Development of a Semi-Inclusive Tagging Algorithm and Implementation of a Continuum Suppression in the Full Event Interpretation for the Belle II Experiment

Gianna Mönig

Masterarbeit in Physik
angefertigt im Physikalischen Institut

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

August 2021

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

Bonn, 30.08.2021
        Date

Signature

1. Gutachter: Prof. Dr. Florian Bernlochner
2. Gutachter: Prof. Dr. Klaus Desch

Contents

Introduction

The Belle II experiment is located at the SuperKEKB accelerator, which collides electrons and positrons at the necessary energy to produce $\Upsilon(4S)$ particles ($b\bar{b}$ resonances). The $\Upsilon(4S)$ subsequently decays into two B mesons over 96 % of the time.

The relatively clean events in the e^+e^- collisions and the hermetic detector make the reconstruction of the full event possible. In addition, the two B mesons are produced in the decay of an $\Upsilon(4S)$ with well known kinematics. Therefore, flavour and momentum information about one B meson make strong constraints on the other B meson. This method is called tagging.

At Belle II, a number of physics analyses include missing energy, e.g. semi-leptonic or leptonic decays. The selection of events with missing final state particles can be quite challenging. In order to improve the selection, a tagging can be performed, by reconstructing the second B meson, referred to as tag-side.

The tagging method can be divided into two different approaches: exclusive and inclusive tagging.

For exclusive tagging the tag-side is reconstructed in specific final state decays. These final states can be either fully hadronic or semi-leptonic. For the hadronic tagging, only events where the tag-side B meson decays fully hadronically are used, as this allows for a full reconstruction of the tag-side. For the semi-leptonic tagging, where the tag-side can decay semi-leptonically, undetected neutrinos from the tag-side make a full reconstruction impossible. The exclusive tagging method employed by Belle II is the Full Event Interpretation [1]. The constraints provided by the reconstruction of the tag-side make it possible to achieve relatively pure signal samples. Nevertheless, the efficiency of the tag-side reconstruction is limited by the sum of the branching fractions of reconstructed decays and by the reconstruction efficiency. In particular analyses of rare decays (like $B \to \mu\nu$, $B \to K^{(*)}\nu\bar{\nu}$) strongly suffer from this drawback due to the limited statistics.

To avoid these losses of efficiency due to the tag-side reconstruction, the inclusive tagging method can be applied. In this approach only the signal-side is reconstructed. All the remaining particles then form the tag-side. This also allows for kinematic constraints on the signal-side, but the quality of the tag-side reconstruction is poorer, leading to a worse resolution on kinematic quantities of interest.

In this thesis a novel hybrid approach to tag-side reconstruction is presented, where the tag-side is only partially reconstructed in order to improve the purity compared to the inclusive tagging, while retaining a high efficiency.

In a second part of the thesis a new feature is added to the Full Event Interpretation to remove background from $e^+e^- \to q\bar{q}$ ($q = u, d, c, s$) events, which is an important background for many Belle

II analyses. The discrimination between $e^+e^- \to \Upsilon(4S)$ and $e^+e^- \to q\bar{q}$ events can be achieved by exploiting the different event kinematics.

30th August 2021 11:19

The Standard Model of Particle Physics

2.1 The Standard Model

The Standard Model (SM) of elementary particles [2] describes all elementary particles, summarised in Figure 2.1, as well as their interactions. It consists of 12 fermions and their corresponding anti-particles, grouped into 3 generations. Each generation is structured in the same way, containing two leptons and two quarks. These massive spin $\frac{1}{2}$ particles interact via three different interactions, which are namely the electromagnetic, the weak and the strong interactions. The coupling strenth under these interactions depends on their charge, which differs for the different fermion types.
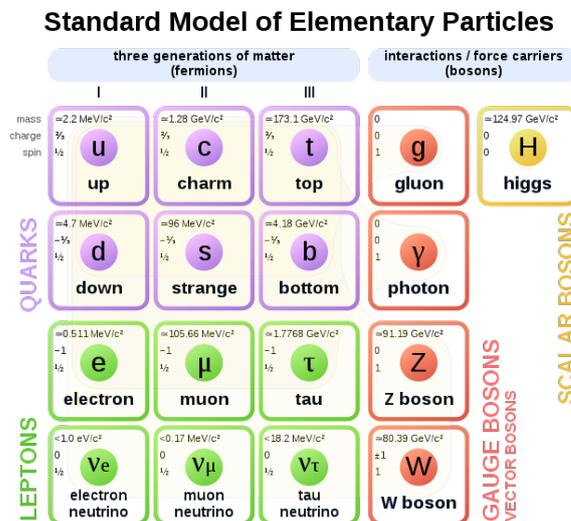
Figure 2.1: Particles contained in the Standard Model with their mass, their electric charge and their spin [3]

The leptons are composed of two types of particles, the electrically charged leptons and the neutrinos. The charged leptons (the electron, the muon and the tau) have an electric charge of $-1e$ where e is the elementary charge, while the neutrinos (the electron-neutrino, the muon-neutrino and the tau-neutrino) are electrically neutral.

The quarks are also composed of two types differing in their electric charge. The up-type quarks (the up, the charm and the top quark) have an electric charge of $\frac{2}{3}e$ while the down-type quarks (the down, the strange and the bottom quark) have a charge of $-\frac{1}{3}e$. Additionally to the electric charge, the quarks have a strong charge called colour. Each quark carries one of the three colours, green, blue or red.

The three interactions are mediated by the gauge bosons with spin 1.

The photon, the interaction particle of the electromagnetic force, is a massless and electrically neutral particle coupling to the electric charge.

The massless gluon, responsible for the strong force, couples to the colour charge. Itself carries a colour and an anti-colour leading to a gluon self-coupling.

Finally the two charged W and the neutral Z boson mediate the weak interaction, by coupling to the weak isospin and the weak hyper-charge. The weak force is experienced by all fermions and couples the charged leptons with the neutrinos and the up-type quarks with the down-type quarks.

The final particle in the Standard Model is the Higgs boson. Through its scalar field the massive gauge bosons (W and Z) acquire their mass.

The charges of the leptons and the quarks do not differ for the three generations, leading to the same coupling to the gauge bosons. For the leptons, this symmetry between the couplings is called the lepton flavour universality (LFU). The only asymmetry between the three families is introduced by the different couplings between the Higgs field and the fermion fields, the Yukawa coupling, which means that the different fermions acquire different masses. In the quark sector, the weak and the mass eigenstates are not identical which leads to transitions between the generations. The Cabibbo-Kobayashi-Maskawa (CKM) matrix describes the strength of these flavour changing interactions in the quark sector.

2.2 Physics beyond the Standard Model

The Standard Model is a well tested and consistent theory which describes many phenomena observed in experiments later on. Nonetheless not all physical observations can be explained by the SM indicating that the theory cannot be complete. Some astronomical observations for example (e.g. the rotation curves of galaxies) cannot be explained without the existence of another form of matter, called dark matter. At the particle level though, no evidence for new particles beyond the SM could be established so far. Nevertheless, some measurements show a discrepancy to the SM expectations, which could become significant with further experimental measurements.

