although B-factories specialise in measuring the products of $\Upsilon(4S)$ decays, many more processes can occur in an e^+e^- collision. These processes are summarised in the Table B.1. Three main categories are the $B\bar{B}$ events, continuum events and low-multiplicity events. $B\bar{B}$ are usually the main decay type of interest. Continuum events correspond to the hadronic (and τ lepton) events. The low-multiplicity events are named as such due to the low number of charged particles present in the final state of such collisions.

Table B.1: The largest production cross-sections at B-factories. The processes are split up into three conventional categories that are often used to refer to the processes. The details on each of the numbers can be found in Ref. [23].

Process	$\sigma(e^+e^- \rightarrow X)$ [nb]
$B\bar{B}$ events	
$\sigma[\Upsilon(4S)]$	1.11
Continuum events	
$\sigma[u\bar{u}(\gamma)]$	1.61
$\sigma[d\bar{d}(\gamma)]$	0.40
$\sigma[s\bar{s}(\gamma)]$	0.38
$\sigma[c\bar{c}(\gamma)]$	1.30
$\sigma[\tau^+\tau^-(\gamma)]$	0.919
Low-multiplicity events	
$\sigma[e^+e^-(\gamma)]$	300
$\sigma[\gamma\gamma(\gamma)]$	4.99
$\sigma[\mu^+\mu^-(\gamma)]$	1.148
$\sigma[e^+e^-e^+e^-]$	39.7
$\sigma[e^+e^-\mu^+\mu^-]$	18.9
$\sigma[\nu\bar{\nu}(\gamma)]$	0.00025

Appendix C

Transformation to the B meson rest frame

Consider an $\Upsilon(4S)$ event, produced at rest, in an e^+e^- collision with a collision energy of $\sqrt{s}$. The $\Upsilon(4S)$ subsequently decays into two B mesons, B and B'. Let's assume the former undergoes a decay $B \to XZ$, where X is some system of particles that cannot be measured, and Z is a particle whose energy is known in the $\Upsilon(4S)$ frame. In such case, the four-momentum of the B meson in the $\Upsilon(4S)$ rest frame is denoted as:

$$p_B = (E_B^*, \vec{p}_B^*), \tag{C.1}$$

and that of the Z particle:

$$p_Z = (E_Z^*, \vec{p}_Z^*). \tag{C.2}$$

Performing a Lorentz boost into the B rest frame from the $\Upsilon(4S)$ rest frame can be done with the factors $\vec{\beta} = -\vec{p}_B^*/E_B^*$ (the negative sign as we are boosting into a frame of reference where B is stationary) and $\gamma = 1/\sqrt{1-\beta^2}$:

$$E_Z^B = \gamma(E_Z^* + \vec{\beta}\vec{p}_Z^*). \tag{C.3}$$

However, the X system, as stated before, cannot be measured and, as a result, the value of β cannot be evaluated from reconstructed decay products. On the other hand, let's assume that the B' is reconstructed, with its total momentum and energy known. Using the fact that $\Upsilon(4S)$ is stationary in its rest frame and the four-momentum relation between B and B', $p_{\Upsilon(4S)} = p_B^* + p_{B'}^*$, one arrives at

$$\sqrt{s} = E_B^* + E_{B'}^*,$$
$$0 = \vec{p}_B^* + \vec{p}_{B'}^*. \tag{C.4}$$

Knowing the precise value of collision energy and the momentum of B' is therefore sufficient to evaluate:

$$\vec{\beta} = \vec{p}_{B'}^*/(\sqrt{s} - E_{B'}^*) \tag{C.5}$$

which can be substituted into Equation (C.3).

Appendix D

FEI signal probability for specific channels of B

In general, FEI signal probability cannot be a good quantity for a quantitative evaluation of reconstruction quality because different classifier chains are necessary to reconstruct B^+ and B^0 candidates. This is discussed in the comparison between feiB$^+$ and feiB^0 modes in Section 6.3.3. However, even for specific channels for given B modes $\mathcal{P}_{\text{FEI}}$ is not a well-calibrated quantity. This is shown in Figure D.1 for B^+ modes and Figure D.2 for B^0 modes. Different distributions have strongly differing shapes, which translate to different selection efficiency and/or purity. Therefore, a general selection on $\mathcal{P}_{\text{FEI}}$ necessarily results in a direct bias to the selected tag-side modes and may not necessarily correspond to the 'best' reconstructed mode.
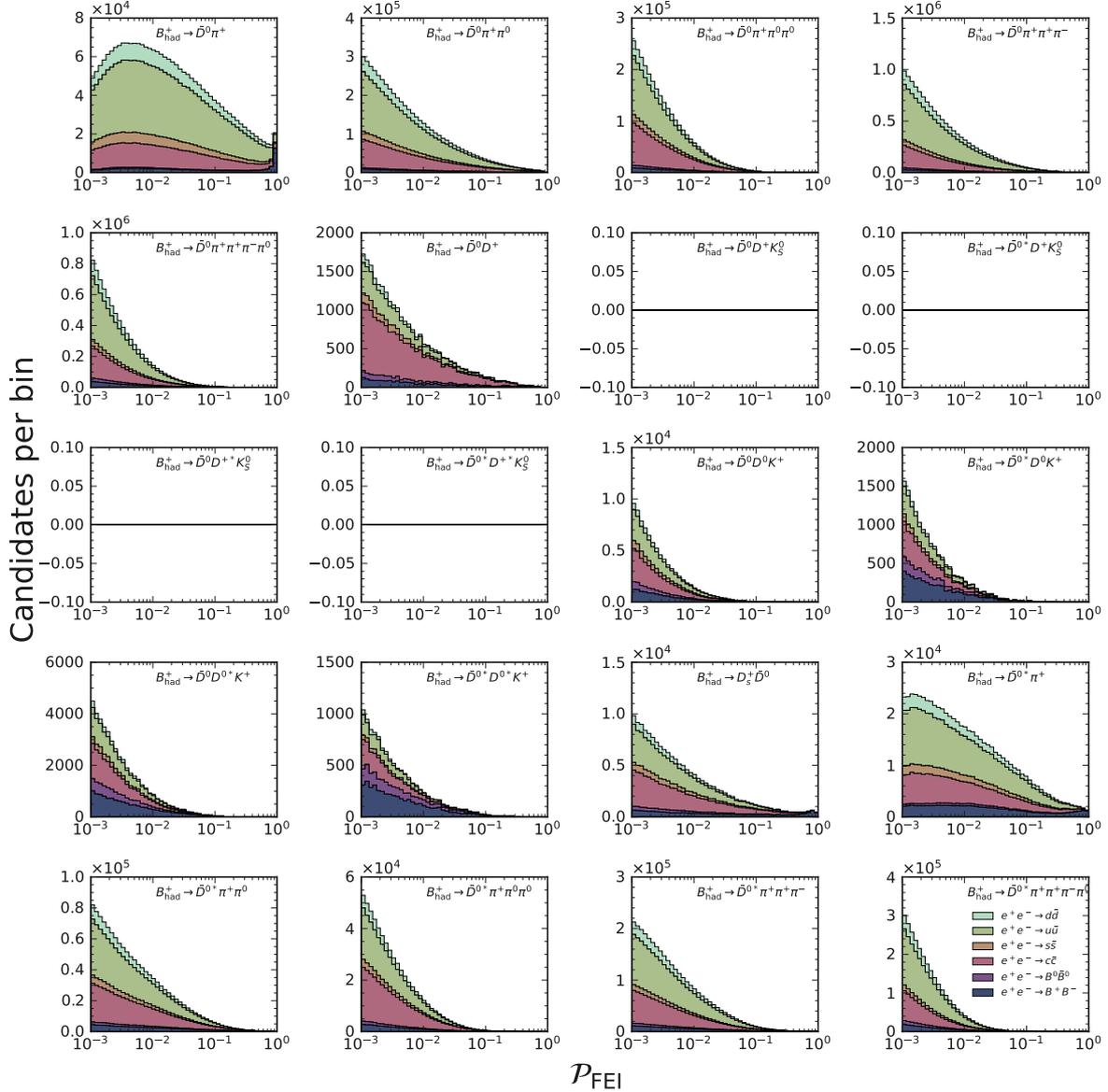
Figure D.1:   FEI signal probabilities for specific modes of B^+ reconstruction after requirements described in Section 6.3. This shows the signal probabilities for the first 20 B^+ modes in Table 6.2. Some Figures are empty because no modes are reconstructed in those channels (either due to insufficient sample size or no training available for said modes.) The legend, y and x axes are shared among all plots.
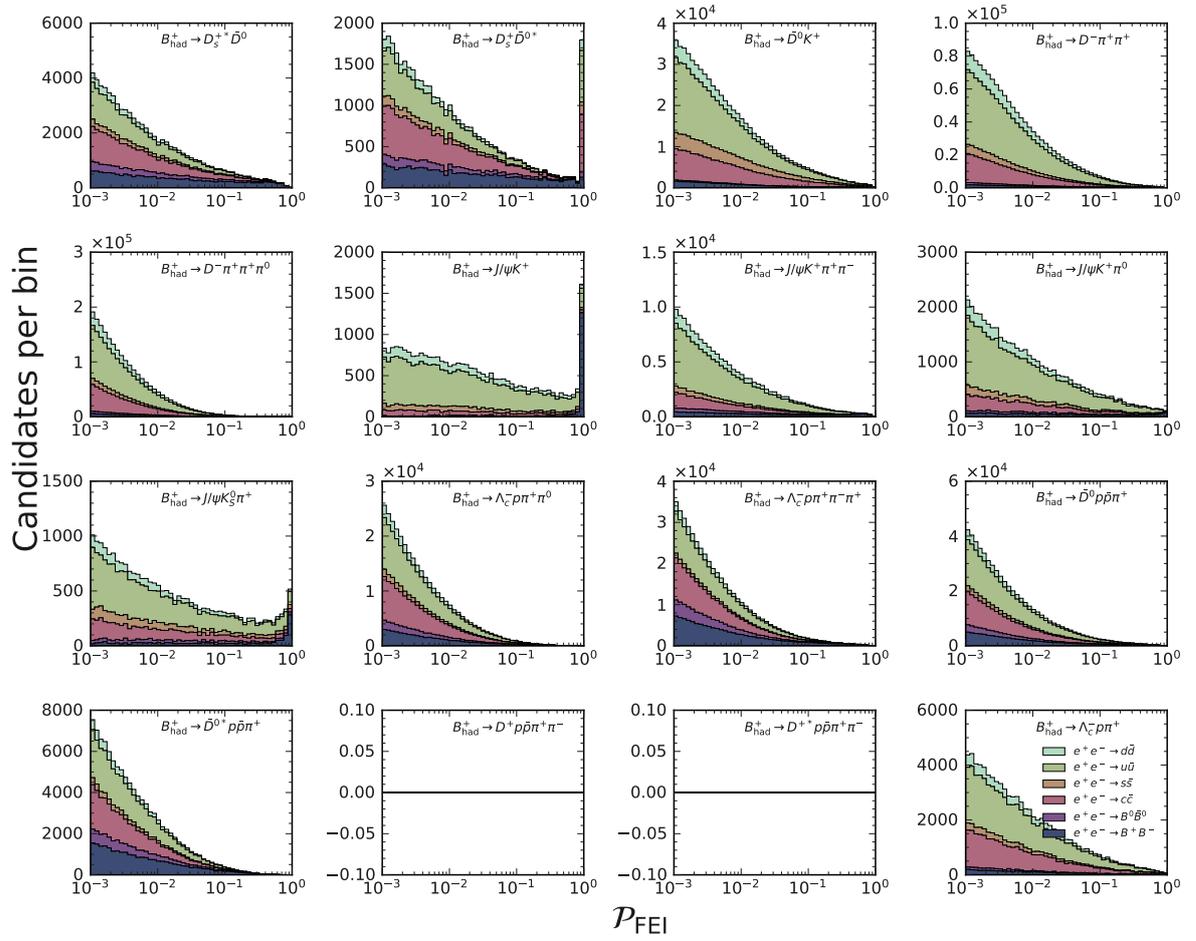
Figure D.1: FEI signal probabilities for specific modes of B^+ reconstruction after requirements described in Section 6.3. This shows the signal probabilities for the modes 20-36 B^+ modes in Table 6.2. Some Figures are empty because no modes are reconstructed in those channels (either due to insufficient sample size or no training available for said modes.) The legend, y and x axes are shared among all plots.
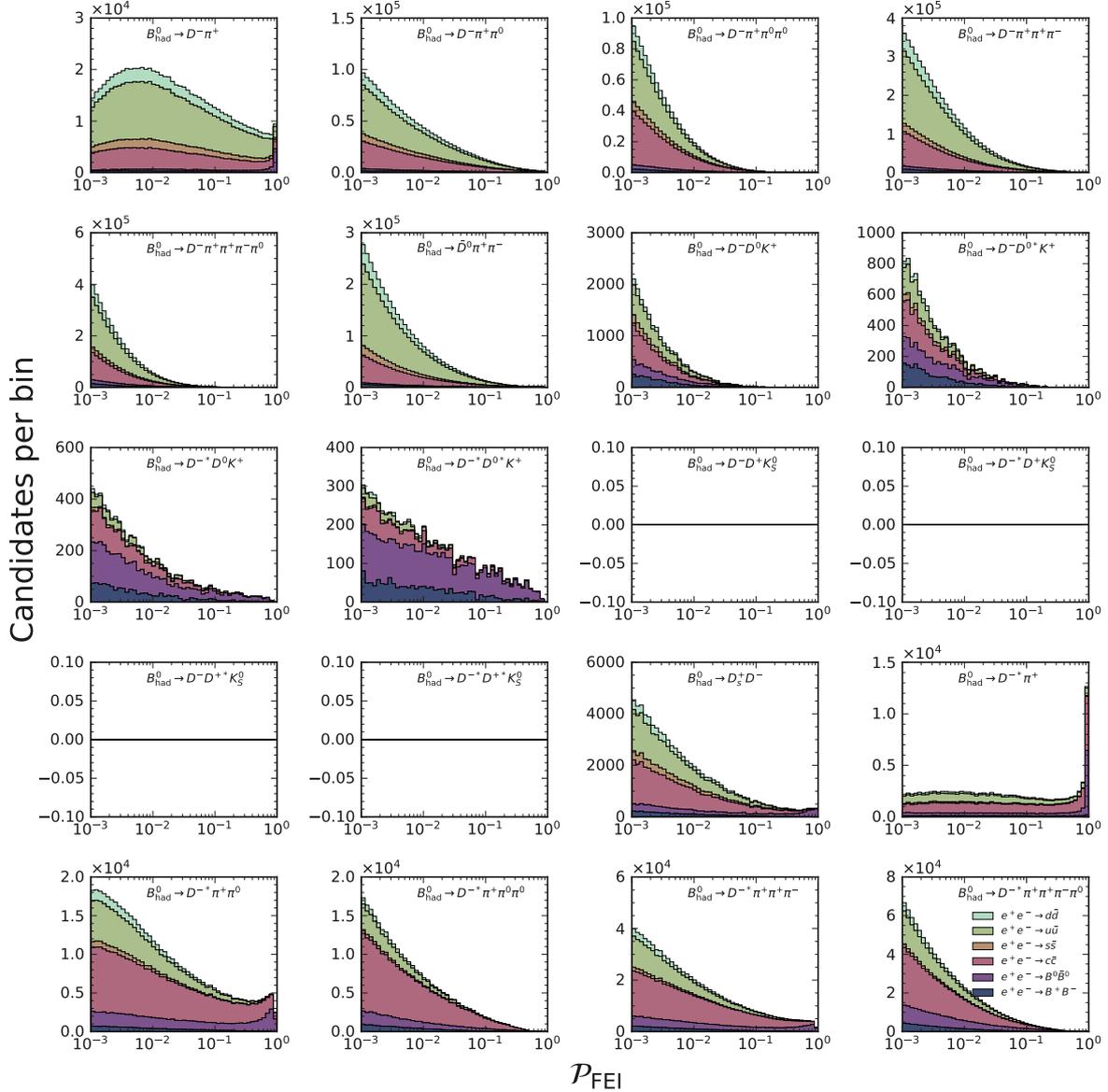
Figure D.2:   FEI signal probabilities for specific modes of B^0 reconstruction after requirements described in Section 6.3. This shows the signal probabilities for the first 20 B^0 modes in Table 6.2. Some Figures are empty because no modes are reconstructed in those channels (either due to insufficient sample size or no training available for said modes.) The legend, y and x axes are shared among all plots.
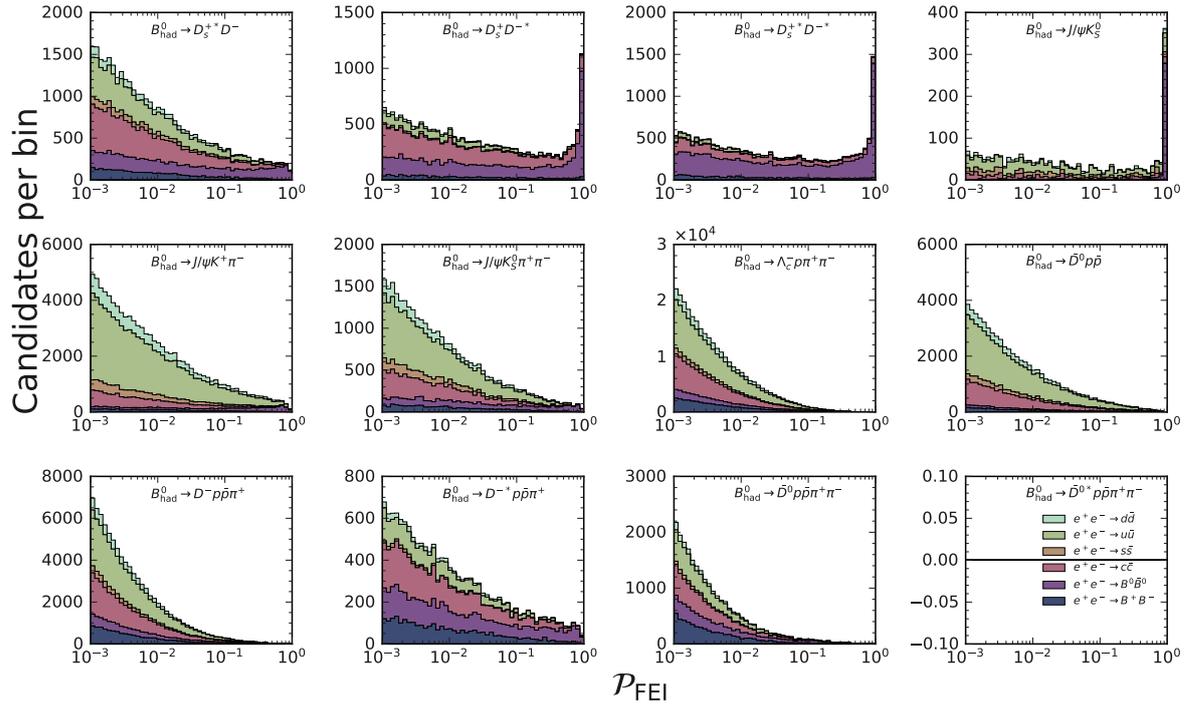
Figure D.2: FEI signal probabilities for specific modes of B^0 reconstruction after requirements described in Section 6.3. This shows the signal probabilities for the modes 20-32 B^0 modes in Table 6.2. Some Figures are empty because no modes are reconstructed in those channels (either due to insufficient sample size or no training available for said modes.) The legend, y and x axes are shared among all plots.

Appendix E

Pre-selection optimisation based on $\frac{S}{\sqrt{S+B}}$

To validate figure-of-merit FOM_2 defined in Equation (6.10), the same optimisation is performed based on the more-standardised FOM_1 defined in Equation (6.9). In general, this would be expected to provide an equivalent result. This is shown in Figure E.1. The Figure is equivalent to Figure 6.16 with a different figure-of-merit in mind. As can be seen, consistent conclusions with Section 6.5.1 can be drawn. Therefore, in this analysis FOM_2 is chosen due to the advantages discussed in the main body of the thesis.
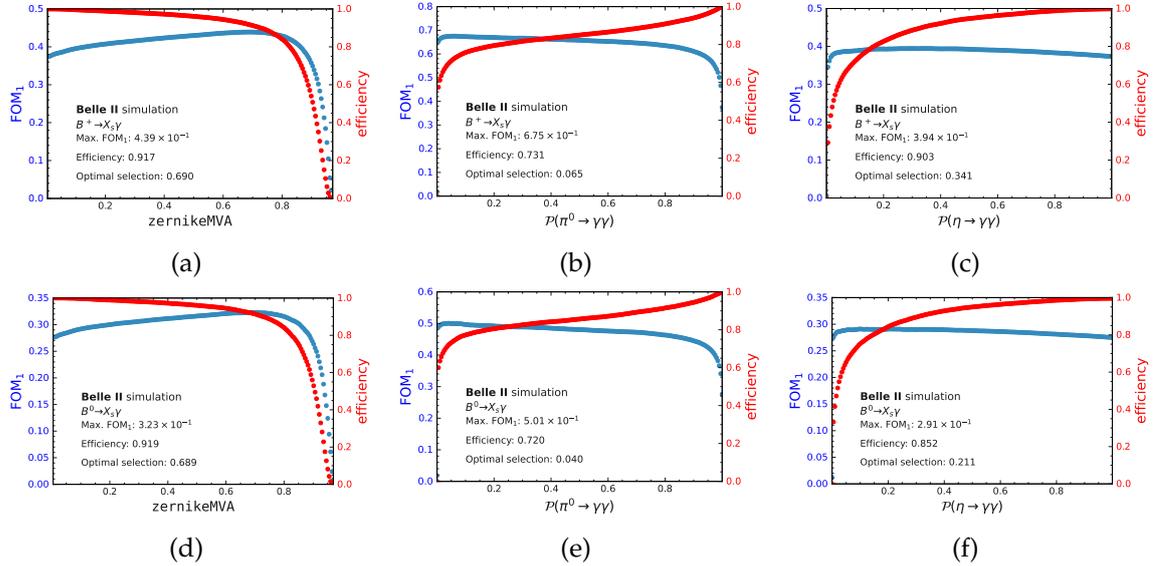
Figure E.1: Optimal selection calculation for observables described in Section 6.4 based on FOM_1 (see Equation (6.9)). For $B^+ \to X_s \gamma$ events the tests are shown in (a) to (c), and for $B^0 \to X_s \gamma$ in (d) to (f) The Figures show the efficiency and FOM_2 score calculated by scanning 200 thresholds of $\mathcal{P}(\pi^0 \to \gamma\gamma)$, $\mathcal{P}(\eta \to \gamma\gamma)$ and `zernikeMVA`. The maximum value of FOM_1, the corresponding threshold and efficiency are shown as well.

Appendix F

Bias testing for continuum BDT training features

In total, 75 features were tested for this analysis as potential candidates for the training features of a BDT to suppress $e^+e^- \rightarrow q\bar{q}$ events. These features are grouped into the following categories:

- Various thrust-based observables (Appendix F.1);

- Sphericity and aplanarity (Appendix F.2);

- Harmonic moments (Appendix F.3);

- Fox-Wolfram moments (Appendix F.4);

- Modified Fox-Wolfram moments (Appendix F.5);

- CLEO cones (Appendix F.6);

- Tag-B meson vertex observables (Appendix F.7);

- Flavour tagger output for the tag-B meson (Appendix F.8).

In this Appendix, they will be discussed and presented, and results for the **Test 1** will be shown for every tested observable. The tests in Appendices F.1 to F.8 provide a sample of 29 variables which then undergo the additional agreement test requirements, referred to as **Test 2** (see Appendix G).

F.1 Thrust-based observables

Thrust axis $\vec{T}$ is defined in terms of N momenta $\vec{p}_i$ ($i \in \{1, 2, .., N\}$). It is the unit vector, which maximises the projection of the $\sum_i \vec{p}_i$. The scalar observable known simply as *thrust* is then defined as [96]:

$$T = \frac{\sum_{i=1}^{N} |\vec{T} \cdot \vec{p}_i|}{\sum_{i=1}^{N} |\vec{p}_i|}.$$ (F.1)

'Thrust'-like observables can be utilised to distinguish $e^+e^- \to q\bar{q}$ and $B \to X_s\gamma$ events following the same argumentation as the one sketched in Figure 6.15. The decay particles of a B tend to be spherically distributed in the detector with a uniformly distributed $T \in (0, 1)$. For $q\bar{q}$ events, their decay particles tend to be directional, therefore, T tends to unity.

Based on the definitions of $\vec{T}$ more thrust-related distributions can be defined. Six thrust-related variables are tested in this analysis:

- $\cos\theta_{\mathrm{TB\wedge TO}}$: the cossine of the angle between the thrust axis of the tag candidate B meson (B meson decay particle momenta evaluated in the collision center-of-mass frame), and the thrust axis of all the other particles (Figure F.1a);

- $\cos\theta_{\mathrm{TB\wedge z}}$: the cossine of the angle between the thrust axis of the tag candidate B meson, and the z-axis of the detector (Figure F.1b);

- T_{B}: the thrust of the tag candidate B meson (Figure F.1c);

- T_{O}: the thrust of all particles *except* the tag candidate B meson (Figure F.1d);

- T: the thrust of all particles in the event (Figure F.1e);

- $\cos\theta_{\mathrm{T}}$: the polar angle component of $\vec{T}$ (Figure F.1f).

The results of **Test 1** for these variables are shown in Figures F.1a to F.1f. Unsurprisingly, variables that include momenta information of the X_s system show strong bias of the E_γ^B spectrum. On the other hand, all tag-side variables are suitable and minimally-biasing.
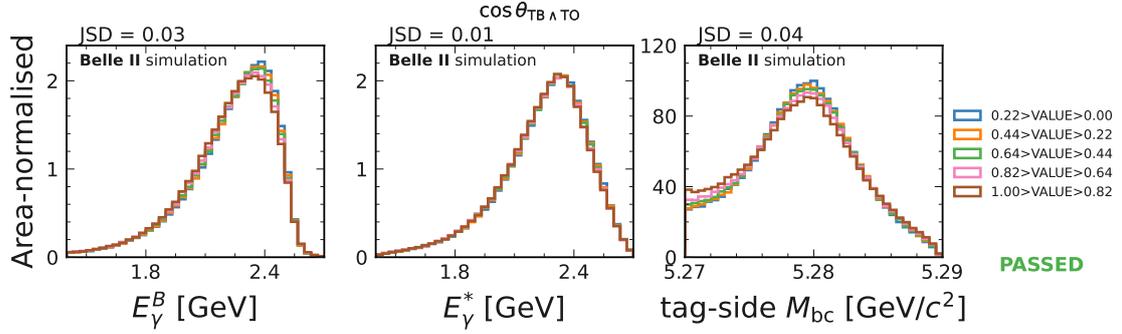
(a)
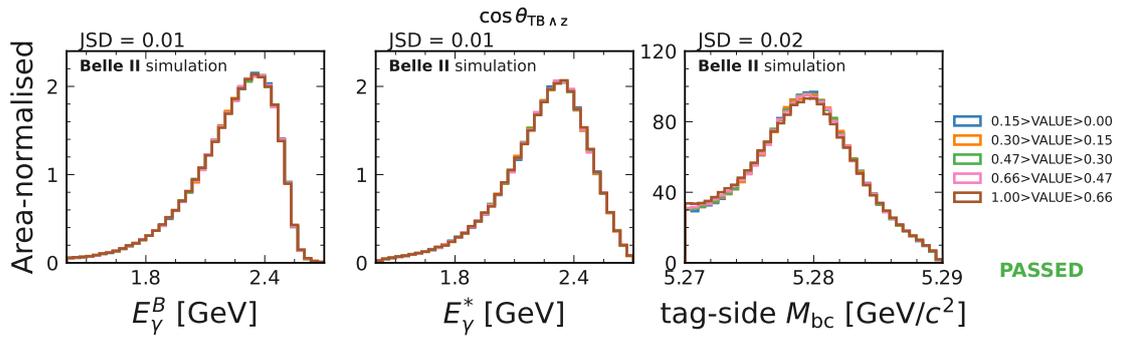
(b)
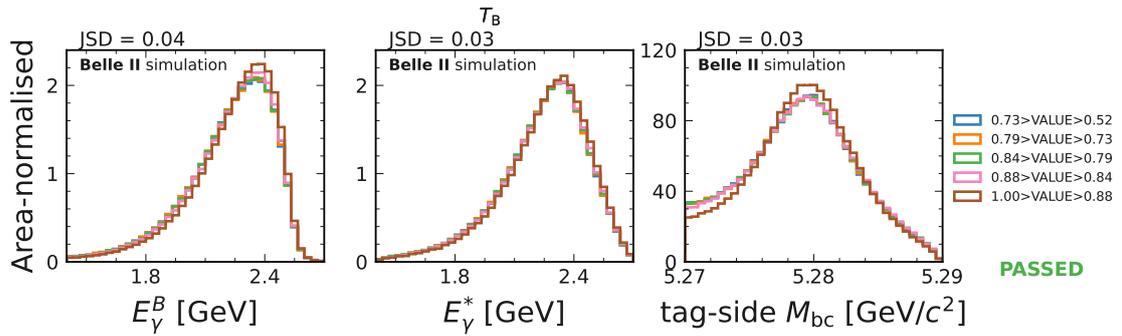
(c)
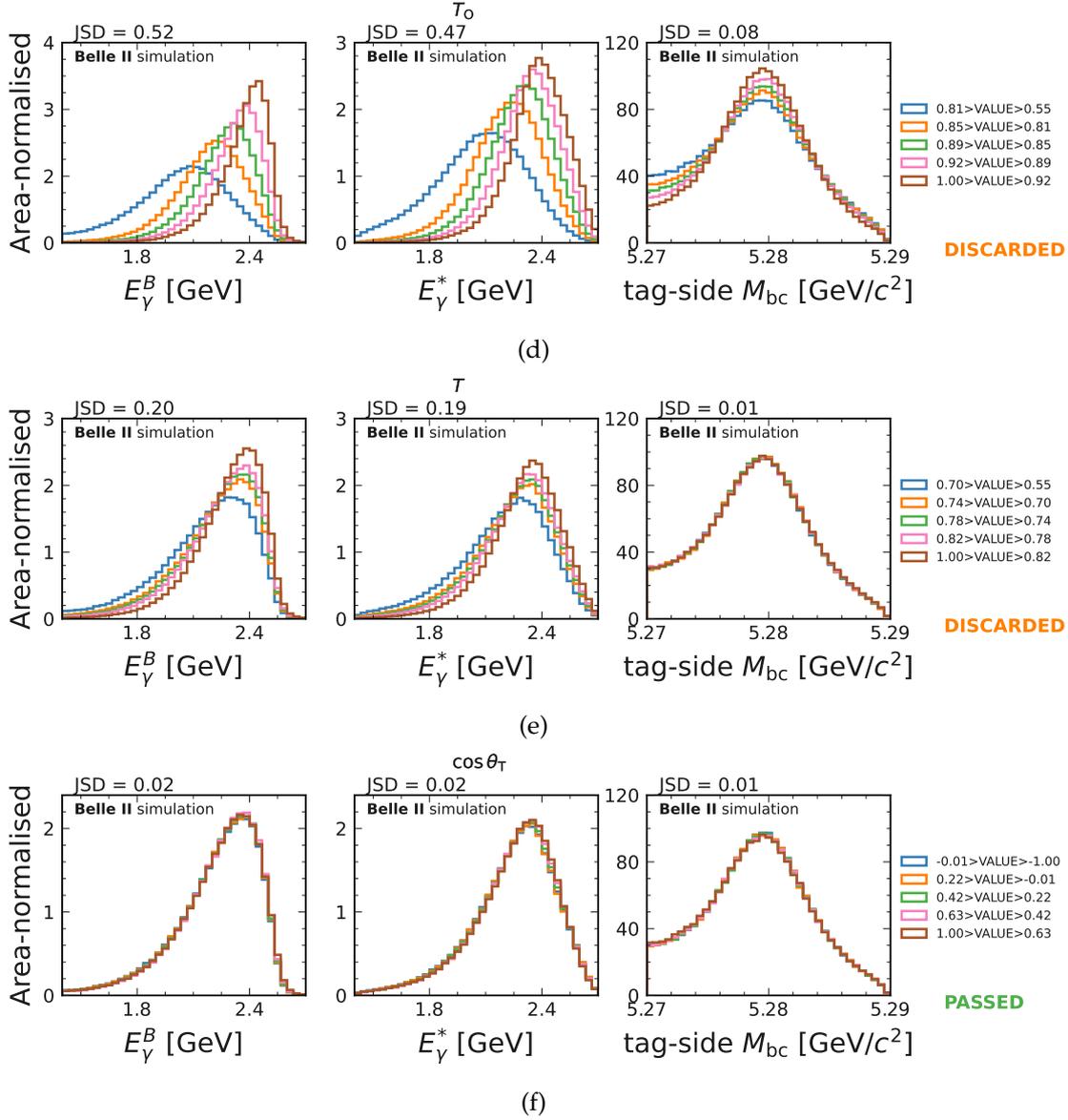
Figure F.1:   The bias-test on E_γ^B, E_γ^* and M_{bc} for thrust-based observables. The test is performed based on **Test 1** strategy, defined in Section 6.5.2. Variable definitions are given in text. The Jensen Shannon distance, as introduced in Equation (6.11), is given for each distribution.

F.2 Sphericity and aplanarity

Sphericity matrix is defined using a collection of momenta $\vec{p}_i$, as [96]:

$$S^{\alpha,\beta} = \frac{\sum_{i=1}^{N} p_i^{\alpha} p_i^{\beta}}{\sum_{i=1}^{N} |\vec{p}_i|^2}, \tag{F.2}$$

where $\alpha, \beta \in \{x, y, z\}$. For an isotropic distribution its three eigenvalues, λ_{1-3}, are expected to be of similar size. On the other hand, collimated distributions tend to have one of the values significantly smaller. Therefore two values are tested in this analysis as continuum suppression features:

- sphericity $\equiv \frac{3}{2}(\lambda_2 + \lambda_3) \in (0, 1)$ (Figure F.2a);

- aplanarity $\equiv \frac{3}{2}\lambda_3 \in (0, 1)$ (Figure F.2b).

In these definitions, $\lambda_{3(2)}$ is the (second-)smallest eigenvalue of the sphericity matrix. A spherical event will have a sphericity close to 1 and aplanarity close to 1/2. As the definitions of the sphericity matrix include momentum of the X_s system, these variables turn out to introduce a significant bias to the photon energy spectrum and are therefore not used.
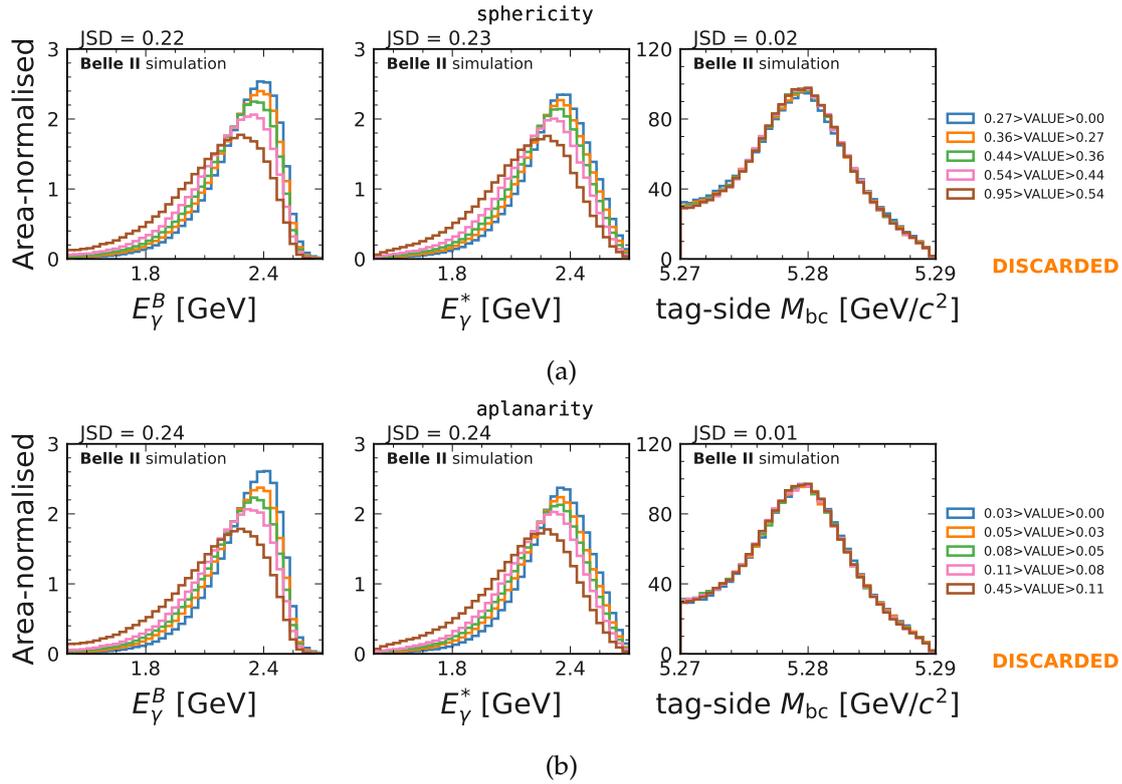
Figure F.2: The bias-test on E_γ^B, E_γ^* and M_{bc} for sphericity and aplanarity. The test is performed based on **Test 1** strategy, defined in Section 6.5.2. Variable definitions are given in the text. The Jensen Shannon distance, as introduced in Equation (6.11), is given for each distribution.

F.3 Harmonic moments

Harmonic moments (or multipole moments) are defined for a collection of momenta $\vec{p}_i$, with respect to an axis $\vec{A}$ as [165]:

$$B_l \equiv \sum_i \frac{|\vec{p}_i|}{\sqrt{s}} P_l(\cos \alpha_i), \qquad (F.3)$$

where $\sqrt{s}$ is the collision center-of-mass energy, α_i is the angle between the particles in the event and $\vec{A}$, and P_l are Legendre polynomials. In this analysis, the first 5 harmonic moments ($l \in \{1, 2, 3, 4, 5\}$) with respect to the thrust axis (see Appendix F.1) are considered. The distributions for **Test 1** are shown in Figures F.3a to F.3e.
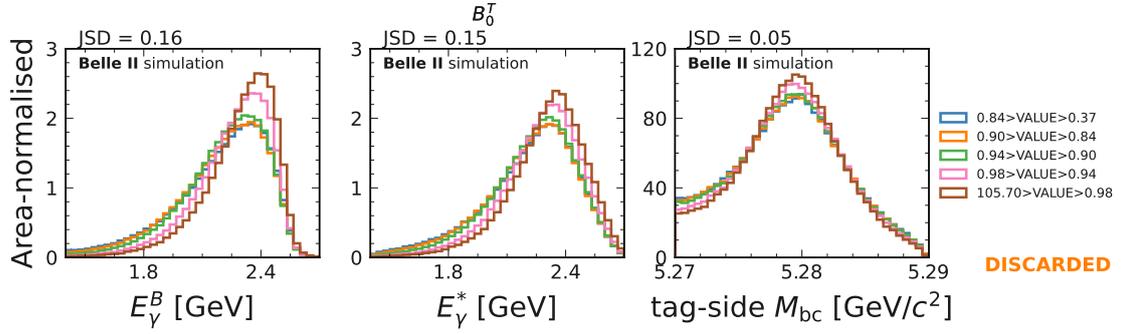
(a)
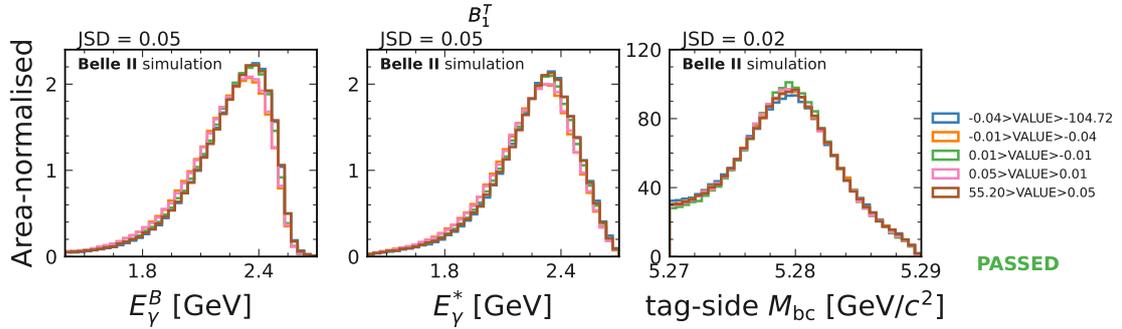
(b)
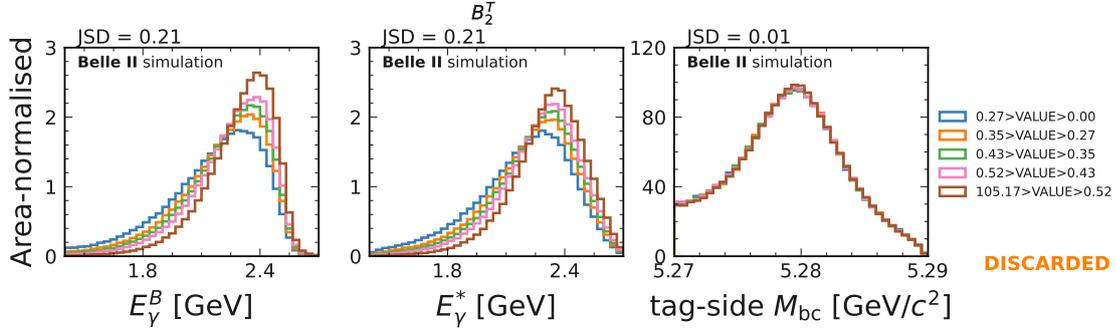
(c)
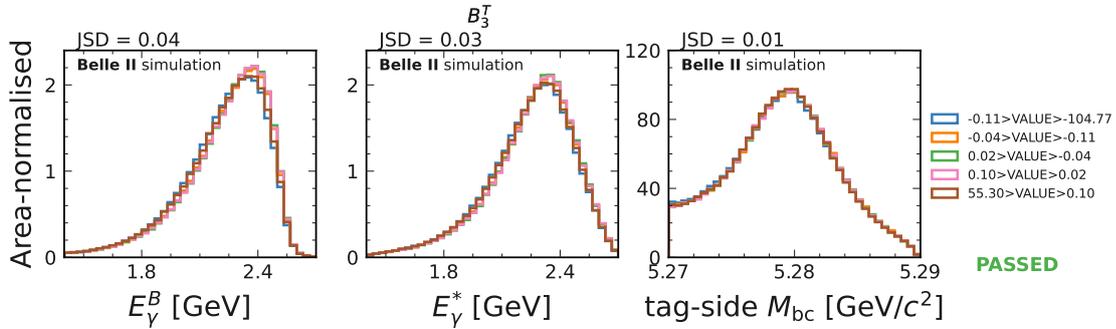
(d)
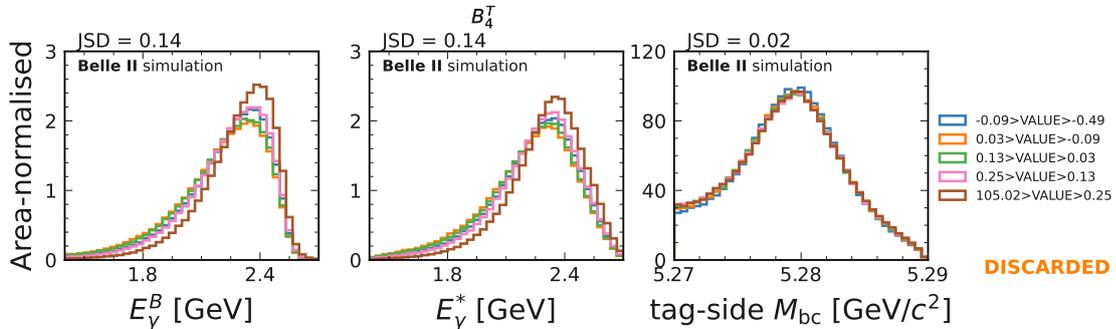
(e)

Figure F.3:   The bias-test on E_γ^B, E_γ^* and M_{bc} for harmonic thrust observables. The test is performed based on **Test 1** strategy, defined in Section 6.5.2. Variable definitions are given in the text. The Jensen Shannon distance as introduced in Equation (6.11) is given for each distribution.

F.4 Fox-Wolfram moments

Fox-Wolfram moments can be used to parametrise the momentum and energy flow distributions and were introduced in Ref. [165]. They are defined for a collection of momentum, $\vec{p}_i$, as:

$$H_l \equiv \sum_{i,j} |\vec{p}_i||\vec{p}_j|P_l(\cos\theta_{ij}), \tag{F.4}$$

where θ_{ij} is the angle between p_i and p_j, and P_l are the Legendre polynomials. Normalised Fox-Wolfram moments,

$$R_l \equiv \frac{H_l}{H_0}, \tag{F.5}$$

are often used, as strongly collimated sets of momenta tend to zero for l_{odd} and to one for l_{even} [96]. Conventionally, R_{1-4} are considered. In addition, in this study, we calculate R_2^B, which only includes momenta of particles used for reconstruction of the tag-B meson candidate. The distributions for **Test 1** are shown in Figures F.4a to F.4e. All these observables prove to be strongly correlated with the photon energy spectrum.

(a)

(b)
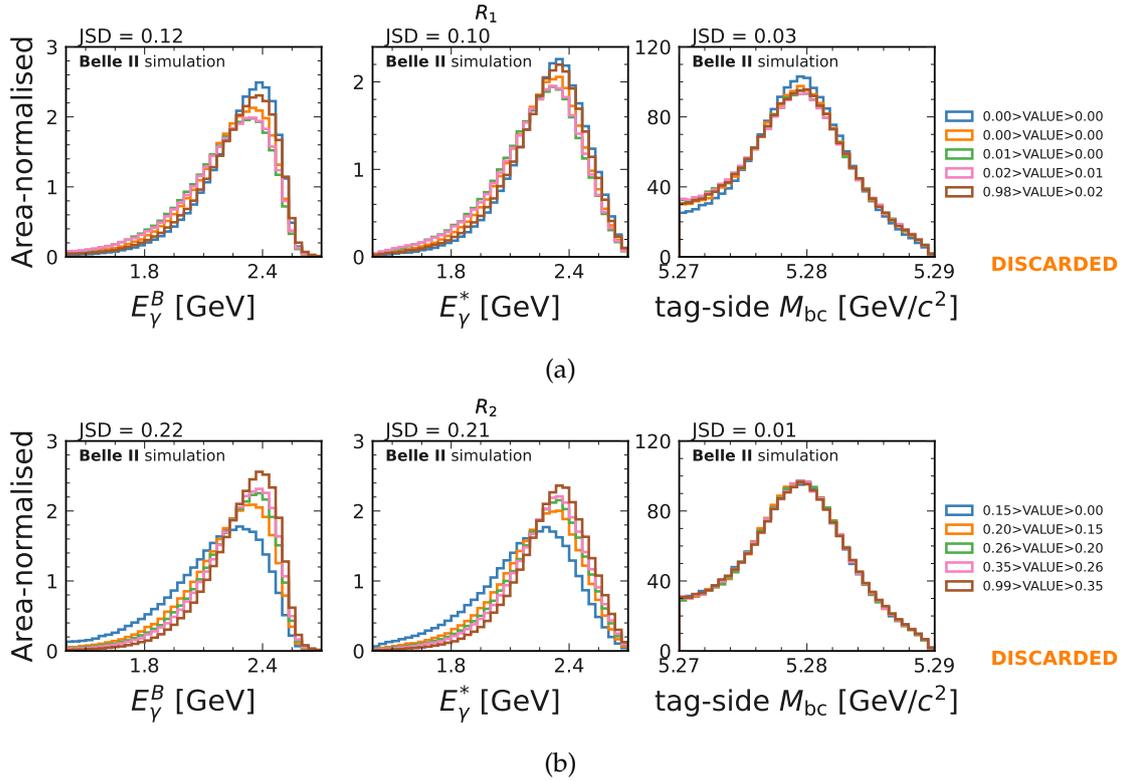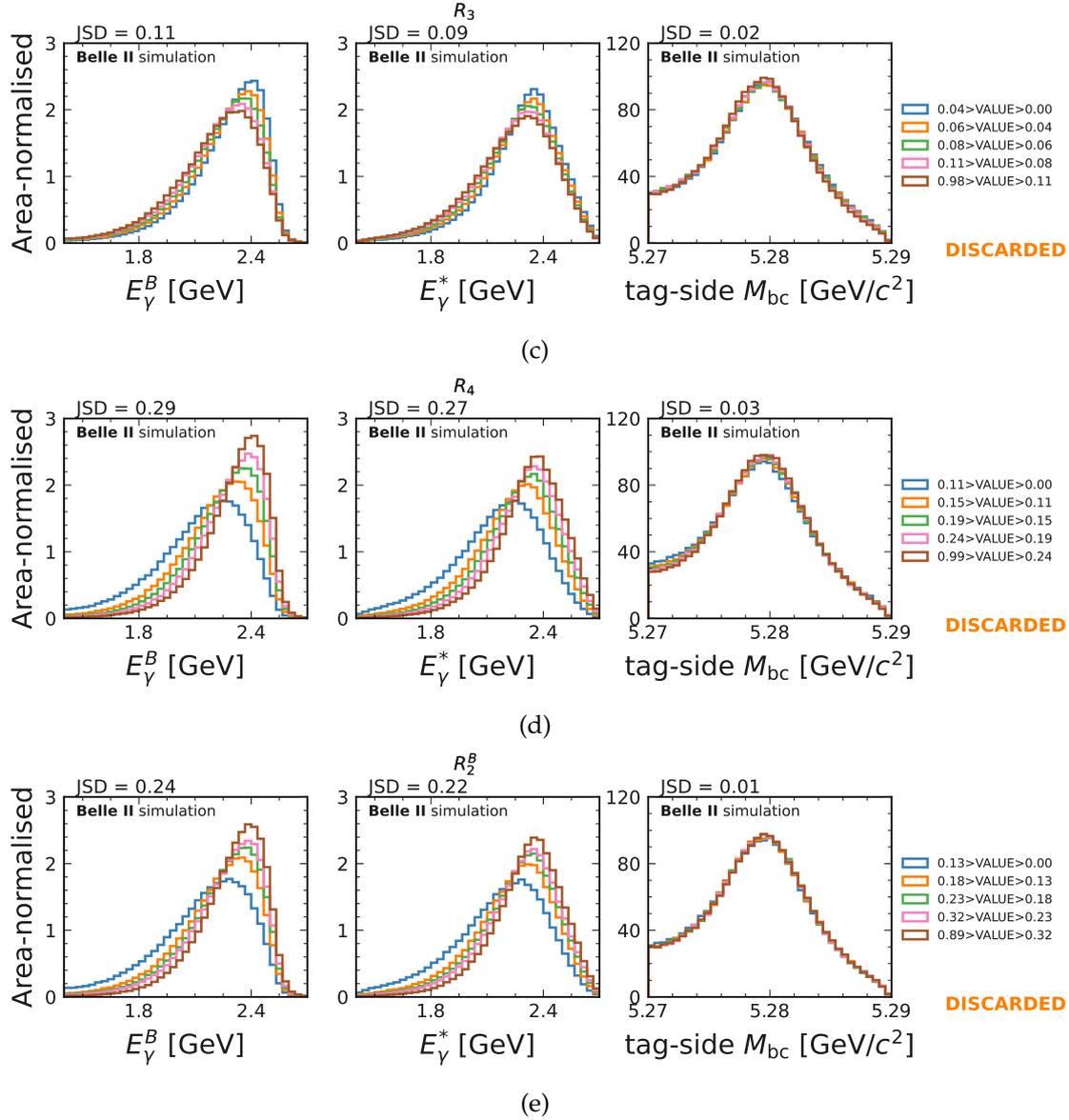
Figure F.4: The bias-test on E_γ^B, E_γ^* and M_{bc} for Fox-Wolfram moments. The test is performed based on **Test 1** strategy, defined in Section 6.5.2. Variable definitions are given in the text. The Jensen Shannon distance as introduced in Equation (6.11) is given for each distribution.
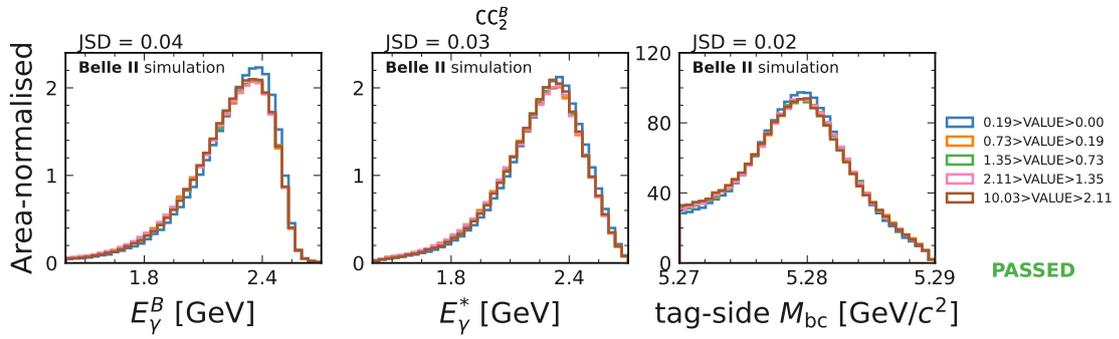
F.5 Modified Fox-Wolfram moments

The Fox-Wolfram moments, introduced in Appendix F.5 were shown to bias the photon-energy spectrum. This issue, and the fact that a better separation can be calculated if considering only certain subsets of particle momenta and energies in the event, motivated the introduction of modified Fox-Wolfram moments, also known as Kakuno-Super-Fox-Wolfram moments. They were originally introduced by the Belle collaboration in 2003 [166] for charmless B decay studies. Modified Fox-Wolfram moments are defined in terms of a set of momenta $\vec{p}_i^s$ that belong to a B meson candidate. Another subset of momenta for other particles not used in the reconstruction $\vec{p}_{jx}^o$, where x represents a category. The categories in question are charged particles (c), neutral particles (n), and missing momentum (m). With this in mind, linear modified Fox-Wolfram moments

$$H_{x,l}^{so} = \frac{1}{N} \sum_i \sum_{jx} Q_{ij(x)}^l |p_{jx}^o| P_l(cos\theta_{i,jx}), \tag{F.6}$$

and quadratic modified Fox-Wolfram moments

$$R_l^{oo} = \frac{1}{N^2} \sum_i \sum_j Q_{ij}^l |p_i^o||p_j^o| P_l(\cos\theta_{i,j}), \tag{F.7}$$

are defined. Here,

- $N \equiv 2 \cdot (\sqrt{s} - E_B^*)$ is a normalisation factor defined in terms of the collisions centre-of-mass energy and the B meson energy in the center-of-mass frame;

- $\sum_i$ runs over all particles used in the reconstruction of the B candidate;

- $\sum_{j(x)}$ runs over all other (or a category $x \in \{c, n, m\}$ of) particles in the event;

- Q_{ij}^l is the product of the charges of candidates corresponding to i and j if l is odd, otherwise 1;

- $\theta_{i,j(x)}$ is the angle between $\vec{p}_i$ and $\vec{p}_{j(x)}$;

- P_l are the Legendre polynomials.

As the charge for $x \in \{n, m\}$ is considered 0, they do not have odd H_ℓ^{so} moments. Two additional variables are included: the transverse momentum of the event, $\vec{p}_T$, and the missing mass of the event, m_{miss}^2, yielding a total of 18 variables suitable for continuum suppression. In this analysis, they are calculated in terms of the tag-B meson candidate and shown in Figures F.5a to F.5r. Six[1] moments pass the requirements of **Test 1**.