One measurement in tension to the SM is the value of $R(D^*)$. It is defined as:

$$R(D^*) = \frac{\mathcal{B}(B \to D^* \tau \nu)}{\mathcal{B}(B \to D^* \ell \nu)}$$

where $\mathcal{B}(B \to D^* \tau \nu)$ is the branching fraction of the $B \to D^* \tau \nu$ decay and $\mathcal{B}(B \to D^* \ell \nu)$ the branching fraction of $B \to D^* \ell \nu$ with $\ell = e, \mu$. As hadronic effects cancel in the ratio of the two branching fractions, the value of $R(D^*)$ is predicted to a very high precision in the SM. In a new physics scenario in which the LFU would be broken, the precise measurement of this value could therefore lead to the discovery of physics beyond the SM. While the SM predicts a value of $R_{D^*} = 0.258 \pm 0.005$, the world average measured value is of $R(D^*) = 0.295 \pm 0.011 \pm 0.008$, which exceeds the SM prediction by 2.5σ. If also $R(D)$ is considered, the total difference to the SM value is about 3.08σ [4] as shown

in Figure 2.2. By improving the experimental techniques and increasing the data statistics, this value should either become closer to the SM expectation or the deviation would become more significant. New physics searches, looking for physics beyond the SM are therefore still extremely important.
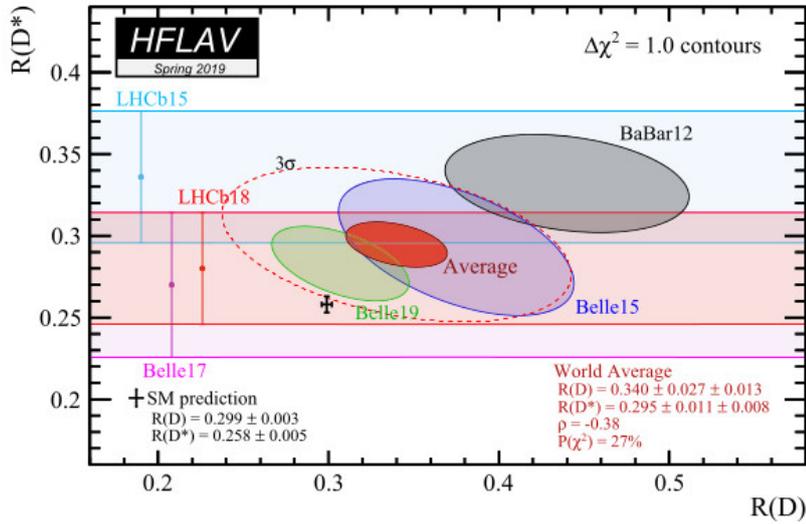
Figure 2.2: Current measurements and SM expectation value of $R(D)$ and $R(D^*)$ [4]

Another aspect which needs further investigations is that the SM includes 18 free parameters which can only be determined by measurements. The matrix elements of the CKM matrix for example cannot be predicted. Hence many experiments, e.g. the Belle II experiment, aim to determine these parameters as precise as possible.

The Belle II Experiment

The Belle II experiment is located at the SuperKEKB accelerator in Tsukuba, Japan. This chapter describes SuperKEKB as well as the Belle II detector.

3.1 SuperKEKB

The SuperKEKB accelerator (shown in Figure 3.1) is an electron-positron collider with a circumference of about 3 km. In two distinct beam pipes, electrons and positrons are accelerated up to momenta of 7 TeV c^{-1} and 4 TeV c^{-1} respectively. At the interaction point (IP) the particle bunches collide with a centre of mass energy of $\sqrt{s} = 10.580$ GeV, corresponding to the invariant mass of the $\Upsilon(4S)$ resonance ($b\bar{b}$), which is slightly above the $B\bar{B}$ threshold. The subsequent decays of $\Upsilon(4S) \rightarrow B^+B^-$ and $\Upsilon(4S) \rightarrow B^0\bar{B}^0$ with a branching fraction close to 100 % allow for the production of a high number of B mesons. Hence the SuperKEKB collider is often called a B factory.
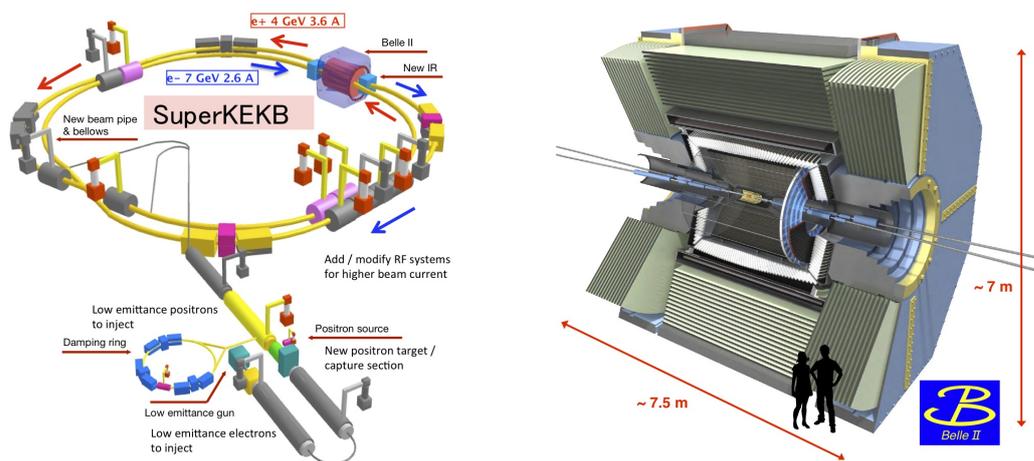
Figure 3.1: The SuperKEKB accelerator complex and the Belle II detector [5]

SuperKEKB was upgraded from its predecessor KEKB in order to significantly increase the instantaneous luminosity. As shown in Figure 3.2, from the targeted luminosity of 6.5×10^{35} cm^{-2} s^{-1}

so far $3.1 \times 10^{34}\,\mathrm{cm^{-2}\,s^{-1}}$ were reached, corresponding to the current world record. The extremely high design luminosity as well as the continuous beam injection provides a high amount of physics events.
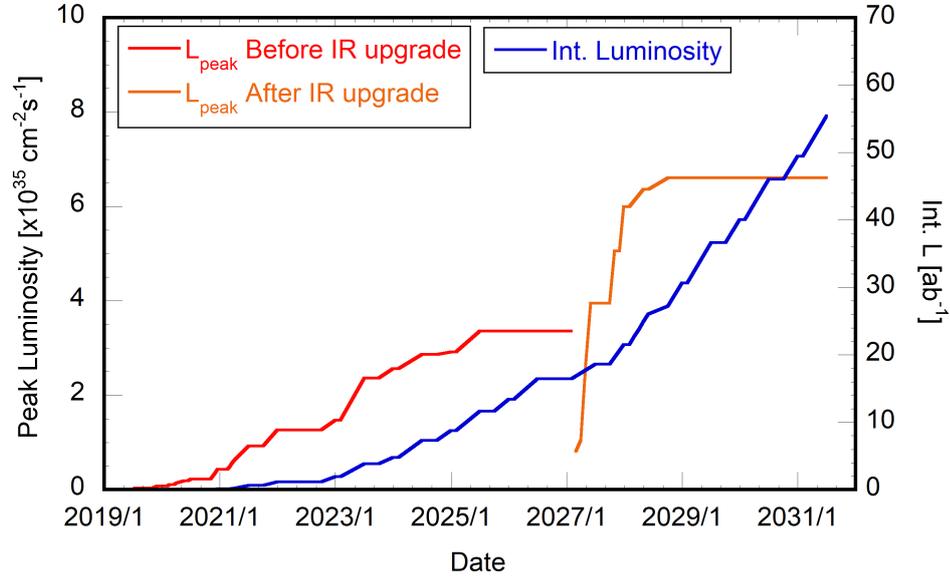
Figure 3.2: Projection plot of the instantaneous luminosity of SuperKEKB with a design luminosity of $6.5 \times 10^{35}\,\mathrm{cm^{-2}\,s^{-1}}$ [6]

3.2 Belle II

The electrons and positrons are collided within the Belle II detector (see Figure 3.1) which measures particles produced in the collisions. The aim of the Belle II experiment is to collect a data set with an integrated luminosity of $50\,\mathrm{ab^{-1}}$ at $\sqrt{s} = m_{\Upsilon(4S)}$, which would correspond to about 50 times the data set of the predecessor Belle. This larger data set would allow many precision measurements of the CKM sector of the SM and enable a number of searches for new physics at the high intensity/high precision frontier. Up to now, about $213\,\mathrm{fb^{-1}}$ of data were recorded (see Figure 3.3), whereof the first $34.6\,\mathrm{fb^{-1}}$ are used in this thesis.

Besides the great benefits of the high instantaneous luminosity of SuperKEKB also many challenges for the detector, such as high beam backgrounds, come along. Therefore, an upgrade of the Belle detector was needed, before Belle II could start its official data taking in the beginning of 2019.

3.2.1 The Detector

The Belle II detector is located cylindrically around the beam pipe of the accelerator, with the IP at the centre. The detector is divided into several sub components, which are briefly introduced in this section and are shown in Figure 3.4. Further details can be found in the Technical Design Report [7].

- The **VerteX Detector** (short VXD) is the inner part of the detector. It is composed of two sub detectors: the PiXel vertex Detector (PXD) and the Silicon Vertex Detector (SVD). With
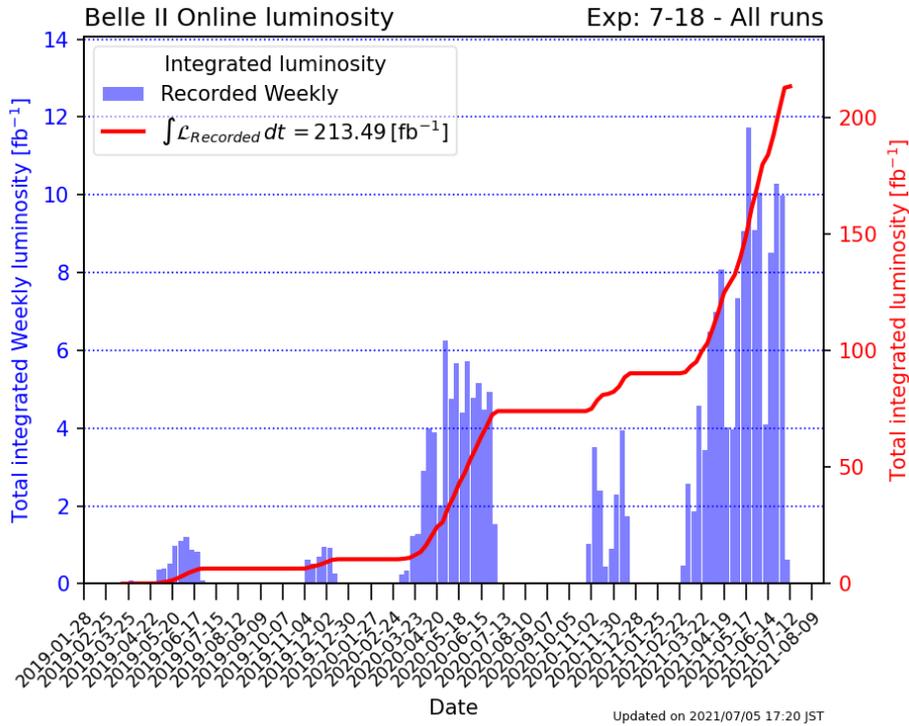
Figure 3.3: Integrated luminosity recorded by the Belle II detector since the beginning of 2019 [6]

its high number of DEPFET (DEpleted P-channel Field Effect Transistor) pixel sensors the PXD is able to deal with the high occupancy very close to the beam pipe. Around the PXD, the double-sided micro strips of the SVD are located. These two detector components enable precision determinations of particle vertices and the IP.

- After the VXD comes the **Central Drift Chamber** (CDC), which is a wire chamber. The CDC is essential in the reconstruction of tracks from the recorded hits. As it is located in a magnetic field, inducing the curvature of charged tracks, the momentum of these tracks can also be determined.

  The particle identification, which is essential for Belle II physics analyses, is performed by the two following detector components.

- The **Time-Of-Propagation counter** (TOP) is installed in the central barrel region. In the bars of fused silica the traversing particles produce Cherenkov light which is then guided to the read-out. The momenta of the particles are then extracted from the time of propagation of the light in the radiator.

- At the end caps an **Aerogel Ring-Imaging Cherenkov detector** (ARICH) is placed. The Cherenkov light produced when the particles traverse the aerogel radiator is detected by the RICH counter. Using the geometry of the Cherenkov rings the momenta of the particles can be determined.
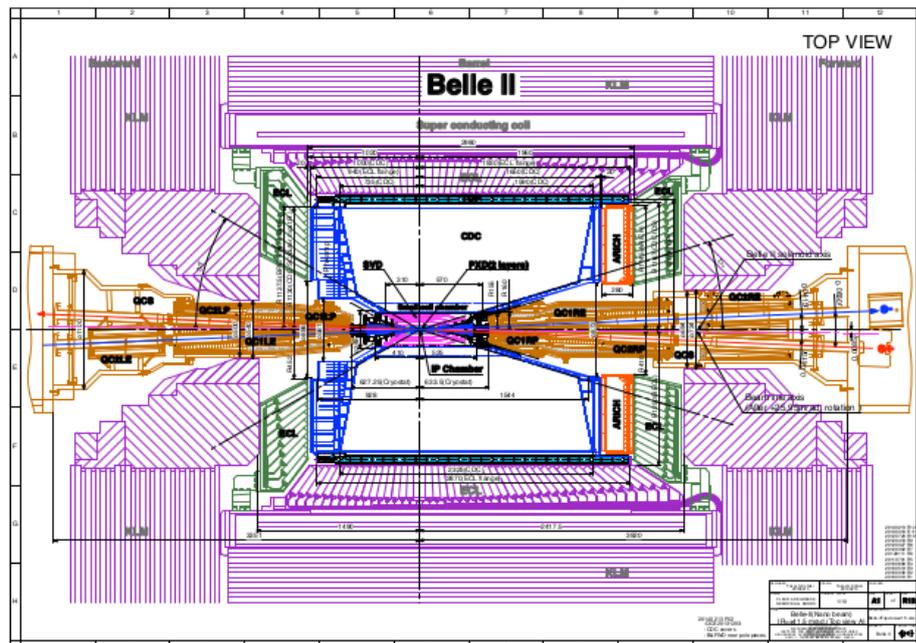
Figure 3.4: Technical Drawing of the Belle II Experiment [7]

- As about one third of the particles produced in the *B* decays are neutral particles, the detection of these is also very important. This is performed by the **Electromagnetic CaLorimeter** (ECL) which is composed of CsI(TI) scintillation crystals and read-out photodiodes. The recorded electromagnetic (EM) showers enable energy measurements of the photons and the electrons.

- Finally the outermost part of the detector is the K_L^0/**Muon Detector** (KLM) which consists of alternating iron plates and active detector elements (resistive plates for the barrel region and plastic scintillators for the end cap region). Additionally to the ECL, the iron plates provide material for the K_L^0 to shower. These showers are then detected by the active material. At the same time, the resistive plates also collect hits left by the travesing muons. These hits allow to distinguish the muons from hadrons by combining them with the tracks recorded in the CDC.

The Principle of the Tag-Side Reconstruction

The main events of interest at the Belle II experiments are the $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ events. For many measurements, one specific decay mode of the B meson must be considered. When a B meson is reconstructed in this relevant decay mode, the B corresponds to the signal-side. In order to select the signal events, cuts can be applied on the kinematics of the signal-side. However, the selection of the signal events suffers if the signal-side contains leptonic or semi-leptonic decays, as missing energy is then included. In order to counteract this loss of information, the fact can be used, that the whole event can be reconstructed because the e^+e^- collisions are relatively clean.

As the energy and the momentum of the e^+e^- system is well known and exactly two B mesons are created, flavour and kinematic information of the second B meson make strong constraints on the signal-side. Therefore, the purity of the signal sample can be further improved by reconstructing the second B meson. This principle is called "tagging". Therefore, in the following, this second B meson will be called the tag-side B or short B_{tag}.

There are two contrary longstanding approaches for the tag-side reconstruction: one in which the tag-side is explicitly reconstructed, in order to gain the most information possible, the exclusive tagging, and the other one where one simply defines all remaining particles in the event to be in the tag-side, the inclusive tagging [8]. The principle of these two approaches is schematically depicted in Figure 4.1.

As shown in figure 4.2 the choice of the tagging method represents a choice between the highest possible efficiency (fraction of the correct signal-side events which are tagged) and the highest possible purity (fraction of tagged events with the correct signal-side). For signal decays with very small branching fractions, only a few signal events are expected in the collected data samples. Therefore the efficiency in the order of 0.1 % to 1 % of the exclusive tagging does not allow to measure these decays.

4.1 Inclusive Tagging

In the inclusive tagging only the signal-side is fully reconstructed. Tracks and EM clusters in the remaining event are then assigned to the tag-side.