---

[1]Due to an unfortunate mistake in the analysis code, $H_{c,1}^{so}$ and $H_{c,3}^{so}$ were not included in the BDT training, although they did pass the **Test 1** selection criteria. Therefore, strictly speaking only 4 variables passed the human-error selection.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

(k)

(l)

(m)

(n)

(o)

(p)
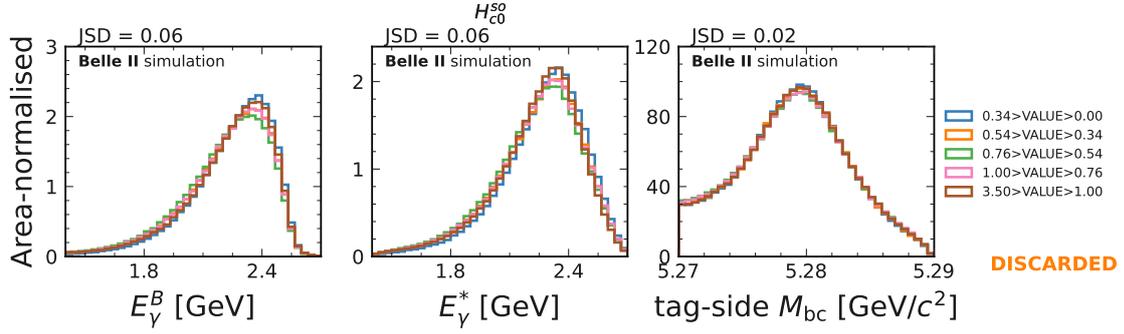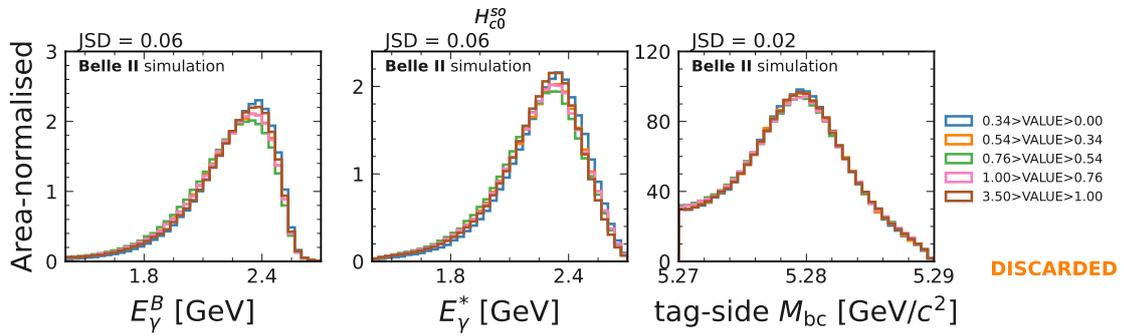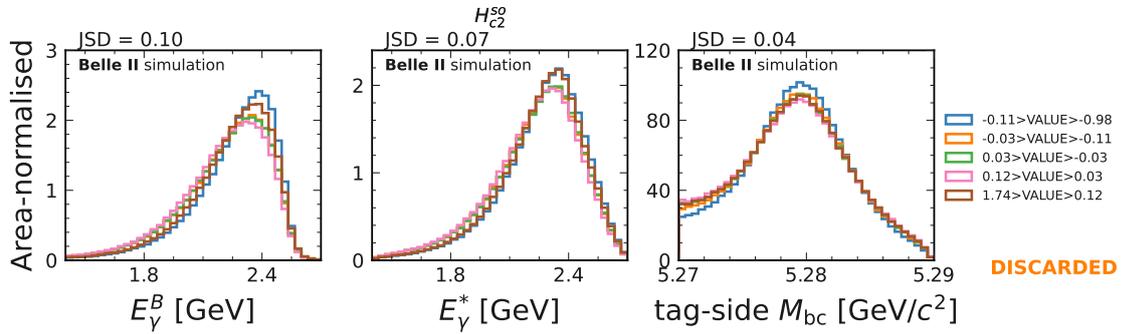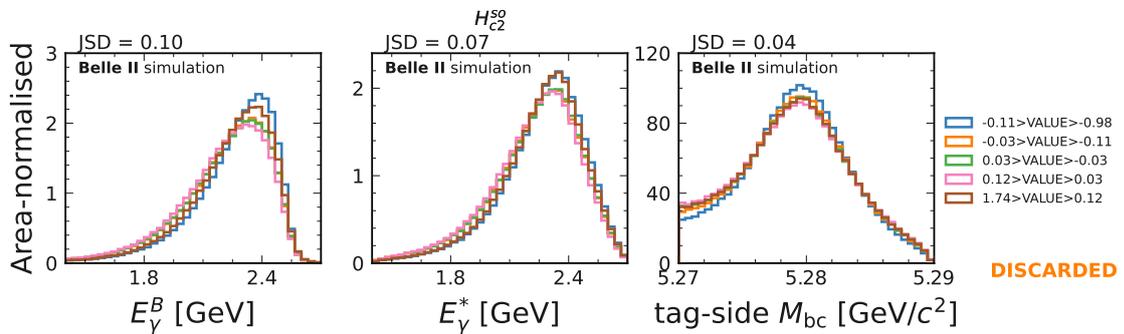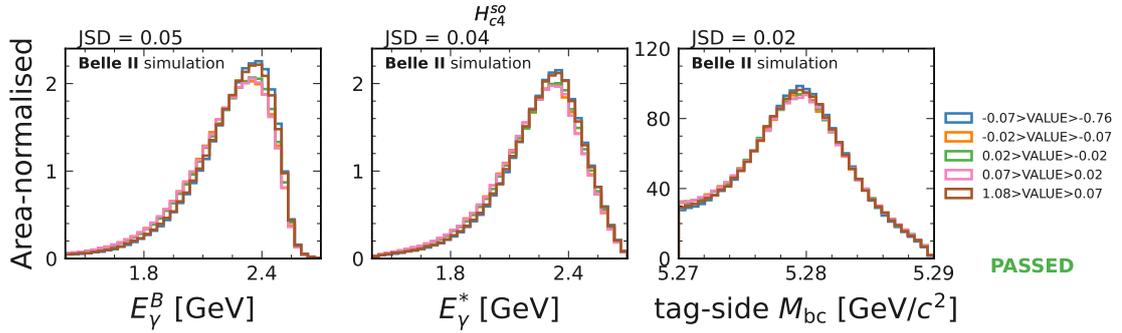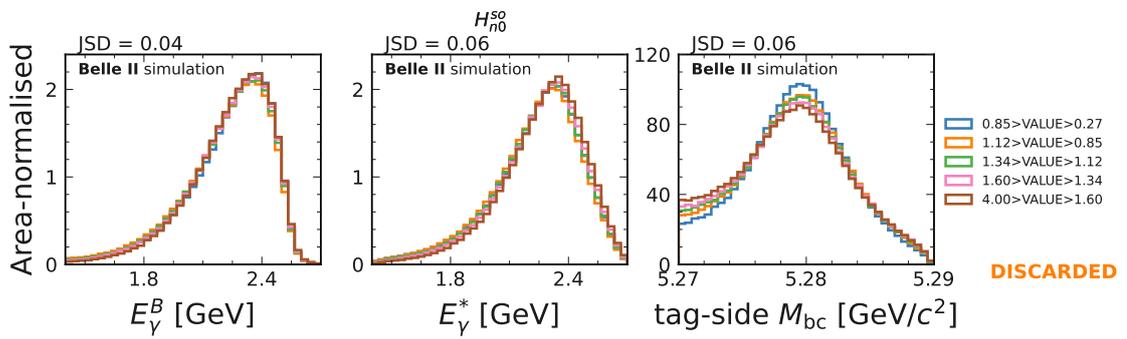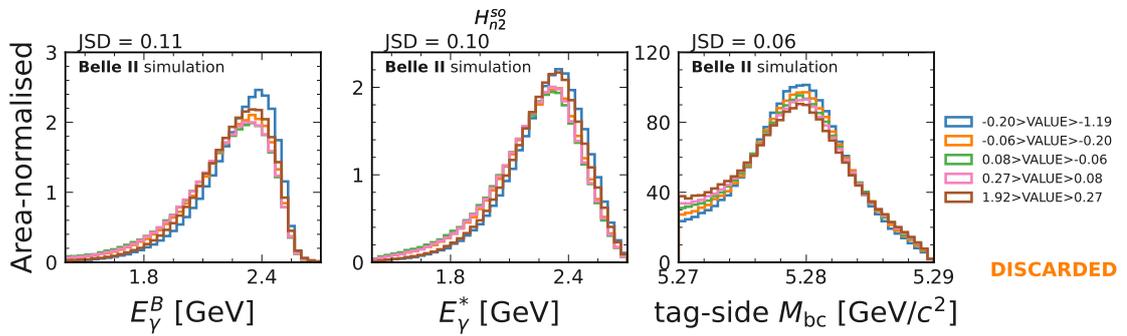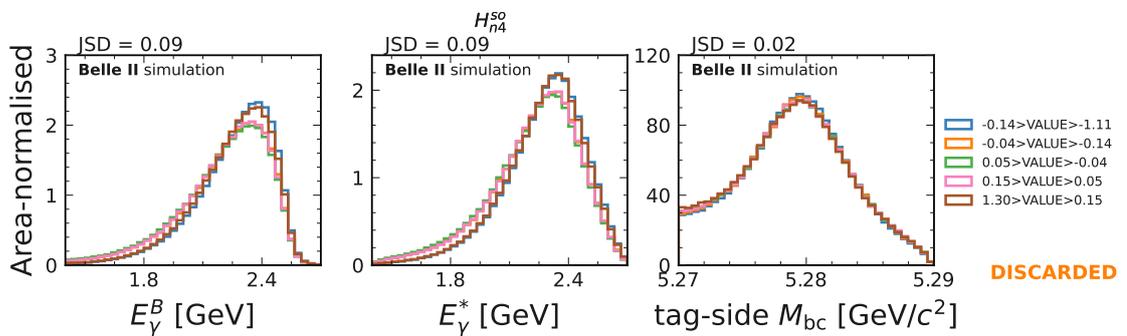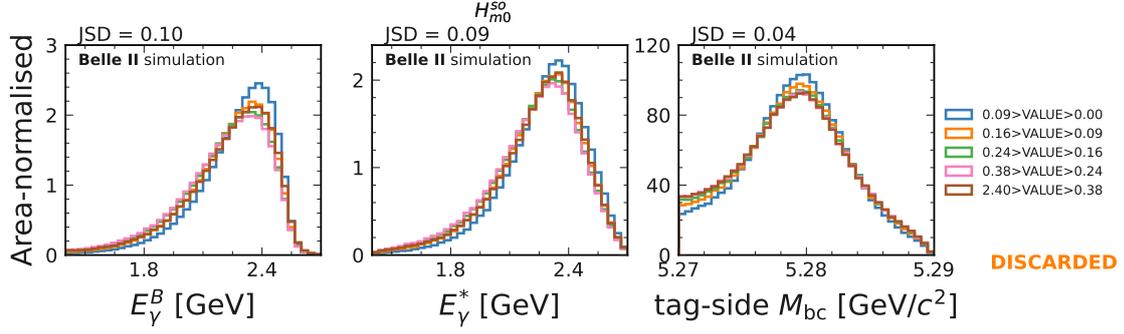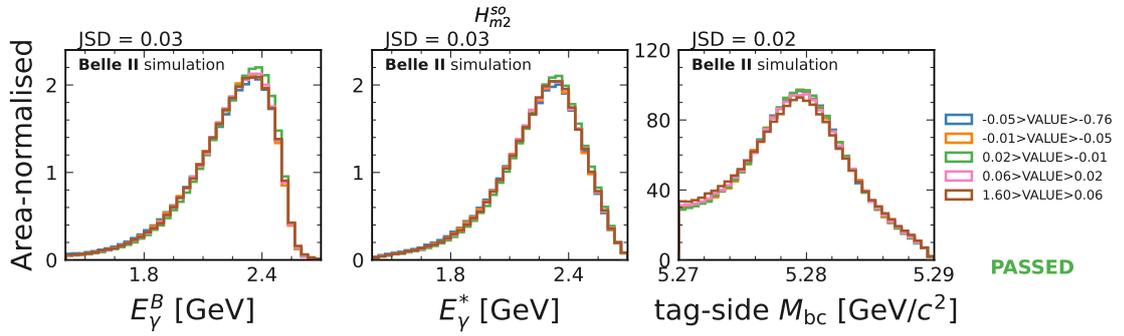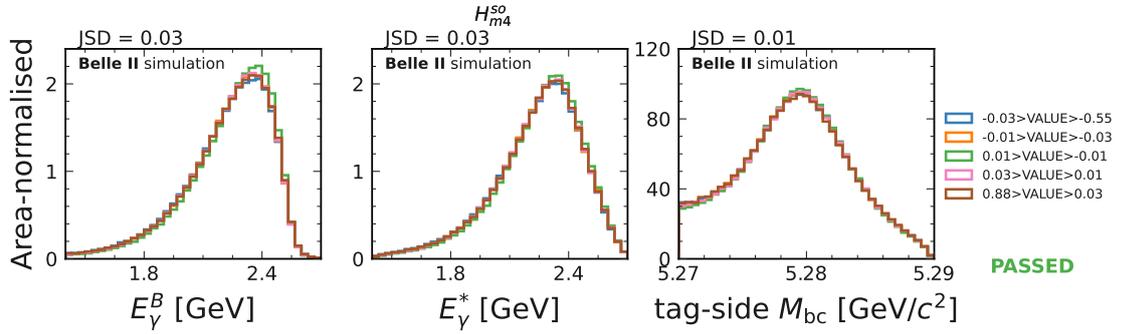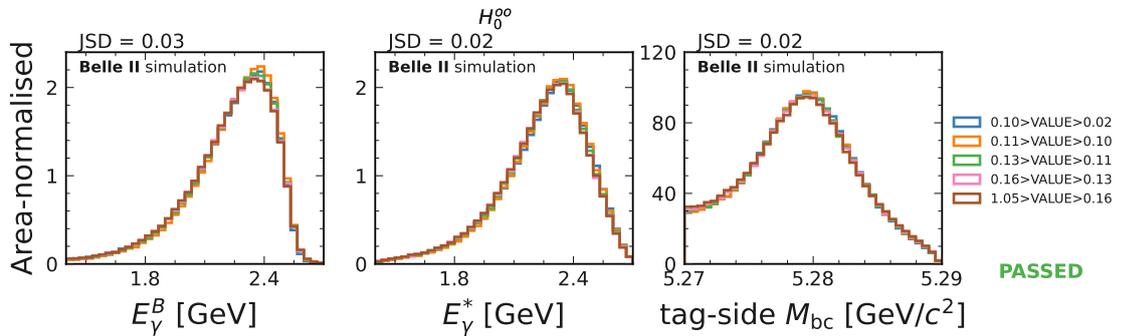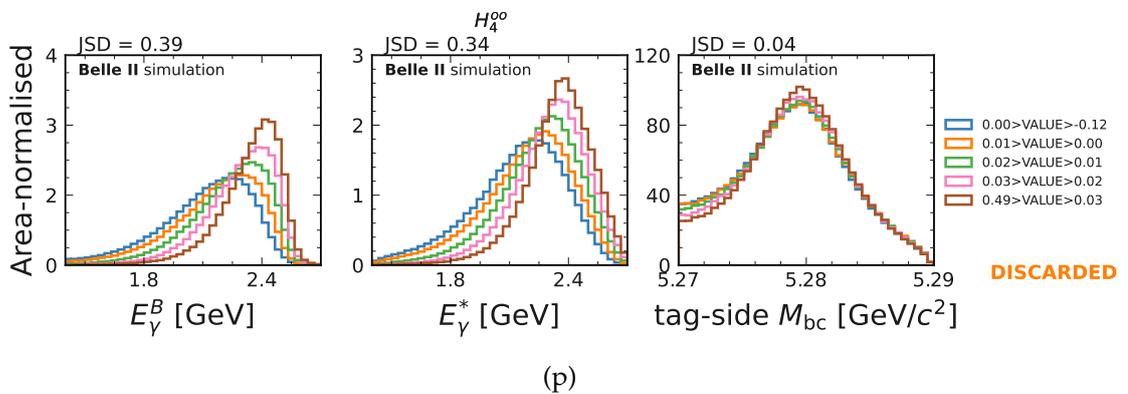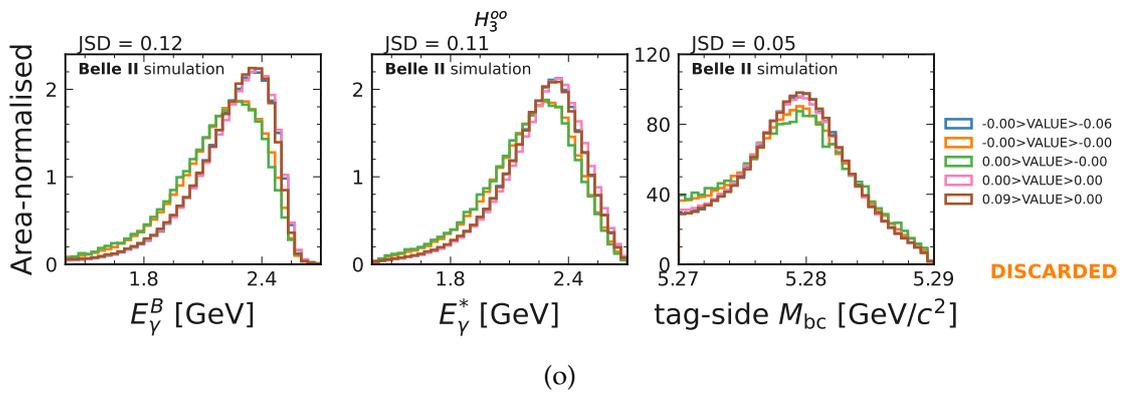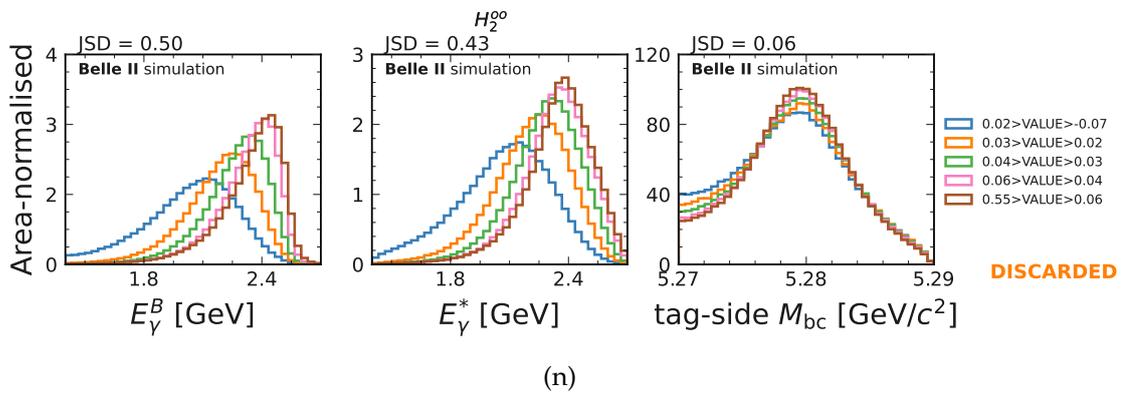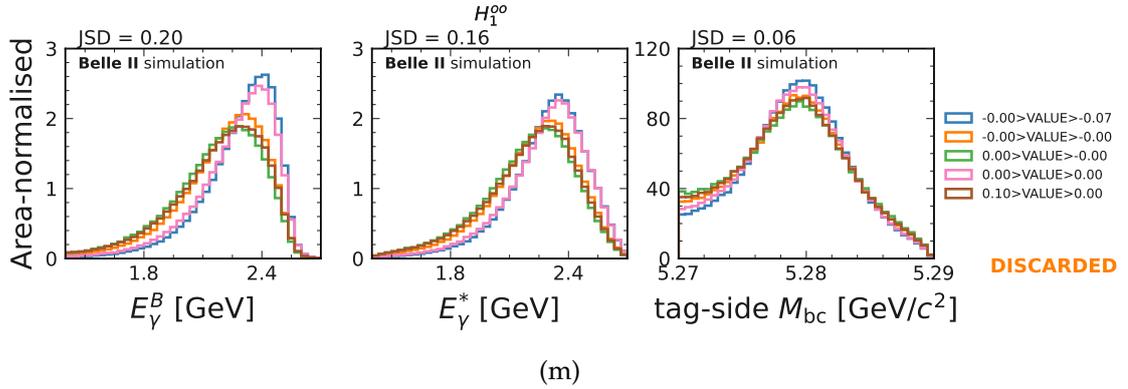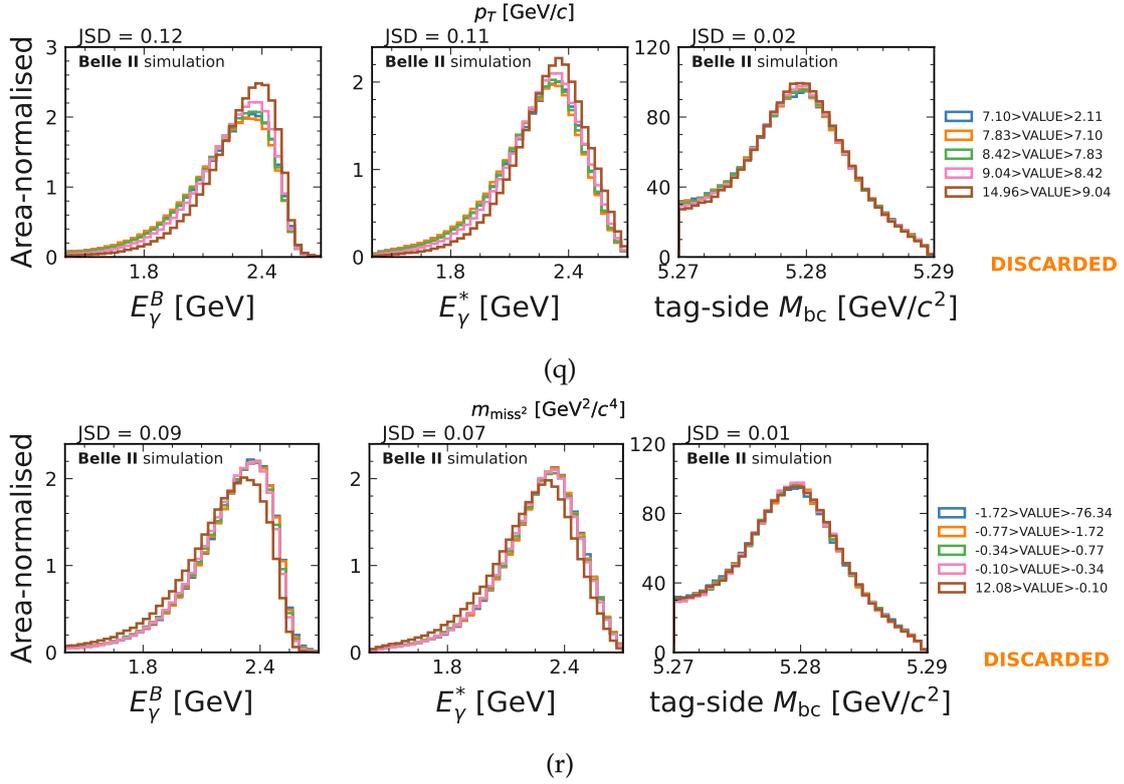
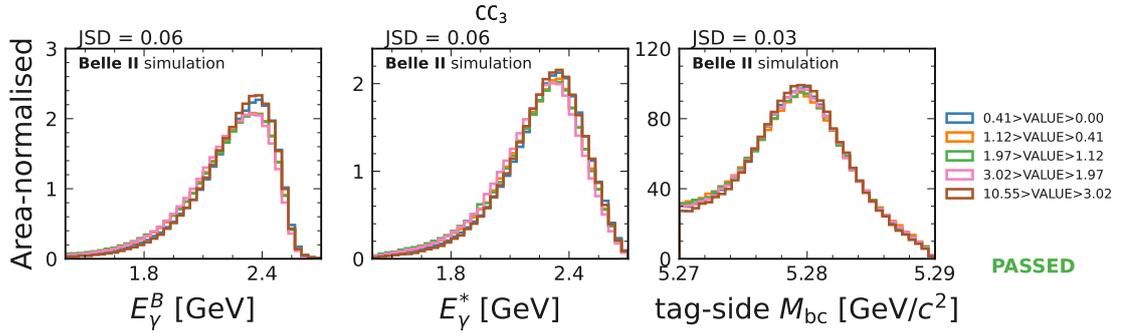Figure F.5: The bias-test on E_γ^B, E_γ^* and M_{bc} for modified Fox-Wolfram moments. The test is performed based on **Test 1** strategy, defined in Section 6.5.2. Variable definitions are given in the text. The Jensen Shannon distance as introduced in Equation (6.11) is given for each distribution.

F.6 CLEO cones

CLEO cones, used by the CLEO collaboration and introduced in Ref. [167], are defined accordingly to momentum flow around an axis $\vec{A}$. Nine cones are defined, with opening angles between 10 and 90 degrees. The energy flux in all of them is measured. Spherical decays tend to have their energy distributed more equally amongst the cones, whereas collinear decays have most of the energy within the first few. In this analysis, 18 CLEO cones are tested: 9 defined with respect to the event thrust axis, CC_i, and 9 defined with respect to the tag-side B meson thrust axis, CC_i^B. The results for **Test 1** are shown in Figures F.6a to F.6r. Outermost cones, generally, show a higher degree of correlation with E_γ^B, E_γ^* and M_{bc}, and only 6 CLEO cones with an opening angle of up to 30 degrees pass the test.

(a)

(b)

(c)

(d)

(e)
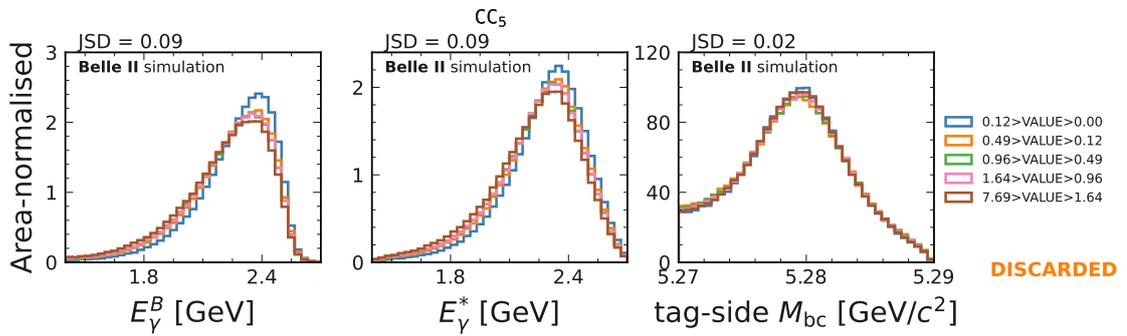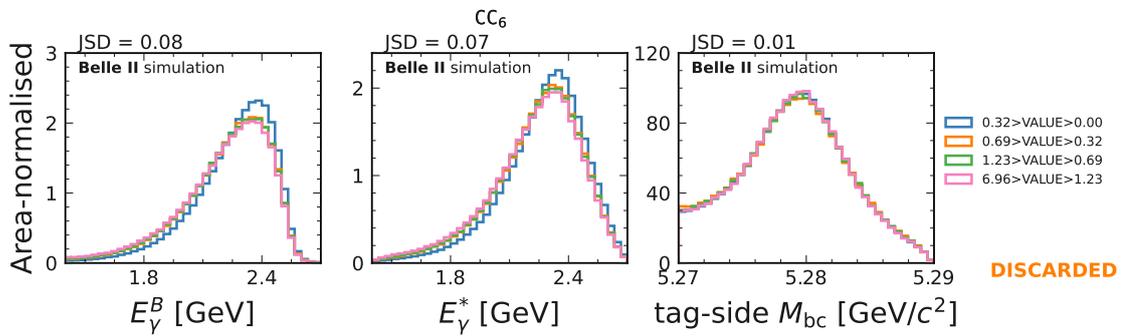

(f)

(g)

(h)
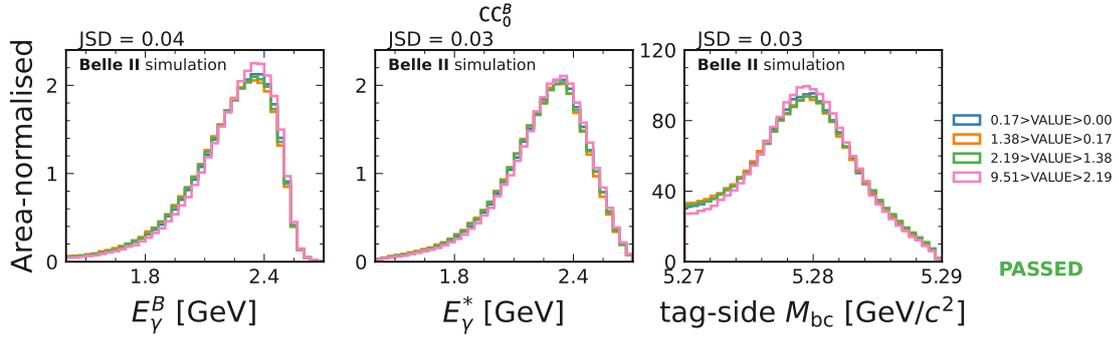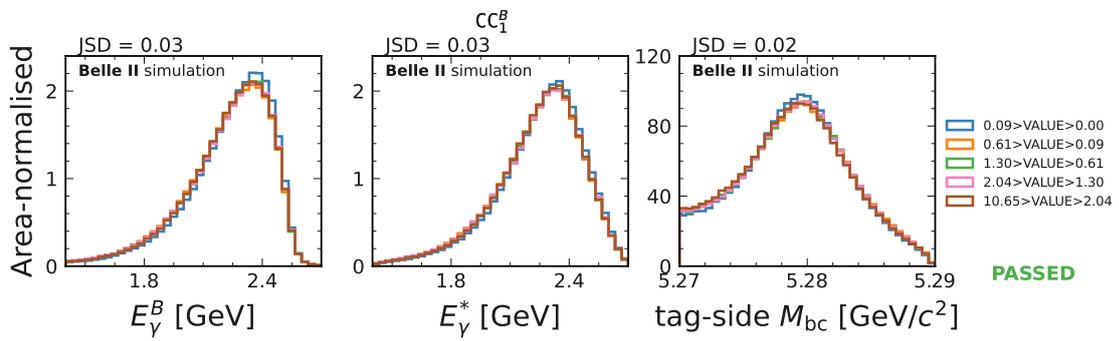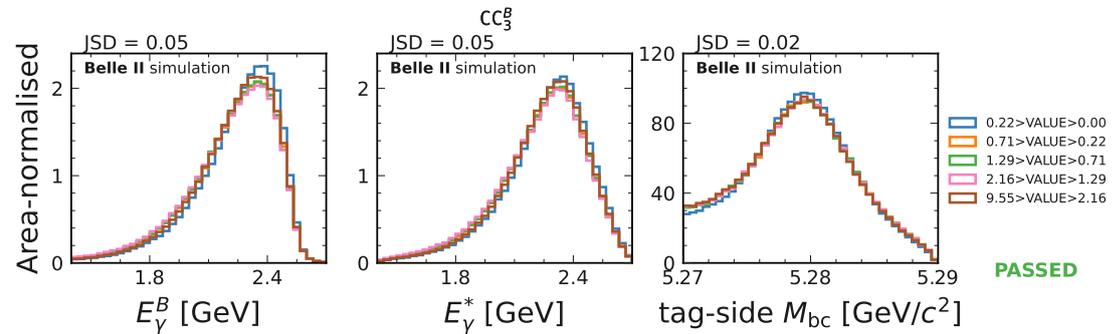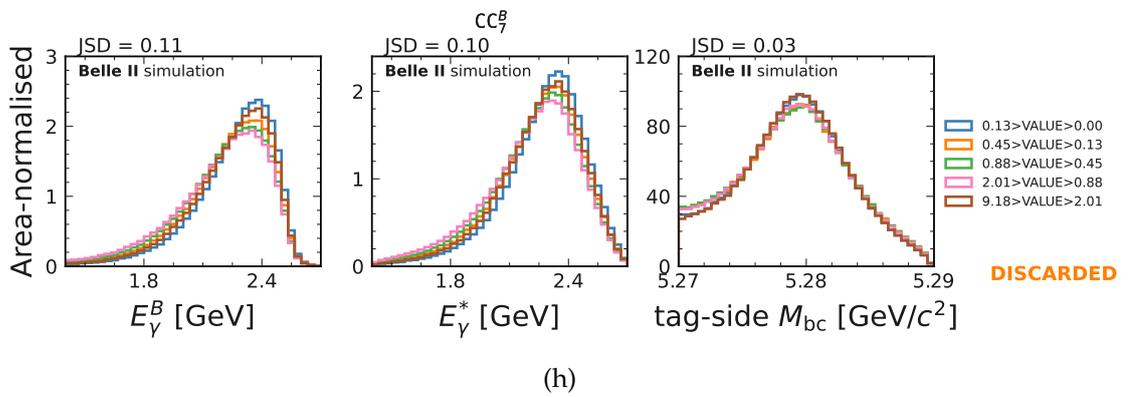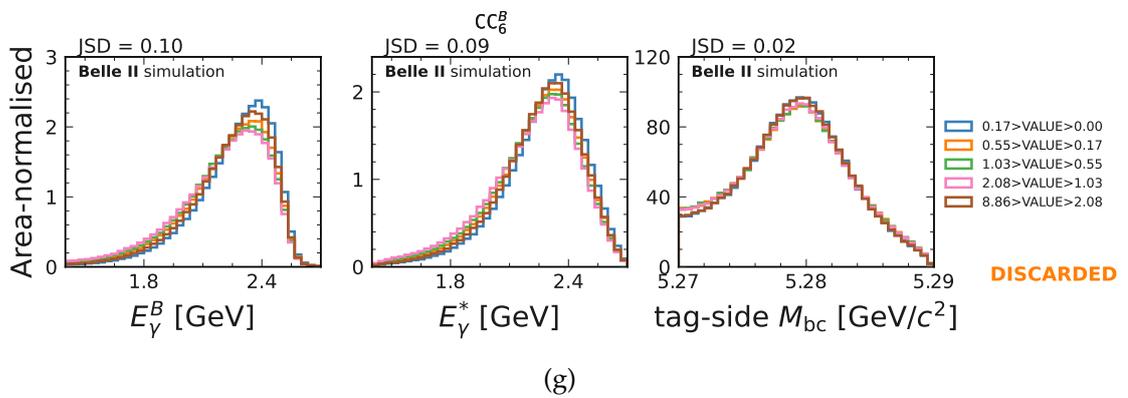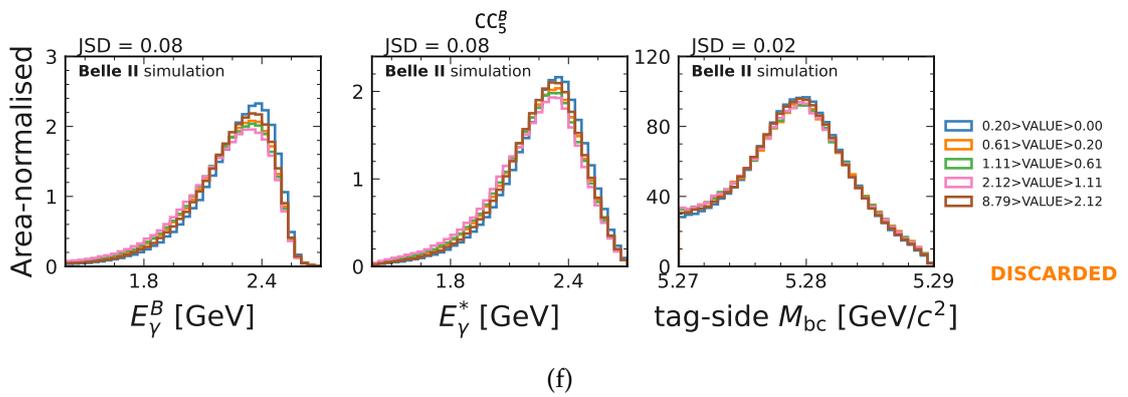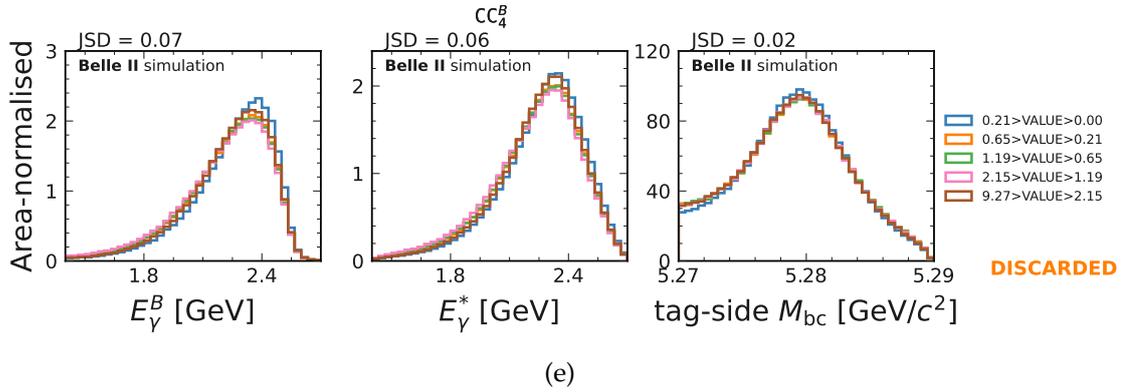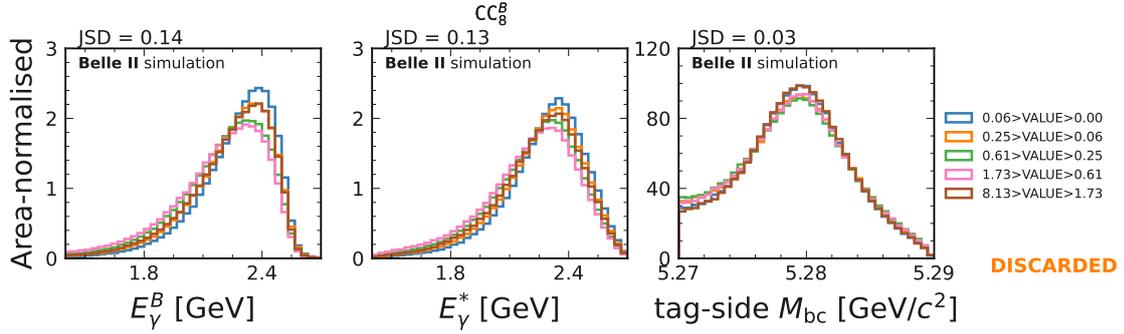
(i)
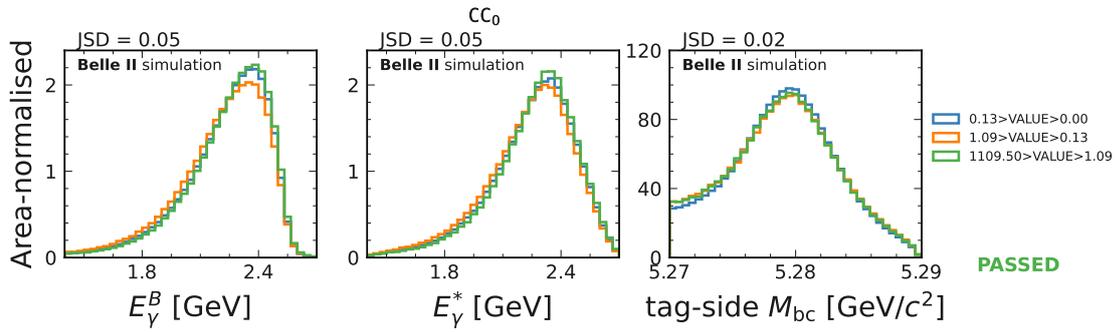
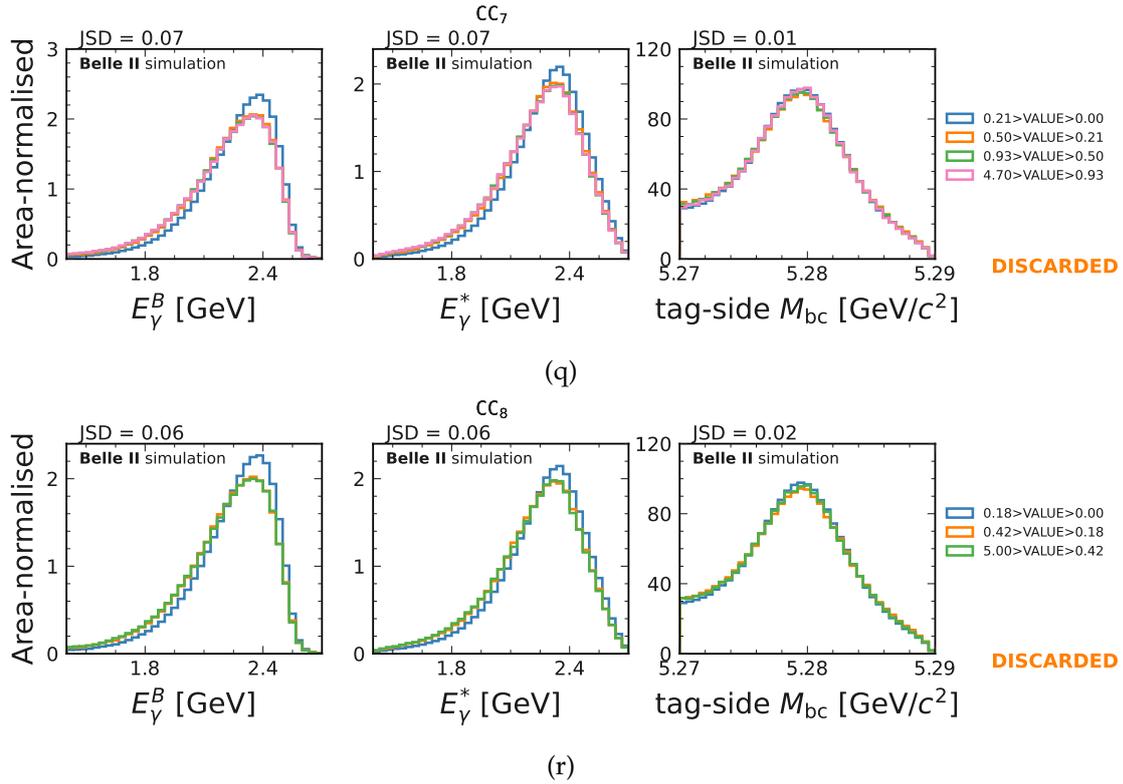(j)
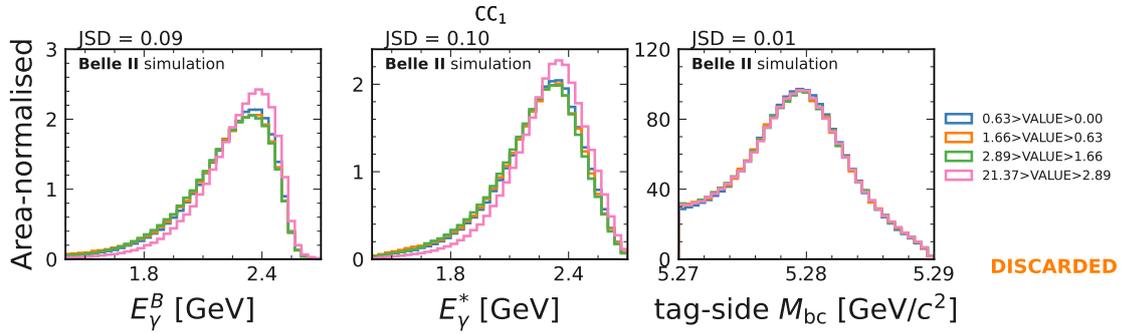
(k)
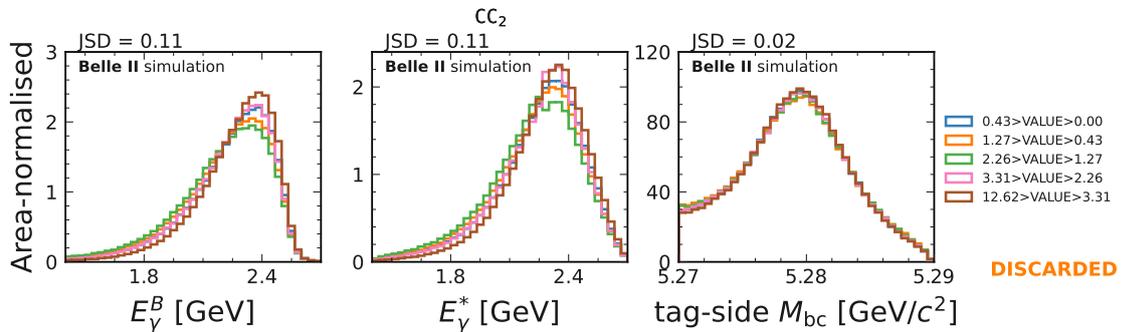
(l)

(m)
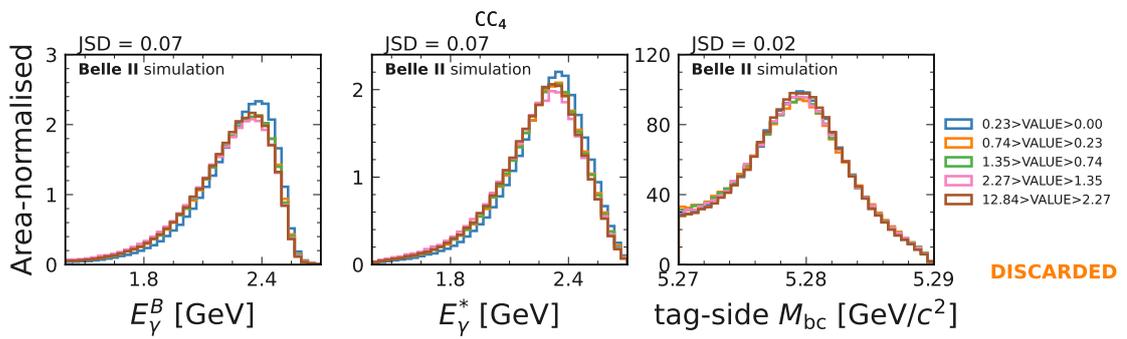
(n)

(o)

(p)

Figure F.6: The bias-test on E_γ^B, E_γ^* and M_{bc} for modified Fox-Wolfram moments. The test is performed based on **Test 1** strategy, defined in Section 6.5.2. Variable definitions are given in the text. The Jensen Shannon distance as introduced in Equation (6.11) is given for each distribution.

F.7 B vertex observables

The kinematic vertex fit of the tag-B meson candidate (see Section 6.3) provides x, y and z coordinates of the decay vertex. True B meson candidates should have successful fits, consistent with a decay near the collision point.

Furthermore, all other tracks originating from particles that were not used in the reconstruction of the tag-B meson have another vertex fit performed. This gives another set of vertex variables, which is checked for kinematic consistency with the decay originating near the interaction point. This fitted B meson candidate is denoted as B_{ROE} and is used only for continuum suppression. In this analysis the following 19 observables are tested:

- x, y and z are tag-B vertex coordinate and their uncertainty distributions;

- χ_V^2 of the tag-B meson vertex fit;

- $x_{B_{\text{ROE}}}$, $y_{B_{\text{ROE}}}$, $z_{B_{\text{ROE}}}$ and their uncertainty distributions;

- B_{ROE} vertex p-value;

- χ_V^2: B_{ROE} vertex χ^2 value;

- $\chi^2_{B_{\text{ROE}};\text{IP}}$: B_{ROE} vertex χ^2 value of the interaction point component;

- $\Delta\tau$: proper decay time difference between tag-B meson and B_{ROE};

- Δz: difference of decay vertex z components between tag-B meson and B_{ROE};

- Δz_B: difference of decay vertex z components between tag-B meson and B_{ROE} in the boost direction.
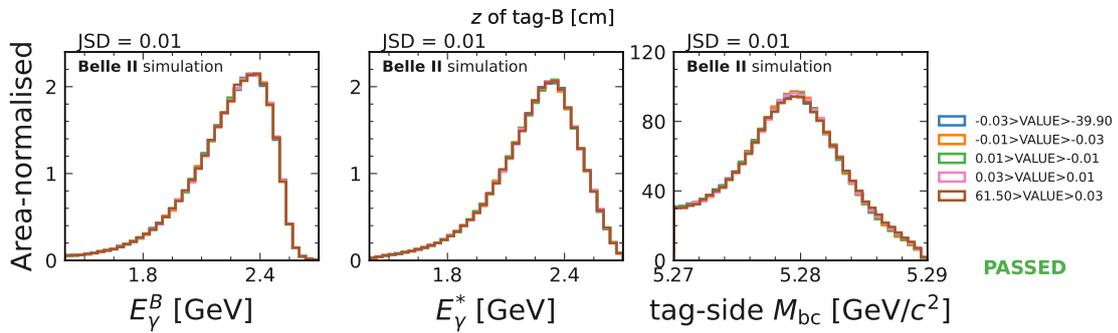
The results for **Test 1** are shown in Figures F.7a to F.7s. Most of these features show no correlation to E_γ^B or E_γ^* as the vertex requirement is a purely physical constraint. The χ^2 of tag-B meson is manually excluded to avoid a bias towards selected FEI modes.
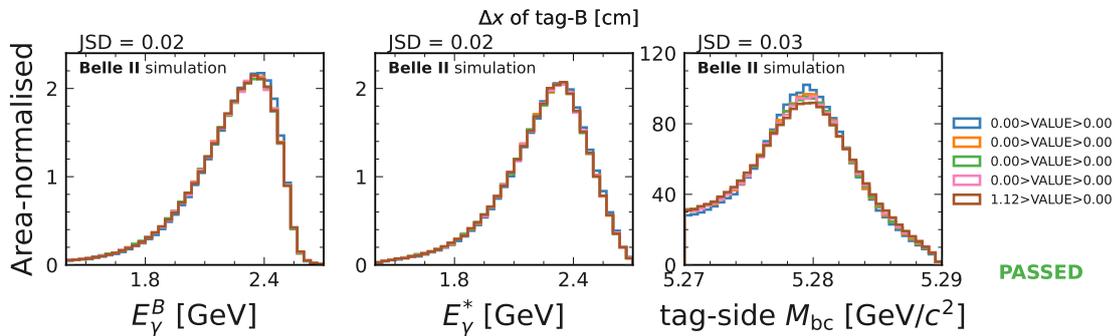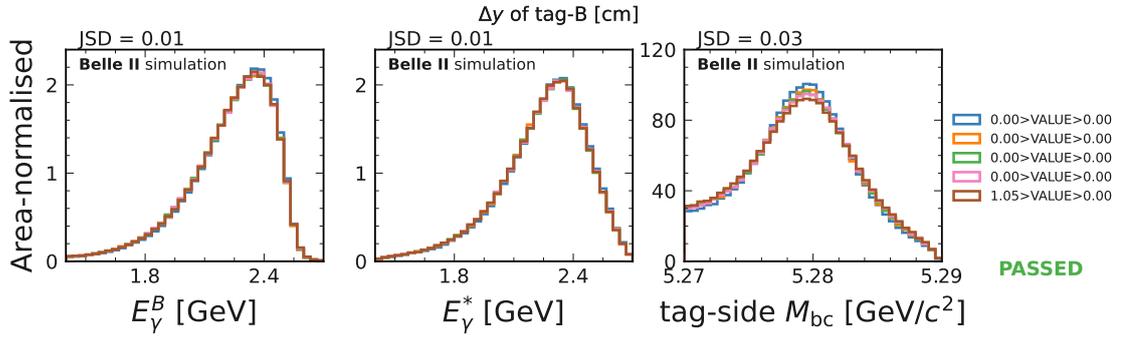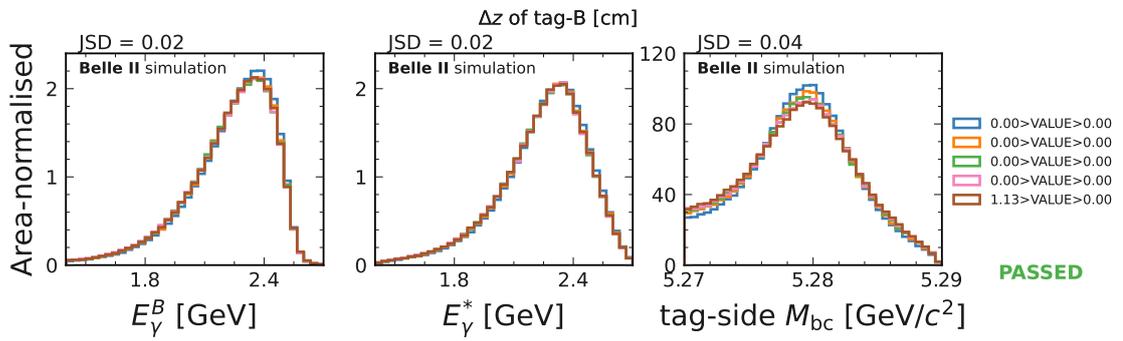
(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

(k)

(l)

(m)

(n)

(o)

(p)
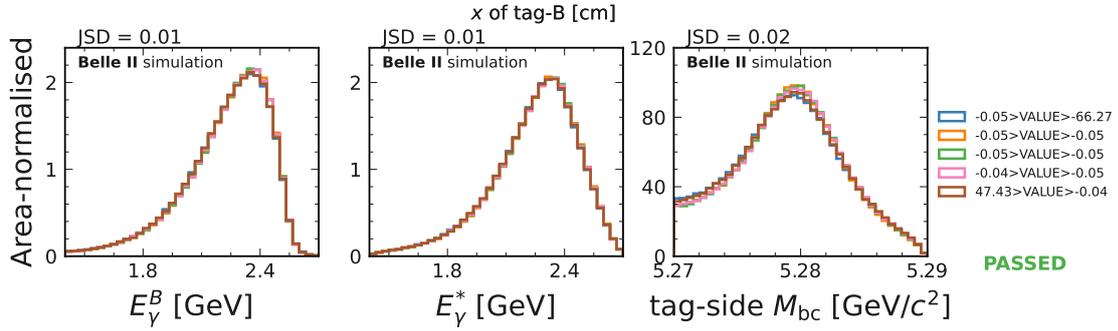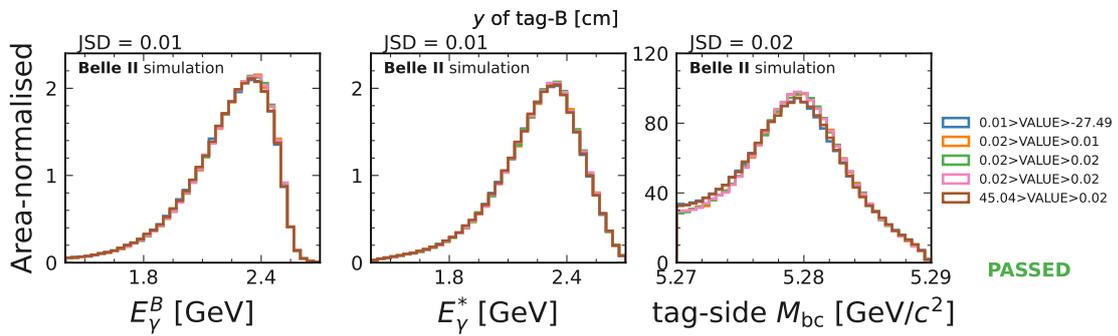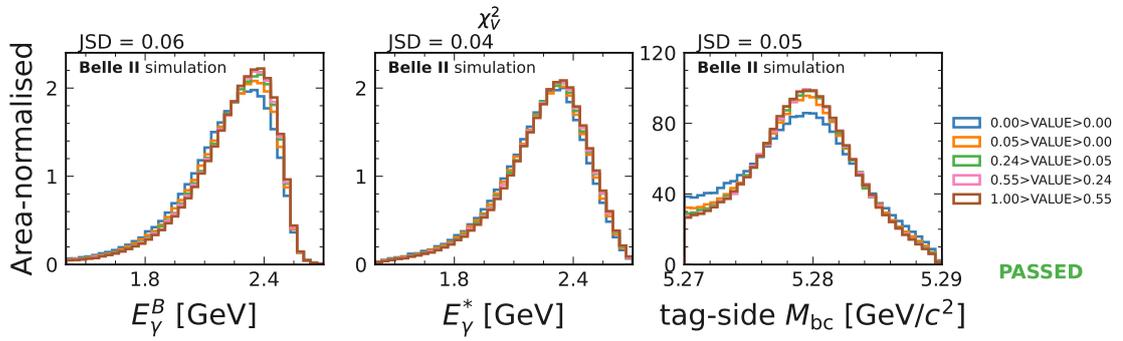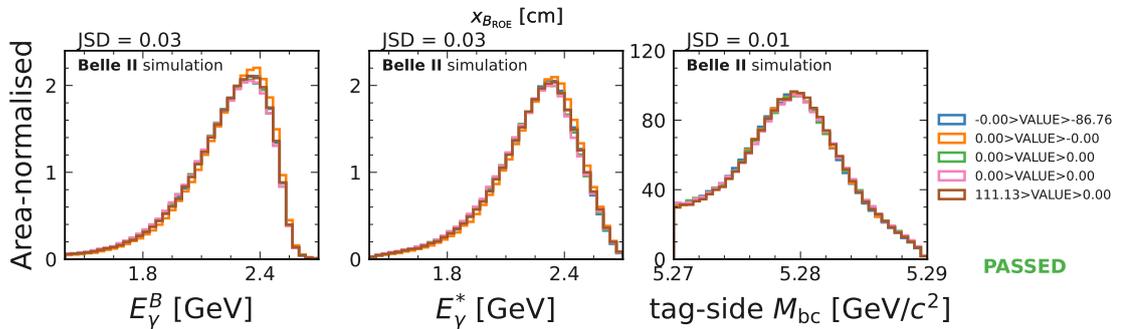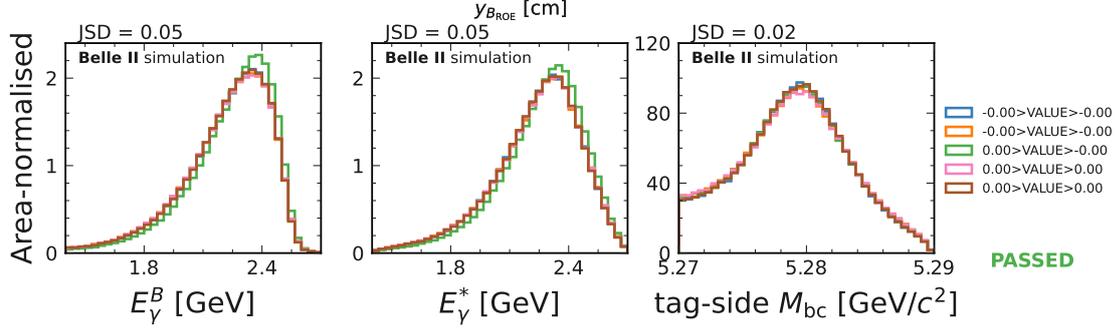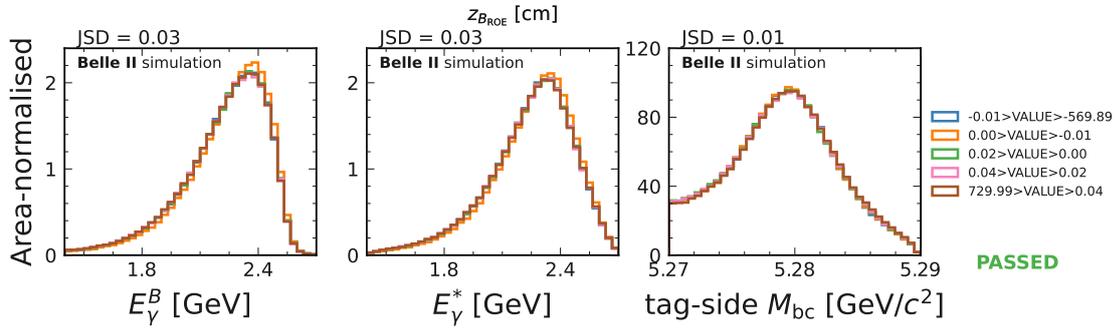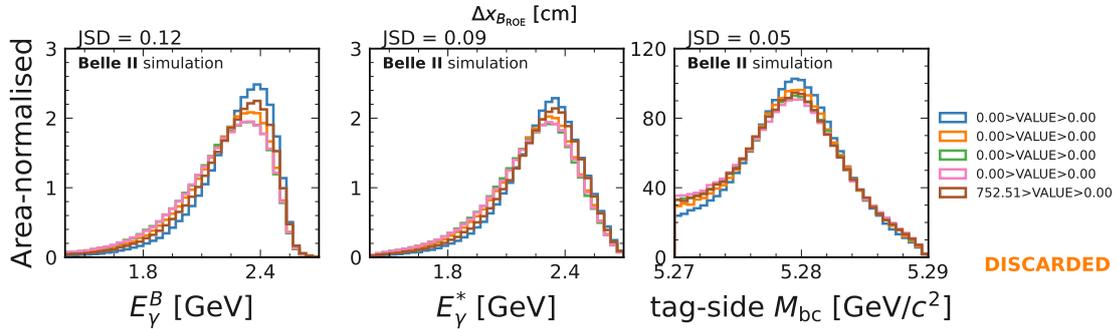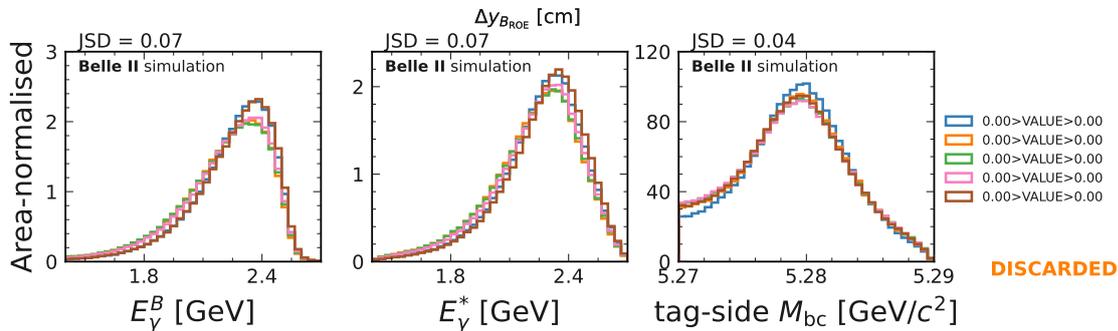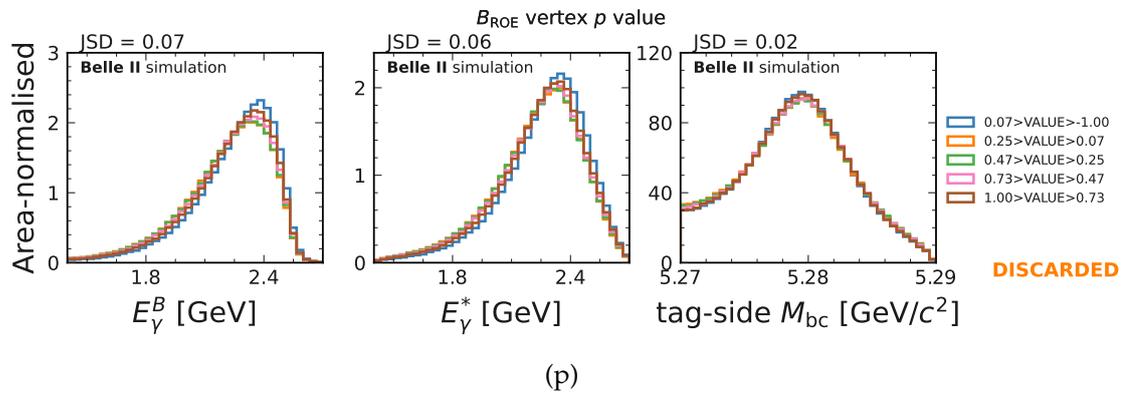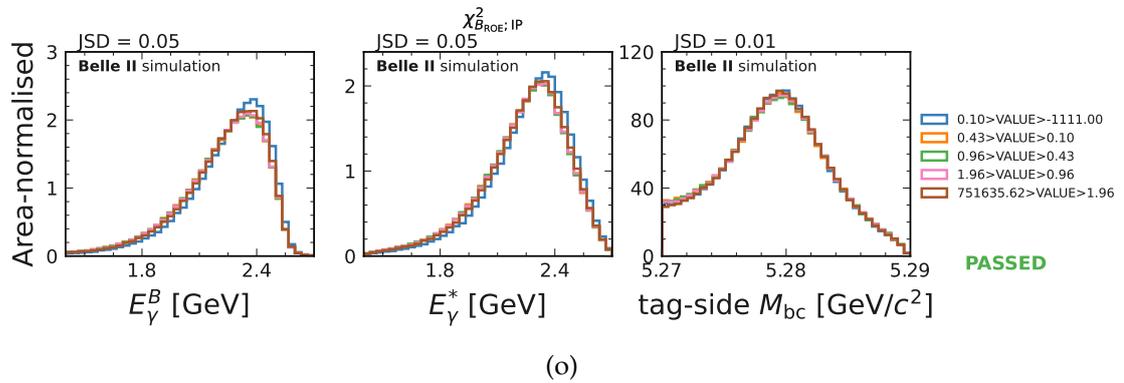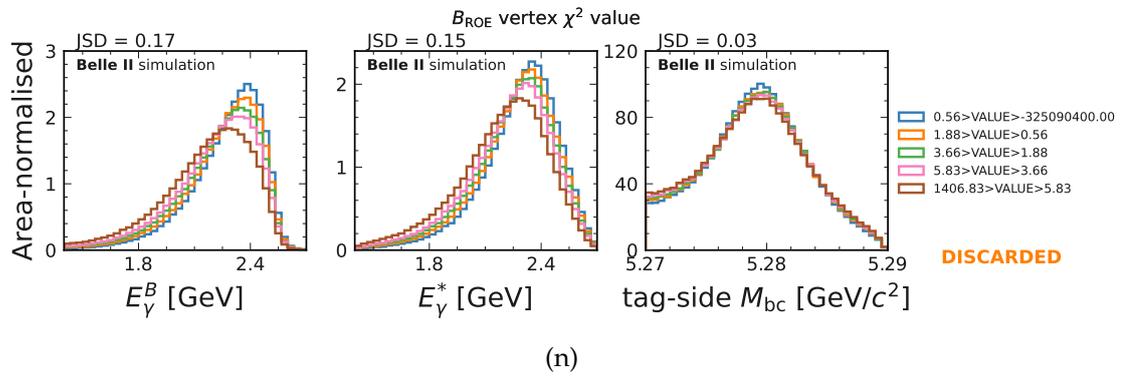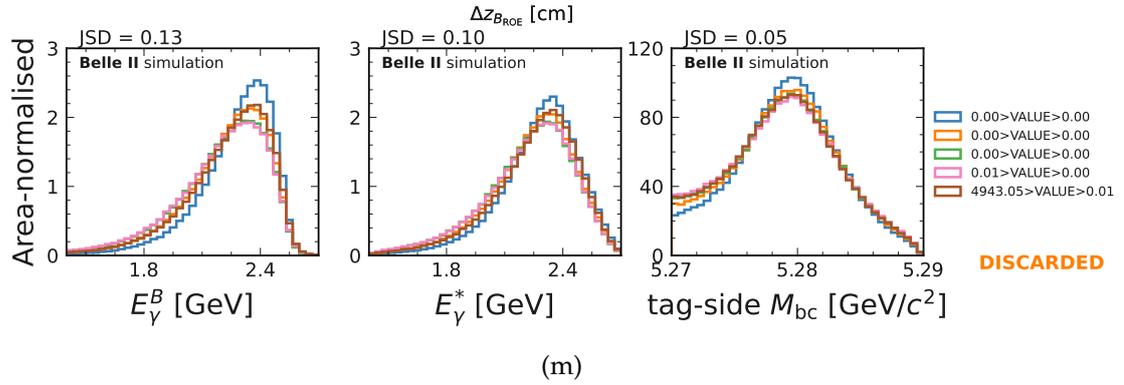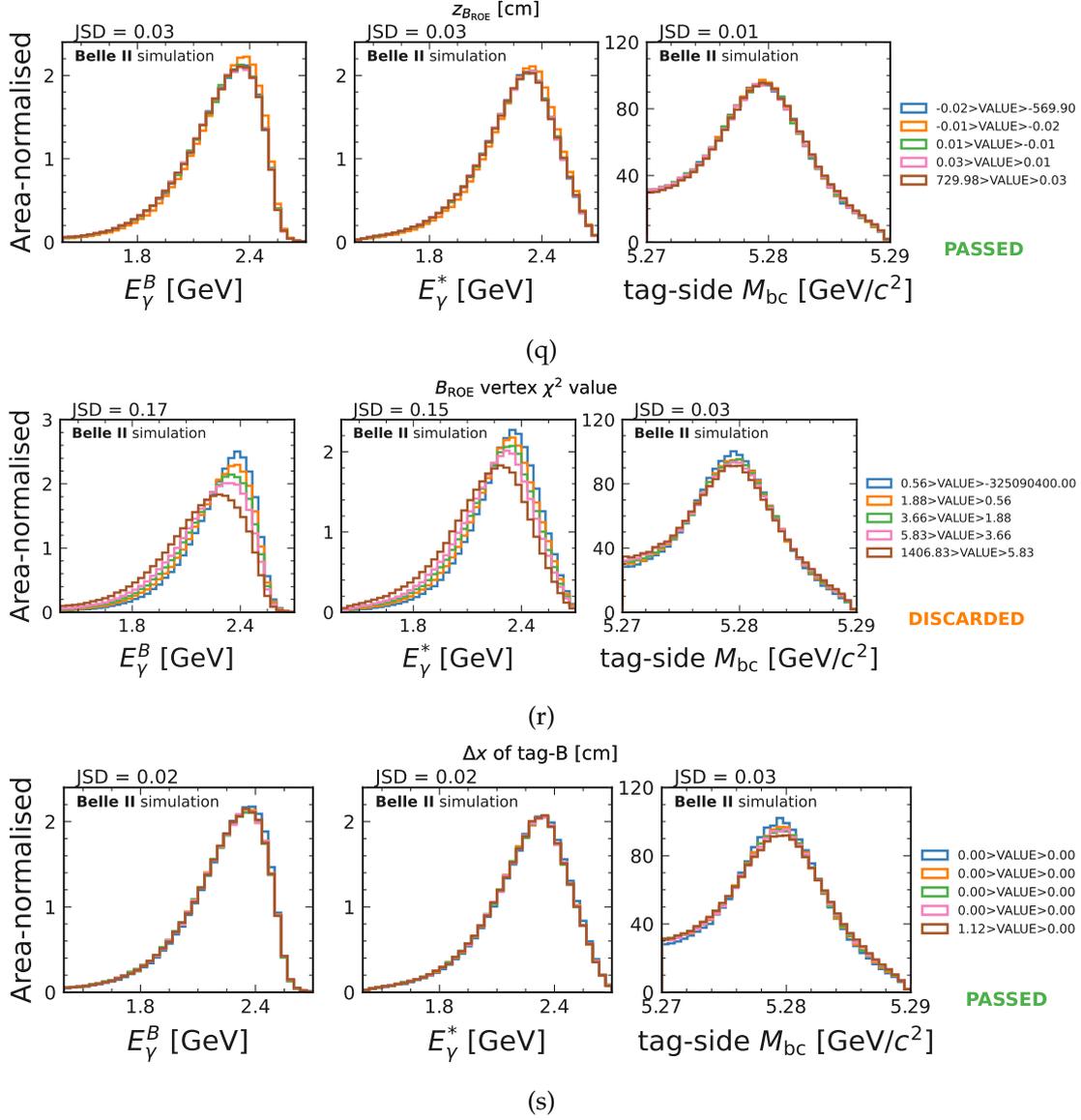
Figure F.7: The bias-test on E_γ^B, E_γ^* and M_{bc} for vertex coordinates of the B mesons. The test is performed based on **Test 1** strategy, defined in Section 6.5.2. Variable definitions are given in the text. The Jensen Shannon distance as introduced in Equation (6.11) is given for each distribution.