This method has the great advantage of a high efficiency, as only the signal-side has to be reconstructed. Nonetheless, the tag-side can often be misreconstructed. This leads to the fact, that cut selections based on the kinematics of the tag-side are imprecise, inducing a relatively low purity of signal events. Additionally the resolution of the signal events in kinematic variables of interest is
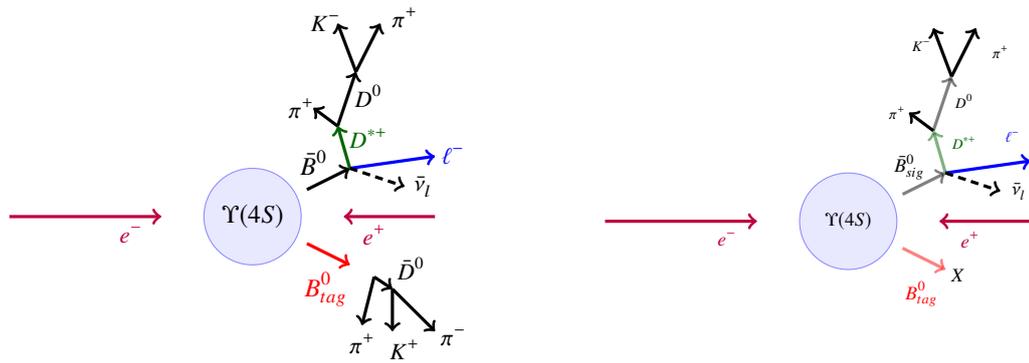
(a) Exclusive: explicit reconstruction B_{tag}

(b) Inclusive: all remaining particles are added to the B_{tag}

Figure 4.1: Schematic representation of the tagging methods, with $B \rightarrow D^* \ell \nu$ as signal-side
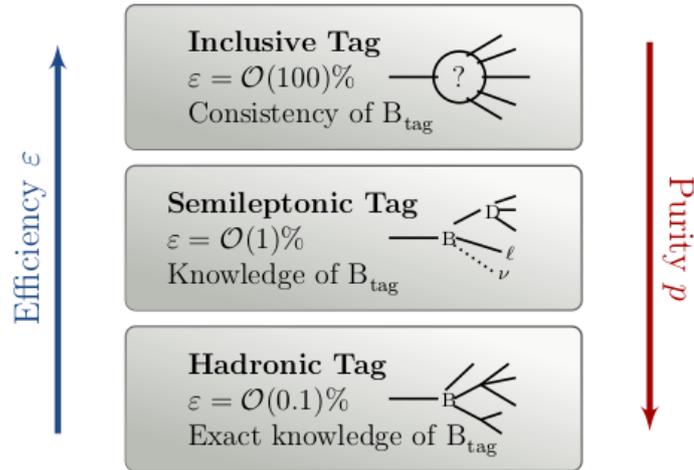
Figure 4.2: Efficiency and purity of the exclusive and the inclusive tagging methods [9]: as semi-leptonic decays include neutrinos, which are not directly detectable in the detector, the exclusive tagging is subdivided into the semi-inclusive and the hadronic tagging

smeared out.

4.2 Exclusive Tagging

As an alternative to the inclusive tagging, the exclusive approach reconstructs the tag-side in specific final state decays. For hadronic modes, the B meson can be explicitly reconstructed, which allows for very strict flavour and kinematic constraints. In semi-leptonic modes the missing neutrinos loosen the constraints which can be made on the kinematics.

As the full $\Upsilon(4S)$ decay chain is explicitly reconstructed, this results in a completeness constraint.

This refers to the requirement of no additionally tracks or clusters in the events, which leads to very pure signal samples. Nonetheless, the total sum of the branching fractions of the implemented possible decay chains of the tag-side B meson and the reconstruction efficiency limit the overall tagging efficiency.

4.2.1 Full Event Interpretation

The Full Event Interpretation [1, 9] (FEI) is an exclusive tagging algorithm developed for the Belle II experiment, where candidates for the tag-side B meson are reconstructed according to predefined decay processes. In order to do so, the algorithm reconstructs the B decay chain in a hierarchical manner in six stages. Each step corresponds to the reconstruction of certain intermediate particle types, using the particles from the previous steps. These different stages as well as the allowed decay processes are shown in figure 4.3.
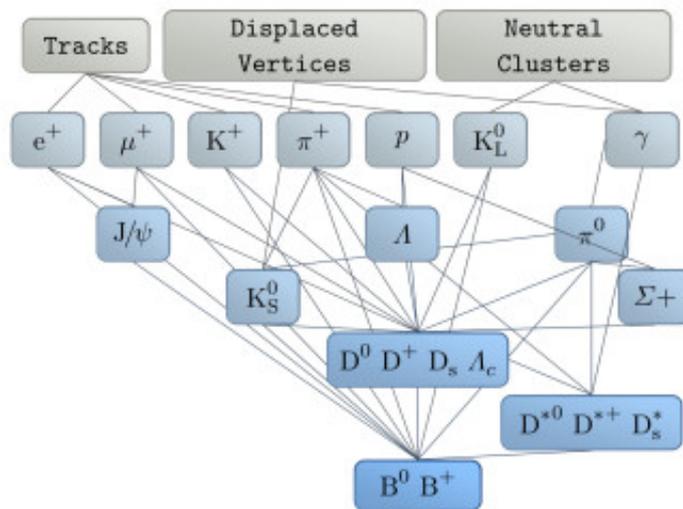
Figure 4.3: Working Principle of the FEI [10]: in six stages the tracks and clusters are assigned to final state particles, which are then assembled to reconstruct different intermediate particles before being finally combined to the tag-side B meson candidates

In the first instance, the tracks and clusters are assigned to particle candidates, which have to pass selections made via pre-cuts and a best candidate selection (BCS). For each final state particle type, a pre-trained boosted decision tree (BDT) is applied which assigns a BDT output to the particle candidates called signal probability. The output can be interpreted as a probability for the candidate to be correctly reconstructed. According to this probability post-cuts are applied and again a BCS is made.

The retained particles are now recombined to reconstruct other intermediate particles. After new pre-cuts and a BCS a vertex fit is performed. BDTs trained for each unique decay of a particle are then applied on the particle decays, where its input is given by the kinematics of the particles as well as by the output of the classifiers before. This process is repeated until the particles are finally combined into B-meson candidates.

As the FEI allows to explicitly reconstruct the tag-side, a high resolution can be obtained on

signal-side quantities computed with information from the tag-side. Nevertheless, the FEI, as an exclusive tagging algorithm, has a relatively low efficiency compared to the inclusive tagging.

Physics Processes in e^+e^- Collisions

As mentioned in chapter 3, SuperKEKB is an e^+e^- collider delivering $e^+e^- \to f\bar{f}$ events, where f is one of the fermions in the Standard Model. The predominant process is Bhabha scattering $(e^+e^- \to e^+e^-(\gamma))$ with a cross section of (300 ± 3) nb [11] at $\sqrt{s} = m_{\Upsilon(4S)}$. This corresponds to about 300 times more Bhabha scattering events than $\Upsilon(4S)$ events. Nevertheless this background source as well as $e^+e^- \to \mu^+\mu^-(\gamma)$ are strongly suppressed by the trigger. At the physics analysis level the main backgrounds to the $\Upsilon(4S)$ events are $e^+e^- \to q\bar{q}$ with $q = u, d, c, s$ and $e^+e^- \to \tau^+\tau^-$. Their cross section are listed in table 5.1.

Process	Cross section nb
$\Upsilon(4S)$	1.110 ± 0.008
$u\bar{u}(\gamma)$	1.61
$d\bar{d}(\gamma)$	0.40
$s\bar{s}(\gamma)$	0.38
$c\bar{c}(\gamma)$	1.30
$\tau^+\tau^-(\gamma)$	0.919

Table 5.1: Production cross section for the relevant $e^+e^- \to f\bar{f}$ events at $\sqrt{s} = 10.58$ GeV [11]

5.1 Signal Processes

The branching fraction of the $\Upsilon(4S)$ to decay into $B\bar{B}$ is greater than 96 %, with (51.4 ± 0.6) % decaying into charged and (48.6 ± 0.6) % into neutral B meson pairs. In the final data set with an integrated luminosity of 50 ab^{-1} which Belle II aims to collect, more than 50 billion $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ events will have been measured [12].

5.1.1 $\bar{B}^0 \to D^*\ell\nu$ Events

In order to develop the algorithm, the well studied decay $\bar{B}^0 \to D^{*+}\ell\nu$, with $\ell = e, \mu$ was used.

The $\bar{B}^0 \to D^{*+}\ell\nu$ decay has a relatively high branching fraction of (5.05 ± 0.14) % [2], which allows for the tag-side reconstruction to be studied with a reasonable statistically sized sample. In addition, $D^{*+}\ell\nu$ can be reconstructed with a relatively high purity.

As shown in the Feynman diagram 5.1, this process gives access to the CKM matrix element V_{cb}, which enters the coupling strength in the vertex with the W boson.

Additionally this decay corresponds to the normalisation of $R(D^*) = \frac{\mathcal{B}(B \to D^*\tau\nu)}{\mathcal{B}(B \to D^*\ell\nu)}$ introduced in chapter 2. Therefore, the precise measurement of the branching fraction of $B \to D^*\ell\nu$ is essential.
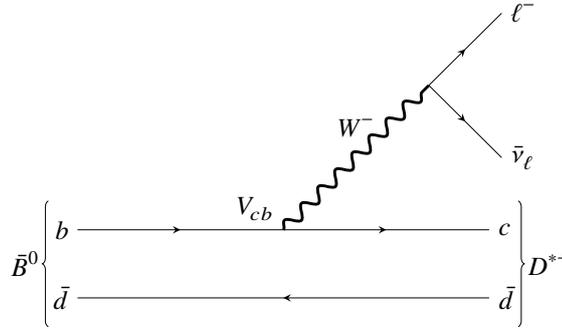
Figure 5.1: Feynman diagram of the $\bar{B}^0 \to D^{*+}\ell\bar{\nu}$ decay

5.1.2 $B^+ \to \mu^+\nu_\mu$ Events

The algorithm is also tested on the $B^+ \to \mu^+\nu_\mu$ decay. This decay could be potentially utilised to measure the CKM matrix element V_{ub}, as can be seen in Figure 5.2. Such an extraction would be theoretically very precise as only the decay constant is required.
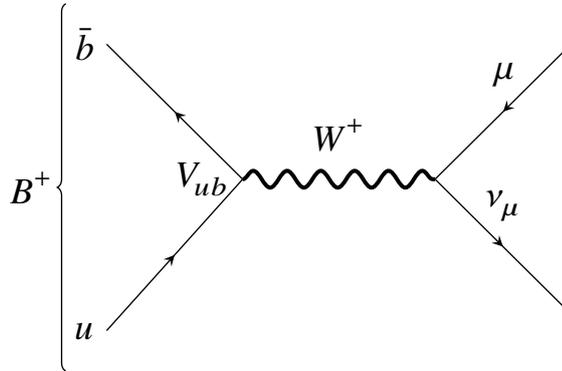
Figure 5.2: Feynman diagram of the $B^+ \to \mu^+\nu_\mu$ decay

However in the SM this decay is suppressed, because of the CKM matrix element and because of helicity. The branching fraction of the decay is given by [13]:

$$\mathcal{B}(B^+ \to \mu^+\nu_\mu) = \frac{G_F^2 m_B m_\mu^2}{8\pi} \left(1 - \frac{m_\mu^2}{m_B^2}\right)^2 f_B^2 |V_{ub}|^2 \tau_B$$

with G_F Fermi's constant, m_B and m_μ the masses of the B meson and the muon respectively, f_B the decay constant and τ_B the lifetime of the B meson. The CKM matrix element V_{ub}, which enters the

branching ratio, is very small as it corresponds to the strength of the transition between the first and the third generation. In addition $\mathcal{B}(B^- \rightarrow \mu^- \nu_\mu)$ is suppressed by the factor m_μ^2. This arises due to the spin being conserved in the interaction. As the B meson has spin 0, the anti-muon and neutrino spin must be opposite. In the rest frame of the B meson both particles are back-to-back. As the neutrino must be left-handed, the anti-muon must have opposite helicity leading to a suppression.

Both facts induce a very small branching fraction of less than 8.6×10^{-7} [2]. Processes beyond the SM could potentially enhance the branching ratio. A charged Higgs as mediator instead of a W boson could for example increase the branching fraction depending on its coupling. Another possibility would be the exotic decay, where a neutral sterile neutrino is produced instead of the SM neutrino [13].

By measuring the branching fraction of the $B^+ \rightarrow \mu^+ \nu_\mu$ decay, limits on the couplings of theses exotic particles to SM particles can be determined.

5.2 Continuum Events

An important background to the $B\bar{B}$ events are the $e^+e^- \rightarrow q\bar{q}$ events, where $q = u, d, c, s$. As shown in Figure 5.3, only about 23.1 % of the $e^+e^- \rightarrow q\bar{q}$ events correspond to the physics events, where $b\bar{b}$ is produced. Thus the events where the other quarks are produced, called continuum events, are predominant.
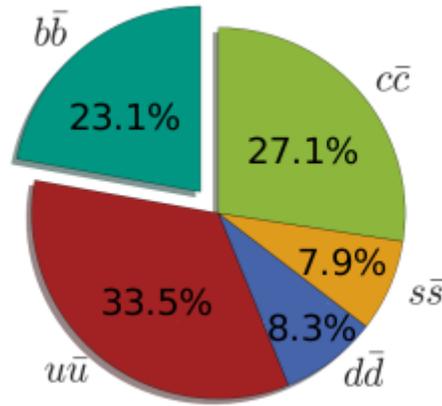
Figure 5.3: Branching fraction of $e^+e^- \rightarrow q\bar{q}$ for the different quark types q at $\sqrt{s} = m_\Upsilon$ [14]

In order to estimate the shape and the amount of this background, e^+e^- collisions with a centre of mass energy slightly below the $\Upsilon(4S)$ resonance can be performed.

Additionally a continuum suppression can be performed, which exploits the fact that the event shape of continuum events differs from that of $B\bar{B}$ events. Therefore variables can be engineered, which provide a good discrimination between $B\bar{B}$ and continuum events. Subsequently, the discriminating power of such features can be combined using a multivariate classifier.

The Semi-Inclusive Tagging Algorithm

The novel tagging method is a semi-inclusive approach, which aims to find a compromise between high efficiency and low purity. As shown in Figure 6.1, the name semi-inclusive stands for the fact, that the tag-side is partially reconstructed. In the following chapter, the exact working principle of the new algorithm is explained.
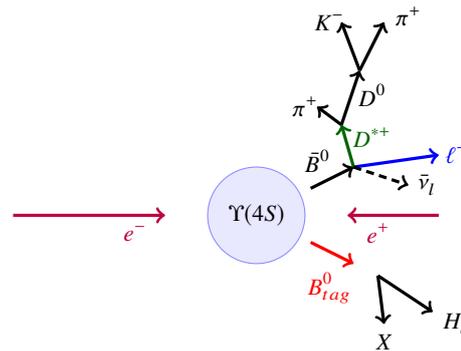
Figure 6.1: Schematic representation of the semi-inclusive tagging method, with $B \to D^* \ell \nu$ as signal-side

6.1 Working Principle of the Semi-Inclusive Tagging Algorithm

As the new approach has an iclusive part, the signal-side must be reconstructed first. It is then required that a charmed hadron H_c can be reconstructed in the remaining event. Once the H_c is reconstructed by the FEI, the remaining particles which neither belong to the signal-side nor to the charmed hadron are assigned to a pseudo X particle. Finally the tag-side B meson can be reconstructed ($B_{tag} \to H_c X$). Because the tag-side consists of a charmed hadron H_c and the pseudo particle X, the new algorithm is called $H_c X$ algorithm.

Having built all possible $\Upsilon(4S)$ candidates from combinations of B_{sig} and B_{tag} candidates ($\Upsilon(4S) \to \bar{B}_{sig} B_{tag}$), a best candidate selection is made, which retains the candidates with the highest signal probability for the H_c. The final $\Upsilon(4S)$ candidate is then chosen to be the candidate with the

smallest absolute value of ΔE (difference between the reconstructed energy and the expected one) of the B_{tag} meson.

6.2 Tag-Side Reconstruction

The reconstruction of the tag-side consists of two separate parts: the reconstruction of the charmed hadron and the inclusive part.