F.8 B **meson flavour tagger outputs**

Belle II flavour tagger is an algorithm that uses MVA methods to determine the flavour of a B that is not reconstructed. In particular, it uses the information of the tag-B meson to infer information about the signal-B meson, in our case the $B \to X_s \gamma$ candidate. The exact procedure is outside of the scope of this thesis, but can be followed up in Ref. [168]. The general reassoning why these observables are tested for $e^+e^- \to q\bar{q}$ suppression is that a correctly reconstructed B meson should, on average, perform better than a combinatorial $q\bar{q}$ candidate. The flavour tagger outputs two distributions: FT$_{\mathrm{BDT}}$ and FT$_{\mathrm{NN}}$ which differ by the internal model that the flavour tagger uses (BDT versus neural network). The results for **Test 1** are given in Figures F.8a and F.8b. While these variables would offer some separation power, the resulting bias is just above the threshold.

(a)

(b)
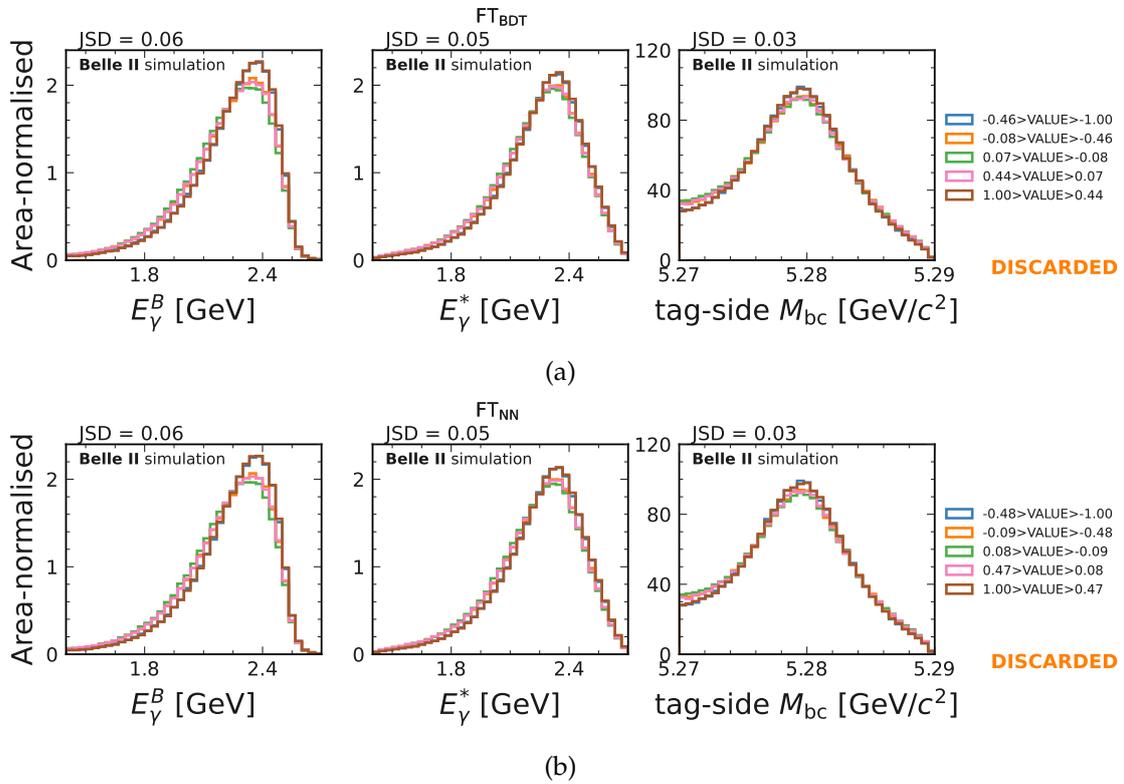
Figure F.8: The bias-test on E_γ^B, E_γ^* and M_{bc} for flavour tagger outputs. The test is performed based on **Test 1** strategy, defined in Section 6.5.2. Variable definitions are given in the text. The Jensen Shannon distance as introduced in Equation (6.11) is given for each distribution.

Appendix G

Data-simulation agreement for BDT training features

All 29 observables that passed the requirements of **Test 1**, introduced in Section 6.5.2, are subsequently tested in a second test. This test, named **Test 2**, aims to test the data-simulation agreement of the features, to ensure that the classifier only learns from features that represent Belle II data adequately.

This is shown, for features that passed **Test 1** (see Appendix F) in Figure G.1. All but three features show good data-simulation agreement, which attests to the high quality of Belle II detector simulation and calibration.

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)

(d)　　　　　　　　　　(e)　　　　　　　　　　(f)
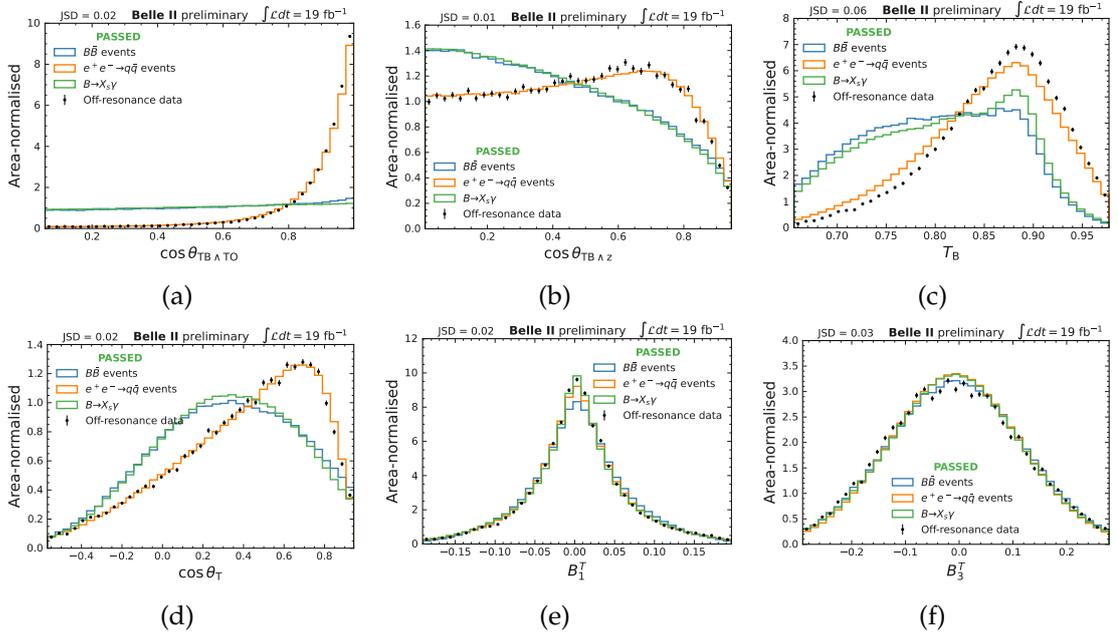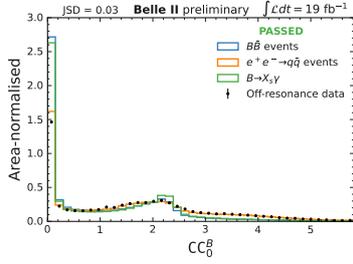
(g)
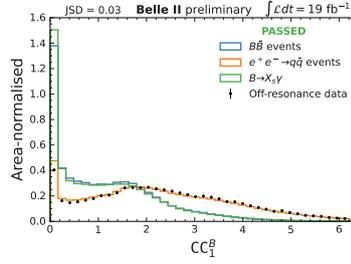
(h)
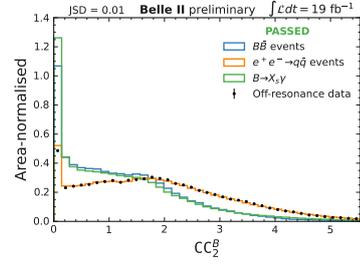
(i)
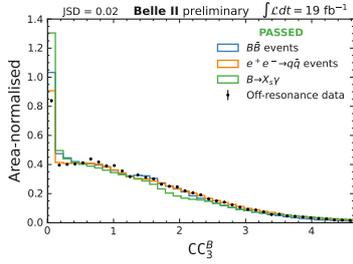
(j)
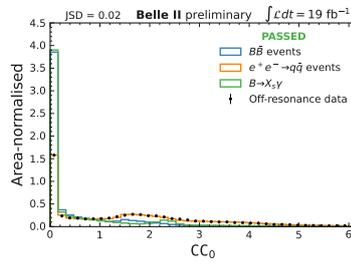
(k)
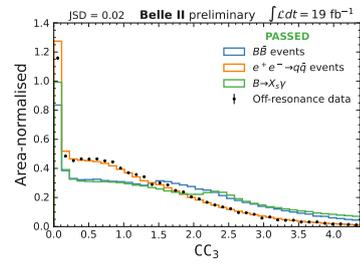
(l)
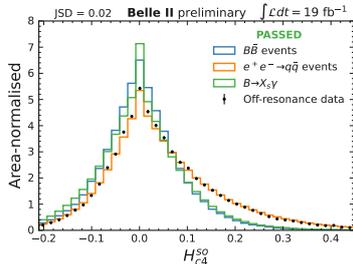
(m)
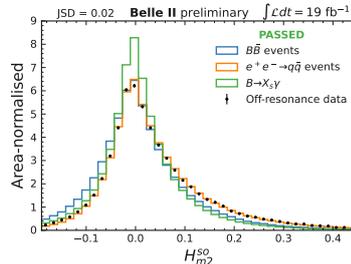
(n)
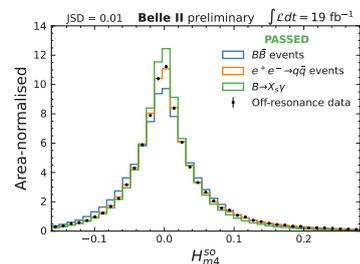
(o)
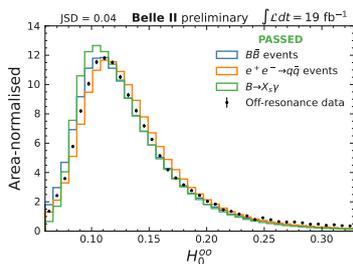
(p)
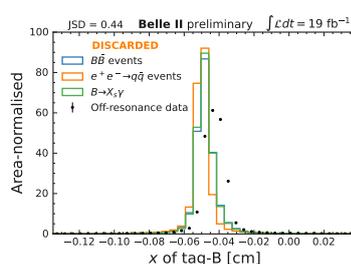
(q)
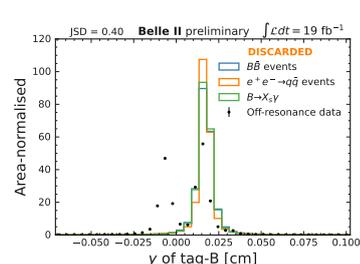
(r)
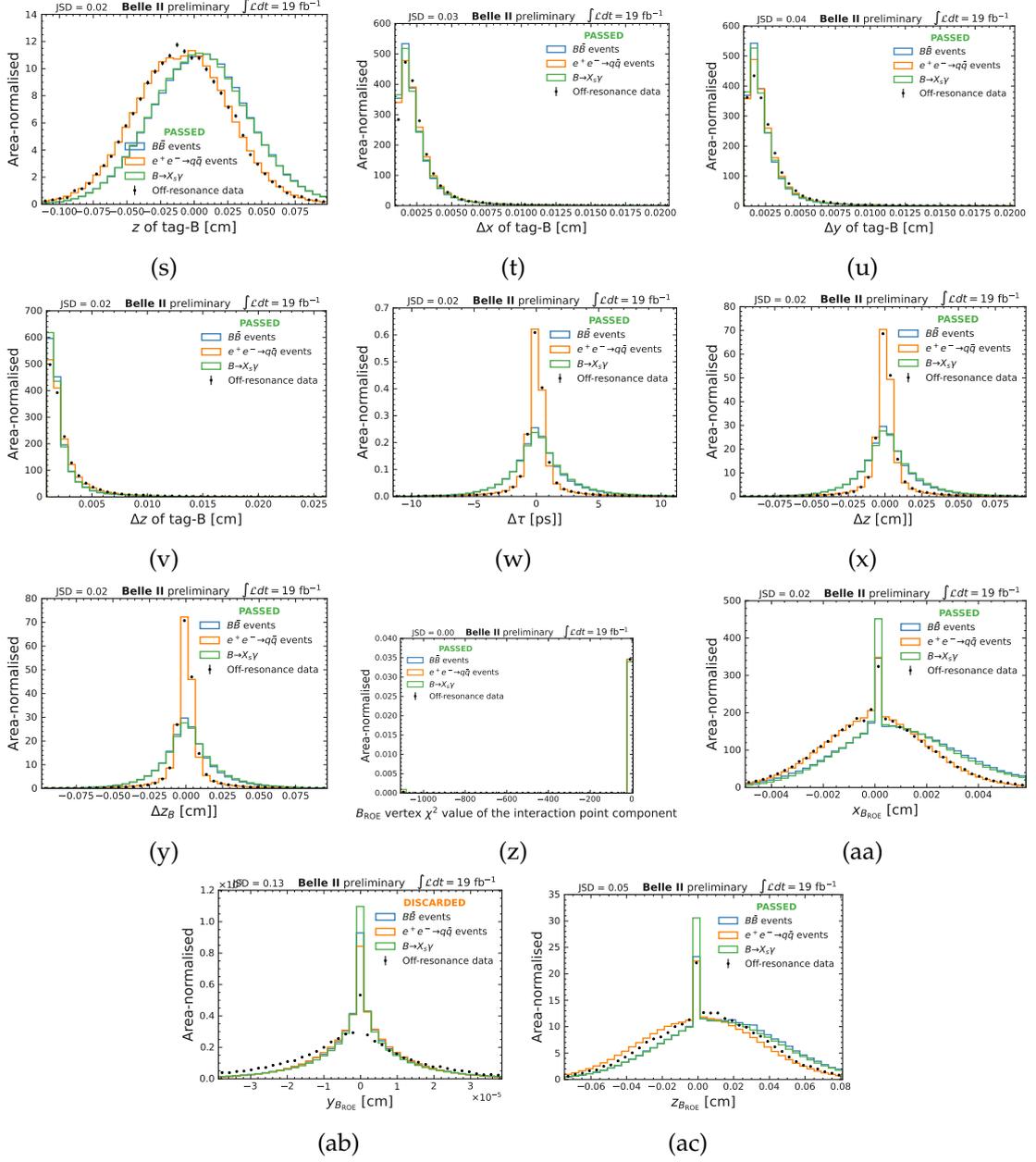
Figure G.1: The data-simulation agreement test between $e^+e^- \to q\bar{q}$ simulation and off-resonance Belle II data. The distributions shown are area-normalised, such that only a shape, but not a normalisation agreement test is performed. The Jensen-Shannon distance is evaluated between simulated $e^+e^- \to q\bar{q}$ and off-resonance data distributions. For reference, also $B \to X_s\gamma$ and generic-$B\bar{B}$ decay distributions are given.

Appendix H

Probability density functions for the M_{bc} fit

In this analysis three probability density functions (PDFs) are used for the description of the M_{bc} distributions. The exact use of the PDFs is discussed broadly in Section 6.8.1. They are defined in the Sections of this Appendix.

H.1   Crystal Ball function

The function used for peaking M_{bc} distribution is the Crystal Ball function, named after the Crystal Ball collaboration, where it was used for calorimeter response and resolution modelling (see e.g. [169]). It is given in terms of four parameters μ, σ, α and n:

$$f(x;\mu,\sigma,\alpha,n) = \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{x-\mu}{\sigma} \geqslant -\alpha \\ A \cdot (B - \frac{x-\mu}{\sigma})^{-n}, & \text{for } \frac{x-\mu}{\sigma} < -\alpha \end{cases} \tag{H.1}$$

with

$$\begin{aligned} A &= \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) \\ B &= \frac{n}{|\alpha|} - |\alpha|. \end{aligned} \tag{H.2}$$

The Crystal Ball function can be understood as a Gaussian peak with a mean μ and width σ, and a polynomial tail of n-th order. The switch from Gaussian to polynomial behaviour occurs based on α.

H.2   ARGUS function

The function to model $e^+e^- \rightarrow q\bar{q}$ distribution is the ARGUS function, introduced by the ARGUS collaboration to model continuum events. It has been ever since been adopted in

most B-factory experiments and is widely used by CLEO, BaBar, Belle, Belle II and others. The function is defined in terms of parameters c and m_0 as [170]:

$$f(m, m_0, c) = m \cdot \sqrt{1 - \left(\frac{m}{m_0}\right)^2} \cdot \exp\left[c \cdot \left(1 - \left(\frac{m}{m_0}\right)^2\right)\right] \tag{H.3}$$

The parameter m_0 can be interpreted as the cut-off: the region where the ARGUS function is 0. It can be understood as the region where M_{bc} distribution becomes kinematically forbidden. The parameter c is the curvature parameter and governs the shape of the ARGUS functions.

H.3   Chebyshev polynomials of the first kind

Chebyshev polynomials of the first kind, named after Pafnuty Chebyshev, are a set of polynomials expressed in terms of cosine and sine functions. They are defined via the following recursive relation:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \tag{H.4}$$

with $T_0 = 1$ and $T_1 = x$. Assuming $i \in \{0, 1, .., N\}$ orders of the Chebyshev are considered, they can be used to approximate a function as:

$$f(x, \vec{k}) = \sum_i^N k_i T_n(x), \tag{H.5}$$

where $\vec{k} = k_0, k_1, k_2, ..., k_N$ are parameters of the fit, understood as scaling factors for the i-th order Chebyshev polynomial. The parameter n_0 is set to 1 and not scaled in this analysis.

Appendix I

Primary M_{bc} fits determining initialisation values

Before performing an M_{bc} fit of the total data set, the PDFs are defined and pre-fitted on the individual components corresponding to the good tag-B mesons, combinatorial $B\bar{B}$ background and continuum background (see Section 6.8.1 for definitions). The M_{bc} fits of each E_{γ}^{B} bin (see Section 6.8.2) are provided for the aforementioned components of the data set in this Chapter. The fitting strategy is introduced in Section 6.8.3. The fits are as follows:

- Crystal Ball PDF fits of the good tag-B meson M_{bc} distribution (Figure I.1).

- ARGUS PDF fits of the continuum event M_{bc} distribution (Figure I.2).

- Chebyshev PDF fits of the combinatorial $B\bar{B}$ tag background distribution (Figure I.3).

Additionally, the ARGUS and Chebyshev PDFs are combined after fitting and a good combined background description is achieved. This is shown in Figure I.4.

Although small inaccuracies of the fitter may be spotted (e.g. $2.4 - 2.6$ GeV region in Figure I.3) The primary goal of the fitter is to describe the Crystal Ball normalisation parameter. Moreover, these fits are prepared on 1.6 ab^{-1}, whereas the measurement is planned on a data set roughly one order of magnitude smaller. Therefore, at this stage, such a description performs sufficiently well for the goal of the analysis.
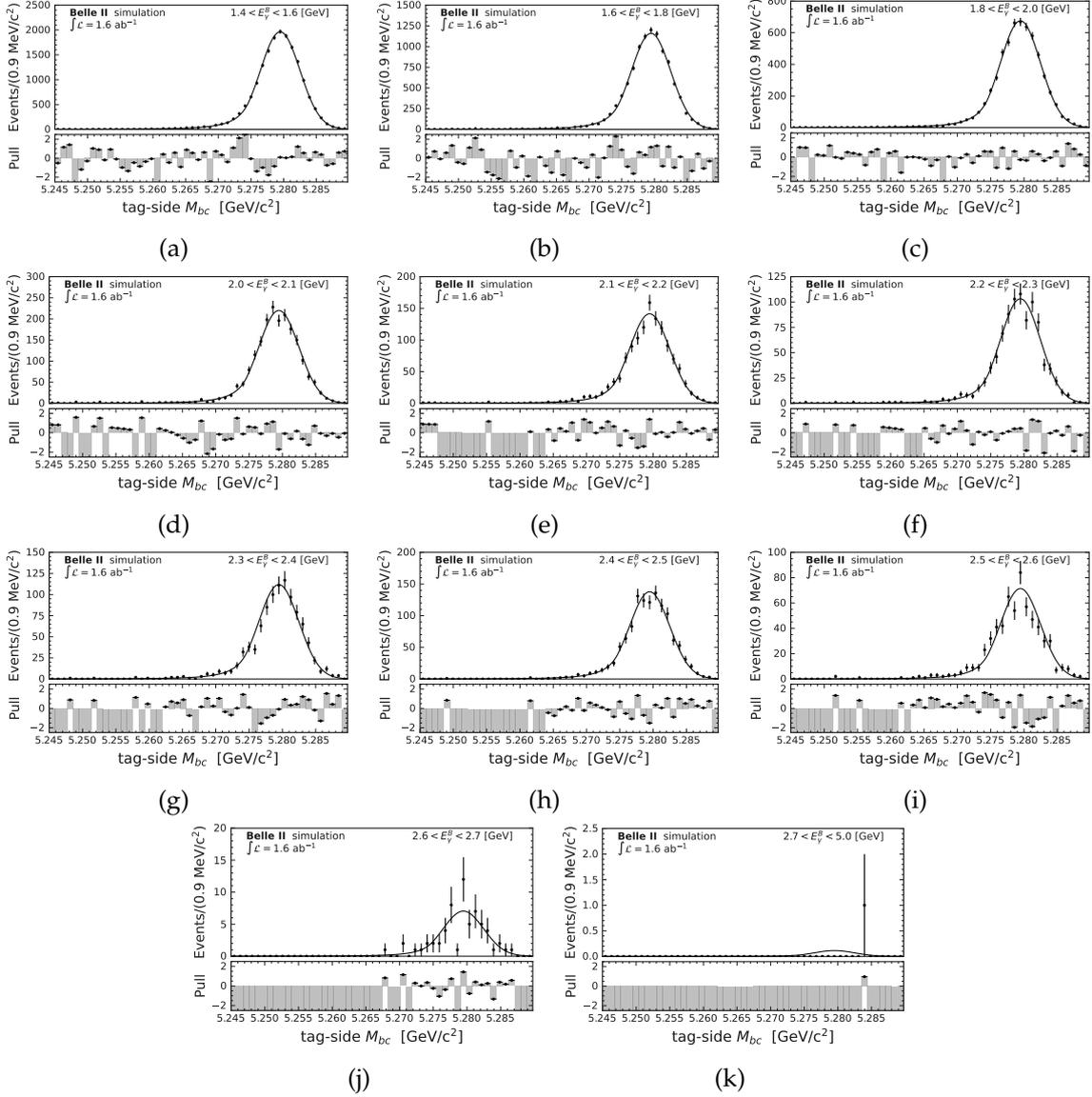
Figure I.1: The fits of the good tag-*B* meson events in generic MC using the Crystal Ball PDF. This fit allows extracting the initialisation values for further fitting of the total data sets. A good description of the M_{bc} distributions can be seen throughout the E_γ^B bins.
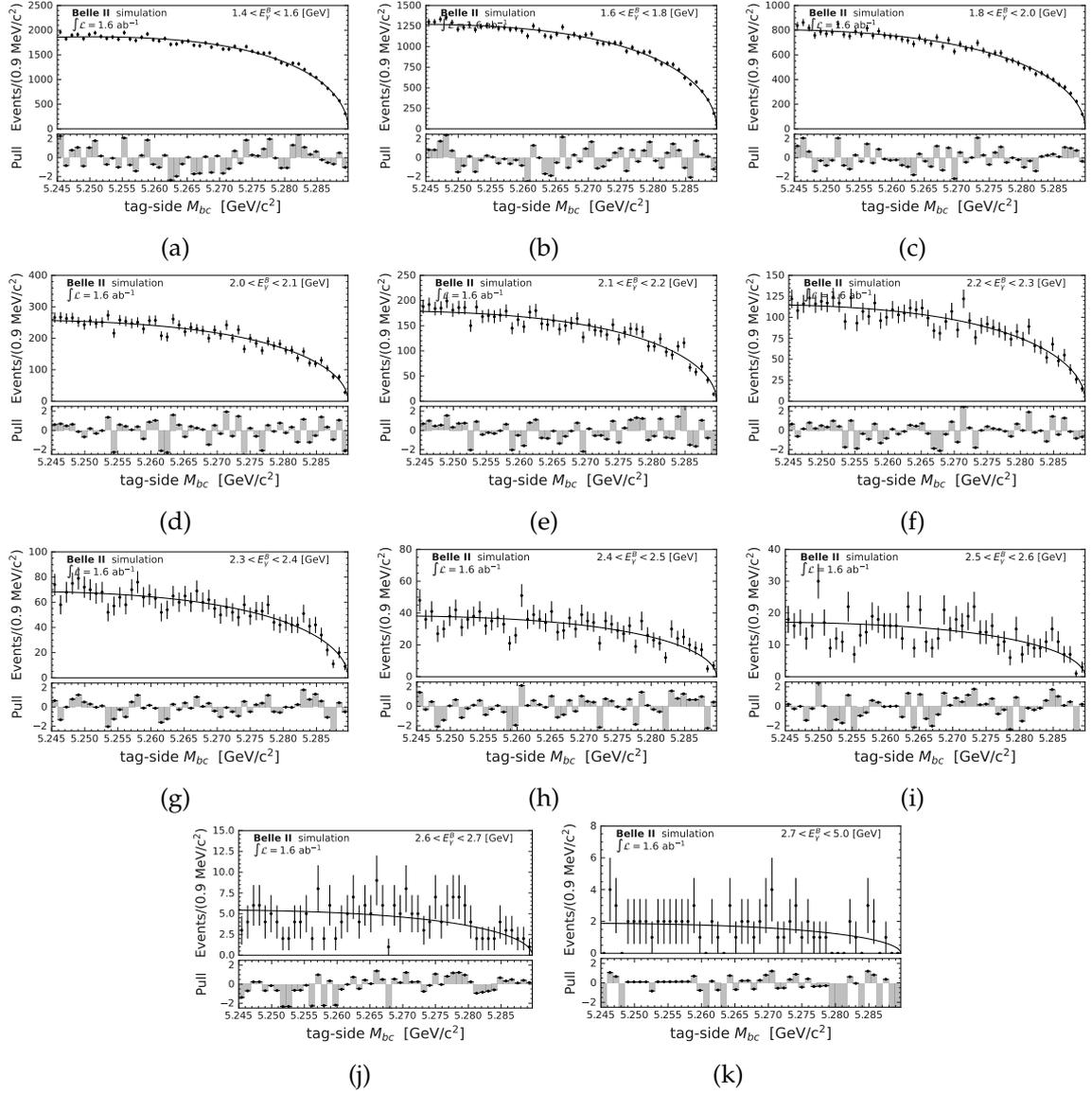
Figure I.2: The fits of the continuum events in generic MC using the ARGUS PDF. This fit allows extracting the initialisation values for further fitting of the total data sets. A good description of the M_{bc} distributions can be seen throughout the E_γ^B bins.
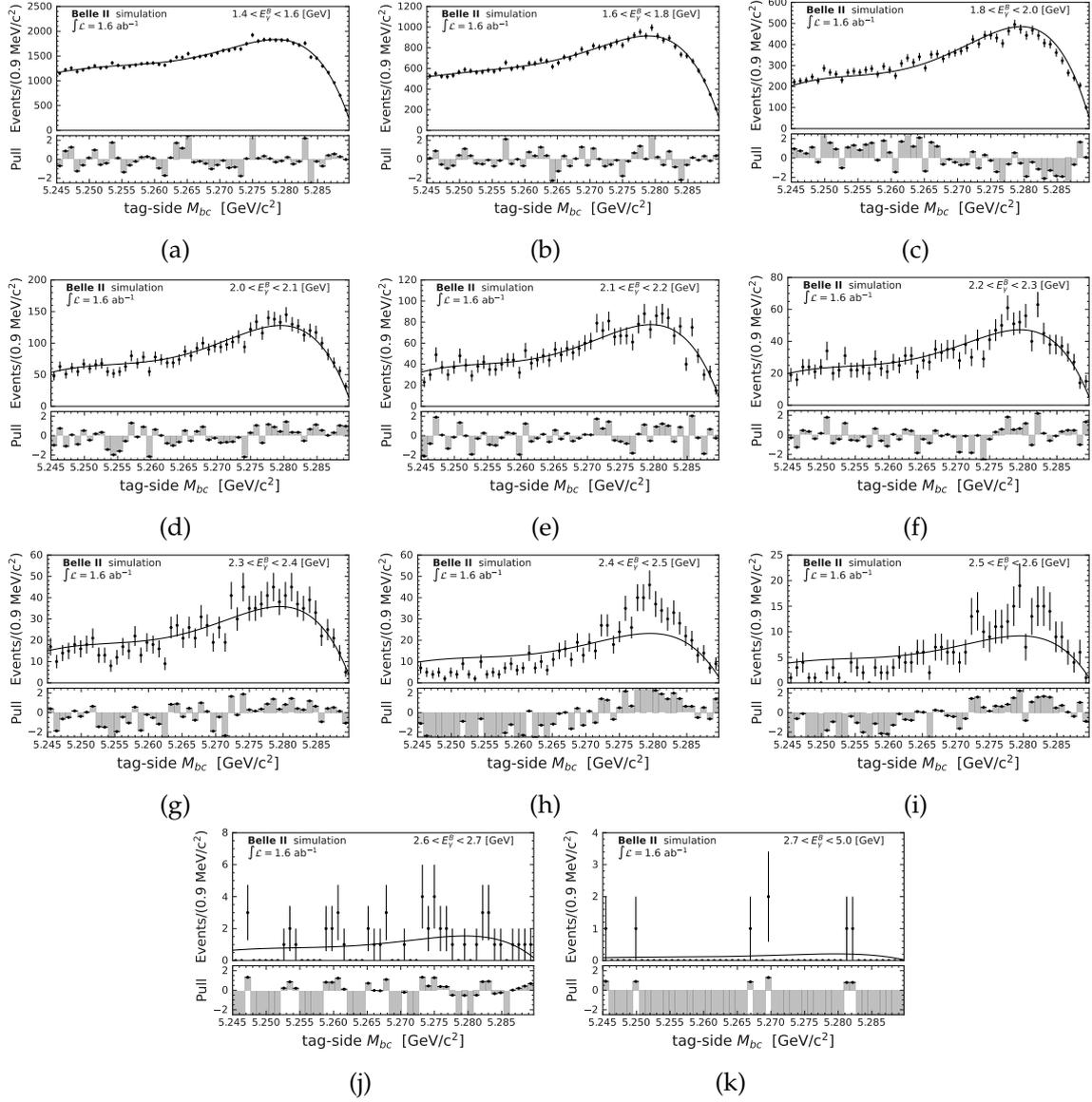
Figure I.3: The fits of events with combinatorial-$B\bar{B}$ tags in generic MC using the Chebyshev PDF. This fit allows extracting the initialisation values for further fitting of the total data sets. A good description of the M_{bc} distributions can be seen throughout the E_γ^B bins.
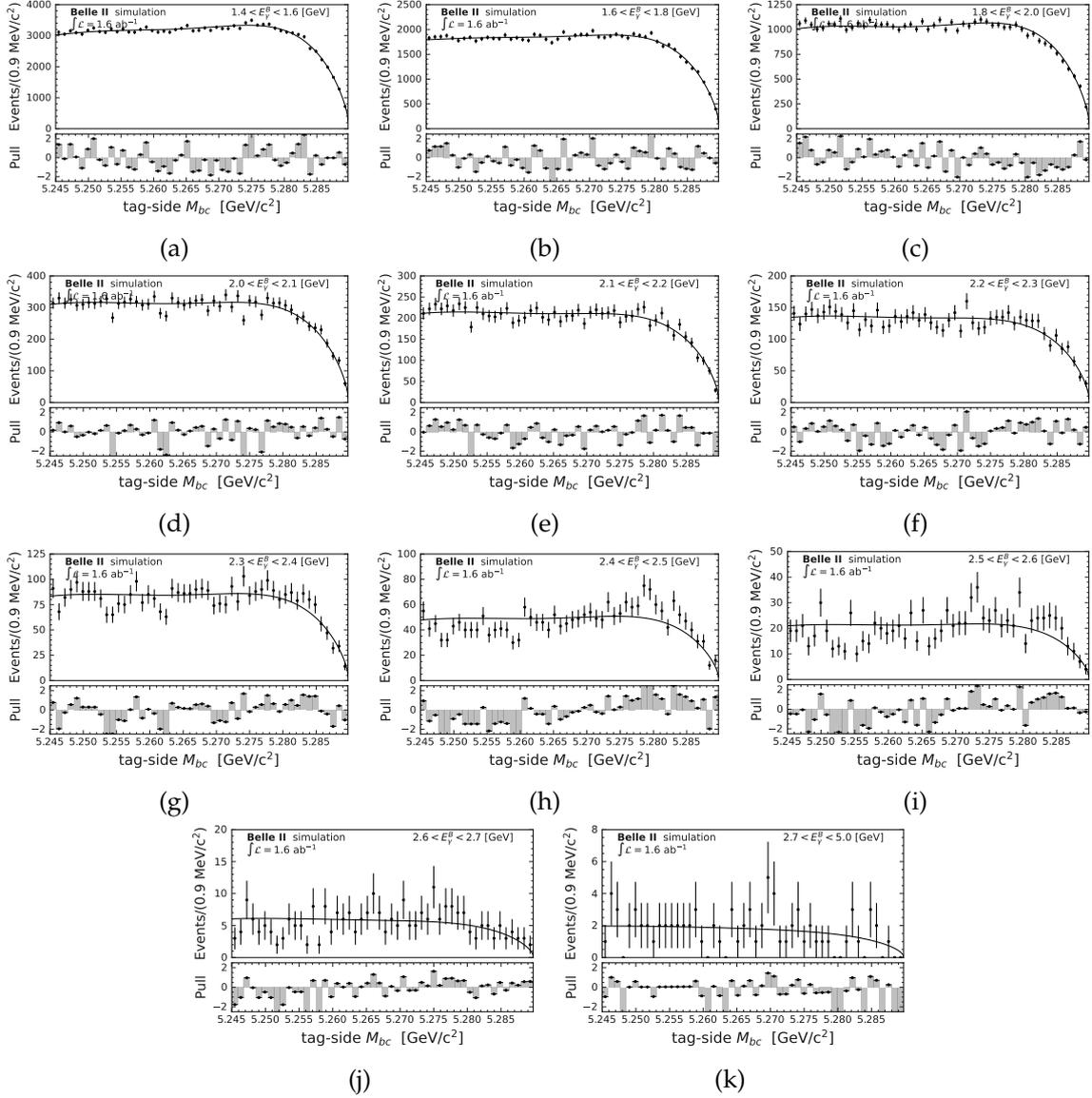
Figure I.4: The combined fits from Figures I.2 and I.3 visualised on the combined combinatorial-$B\bar{B}$ tag and continuum event data sets in generic MC. An excellent overall description of the M_{bc} distributions can be seen, especially taking into account that the Belle II data set for this analysis is roughly one order of magnitude smaller than the amount of simulated data fit here.

Appendix J

Remaining-$B\bar{B}$ background after the M_{bc} fitting

As seen in Section 6.14.1, a significant amount of non-$B \rightarrow X_s\gamma$ events are captured in the estimated numbers of good tag-B mesons. This is related to the fact that an irreducible background component remains in the analysis due to the inclusive treatment of the X_s system. The composition of this component for B^+ and B^0 decays is shown in Tables J.1 and J.2. The results are calculated based on Belle II MC and show the fraction of B decays that contribute within a given E_γ^B interval. Only decay channels which contribute by at least 1% in at least one of the E_γ^B intervals are listed. Particularly, in low-E_γ^B regions thousands of decay channels may contribute to the background, therefore, the decay channels listed in the tables only cover a fraction of them.

Table J.1: B^0 decays that contribute to the good tag-B meson background. The percentages are of total B^0 background in a given interval.

E_γ^B interval	1.4 – 1.6 GeV	1.6 – 1.8 GeV	1.8 – 2.0 GeV	2.0 – 2.1 GeV	2.1 – 2.2 GeV	2.2 – 2.3 GeV	2.3 – 2.4 GeV	2.4 – 2.5 GeV	2.5 – 2.6 GeV	2.6 – 2.7 GeV	2.7 – 5.0 GeV
$B^0 \to D^*(2010)^- e^+ \nu_e$	0.13	0.14	0.09	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to D^- e^+ \nu_e$	0.06	0.04	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to \rho(770)^+ D^*(2010)^-$	0.04	0.08	0.14	0.20	0.19	0.02	0.04	0.00	0.00	0.00	0.00
$B^0 \to \rho(770)^+ D^-$	0.04	0.09	0.17	0.26	0.31	0.14	0.08	0.00	0.00	0.00	0.00
$B^0 \to D^*(2010)^- \pi^+ \pi^0$	0.03	0.04	0.06	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to D^- \mu^+ \nu_\mu$	0.02	0.02	0.01	0.02	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to D_2^*(2460)^- \rho(770)^+$	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to D^- D_s^+$	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to X_u e^+ \nu_e$	0.01	0.00	0.01	0.01	0.01	0.00	0.04	0.00	0.11	0.00	0.00
$B^0 \to D_1(2420)^- e^+ \nu_e$	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to D^*(2010)^- \pi^+ \eta$	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to D^- \pi^+ \eta$	0.00	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to a_1(1260)^+ D^-$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to D^*(2010)^- \eta \pi^+$	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^0 \to D^- \eta \pi^+$	0.00	0.00	0.01	0.02	0.01	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to \overline{D}^0 \eta$	0.00	0.00	0.00	0.01	0.02	0.08	0.04	0.00	0.00	0.00	0.00
$B^0 \to \overline{D}^*(2007)^0 \eta$	0.00	0.00	0.00	0.01	0.02	0.02	0.00	0.00	0.00	0.00	0.00
$B^0 \to \overline{D}^*(2007)^0 \pi^0$	0.00	0.00	0.00	0.00	0.01	0.04	0.00	0.00	0.00	0.00	0.00
$B^0 \to \overline{D}^0 \pi^0$	0.00	0.00	0.00	0.00	0.02	0.09	0.08	0.00	0.00	0.00	0.00
$B^0 \to \rho(770)^- e^+ \nu_e$	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.11	0.00	0.00	0.00
$B^0 \to K^*(1410)^0 \pi^0$	0.00	0.00	0.00	0.00	0.00	0.03	0.04	0.00	0.11	0.00	0.00
$B^0 \to X_u \mu^+ \nu_\mu$	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
$B^0 \to \eta K_2^*(1430)^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.11	0.00	0.00	0.00
$B^0 \to \pi^- \eta \pi^+ K^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to J/\psi(1S) K_L^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to \overline{\Sigma}_c^- \pi^0 n$	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to \chi_{c0}(1P) K_S^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to \eta_c(1S) K_S^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to \psi(2S) K^+ \pi^-$	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to K^*(892)^0 \eta$	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.22	0.11	0.00	0.00
$B^0 \to \eta_c(1S) K_L^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to \rho(770)^+ \rho(770)^-$	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$B^0 \to K^0 \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.11	1.00	0.00
$B^0 \to \pi^+ \pi^- \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.11	0.00	0.00
$B^0 \to J/\psi(1S) K_S^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00
$B^0 \to J/\psi(1S) \overline{K}^*(892)^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00
$B^0 \to \eta \omega(782)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00
$B^0 \to \pi^0 \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00
$B^0 \to \eta'(958) K_L^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00
Total	0.38	0.47	0.58	0.66	0.65	0.44	0.81	1.00	0.89	1.00	0.00

Table J.2: B^+ decays that contribute to the good tag-B meson background. The percentages are of total B^+ background in a given interval.

E_γ^B interval	1.4 − 1.6 GeV	1.6 − 1.8 GeV	1.8 − 2.0 GeV	2.0 − 2.1 GeV	2.1 − 2.2 GeV	2.2 − 2.3 GeV	2.3 − 2.4 GeV	2.4 − 2.5 GeV	2.5 − 2.6 GeV	2.6 − 2.7 GeV	2.7 − 5.0 GeV
$B^+ \to \overline{D}^*(2007)^0 e^+ \nu_e$	0.13	0.11	0.08	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \rho(770)^+ \overline{D}^0$	0.09	0.16	0.30	0.45	0.54	0.45	0.06	0.07	0.00	0.00	0.00
$B^+ \to \overline{D}^0 e^+ \nu_e$	0.07	0.05	0.04	0.02	0.02	0.07	0.03	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}^*(2007)^0 \rho(770)^+$	0.05	0.11	0.19	0.27	0.17	0.02	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}^0 \mu^+ \nu_\mu$	0.03	0.02	0.02	0.01	0.02	0.00	0.03	0.00	0.00	0.00	0.00
$B^+ \to D^*(2010)^- \pi^0 \pi^+ \pi^+$	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}_2^*(2460)^0 \rho(770)^+$	0.02	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}^0 D_s^+$	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}^*(2007)^0 a_1(1260)^+$	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}_1(2420)^0 e^+ \nu_e$	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}_1(2430)^0 \rho(770)^+$	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}^0 \rho(770)^+ \pi^0$	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}^0 \pi^+ \eta$	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}^0 \eta \pi^+$	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}^0 \pi^+$	0.00	0.00	0.00	0.01	0.01	0.05	0.03	0.00	0.00	0.00	0.00
$B^+ \to J/\psi(1S) K^*(892)^+$	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.03	0.00	0.00	0.00
$B^+ \to Xu0 \, e^+ \nu_e$	0.00	0.00	0.00	0.00	0.01	0.00	0.06	0.00	0.00	0.00	0.00
$B^+ \to D_s^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00
$B^+ \to \overline{D}_0^*(2300)^0 \mu^+ \nu_\mu$	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
$B^+ \to \rho(770)^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.03	0.06	0.00	0.00
$B^+ \to J/\psi(1S) K^+$	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.07	0.00	0.00	0.00
$B^+ \to \eta K_0^*(1430)^+$	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.12	0.00	0.00
$B^+ \to f_0(980) \pi^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
$B^+ \to K^*(1410)^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.03	0.00	0.00	0.00
$B^+ \to \pi^0 e^+ \nu_e$	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.03	0.00	0.00	0.00
$B^+ \to a_1(1260)^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.03	0.06	0.00	0.00
$B^+ \to \eta \pi^+$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.50	0.00
$B^+ \to \rho(770)^+ \eta$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
$B^+ \to \eta_c(1S) K^+$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
$B^+ \to \omega(782) K^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
$B^+ \to \pi^0 \mu^+ \nu_\mu$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
$B^+ \to \rho(1450)^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00
$B^+ \to \phi(1020) K^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
$B^+ \to \rho(770)^+ K^*(892)^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00
$B^+ \to f_2(1270) K^*(892)^+$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
$B^+ \to \pi^+ J/\psi(1S) \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
$B^+ \to \pi^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.06	0.00	0.00
$B^+ \to K^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.12	0.50	0.00
$B^+ \to b_1(1235)^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.06	0.00	0.00
$B^+ \to K^*(892)^+ \eta$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.06	0.00	0.00
$B^+ \to K^+ \pi^0 \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00
$B^+ \to \rho(770)^+ K^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
$B^+ \to \rho(770)^+ \rho(770)^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
$B^+ \to K_0^*(1430)^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
$B^+ \to \omega(782) \pi^+ \eta$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
$B^+ \to b_1(1235)^+ \eta$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
$B^+ \to \psi(2S) K^+$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
$B^+ \to \eta K_2^*(1430)^+$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
$B^+ \to \eta'(958) K^+$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.06	0.00	0.00
$B^+ \to \omega(782) \pi^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
$B^+ \to \eta e^+ \nu_e$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00
$B^+ \to K^*(892)^+ \pi^0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00
$B^+ \to \eta K^+$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00
Total	0.46	0.53	0.68	0.79	0.82	0.72	0.90	0.80	0.88	1.00	0.00

Appendix K

Photon energy resolution estimation

Each photon has a true energy, $\tilde{E}^B_\gamma$, which is the real energy it was produced at (in data) or generated (simulation). $\tilde{E}^B_\gamma$ is not generally the same as the measured energy, commonly denoted as E^B_γ in this thesis.

Based on these two quantities, the resolution of the photon energy measured in the signal-B meson rest frame is evaluated. The quantity from Equation (6.32):

$$\tilde{E}^B_\gamma - E^B_\gamma \qquad\qquad (K.1)$$

is fitted with a double-sided Crystal Ball function. A double-sided Crystal Ball is a generalised version of Equation (H.1), where an exponential tail is attached to both sides of the central Gaussian. This replaced the parameters $\{\alpha, n\}$ with $\{\alpha_L, \alpha_R, n_R, n_L\}$, where indices R and L denote that these parameters represent, correspondingly, 'right' or 'left' side of the Gaussian. The main parameter of interest when measuring the resolution is σ, which corresponds to the width of the central Gaussian and is directly interpreted as the resolution.

The unbinned negative log-likelihood fits are performed in intervals of $\tilde{E}^B_\gamma$. Although all double-sided Crystal Ball parameters are estimated by the fitter, only σ is focused on for this study. The hybrid signal-model samples are fitted, where a good tag-B meson (based on Figure 6.28, selected with maximum $\mathcal{P}_{\mathrm{FEI}}$) is used to evaluate the E^B_γ. For comparison, the same is done where events are only taken if a 'perfect' reconstruction of the tag-B meson is achieved. The former is visualised in Figure K.1, whereas the latter is in Figure K.2. The parameters σ and their corresponding uncertainties estimated by the fitter are then summarised in Figure 6.67. In Section 6.13.3 further discussion follows.
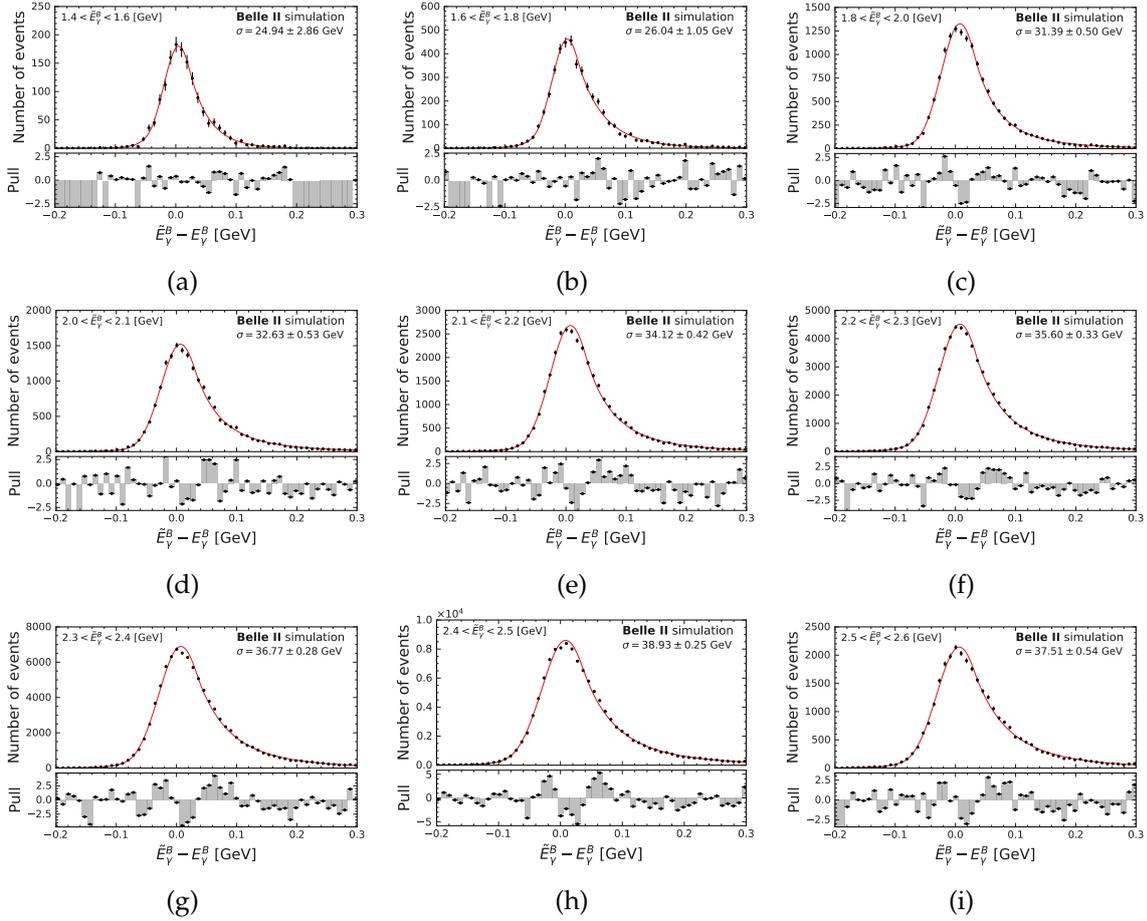
Figure K.1: The double-sided Crystal Ball fits, based on Equation (6.32), on the hybrid signal-model data set where good tag-B mesons are used for E_γ^B reconstruction. The parameter σ, corresponding to the width of the central Gaussian part is evaluated. This parameter is equated to the resolution of E_γ^B in this analysis. The fits are performed in intervals of $\tilde{E}_\gamma^B$ and no $B \to X_s \gamma$ photons can be produced with $\tilde{E}_\gamma^B \gtrsim 2.6$ GeV, due to kinematic constraints.
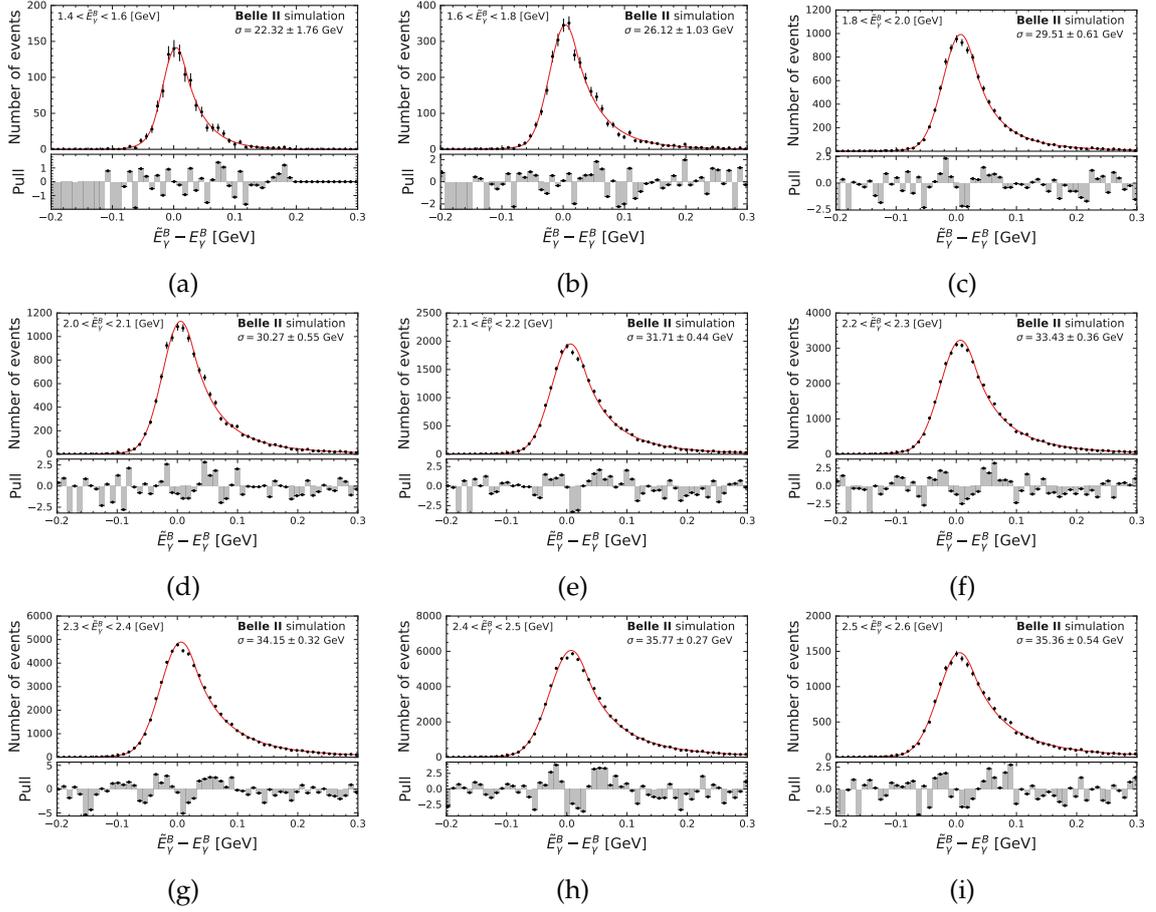
Figure K.2: The double-sided Crystal Ball fits, based on Equation (6.32), on the hybrid signal-model data set where only perfectly reconstructed tag-B mesons are used for $E_\gamma^{\check B}$ reconstruction. The parameter σ, corresponding to the width of the central Gaussian part is evaluated. This parameter is equated to the resolution of E_γ^B in this analysis. Note that the fits are performed in intervals of $\tilde E_\gamma^B$, and no $B \to X_s\gamma$ photons can be produced with $\tilde E_\gamma^B \gtrsim 2.6$ GeV, due to kinematic constraints.