The charmed hadron is reconstructed fully hadronically with the FEI. The following H_c particle types $H_c = D^{*+}, D^{*0}, D^+, D^0, D_s, \Lambda_c$ and J/ψ are reconstructed.

The inclusive part X consists of remaining particles in the event, called rest of event (ROE) after the reconstruction of the signal-side and the H_c candidate.

6.2.1 Tag-Side Selections

ROE Selections

The ROE is built out of π^+, K^+, p^+, γ, e and μ lists.

Used candidates must pass the cuts on the particle ID variables presented in table 6.1. These variables correspond to the identification probability of a particle type under its own hypothesis.

Particle	Cuts				
π^+	$\pi_{ID} > 0.6$	$K_{ID} < 0.6$	$e_{ID} < 0.9$	$\mu_{ID} < 0.9$	
K^+	$\pi_{ID} < 0.6$	$K_{ID} > 0.6$	$e_{ID} < 0.9$	$\mu_{ID} < 0.9$	
p^+	$\pi_{ID} < 0.6$	$K_{ID} < 0.6$	$e_{ID} < 0.9$	$\mu_{ID} < 0.9$	$p_{ID} > 0.6$
e			$e_{ID} > 0.9$		
μ			$e_{ID} < 0.9$	$\mu_{ID} > 0.9$	

Table 6.1: Selection cuts on the particle ID for the ROE

Additionally a selection on the track quality of the leptons is applied: the distance to the point of closest approach (POCA) in the $r\phi$ plane must be smaller than 0.5 cm ($|d0| < 0.5$ cm), the POCA in respect to the interaction point (IP) $|dz| < 2$ cm, the transverse momentum $p_t > 0.3$ GeV and the polar angle $0.297 < \theta < 2.618$.

Application of the $H_c X$ Method on Monte Carlo Generated Data with $B \to D^* \ell \nu$ as Signal-Side

As the semi-inclusive tagging approach cannot be applied without reconstructing the signal-side, a specific signal-side decay must be chosen in order to test the performance of the new algorithm. In the following chapter the $H_c X$ method is applied on the $B \to D^* \ell \nu$ decay. The distributions of some important tag-side and signal-side kinematic variables are compared between events tagged with the $H_c X$ algorithm and events tagged by the FEI. Additionally, the efficiency of the two methods as well as the purity of the tagged events are considered.

7.1 MC Samples for the Evaluation of the Performance of the $H_c X$ Algorithm

The first studies of the algorithm are made with $100 \, \mathrm{fb}^{-1}$ of Monte Carlo events. The used batches are indicated in table 7.1.

Sample	Number of Events ($\times 10^6$)
$\Upsilon(4S) \to B^0 \bar{B}^0$	51
$\Upsilon(4S) \to B^+ B^-$	54
$c\bar{c}$	132.9
$u\bar{u}$	160.5
$d\bar{d}$	40.1
$s\bar{s}$	38.3

Table 7.1: Summary of simulated samples which are used to evaluate the performance of the $H_c X$ tagging algorithm

7.2 Signal-Side Reconstruction

The signal-side consists of the $\bar{B}^0 \to D^{*+} \ell \bar{\nu}$ decay, where the D^{*+} subsequently decays into $D^0 \pi^+$.

The D^0 is reconstructed in several final states, which include $D^0 \to K^- \pi^+$, $D^0 \to K^- \pi^+ \pi^- \pi^+$, $D^0 \to K^- \pi^+ \pi^0$, $D^0 \to K^- \pi^+ \pi^- \pi^+ \pi^0$. The complex conjugated mode $B^0 \to D^{*-} \ell^+ \nu$ is also reconstructed.

7.2.1 Signal-Side Selections

For the signal-side the following selections for tracks are applied: the distance to the POCA in the $r\phi$ plane must be smaller than $0.5\,$cm ($|d0| < 0.5\,$cm) and the z coordinate of the POCA smaller than $2\,$cm ($|z0| < 2\,$cm), the track must be within the CDC angular acceptance (thetaInCDCAcceptance == 1) and the track must have left tracks in the CDC (nCDCHits> 0).

During the reconstruction, cuts are made on the reconstructed mass of the D^0 ($1.8\,$GeV $< M <$ $1.95\,$GeV) and on the mass difference between the D^{*+} and the D^0 ($0.139\,$GeV $< \Delta m < 0.16\,$GeV).

Finally a selection is applied on the track quality of the lepton: $|d0| < 0.5\,$cm, POCA in respect to the interaction point (IP) $|dz| < 2\,$cm, the transverse momentum $p_t > 0.3\,$GeV, the momentum in the centre of mass frame $p^{CM} > 1\,$GeV and the polar angle $0.297 < \theta < 2.618$. An electron is kept if $e_{ID} > 0.9$ and a muon if $\mu_{ID} > 0.9$ and $e_{ID} < 0.9$.

7.3 Offline Cuts

A final event selection is made offline. Cuts on the tag-side are made on the beam constrained B meson mass $m_{bc} > 5.26\,$GeV and on the energy difference between the expected and the reconstructed B $-0.2\,$GeV $< \Delta E < 0.2\,$GeV. In addition the signal probability of H_c must be above 0.001. On the signal-side only events with $-1.5 < \cos(\theta_{BY}) < 1.5$ (θ_{BY} being the angle between B_{sig} and the $D^* \ell$ system), $p_T^\ell > 1\,$GeV and $p_T^{D^*} < 2.4\,$GeV are kept. Concerning the whole event, a cut on the second Fox Wolfram moment foxWolframR2< 0.3 is made in order to suppress continuum background.

Finally a fully hadronic decay of the tag-side B-meson is required, by vetoing events, wehere the tag-side contains at least one lepton.

For the FEI exactly the same cuts are applied except the signal probability cut which is applied on B_{tag} and not on H_c.

7.4 Distributions of Kinematic Tag-Side Variables

The resolution of some of the important kinematic variables are investigated, as they strongly depend on the tag-side reconstruction.

The first variable is the beam constrained B mass, which is defined as [11]:

$$m_{bc} = \left(\left(\frac{\sqrt{s}}{2} \right)^2 - (\vec{p}_{B_{tag}}^{CM})^2 \right)^{\frac{1}{2}}$$

where $\vec{p}_{B_{tag}}^{CM}$ is the momentum vector of the B_{tag} in the collision centre of mass frame. As the energy of the B meson is simply given by $\frac{\sqrt{s}}{2}$ in the CM frame, the computed m_{bc} should peak at $(5.27934 \pm 0.00012)\,$GeV, the mass of the B meson.

The other important variable is ΔE. It is defined as [11]:

$$\Delta E = E^{CM}_{B_{tag}} - \frac{\sqrt{s}}{2}$$

where $E^{CM}_{B_{tag}}$ is the energy of the reconstructed B_{tag} in the centre of mass frame of the collision. ΔE should peak at $0\,\text{GeV}$ as both B mesons have $E = \frac{\sqrt{s}}{2}$ in the CM frame.

In Figure 7.1, the distribution of m_{bc} is shown for the H_cX tagging as well as for the FEI. The definition of a signal event is based on the MC truth information. If the truth information for one of the B mesons is a $D^*\ell\nu$ decay, the event is considered to be signal, independent of whether or not the actual reconstruction was correct.

(a) FEI
(b) H_cX
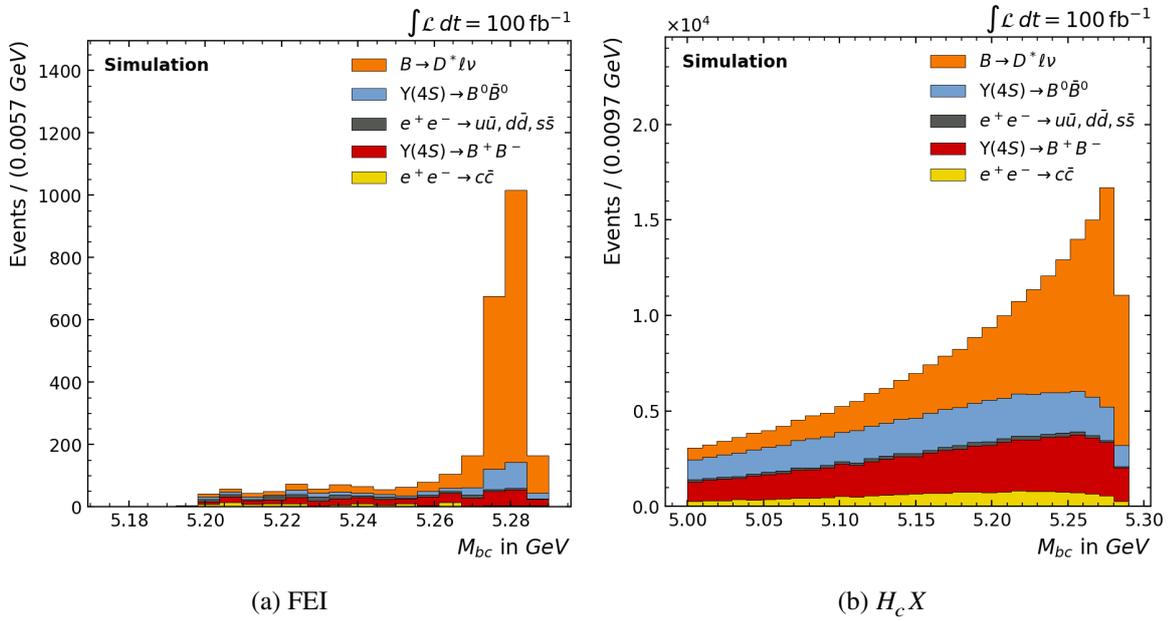
Figure 7.1: Comparison of the m_{bc} distributions of the FEI and H_cX tagging method

In the m_{bc} distribution the signal peaks at the real B meson mass for both tagging methods. The peak for the H_cX tagging has clearly a much larger resolution than that of the FEI. Additionally the background slightly peaks at the B meson mass. Nevertheless, m_{bc} is still a good cut variable for the H_cX tagging, as it allows for a good discrimination between signal and background.

Figure 7.2 shows the ΔE distribution of both tagging methods. While the FEI distribution peaks at $0\,\text{GeV}$, the H_cX distribution has a second peak at about $-0.5\,\text{GeV}$. This indicates, that some energy is missing in the reconstruction of the tag-side. This matches the expectation. While the FEI discards events where particles are missing, the H_cX retains them, leading to this wider distribution. Nevertheless, the second peak seems to strongly depend on the cuts applied to the particles added to the X system, especially on those applied to the photons. The cut on the photons was therefore chosen to be very loose in order to obtain the best ΔE distribution.

(a) FEI

(b) $H_c X$
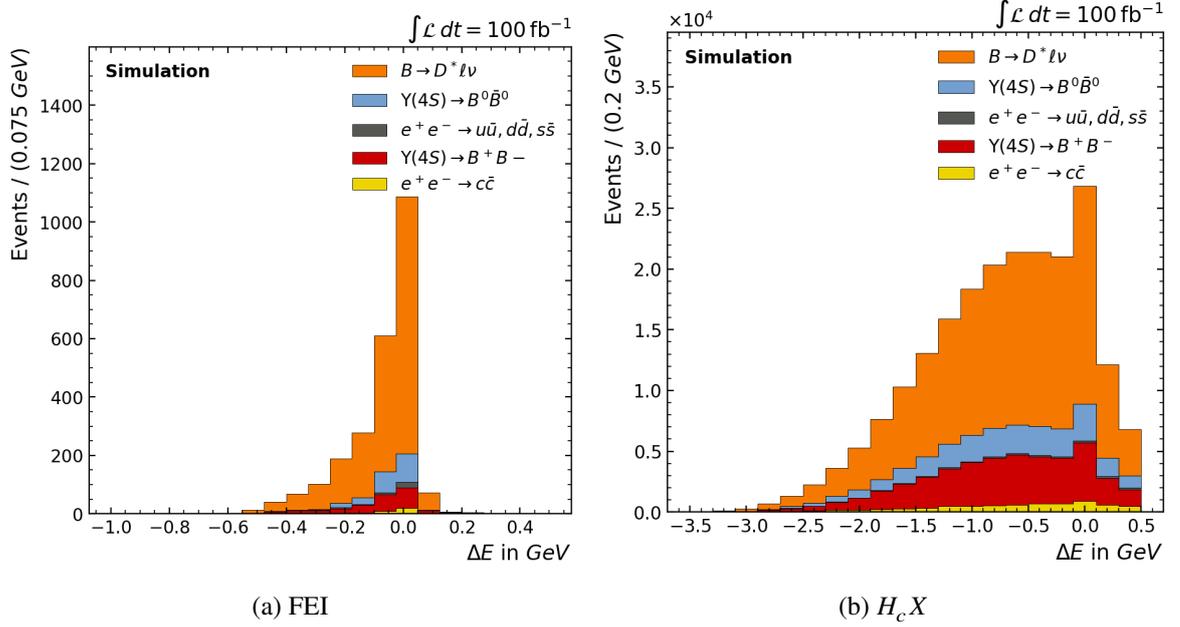
Figure 7.2: Comparison of the ΔE distributions of the FEI and $H_c X$ tagging method

7.5 m^2_{miss} Distribution

The main signal variable of interest is missing mass squared (m^2_{miss}), defined as:

$$\left(\frac{\sqrt{s}}{2} - E^{CM}_{B_{sig}} \right)^2 - \left(\vec{p}^{CM}_{x,B_{tag}} + \vec{p}^{CM}_{x,B_{sig}} \right)^2$$

where E^{CM}_{sig} is the energy of B_{sig} in the CM frame and $\vec{p}^{CM}_{B_{sig}}$ and $\vec{p}^{CM}_{B_{tag}}$ are the momenta of B_{sig} and B_{tag} respectively in the CM frame. As the tag-side is requested to be fully hadronic and the decay $B \to D^* \ell \nu$ contains one neutrino, m^2_{miss} should peak at $m^2_\nu \sim 0\,\text{GeV}^2$ for correctly reconstructed signal events. The variable is used as a fit variable to determine the number of signal $B \to D^* \ell \nu$ decays in data and hence assess the performance of the new tagging approach. The distribution of m^2_{miss} is shown for the $H_c X$ tagging and the FEI in Figure 7.3.

The signal distribution of the $H_c X$ tagging clearly peaks at the expected value of $0\,\text{GeV}^2$. Nevertheless the resolution of m^2_{miss} in $H_c X$ is larger than for that of the FEI. The standard deviation of the signal distribution is $0.62\,\text{GeV}^2$ ($H_c X$) compared to $0.38\,\text{GeV}^2$ (FEI). This loss in resolution is not unexpected as the requirements for the tag-side are much smaller for the semi-inclusive approach than for the exclusive one. For the $H_c X$ tagging, background hits and clusters can wrongly be added to the X system, which would lead to the broadening of the signal peak.

The main concern which can be seen in Figure 7.3, is that the background also peaks at $0\,\text{GeV}^2$ for the $H_c X$ tagging. This results in a poorer discrimination between signal and background in data. In case of the misreconstruction of the signal-side, all the remaining particles belonging to the B meson qualified as signal are added to the tag-side, which can lead to the observed sculpting of the
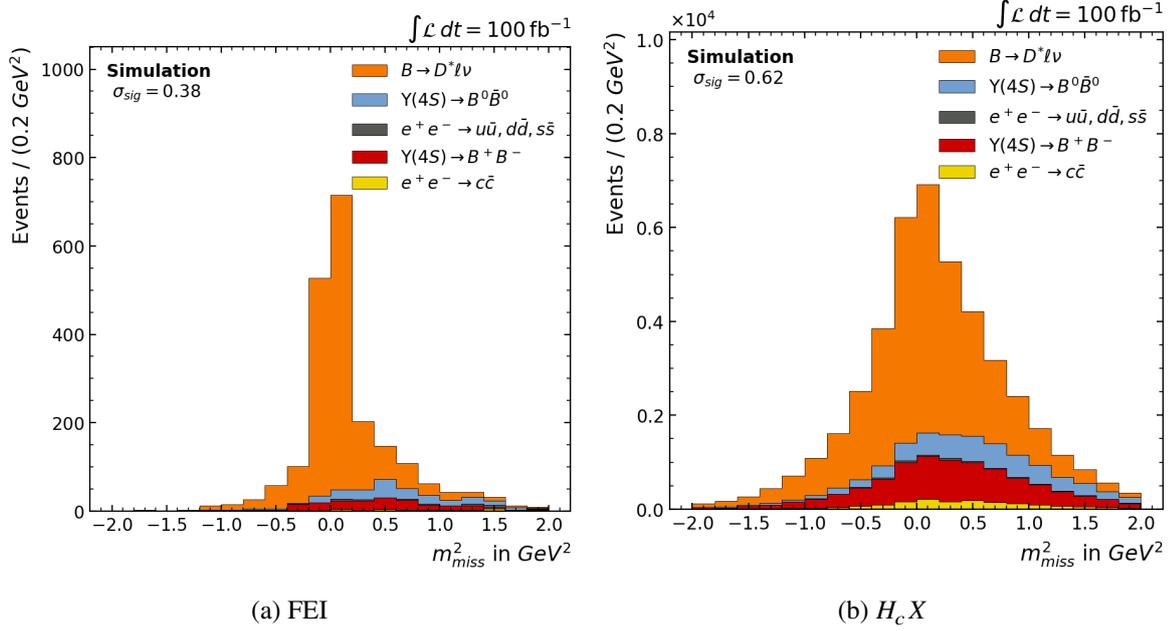
Figure 7.3: Comparison of the m_{miss}^2 distributions of the FEI and H_cX tagging method