Glossary

2HDM Two-Higgs doublet model. 28, 29, 30

ARICH Aerogel ring-imaging Cherenkov counter. 44, 45, 46

AUC Area-under-curve. 56, 89, 92, 93

basf2 Belle II analysis software framework. 47, 48, 49, 70, 77, 80, 255

BDT Boosted decision tree. 10, 55, 57, 58, 64, 70, 78, 84, 85, 86, 87, 88, 89, 94, 95, 166, 189, 190, 192, 194, 196, 198, 199, 200, 202, 204, 206, 208, 210, 212, 214, 216, 217, 218, 219, 220, 222

BSM Beyond-Standard-Model. 19, 28

CDC Central drift chamber. 44, 45, 48, 72, 73

CKM Cabibo-Kobayashi-Maskawa. 16, 17, 18, 20

DEPFET Depleted p-channel field effect transistor. 43

ECL Electromagnetic calorimeter. 46, 47, 48, 49, 50, 70, 72, 77, 78, 80, 125, 126, 127

FEI Full event interpretation algorithm. 8, 9, 10, 70, 71, 72, 73, 74, 77, 84, 89, 97, 100, 104, 113, 120, 121, 122, 131, 132, 133, 136, 145, 147, 148, 149, 152, 153, 155, 166, 167, 181, 182, 183, 184, 185, 186, 211, 255

ISR Initial-state radiation. 125, 126, 127

KLM K_L and μ detector. 46, 47

MC Monte-Carlo method based simulation. 65, 66, 67, 70, 73, 74, 75, 76, 78, 79, 81, 82, 85, 87, 88, 89, 96, 97, 99, 100, 101, 104, 107, 109, 110, 111, 112, 113, 114, 115, 116, 117, 120, 121, 123, 124, 127, 128, 130, 131, 132, 133, 135, 136, 137, 138, 140, 141, 142, 143, 144, 145, 147, 151, 152, 153, 155, 156, 157, 158, 159, 161, 226, 227, 228, 229, 231

MVA Multivariate algorithm. 55, 56, 57, 58, 77, 80, 217

PDF Probability density function. 51, 52, 107, 108, 109, 113, 115, 116, 118, 120, 121, 123, 124, 137, 153, 154, 166, 223, 225, 226, 227, 228

PXD Pixel detector. 43, 45, 48, 72

QCD Quantum chromodynamics. 13, 22

ROC Receiver operating characteristic. 56, 92

SM The Standard Model. 13, 15, 16, 17, 18, 19, 20, 22, 23, 28, 29, 165

SVD Silicon vertex detector. 43, 44, 45, 48, 72

TOP Time-of-propagation detector. 44, 45, 46, 72

Bibliography

[1] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 16 Oct. 1964. DOI: 10.1103/PhysRevLett.13.508 (cit. on pp. 11, 13).

[2] G. Aad *et al.*, ATLAS collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B*, vol. 716, pp. 1–29, 2012. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on p. 11).

[3] S. Chatrchyan *et al.*, CMS collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys. Lett. B*, vol. 716, pp. 30–61, 2012. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex] (cit. on p. 11).

[4] S. Abachi, B. Abbott, M. Abolins, *et al.*, D0 collaboration, "Observation of the Top Quark," *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, 14 Apr. 1995. DOI: 10.1103/PhysRevLett.74.2632 (cit. on p. 11).

[5] F. Abe, H. Akimoto, A. Akopian, *et al.*, CDF collaboration, "Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab," *Phys. Rev. Lett.*, vol. 74, pp. 2626–2631, 14 Apr. 1995. DOI: 10.1103/PhysRevLett.74.2626 (cit. on p. 11).

[6] H. Harari, "A new quark model for hadrons," *Phys. Lett. B*, vol. 57, no. 3, pp. 265–269, 1975. DOI: https://doi.org/10.1016/0370-2693(75)90072-6 (cit. on p. 11).

[7] B. Abi, T. Albahri, S. Al-Kilani, *et al.*, Muon $g-2$ collaboration, "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm," *Phys. Rev. Lett.*, vol. 126, p. 141 801, 14 Apr. 2021. DOI: 10.1103/PhysRevLett.126.141801 (cit. on p. 11).

[8] D. Hanneke, S. Fogwell, and G. Gabrielse, "New Measurement of the Electron Magnetic Moment and the Fine Structure Constant," *Phys. Rev. Lett.*, vol. 100, p. 120 801, 12 Mar. 2008. DOI: 10.1103/PhysRevLett.100.120801 (cit. on p. 11).

[9] L. Morel, Z. Yao, P. Cladé, *et al.*, "Determination of the fine-structure constant with an accuracy of 81 parts per trillion," *Nature*, vol. 588, no. 7836, pp. 61–65, 2020. DOI: 10.1038/s41586-020-2964-7 (cit. on p. 11).

[10]   S. Li, Y. Xiao, and J. M. Yang, "Can electron and muon $g - 2$ anomalies be jointly explained in SUSY?" *Eur. Phys. J. C*, vol. 82, no. 3, p. 276, 2022. DOI: 10.1140/epjc/s10052-022-10242-y. arXiv: 2107.04962 [hep-ph] (cit. on p. 11).

[11]   J. R. Ellis, "Limits of the standard model," in *PSI Zuoz Summer School on Exploring the Limits of the Standard Model*, Nov. 2002. arXiv: hep-ph/0211168 (cit. on p. 11).

[12]   M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, 1995, ISBN: 978-0-201-50397-5. DOI: 10.1201/9780429503559 (cit. on pp. 13, 20, 175).

[13]   M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 978-1-107-03426-6. DOI: 10.1017/CBO9781139525367 (cit. on p. 13).

[14]   D. Griffiths, *Introduction to elementary particles*, Wiley-VCH, 2008, ISBN: 978-3-527-40601-2. DOI: 10.1002/9783527618460 (cit. on p. 13).

[15]   S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967. DOI: 10.1103/PhysRevLett.19.1264 (cit. on p. 13).

[16]   R. L. Workman, `Particle Data Group` collaboration, "Review of Particle Physics," *Prog. Theor. Exp. Phys.*, vol. 2022, p. 083C01, 2022. DOI: 10.1093/ptep/ptac097 (cit. on pp. 14, 15, 17, 18, 22, 31–33, 39, 67, 128, 129, 149, 166, 167, 172).

[17]   C. Burgard, *Standard model of physics*, http://www.texample.net/tikz/examples/model-physics/, Accessed: 2023-01-20 (cit. on p. 14).

[18]   S. M. Bilenky, "Neutrinos: Majorana or Dirac?" *Universe*, vol. 6, no. 9, p. 134, 2020. DOI: 10.3390/universe6090134 (cit. on p. 16).

[19]   C. S. Wu, E. Ambler, R. W. Hayward, *et al.*, "Experimental Test of Parity Conservation in β Decay," *Phys. Rev.*, vol. 105, pp. 1413–1414, 1957. DOI: 10.1103/PhysRev.105.1413 (cit. on p. 16).

[20]   J. H. Christenson, J. W. Cronin, V. L. Fitch, *et al.*, "Evidence for the 2π Decay of the K_2^0 Meson," *Phys. Rev. Lett.*, vol. 13, pp. 138–140, 1964. DOI: 10.1103/PhysRevLett.13.138 (cit. on p. 16).

[21]   M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction," *Prog. Theor. Phys.*, vol. 49, pp. 652–657, 1973. DOI: 10.1143/PTP.49.652 (cit. on p. 16).

[22]   N. Cabibbo, "Unitary Symmetry and Leptonic Decays," *Phys. Rev. Lett.*, vol. 10, pp. 531–533, 1963. DOI: 10.1103/PhysRevLett.10.531 (cit. on p. 16).

[23]   W. Altmannshofer *et al.*, `Belle II` collaboration, "The Belle II Physics Book," *Prog. Theor. Exp. Phys.*, vol. 2019, no. 12, E. Kou and P. Urquijo, Eds., p. 123C01, 2019, [Erratum: PTEP 2020, 029201 (2020)]. DOI: 10.1093/ptep/ptz106. arXiv: 1808.10567 [hep-ex] (cit. on pp. 18, 37, 38, 41, 177).

[24] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak Interactions with Lepton-Hadron Symmetry," *Phys. Rev. D*, vol. 2, pp. 1285–1292, 1970. DOI: 10.1103/PhysRevD.2.1285 (cit. on p. 19).

[25] T. Mannel, "Theoretical aspects of $b \to s\gamma$ transitions," in *4th International Conference on B Physics and CP Violation (BCP 4)*, Mar. 2001, pp. 75–80. DOI: 10.1142/9789812811219_0015. arXiv: hep-ph/0103310 (cit. on p. 19).

[26] M. Misiak, A. Rehman, and M. Steinhauser, "Towards $\overline{B} \to X_s\gamma$ at the NNLO in QCD without interpolation in m_c," *J. High Energy Phys.*, vol. 06, p. 175, 2020. DOI: 10.1007/JHEP06(2020)175. arXiv: 2002.01548 [hep-ph] (cit. on pp. 19–21, 23, 29).

[27] M. Kaminski, M. Misiak, and M. Poradzinski, "Tree-level contributions to $B \to X_s\gamma$," *Phys. Rev. D*, vol. 86, p. 094004, 2012. DOI: 10.1103/PhysRevD.86.094004. arXiv: 1209.0965 [hep-ph] (cit. on pp. 19, 175).

[28] M. Misiak *et al.*, "Updated NNLO QCD predictions for the weak radiative B-meson decays," *Phys. Rev. Lett.*, vol. 114, no. 22, p. 221801, 2015. DOI: 10.1103/PhysRevLett.114.221801. arXiv: 1503.01789 [hep-ph] (cit. on pp. 19, 20, 23, 175).

[29] K. G. Chetyrkin, M. Misiak, and M. Munz, "Weak radiative B meson decay beyond leading logarithms," *Phys. Lett. B*, vol. 400, pp. 206–219, 1997, [Erratum: Phys.Lett.B 425, 414 (1998)]. DOI: 10.1016/S0370-2693(97)00324-9. arXiv: hep-ph/9612313 (cit. on pp. 20, 175).

[30] J. Charles *et al.*, "Current status of the Standard Model CKM fit and constraints on $\Delta F = 2$ New Physics," *Phys. Rev. D*, vol. 91, no. 7, p. 073007, 2015. DOI: 10.1103/PhysRevD.91.073007. arXiv: 1501.05013 [hep-ph] (cit. on p. 20).

[31] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, "Weak decays beyond leading logarithms," *Rev. Mod. Phys.*, vol. 68, pp. 1125–1144, 1996. DOI: 10.1103/RevModPhys.68.1125. arXiv: hep-ph/9512380 (cit. on p. 20).

[32] A. J. Buras, "Weak Hamiltonian, CP violation and rare decays," in *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*, Jun. 1998, pp. 281–539. arXiv: hep-ph/9806471 (cit. on pp. 20, 22).

[33] M. Misiak, "$\overline{B} \to X_s\gamma$: current status," *Acta Phys. Polon. B*, vol. 40, J. Gluza, K. Kolodziej, and M. Zralek, Eds., pp. 2987–2996, 2009. arXiv: 0911.1651 [hep-ph] (cit. on p. 20).

[34] M. Benzke, S. J. Lee, M. Neubert, *et al.*, "Factorization at Subleading Power and Irreducible Uncertainties in $\overline{B} \to X_s\gamma$ Decay," *J. High Energy Phys.*, vol. 08, p. 099, 2010. DOI: 10.1007/JHEP08(2010)099. arXiv: 1003.5012 [hep-ph] (cit. on p. 20).

[35] A. Gunawardana and G. Paz, "Reevaluating uncertainties in $\overline{B} \to X_s\gamma$ decay," *J. High Energy Phys.*, vol. 11, p. 141, 2019. DOI: 10.1007/JHEP11(2019)141. arXiv: 1908.02812 [hep-ph] (cit. on p. 20).

[36]   A. J. Buras, M. Misiak, M. Munz, *et al.*, "Theoretical uncertainties and phenomeno-
logical aspects of $B \rightarrow X_s \gamma$ decay," *Nucl. Phys. B*, vol. 424, pp. 374–398, 1994. DOI:
`10.1016/0550-3213(94)90299-2`. arXiv: `hep-ph/9311345` (cit. on pp. 21, 22).

[37]   H. M. Asatrian, T. Ewerth, H. Gabrielyan, *et al.*, "Charm quark mass dependence of
the electromagnetic dipole operator contribution to $\bar{B} \rightarrow X_s \gamma$ at $O(\alpha_s^2)$," *Phys. Lett.
B*, vol. 647, pp. 173–178, 2007. DOI: `10.1016/j.physletb.2007.02.027`. arXiv: `hep-
ph/0611123` (cit. on p. 21).

[38]   H. M. Asatrian, T. Ewerth, A. Ferroglia, *et al.*, "Complete (O_7, O_8) contribution to $B \rightarrow
X_s \gamma$ at order α_s^2," *Phys. Rev. D*, vol. 82, p. 074 006, 2010. DOI: `10.1103/PhysRevD.82.
074006`. arXiv: `1005.5587 [hep-ph]` (cit. on p. 21).

[39]   R. Boughezal, M. Czakon, and T. Schutzmeier, "NNLO fermionic corrections to the
charm quark mass dependent matrix elements in $\bar{B} \rightarrow X_s \gamma$," *J. High Energy Phys.*,
vol. 09, p. 072, 2007. DOI: `10.1088/1126-6708/2007/09/072`. arXiv: `0707.3090 [hep-
ph]` (cit. on p. 21).

[40]   A. Ferroglia and U. Haisch, "Chromomagnetic Dipole-Operator Corrections in $\bar{B} \rightarrow
X_s \gamma$ at $O(\beta_0 \alpha_s^2)$," *Phys. Rev. D*, vol. 82, p. 094 012, 2010. DOI: `10.1103/PhysRevD.82.
094012`. arXiv: `1009.2144 [hep-ph]` (cit. on p. 21).

[41]   M. Misiak and M. Poradzinski, "Completing the Calculation of BLM corrections to
$\bar{B} \rightarrow X_s \gamma$," *Phys. Rev. D*, vol. 83, p. 014 024, 2011. DOI: `10.1103/PhysRevD.83.014024`.
arXiv: `1009.5685 [hep-ph]` (cit. on p. 21).

[42]   T. Huber, M. Poradziński, and J. Virto, "Four-body contributions to $\overline{B} \rightarrow X_s \gamma$ at NLO,"
J. High Energy Phys., vol. 01, p. 115, 2015. DOI: `10.1007/JHEP01(2015)115`. arXiv:
`1411.7677 [hep-ph]` (cit. on p. 21).

[43]   M. Czakon, U. Haisch, and M. Misiak, "Four-Loop Anomalous Dimensions for Ra-
diative Flavour-Changing Decays," *J. High Energy Phys.*, vol. 03, p. 008, 2007. DOI: `10.
1088/1126-6708/2007/03/008`. arXiv: `hep-ph/0612329` (cit. on p. 21).

[44]   M. Misiak, A. Rehman, and M. Steinhauser, "On charm-mass dependent NNLO cor-
rections to $\mathcal{B}(\bar{B} \rightarrow X_s \gamma)$," *PoS*, vol. RADCOR2019, p. 033, 2019. DOI: `10.22323/1.375.
0033`. arXiv: `2002.03021 [hep-ph]` (cit. on p. 21).

[45]   P. Gambino and M. Misiak, "Quark mass effects in $\bar{B} \rightarrow X_s \gamma$," *Nucl. Phys. B*, vol. 611,
pp. 338–366, 2001. DOI: `10.1016/S0550-3213(01)00347-9`. arXiv: `hep-ph/0104034`
(cit. on pp. 21, 23).

[46]   A. Alberti, P. Gambino, K. J. Healey, *et al.*, "Precision Determination of the Cabibbo-
Kobayashi-Maskawa Element V_{cb}," *Phys. Rev. Lett.*, vol. 114, no. 6, p. 061 802, 2015. DOI:
`10.1103/PhysRevLett.114.061802`. arXiv: `1411.6560 [hep-ph]` (cit. on p. 22).

[47]   S. Bertolini, F. Borzumati, and A. Masiero, "QCD Enhancement of Radiative b Decays,"
Phys. Rev. Lett., vol. 59, p. 180, 1987. DOI: `10.1103/PhysRevLett.59.180` (cit. on p. 22).

[48] M. Neubert, "Soft-collinear factorization and the calculation of the $B \to X_s\gamma$ rate," *AIP Conf. Proc.*, vol. 756, no. 1, J. Tran Thanh Van, Ed., pp. 91–100, 2005. DOI: 10.1063/1.1920936. arXiv: hep-ph/0408208 (cit. on p. 23).

[49] Z. Ligeti, I. W. Stewart, and F. J. Tackmann, "Treating the *b* quark distribution function with reliable uncertainties," *Phys. Rev. D*, vol. 78, p. 114014, 2008. DOI: 10.1103/PhysRevD.78.114014. arXiv: 0807.1926 [hep-ph] (cit. on pp. 23–25, 27).

[50] A. Limosani *et al.*, Belle collaboration, "Measurement of Inclusive Radiative *B*-meson Decays with a Photon Energy Threshold of 1.7 GeV," *Phys. Rev. Lett.*, vol. 103, p. 241801, 2009. DOI: 10.1103/PhysRevLett.103.241801. arXiv: 0907.1384 [hep-ex] (cit. on pp. 24, 33, 35, 37, 38).

[51] M. Neubert, "Analysis of the photon spectrum in inclusive $B \to X_s\gamma$ decays," *Phys. Rev. D*, vol. 49, pp. 4623–4633, 1994. DOI: 10.1103/PhysRevD.49.4623. arXiv: hep-ph/9312311 (cit. on pp. 24, 25).

[52] I. I. Y. Bigi, M. A. Shifman, N. G. Uraltsev, *et al.*, "On the motion of heavy quarks inside hadrons: Universal distributions and inclusive decays," *Int. J. Mod. Phys. A*, vol. 9, pp. 2467–2504, 1994. DOI: 10.1142/S0217751X94000996. arXiv: hep-ph/9312359 (cit. on p. 24).

[53] M. Neubert, "Subleading shape functions and the determination of $|V_{ub}|$," *Phys. Lett. B*, vol. 543, pp. 269–275, 2002. DOI: 10.1016/S0370-2693(02)02462-0. arXiv: hep-ph/0207002 (cit. on p. 25).

[54] T. Mannel and M. Neubert, "Resummation of nonperturbative corrections to the lepton spectrum in inclusive $B \to X_q\ell\bar{\nu}$ decays," *Phys. Rev. D*, vol. 50, pp. 2037–2047, 3 Aug. 1994. DOI: 10.1103/PhysRevD.50.2037 (cit. on p. 25).

[55] C. W. Bauer, "Corrections to moments of the photon spectrum in the inclusive decay $B \to X_s\gamma$," *Phys. Rev. D*, vol. 57, pp. 5611–5619, 1998, [Erratum: Phys.Rev.D 60, 099907 (1999)]. DOI: 10.1103/PhysRevD.60.099907. arXiv: hep-ph/9710513 (cit. on p. 26).

[56] A. Kapustin, Z. Ligeti, and H. D. Politzer, "Leading logarithms of the *b* quark mass in inclusive $B \to X_s\gamma$ decay," *Phys. Lett. B*, vol. 357, pp. 653–658, 1995. DOI: 10.1016/0370-2693(95)00962-K. arXiv: hep-ph/9507248 (cit. on p. 26).

[57] D. Benson, I. I. Bigi, and N. Uraltsev, "On the photon energy moments and their 'bias' corrections in $B \to X_s\gamma$," *Nucl. Phys. B*, vol. 710, pp. 371–401, 2005. DOI: 10.1016/j.nuclphysb.2004.12.035. arXiv: hep-ph/0410080 (cit. on p. 26).

[58] B. O. Lange, M. Neubert, and G. Paz, "Theory of charmless inclusive *B* decays and the extraction of $|V_{ub}|$," *Phys. Rev. D*, vol. 72, p. 073006, 2005. DOI: 10.1103/PhysRevD.72.073006. arXiv: hep-ph/0504071 (cit. on p. 26).

[59] J. R. Andersen and E. Gardi, "Inclusive spectra in charmless semileptonic *B* decays by dressed gluon exponentiation," *J. High Energy Phys.*, vol. 01, p. 097, 2006. DOI: 10.1088/1126-6708/2006/01/097. arXiv: hep-ph/0509360 (cit. on p. 26).

[60] P. Gambino, P. Giordano, G. Ossola, *et al.*, "Inclusive semileptonic B decays and the determination of $|V_{ub}|$," *J. High Energy Phys.*, vol. 10, p. 058, 2007. DOI: 10.1088/1126-6708/2007/10/058. arXiv: 0707.2493 [hep-ph] (cit. on p. 26).

[61] U. Aglietti, F. Di Lodovico, G. Ferrera, *et al.*, "Inclusive measure of $|V_{ub}|$ with the analytic coupling model," *Eur. Phys. J. C*, vol. 59, pp. 831–840, 2009. DOI: 10.1140/epjc/s10052-008-0817-x. arXiv: 0711.0860 [hep-ph] (cit. on p. 26).

[62] F. U. Bernlochner, H. Lacker, Z. Ligeti, *et al.*, SIMBA collaboration, "Precision Global Determination of the $B \to X_s \gamma$ Decay Rate," *Phys. Rev. Lett.*, vol. 127, no. 10, p. 102 001, 2021. DOI: 10.1103/PhysRevLett.127.102001. arXiv: 2007.04320 [hep-ph] (cit. on pp. 26, 67, 168, 169, 172).

[63] A. L. Kagan and M. Neubert, "QCD anatomy of $B \to X_s \gamma$ decays," *Eur. Phys. J. C*, vol. 7, pp. 5–27, 1999. DOI: 10.1007/s100529800959. arXiv: hep-ph/9805303 (cit. on pp. 26, 27, 66).

[64] A. Ryd, D. Lange, N. Kuznetsova, *et al.*, "EvtGen: A Monte Carlo Generator for *B*-Physics," May 2005 (cit. on pp. 26, 65, 66).

[65] F. Borzumati and C. Greub, "2HDMs predictions for $\bar{B} \to X_s \gamma$ in NLO QCD," *Phys. Rev. D*, vol. 58, p. 074 004, 1998. DOI: 10.1103/PhysRevD.58.074004. arXiv: hep-ph/9802391 (cit. on p. 28).

[66] C. Bobeth, M. Misiak, and J. Urban, "Matching conditions for $b \to s \gamma$ and $b \to s$ gluon in extensions of the standard model," *Nucl. Phys. B*, vol. 567, pp. 153–185, 2000. DOI: 10.1016/S0550-3213(99)00688-4. arXiv: hep-ph/9904413 (cit. on p. 28).

[67] T. Hermann, M. Misiak, and M. Steinhauser, "$\bar{B} \to X_s \gamma$ in the Two Higgs Doublet Model up to Next-to-Next-to-Leading Order in QCD," *J. High Energy Phys.*, vol. 11, p. 036, 2012. DOI: 10.1007/JHEP11(2012)036. arXiv: 1208.2788 [hep-ph] (cit. on pp. 28, 29).

[68] F. Borzumati, C. Greub, and Y. Yamada, "Beyond leading order corrections to $\bar{B} \to X_s \gamma$ at large $\tan \beta$: The Charged Higgs contribution," *Phys. Rev. D*, vol. 69, p. 055 005, 2004. DOI: 10.1103/PhysRevD.69.055005. arXiv: hep-ph/0311151 (cit. on p. 28).

[69] G. Degrassi, P. Gambino, and P. Slavich, "QCD corrections to radiative *B* decays in the MSSM with minimal flavor violation," *Phys. Lett. B*, vol. 635, pp. 335–342, 2006. DOI: 10.1016/j.physletb.2006.02.067. arXiv: hep-ph/0601135 (cit. on p. 28).

[70] A. Freitas, E. Gasser, and U. Haisch, "Supersymmetric large $\tan \beta$ corrections to $\Delta M_{(d,s)}$ and $B_{(d,s)} \to \mu^+ \mu^-$ revisited," *Phys. Rev. D*, vol. 76, p. 014 016, 2007. DOI: 10.1103/PhysRevD.76.014016. arXiv: hep-ph/0702267 (cit. on p. 28).

[71] M. Ciuchini, A. Masiero, P. Paradisi, *et al.*, "Soft SUSY breaking grand unification: Leptons versus quarks on the flavor playground," *Nucl. Phys. B*, vol. 783, pp. 112–142, 2007. DOI: 10.1016/j.nuclphysb.2007.05.032. arXiv: hep-ph/0702144 (cit. on p. 28).

[72] A. J. Buras, A. Poschenrieder, M. Spranger, *et al.*, "The Impact of universal extra dimensions on $B \to X_s \gamma$, $B \to X_s$gluon, $B \to X_s \mu^+ \mu^-$, $K_L^0 \to \pi^0 e^+ e^-$ and ϵ' / ϵ," *Nucl. Phys. B*, vol. 678, pp. 455–490, 2004. DOI: 10.1016/j.nuclphysb.2003.11.010. arXiv: hep-ph/0306158 (cit. on p. 28).

[73] K. Agashe, G. Perez, and A. Soni, "Flavor structure of warped extra dimension models," *Phys. Rev. D*, vol. 71, p. 016 002, 2005. DOI: 10.1103/PhysRevD.71.016002. arXiv: hep-ph/0408134 (cit. on p. 28).

[74] U. Haisch and A. Weiler, "Bound on minimal universal extra dimensions from $\bar{B} \to X_s \gamma$," *Phys. Rev. D*, vol. 76, p. 034 014, 2007. DOI: 10.1103/PhysRevD.76.034014. arXiv: hep-ph/0703064 (cit. on p. 28).

[75] A. Freitas and U. Haisch, "$\bar{B} \to X_s \gamma$ in two universal extra dimensions," *Phys. Rev. D*, vol. 77, p. 093 008, 2008. DOI: 10.1103/PhysRevD.77.093008. arXiv: 0801.4346 [hep-ph] (cit. on p. 28).

[76] A. J. Buras, A. Poschenrieder, S. Uhlig, *et al.*, "Rare *K* and *B* Decays in the Littlest Higgs Model without T^- Parity," *J. High Energy Phys.*, vol. 11, p. 062, 2006. DOI: 10.1088/1126-6708/2006/11/062. arXiv: hep-ph/0607189 (cit. on p. 28).

[77] M. Blanke, A. J. Buras, A. Poschenrieder, *et al.*, "Particle-Antiparticle Mixing, ϵ_K, $\Delta\Gamma_q$, A_{SL}^q, $A_{CP}(B_d \to \psi K_S^0)$, $A_{CP}(B_s \to \psi\phi)$ and $B \to X_{s/d}\gamma$ in the Littlest Higgs Model with T-Parity," *J. High Energy Phys.*, vol. 12, p. 003, 2006. DOI: 10.1088/1126-6708/2006/12/003. arXiv: hep-ph/0605214 (cit. on p. 28).

[78] C. Promberger, S. Schatt, F. Schwab, *et al.*, "Bounding the Minimal 331 Model through the Decay $B \to X_s\gamma$," *Phys. Rev. D*, vol. 77, p. 115 022, 2008. DOI: 10.1103/PhysRevD.77.115022. arXiv: 0802.0949 [hep-ph] (cit. on p. 28).

[79] G. C. Branco, P. M. Ferreira, L. Lavoura, *et al.*, "Theory and phenomenology of two-Higgs-doublet models," *Phys. Rept.*, vol. 516, pp. 1–102, 2012. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph] (cit. on p. 28).

[80] J. L. Feng and T. Moroi, "Determining $\tan\beta$ from the SUSY Higgs sector at future e^+e^- colliders," *Phys. Rev. D*, vol. 56, pp. 5962–5980, 1997. DOI: 10.1103/PhysRevD.56.5962. arXiv: hep-ph/9612333 (cit. on p. 29).

[81] M. Ciuchini, G. Degrassi, P. Gambino, *et al.*, "Next-to-leading QCD corrections to $B \to X_s\gamma$: Standard model and two Higgs doublet model," *Nucl. Phys. B*, vol. 527, pp. 21–43, 1998. DOI: 10.1016/S0550-3213(98)00244-2. arXiv: hep-ph/9710335 (cit. on p. 29).

[82] M. Misiak and M. Steinhauser, "Weak radiative decays of the *B* meson and bounds on $M_{H^\pm}$ in the Two-Higgs-Doublet Model," *Eur. Phys. J. C*, vol. 77, no. 3, p. 201, 2017. DOI: 10.1140/epjc/s10052-017-4776-y. arXiv: 1702.04571 [hep-ph] (cit. on pp. 29, 30).

[83] O. Atkinson, M. Black, A. Lenz, *et al.*, "Cornering the Two Higgs Doublet Model Type II," *J. High Energy Phys.*, vol. 04, p. 172, 2022. DOI: 10.1007/JHEP04(2022)172. arXiv: 2107.05650 [hep-ph] (cit. on p. 29).

[84] Y. Amhis *et al.*, "Averages of *b*-hadron, *c*-hadron, and τ-lepton properties as of 2021," Jun. 2022. arXiv: 2206.07501 [hep-ex] (cit. on pp. 31–33, 67).

[85] V. Bellée, P. Pais, A. Puig Navarro, *et al.*, "Using an amplitude analysis to measure the photon polarisation in $B \to K\pi\pi\gamma$ decays," *Eur. Phys. J. C*, vol. 79, no. 7, p. 622, 2019. DOI: 10.1140/epjc/s10052-019-7127-3. arXiv: 1902.09201 [hep-ph] (cit. on p. 33).

[86] O. Buchmuller and H. Flacher, "Fit to moment from $B \to X_c\ell\bar{\nu}$ and $B \to X_s\gamma$ decays using heavy quark expansions in the kinetic scheme," *Phys. Rev. D*, vol. 73, p. 073 008, 2006. DOI: 10.1103/PhysRevD.73.073008. arXiv: hep-ph/0507253 (cit. on pp. 33, 162, 163).

[87] S. Chen *et al.*, CLEO collaboration, "Branching fraction and photon energy spectrum for $b \to s\gamma$," *Phys. Rev. Lett.*, vol. 87, p. 251 807, 2001. DOI: 10.1103/PhysRevLett.87.251807. arXiv: hep-ex/0108032 (cit. on pp. 33, 35).

[88] B. Aubert *et al.*, BaBar collaboration, "Measurement of the $B \to X_s\gamma$ branching fraction and photon energy spectrum using the recoil method," *Phys. Rev. D*, vol. 77, p. 051 103, 2008. DOI: 10.1103/PhysRevD.77.051103. arXiv: 0711.4889 [hep-ex] (cit. on pp. 33, 37, 38, 63, 105, 166).

[89] J. P. Lees *et al.*, BaBar collaboration, "Precision Measurement of the $B \to X_s\gamma$ Photon Energy Spectrum, Branching Fraction, and Direct CP Asymmetry $A_{CP}(B \to X_{s+d}\gamma)$," *Phys. Rev. Lett.*, vol. 109, p. 191 801, 2012. DOI: 10.1103/PhysRevLett.109.191801. arXiv: 1207.2690 [hep-ex] (cit. on pp. 33, 37).

[90] J. P. Lees *et al.*, BaBar collaboration, "Exclusive Measurements of $b \to s\gamma$ Transition Rate and Photon Energy Spectrum," *Phys. Rev. D*, vol. 86, p. 052 012, 2012. DOI: 10.1103/PhysRevD.86.052012. arXiv: 1207.2520 [hep-ex] (cit. on p. 33).

[91] T. Saito *et al.*, Belle collaboration, "Measurement of the $\bar{B} \to X_s\gamma$ Branching Fraction with a Sum of Exclusive Decays," *Phys. Rev. D*, vol. 91, no. 5, p. 052 004, 2015. DOI: 10.1103/PhysRevD.91.052004. arXiv: 1411.7198 [hep-ex] (cit. on pp. 33–35).

[92] A. Abdesselam *et al.*, Belle collaboration, "Measurement of the inclusive $B \to X_{s+d}\gamma$ branching fraction, photon energy spectrum and HQE parameters," in *38th International Conference on High Energy Physics*, Aug. 2016. arXiv: 1608.02344 [hep-ex] (cit. on pp. 33, 37).

[93] J. P. Lees *et al.*, BaBar collaboration, "Measurements of direct CP asymmetries in $B \to X_s\gamma$ decays using sum of exclusive decays," *Phys. Rev. D*, vol. 90, no. 9, p. 092 001, 2014. DOI: 10.1103/PhysRevD.90.092001. arXiv: 1406.0534 [hep-ex] (cit. on p. 34).

[94] P. del Amo Sanchez *et al.*, BaBar collaboration, "Study of $B \to X\gamma$ Decays and Determination of $|V_{td}/V_{ts}|$," *Phys. Rev. D*, vol. 82, p. 051 101, 2010. DOI: 10.1103/PhysRevD.82.051101. arXiv: 1005.4087 [hep-ex] (cit. on p. 34).

[95] "Observation of inclusive $B \to X_s\gamma$ decays in early Phase III data (untagged)," no. BELLE2-NOTE-PH-2022-035, Mar. 2021 (cit. on pp. 35, 36).

[96] A. J. Bevan *et al.*, `BaBar`, `Belle` collaboration, "The Physics of the B Factories," *Eur. Phys. J. C*, vol. 74, p. 3026, 2014. DOI: `10.1140/epjc/s10052-014-3026-9`. arXiv: `1406.6311 [hep-ex]` (cit. on pp. 37, 40, 72, 189, 193, 197).

[97] B. Aubert *et al.*, `BaBar` collaboration, "Search for the $B^+ \to K^+ \nu \bar{\nu}$ Decay Using Semi-Leptonic Tags," Nov. 2009. arXiv: `0911.1988 [hep-ex]` (cit. on p. 37).

[98] L. Aggarwal *et al.*, `Belle II` collaboration, "Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond," Jul. 2022. arXiv: `2207.06307 [hep-ex]` (cit. on pp. 38, 167, 255).

[99] D. Boutigny *et al.*, `BaBar` collaboration, "BaBar technical design report," Mar. 1995 (cit. on p. 39).

[100] A. Abashian *et al.*, `Belle` collaboration, "The Belle Detector," *Nucl. Instrum. Meth. A*, vol. 479, pp. 117–232, 2002. DOI: `10.1016/S0168-9002(01)02013-7` (cit. on p. 39).

[101] A. A. Alves Jr. *et al.*, `LHCb` collaboration, "The LHCb Detector at the LHC," *J. Instrum.*, vol. 3, S08005, 2008. DOI: `10.1088/1748-0221/3/08/S08005` (cit. on p. 39).

[102] D. Andrews *et al.*, `CLEO` collaboration, "The CLEO Detector," *Nucl. Instrum. Meth.*, vol. 211, p. 47, 1983. DOI: `10.1016/0167-5087(83)90556-2` (cit. on p. 39).

[103] H. Albrecht *et al.*, `ARGUS` collaboration, "ARGUS: A Universal Detector at DORIS-II," *Nucl. Instrum. Meth. A*, vol. 275, pp. 1–48, 1989. DOI: `10.1016/0168-9002(89)90334-3` (cit. on p. 39).

[104] K. Akai, K. Furukawa, and H. Koiso, `SuperKEKB` collaboration, "SuperKEKB Collider," *Nucl. Instrum. Meth. A*, vol. 907, pp. 188–199, 2018. DOI: `10.1016/j.nima.2018.08.017`. arXiv: `1809.01958 [physics.acc-ph]` (cit. on pp. 39, 40).

[105] K. Oide, "KEKB *B*-factory, the luminosity frontier," *Prog. Theor. Phys.*, vol. 122, pp. 69–80, 2009. DOI: `10.1143/PTP.122.69` (cit. on p. 39).

[106] Y. Funakoshi *et al.*, "The SuperKEKB Has Broken the World Record of the Luminosity," *JACoW*, vol. IPAC2022, pp. 1–5, 2022. DOI: `10.18429/JACoW-IPAC2022-MOPLXGD1` (cit. on p. 41).

[107] T. Abe *et al.*, `Belle II` collaboration, "Belle II Technical Design Report," Nov. 2010. arXiv: `1011.0352 [physics.ins-det]` (cit. on pp. 41, 43, 46, 47).

[108] Belle II collaboration, *Electrons and positrons collide for the first time in the SuperKEKB accelerator*, `https://www.kek.jp/en/newsroom/2018/04/26/0700/`, accessed on 16/01/23, Apr. 2018 (cit. on p. 42).

[109] J. Kemmer and G. Lutz, "New detector concepts," *Nucl. Instrum. Meth. A*, vol. 253, pp. 365–377, 1987. DOI: `10.1016/0168-9002(87)90518-3` (cit. on p. 43).

[110] B. Wang *et al.*, `Belle II DEPFET`, `PXD` collaboration, "Operational experience of the Belle II pixel detector," *Nucl. Instrum. Meth. A*, vol. 1032, p. 166 631, 2022. DOI: `10.1016/j.nima.2022.166631` (cit. on p. 43).

[111] C. Irmler *et al.*, `Belle II SVD` collaboration, "The silicon vertex detector of the Belle II experiment," *Nucl. Instrum. Meth. A*, vol. 1045, p. 167 578, 2023. DOI: `10.1016/j.nima.2022.167578` (cit. on pp. 43, 44).

[112] N. Taniguchi, `Belle II` collaboration, "Central Drift Chamber for Belle II," *J. Instrum.*, vol. 12, no. 06, L. Shekhtman, Ed., p. C06014, 2017. DOI: `10.1088/1748-0221/12/06/C06014` (cit. on p. 44).

[113] J. Kandra, T. Bilka, L. Kapitanova, *et al.*, "Calibration and alignment of the Belle II tracker," in *Connecting the Dots and Workshop on Intelligent Trackers*, Oct. 2019. arXiv: `1910.06289 [physics.ins-det]` (cit. on p. 44).

[114] V. Bertacchi *et al.*, `Belle II Tracking Group` collaboration, "Track finding at Belle II," *Comput. Phys. Commun.*, vol. 259, p. 107 610, 2021. DOI: `10.1016/j.cpc.2020.107610`. arXiv: `2003.12466 [physics.ins-det]` (cit. on pp. 45, 47–49, 255).

[115] Y. Yusa, `Belle II` collaboration, "ARICH for Belle II," *J. Instrum.*, vol. 9, no. 10, p. C10015, 2014. DOI: `10.1088/1748-0221/9/10/C10015` (cit. on pp. 45, 46).

[116] J. Fast, `Belle IIBarrel Particle Identification Group` collaboration, "The Belle II imaging Time-of-Propagation (iTOP) detector," *Nucl. Instrum. Meth. A*, vol. 876, P. Krizan, S. Korpar, G. Hallewell, *et al.*, Eds., pp. 145–148, 2017. DOI: `10.1016/j.nima.2017.02.045` (cit. on pp. 45, 46).

[117] M. Yonenaga *et al.*, "Performance evaluation of the aerogel RICH counter for the Belle II spectrometer using early beam collision data," *PTEP*, vol. 2020, no. 9, 093H01, 2020. DOI: `10.1093/ptep/ptaa090`. arXiv: `2008.06251 [physics.ins-det]` (cit. on p. 46).

[118] K. Kojima, `Belle II TOP group` collaboration, "The operation and performance of the TOP detector at the Belle II experiment," *PoS*, vol. EPS-HEP2021, p. 803, 2022. DOI: `10.22323/1.398.0803` (cit. on p. 46).

[119] K. Miyabayashi, "Belle II electromagnetic calorimeter and its performance during early SuperKEKB operation," *J. Instrum.*, vol. 15, no. 10, p. C10016, 2020. DOI: `10.1088/1748-0221/15/10/C10016` (cit. on pp. 46, 47).

[120] V. Aulchenko *et al.*, "Electromagnetic calorimeter for Belle II," *J. Phys. Conf. Ser.*, vol. 587, no. 1, p. 012 045, 2015. DOI: `10.1088/1742-6596/587/1/012045` (cit. on p. 46).

[121] T. Aushev *et al.*, "A scintillator based endcap K_L and muon detector for the Belle II experiment," *Nucl. Instrum. Meth. A*, vol. 789, pp. 134–142, 2015. DOI: `10.1016/j.nima.2015.03.060`. arXiv: `1406.3267 [physics.ins-det]` (cit. on pp. 47, 48).

[122] J.-F. Krohn, "K_L^0 Identification Studies for BELLE II," University of Hamburg, M.S. thesis, University of Hamburg, 2016, 98 p. DOI: `10.3204/PUBDB-2017-00968` (cit. on pp. 47, 48).

[123] T. Kuhr, C. Pulvermacher, M. Ritter, *et al.*, `Belle IIFramework Software Group` collaboration, "The Belle II Core Software," *Comput. Softw. Big Sci.*, vol. 3, no. 1, p. 1, 2019. DOI: `10.1007/s41781-018-0017-9`. arXiv: `1809.04299` [`physics.comp-ph`] (cit. on p. 47).

[124] M. Aaboud *et al.*, `ATLAS` collaboration, "Measurement of the Higgs boson mass in the $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels with $\sqrt{s}$ = 13 TeV pp collisions using the ATLAS detector," *Phys. Lett. B*, vol. 784, pp. 345–366, 2018. DOI: `10.1016/j.physletb.2018.07.050`. arXiv: `1806.00242` [`hep-ex`] (cit. on p. 51).

[125] A. M. Sirunyan *et al.*, `CMS` collaboration, "A measurement of the Higgs boson mass in the diphoton decay channel," *Phys. Lett. B*, vol. 805, p. 135 425, 2020. DOI: `10.1016/j.physletb.2020.135425`. arXiv: `2002.06398` [`hep-ex`] (cit. on p. 51).

[126] O. Behnke, K. Kröninger, T. Schörner-Sadenius, *et al.*, Eds., *Data analysis in high energy physics: A practical guide to statistical methods*, Wiley-VCH, 2013, ISBN: 978-3-527-41058-3. DOI: `10.1002/9783527653416` (cit. on pp. 51, 55, 58).

[127] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse*, Vieweg+Teubner, 1998, ISBN: 978-3-663-05690-4. DOI: `10.1007/978-3-663-05690-4` (cit. on pp. 51, 58, 59).

[128] G. Bohm and G. Zech, *Introduction to statistics and data analysis for physicists*, DESY, 2014, ISBN: 978-3-935702-88-1. DOI: `10.3204/DESY-BOOK/statistics` (cit. on p. 51).

[129] F. James, *Statistical methods in experimental physics (2nd edition)*, en, World Scientific Publishing, 2006, ISBN: 978-981-270-527-3. DOI: `10.1142/6096` (cit. on pp. 51, 52).

[130] R. J. Barlow, "Extended maximum likelihood," *Nucl. Instrum. Meth. A*, vol. 297, pp. 496–506, 1990. DOI: `10.1016/0168-9002(90)91334-8` (cit. on p. 51).

[131] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *Annals Math. Statist.*, vol. 9, no. 1, pp. 60–62, 1938. DOI: `10.1214/aoms/1177732360` (cit. on p. 53).

[132] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations," *Comput. Phys. Commun.*, vol. 10, pp. 343–367, 1975. DOI: `10.1016/0010-4655(75)90039-9` (cit. on p. 53).

[133] F. James, "MINUIT: Function Minimization and Error Analysis Reference Manual," 1998, CERN Program Library Long Writeups (cit. on p. 53).

[134] J. Eschle, A. Puig Navarro, R. Silva Coutinho, *et al.*, "zfit: Scalable pythonic fitting," *SoftwareX*, vol. 11, p. 100 508, 2020. DOI: `https://doi.org/10.1016/j.softx.2020.100508` (cit. on p. 54).

[135] H. Dembinski and P. O. et al., "scikit-hep/iminuit," Dec. 2020. DOI: `10.5281/zenodo.3949207` (cit. on p. 54).

[136] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference and prediction*, en, New York, NY: Springer, 2001, ISBN: 9780387952840. DOI: 10.1007/978-0-387-84858-7 (cit. on p. 55).

[137] C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2016 (cit. on p. 55).

[138] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001. DOI: 10.1214/aos/1013203451 (cit. on p. 58).

[139] J. H. Friedman, "Stochastic gradient boosting," *Comput. Stat. Data. Anal.*, vol. 38, no. 4, pp. 367–378, 2002, Nonlinear Methods and Data Mining. DOI: 10.1016/S0167-9473(01)00065-2 (cit. on p. 58).

[140] T. Keck, "FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment," *Comput. Softw. Big Sci.*, vol. 1, no. 1, p. 2, 2017. DOI: 10.1007/s41781-017-0002-8 (cit. on pp. 58, 92).

[141] S. Schmitt, "Data Unfolding Methods in High Energy Physics," *EPJ Web Conf.*, vol. 137, Y. Foka, N. Brambilla, and V. Kovalenko, Eds., p. 11 008, 2017. DOI: 10.1051/epjconf/201713711008. arXiv: 1611.01927 [physics.data-an] (cit. on pp. 58, 60).

[142] G. Cowan, "A survey of unfolding methods for particle physics," *Conf. Proc. C*, vol. 0203181, M. R. Whalley and L. Lyons, Eds., pp. 248–257, 2002 (cit. on p. 58).

[143] L. Brenner, R. Balasubramanian, C. Burgard, *et al.*, "Comparison of unfolding methods using RooFitUnfold," *Int. J. Mod. Phys. A*, vol. 35, no. 24, p. 2 050 145, 2020. DOI: 10.1142/S0217751X20501456. arXiv: 1910.14654 [physics.data-an] (cit. on pp. 58, 61).

[144] J. Walsh and L. M. Delves, *Numerical solution of integral equations*, en, Oxford University Press, 1974, ISBN: 9780198533429 (cit. on p. 59).

[145] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics," *J. Instrum.*, vol. 7, T10003, 2012. DOI: 10.1088/1748-0221/7/10/T10003. arXiv: 1205.6201 [physics.data-an] (cit. on p. 61).

[146] G. D'Agostini, "Improved iterative Bayesian unfolding," in *Alliance Workshop on Unfolding and Data Correction*, Oct. 2010. arXiv: 1010.0632 [physics.data-an] (cit. on p. 61).

[147] G. D'Agostini, "A Multidimensional unfolding method based on Bayes' theorem," *Nucl. Instrum. Meth. A*, vol. 362, pp. 487–498, 1995. DOI: 10.1016/0168-9002(95)00274-X (cit. on p. 61).

[148] B. Malaescu, "An Iterative, dynamically stabilized method of data unfolding," Jul. 2009. arXiv: 0907.3791 [physics.data-an] (cit. on p. 61).

[149] A. Hocker and V. Kartvelishvili, "SVD approach to data unfolding," *Nucl. Instrum. Meth. A*, vol. 372, pp. 469–481, 1996. DOI: 10.1016/0168-9002(95)01478-0. arXiv: hep-ph/9509307 (cit. on p. 61).

[150] A. Roodman, "Blind analysis in particle physics," *eConf*, vol. C030908, L. Lyons, R. P. Mount, and R. Reitmeyer, Eds., TUIT001, 2003. arXiv: physics/0312102 (cit. on p. 61).

[151] B. F. L. Ward, S. Jadach, and Z. Was, "Precision calculation for $e^+e^- \to 2f$: The KK MC project," *Nucl. Phys. B Proc. Suppl.*, vol. 116, J. Blumlein, F. Jegerlehner, T. Riemann, *et al.*, Eds., pp. 73–77, 2003. DOI: 10.1016/S0920-5632(03)80147-0. arXiv: hep-ph/0211132 (cit. on p. 65).

[152] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph] (cit. on p. 65).

[153] S. Agostinelli *et al.*, GEANT4 collaboration, "GEANT4 – a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 65).

[154] C. Ramirez, J. F. Donoghue, and G. Burdman, "Semileptonic $b \to u$ decay," *Phys. Rev. D*, vol. 41, p. 1496, 1990. DOI: 10.1103/PhysRevD.41.1496 (cit. on p. 66).

[155] T. Keck, "Machine learning algorithms for the Belle II experiment and their validation on Belle data," Ph.D. dissertation, KIT, Karlsruhe, 2017. DOI: 10.1007/978-3-319-98249-6 (cit. on p. 70).

[156] T. Keck *et al.*, "The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment," *Comput. Softw. Big Sci.*, vol. 3, no. 1, p. 6, 2019. DOI: 10.1007/s41781-019-0021-8. arXiv: 1807.08680 [hep-ex] (cit. on p. 70).

[157] J.-F. Krohn *et al.*, Belle II analysis software group collaboration, "Global decay chain vertex fitting at Belle II," *Nucl. Instrum. Meth. A*, vol. 976, p. 164269, 2020. DOI: 10.1016/j.nima.2020.164269. arXiv: 1901.11198 [hep-ex] (cit. on p. 72).

[158] A. Hershenhorn, "Search for axion like particles with the BaBar detector and photon hadron separation using Zernike moments at Belle II," Ph.D. dissertation, University of British Columbia, 2021. DOI: http://dx.doi.org/10.14288/1.0398055 (cit. on pp. 77, 78).

[159] G. Punzi, "Sensitivity of searches for new signals and its optimization," *eConf*, vol. C030908, L. Lyons, R. P. Mount, and R. Reitmeyer, Eds., MODT002, 2003. arXiv: physics/0308063 (cit. on p. 85).

[160] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, 1991. DOI: 10.1109/18.61115 (cit. on p. 88).

[161] F. Abudinén *et al.*, Belle II collaboration, "A calibration of the Belle II hadronic tag-side reconstruction algorithm with $B \to X\ell\nu$ decays," Aug. 2020. arXiv: 2008.06096 [hep-ex] (cit. on pp. 121, 122).

[162] F. Abudinén *et al.*, Belle II collaboration, "Measurement of the data to MC ratio of photon reconstruction efficiency of the Belle II calorimeter using radiative muon pair events," no. BELLE2-NOTE-PL-2021-008, Sep. 2021 (cit. on pp. 125, 255).

[163] F. Abudinén *et al.*, `Belle II` collaboration, "Measurement of the photon-energy spectrum in inclusive $B \to X_s \gamma$ decays identified using hadronic decays of the recoil B meson in 2019-2021 Belle II data," Oct. 2022. arXiv: `2210.10220 [hep-ex]` (cit. on pp. 165, 171, 255).

[164] J. Haller, A. Hoecker, R. Kogler, *et al.*, "Update of the global electroweak fit and constraints on two-Higgs-doublet models," *Eur. Phys. J. C*, vol. 78, no. 8, p. 675, 2018. DOI: `10.1140/epjc/s10052-018-6131-3`. arXiv: `1803.01853 [hep-ph]` (cit. on p. 169).

[165] G. C. Fox and S. Wolfram, "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes," *Phys. Rev. Lett.*, vol. 41, p. 1581, 1978. DOI: `10.1103/PhysRevLett.41.1581` (cit. on pp. 195, 197).

[166] S. H. Lee *et al.*, `Belle` collaboration, "Evidence for $B^0 \to \pi^0\pi^0$," *Phys. Rev. Lett.*, vol. 91, H. W. K. Cheung and T. S. Pratt, Eds., p. 261 801, 2003. DOI: `10.1103/PhysRevLett.91.261801`. arXiv: `hep-ex/0308040` (cit. on p. 199).

[167] D. M. Asner *et al.*, `CLEO` collaboration, "Search for exclusive charmless hadronic B decays," *Phys. Rev. D*, vol. 53, pp. 1039–1050, 1996. DOI: `10.1103/PhysRevD.53.1039`. arXiv: `hep-ex/9508004` (cit. on p. 205).

[168] F. Abudinén *et al.*, `Belle II` collaboration, "B-flavor tagging at Belle II," *Eur. Phys. J. C*, vol. 82, no. 4, p. 283, 2022. DOI: `10.1140/epjc/s10052-022-10180-9`. arXiv: `2110.00790 [hep-ex]` (cit. on p. 217).

[169] T. Skwarnicki, "A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances," Ph.D. dissertation, Cracow, INP, 1986 (cit. on p. 223).

[170] H. Albrecht *et al.*, `ARGUS` collaboration, "Search for Hadronic $b \to u$ Decays," *Phys. Lett. B*, vol. 241, pp. 278–282, 1990. DOI: `10.1016/0370-2693(90)91293-K` (cit. on p. 224).

Disclaimer

The analysis of $B \to X_s \gamma$ presented in this thesis has been proposed to me by Prof. Dr. Kerstin Tackmann, who supervised my work. Throughout the years I have also received constant feedback from Dr. Arthur Bolz, Dr. Lu Cao, Dr. Markus Röhrken, Dr. Ilya Komarov, Dr. Simon Wehle, and many other Belle II colleagues. The data used in the analysis has been provided by the Belle II collaboration. The electron and positron beams are provided by the SuperKEKB collider. The work presented in Chapter 6 is designed and performed primarily by me, with the support of the aforementioned people. It has been documented in a preprint, available in Ref. [163]. The photon detection efficiency data-to-simulation comparison was performed by me and Dr. Simon Wehle, continuing the work that was set up by Dr. Natalia Kovalchuk and Prof. Dr. Torben Ferber. This has been documented in the Belle II public note in Ref. [162]. Additional studies performed by other members of the Belle II collaboration, necessary for $B \to X_s \gamma$, but not part of the original work of this thesis, are clearly named so in the text. Namely:

- The FEI algorithm, its calibration and inclusion to the `basf2` (Sections 6.3.1 and 6.11.1);

- The `zernikeMVA` training and inclusion to the `basf2` (Section 6.4.3);

- The π^0 and η veto, validation on data and their inclusion to the `basf2` (Sections 6.4.4 and 6.11.2).

The global fit of $B \to X_s \gamma$ experimental results (Section 7.3) was performed by the SIMBA collaboration. The work discussed in Section 7.2 has been performed based on the studies of Chapter 6 and included in a co-authored preprint in Ref. [98]. I have also co-authored publications with the Belle II tracking group, documented in Ref. [114]. As a member of the Belle II collaboration, I co-authored many other publications and conference notes, which are not presented in this thesis[1].

All of the text in this thesis has been worded by me. The figures that have not been produced by me are clearly indicated as such with a credit or a reference to the source.

---

[1] https://inspirehep.net/literature?q=Henrikas%20Svidras%20or%20H.%20Svidras

Acknowledgments

Nothing I have ever achieved would have been possible if not for my family who supported me when I went to study abroad and are always there waiting for me when I return. Ačiū mama, tėti, močiute, Jolita – ačiū jums už viską, ką man suteikiate ir ką dėl manęs padarėte. Ačiū net Boltui ir Deilui, kurie galbūt ir nežino, ką aš veikiu, bet visuomet džiaugiasi mane matydami.

I want to thank my friends who, despite the distance between us, are still people I can rely on, talk with, laugh with, and meet with whenever we like. Ačiū Martynai, Viliau ir Lukai, už neblėstančią draugystę nuo pat mokyklos ir universiteto laikų. Už visas tas valandas kurias praleidžiame kartu ir virtualiai. Be jūsų mano gyvenimas būtų be galo pilkesnis.

I want to thank the amazing people whom I have had the pleasure to work with at DESY. Namely, my supervisor, Kerstin Tackmann, who has provided me with the chance to be a part of the science community and supported me in my research throughout the years. I want to thank the rest of the group: Arthur Bolz, Lu Cao, Markus Röhrken, Simon Wehle, and Ilya Komarov, who guided me to become a better researcher. My deepest thanks go to Sasha Glazov without whose support and advice the work of this thesis would likely not have materialised. I also want to express my gratitude to Carsten Niebuhr for his excellent leadership of the Belle II group at DESY, and Andreas Gellrich who makes sure that the technical side of the work is always on top. I am also grateful to Frank Tackmann, Bahman Dehnadi and Ivan Novikov for their support in the theoretical side of the analysis.

The nicest thing is that I not only worked with people at DESY, but I found friends. I want to thank all of you for helping me stay sane and: Maria Konstantinova, for being someone that I never expected to find here and yet nothing better could have ever happened to me; Sascha Dreyer, for our survival of the pandemic, shift parties and wine evenings; Navid K. Rad, for every meter that we climb up and fall off walls; Anselm Baur, for being a great office neighbour and our trips to Obi; Munira Khan, for all our self-deprecating humour that makes the day; Filippo Dattola, Alberto Martini, per ogni volta che ho parlato con voi in italiano; Cyrille Praz, for introducing us to the Brew Dog evenings; Michel Hernández Villanueva, for some of the best Mexican food ever; Varghese Babu, for the incredible Indian cuisine; Tommy Martinov, for providing some of the funniest banter; Merle Graf-Schreiber, for enduring some of the *best* TV shows together; Marcel Hohmann, for coming over all the way from down under to become the bingo master; Abtin Narimani, for all the fun evenings and life advice; Michael de Nuccio, for the positivity and playfulness I always feel when hanging out; Eldar Ganiev, for your good energy and humour, and for being the first to present my analysis to the wider world; and everyone else who I inevitably forget but appreciate not a tiny bit less.

To everyone who took the time to give me comments on this thesis. I do not need to list you, but I want you to know that without your help I would have never succeeded.

Declaration

I hereby declare upon oath that I have not submitted a dissertation with the same research topic to an academic higher education institution that had been already accepted or evaluated as insufficient in an earlier doctoral procedure.

Ich versichere, dass ich keine Dissertation mit dem gleichen Forschungsthema schon einmal in einem fruheren Promotionsverfahren an einer wissenschaftlichen Hochschule eingereicht habe, die angenommen oder als ungenugend beurteilt worden ist.

I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated in the dissertation.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare upon oath that the dissertation submitted in electronic form and the printed bound copy of the dissertation submitted for archiving are identical.

Ich versichere an Eides statt, dass die in elektronischer Form eingereichte Dissertation und das zur Archivierung eingereichte gedruckte gebundene Exemplar der Dissertation identisch sind.

Henrikas Svidras
Hamburg, 30 March 2023