background.

Despite these two drawbacks, the number of tagged $D^*\ell\nu$ events is greatly improved by the H_cX tagging. The comparison between the m_{miss}^2 distribution for the signal is shown in Figure 7.4. While the signal yield of the FEI is $1\,705 \pm 41$ the one of the H_cX tagging is $29\,471 \pm 172$. This corresponds to a signal yield which is a factor of 17.3 ± 0.4 higher for H_cX tagging than that of the FEI.
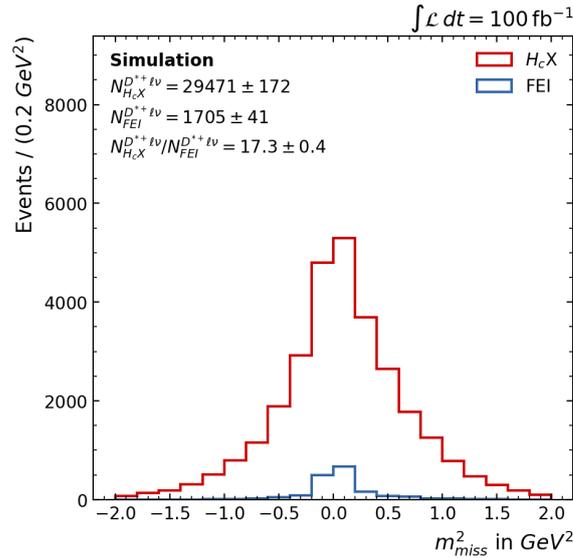
Figure 7.4: Comparison of the signal yield tagged by the FEI and the H_cX tagging methods

7.6 Efficiency and Purity of the H_cX Tagging Algorithm

In order to quantify the performance of the new tagging approach, one can have a look at the efficiency and the purity of the tagging method.

The efficiency is defined as:

$$\epsilon = \frac{N_{H_cX}^{D^*\ell\nu}}{N^{D^*\ell\nu}}$$

where $N_{H_cX}^{D^*\ell\nu}$ is the number of tagged signal events and $N^{D^*\ell\nu}$ the total number of signal events in the simulated data sample. $N^{D^*\ell\nu}$ is calculated by using the branching fraction of $B \rightarrow D^*\ell\nu$. For the H_cX algorithm an efficiency of $(2.74 \pm 0.02)\,\%$ is computed. In comparison, one gets an efficiency of $(0.159 \pm 0.004)\,\%$ for the FEI, which shows that the new approach indeed leads to a strong increase in efficiency.

The purity is given by the fraction of signal events in the total sample of tagged events. The H_cX tagging has a purity of $(69.5 \pm 0.4)\,\%$ compared to $(82 \pm 2)\,\%$ for the FEI. As expected, the purity suffers from the fact that the tag-side is not fully reconstructed. Nevertheless the loss seems relatively small compared to the increase of efficiency.

Finally the behaviour of the efficiency as a function of the purity is shown in Figure 7.5. The different values are obtained by varying the cut on the signal probability of the H_c for H_cX or of B_{tag} for the FEI. It is apparent, that the slope for H_cX is much larger than for the FEI. Tighter cuts on the signal probability lead for both tagging methods to an increase of purity. The effect on the efficiency is however much larger for H_cX. Also for the tightest cut, the efficiency exceeds the one of the FEI but does not reach its purity.
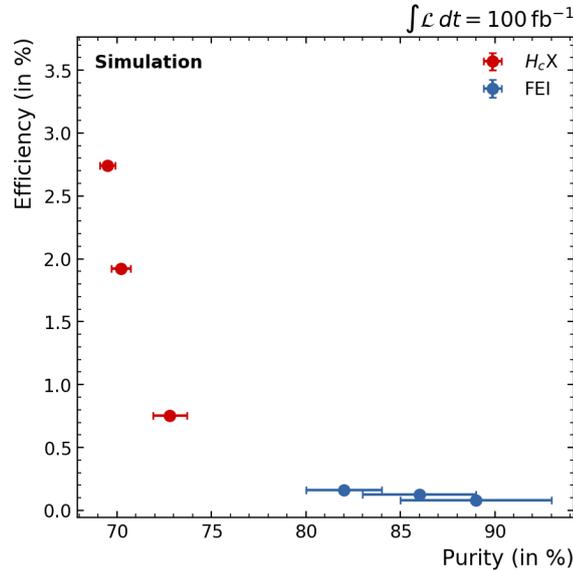
Figure 7.5: Efficiency against the purity of the H_cX and the FEI tagging method for different cuts on the H_c (H_cX) classifier output or the B_{tag} classifier output (FEI)

Calibration of the H_cX Algorithm with Data

The tagging algorithm was so far only tested on MC samples. As MC does not perfectly emulate data, it is not expected that the efficiency on data reaches the one in MC. For example, the performance of the BDTs employed by the FEI can be significantly impacted by mismodelling of certain variables or decays in MC compared to data. The extent of this impact on the efficiency of the tagging method can be expressed by the calibration factor between MC and data which is defined as follows:

$$\frac{N_{data}^{D^*\ell\nu}}{N_{MC}^{D^*\ell\nu} \times \frac{\mathcal{L}_{data}}{\mathcal{L}_{MC}}}$$

where $N_{data}^{D^*\ell\nu}$ and $N_{MC}^{D^*\ell\nu}$ are the number of tagged signal events in data and MC respectively. The factor $\frac{\mathcal{L}_{data}}{\mathcal{L}_{MC}}$ corresponds to the scaling between the two different integrated luminosities of the data and MC samples. A calibration factor of 1 would mean, that the algorithm performs equally well when applied to data as for MC.

It is important to note, that as the H_cX method is a semi-inclusive approach, it is not possible to determine a calibration factor which is independent of the signal-side. Therefore the factor calculated here for the FEI could deviate from the official signal independent one of 0.830 ± 0.029 for tag-side B^0 [10]. Furthermore, a different signal decay is employed to that of [10].

In order to calculate the calibration factor, the algorithm was applied to the first $34.6\,\text{fb}^{-1}$ of the about $200\,\text{fb}^{-1}$ of data collected so far.

8.1 Fit to Data

In the calculation of the calibration factor, the MC sample is required for two purposes: firstly directly in the factor and secondly to extract template shapes to perform the fit. In order to keep the statistical error on the number of tagged signal events in the MC sample and the uncertainty on the fitted yield associated with MC statistics uncorrelated, two different MC samples are used. One sample is used to determine $N_{MC}^{D^*\ell\nu}$ and the other sample as template for the fit.

The fit to the data is made with the binfit package, which performs a binned maximum likelihood fit according to the fit procedure documented in [10]. The fit results are shown in Figure 8.1.

For the H_cX tagging, the fit is performed in a m_{miss}^2 range between $-2.05\,\text{GeV}^2$ and $2.05\,\text{GeV}^2$, but
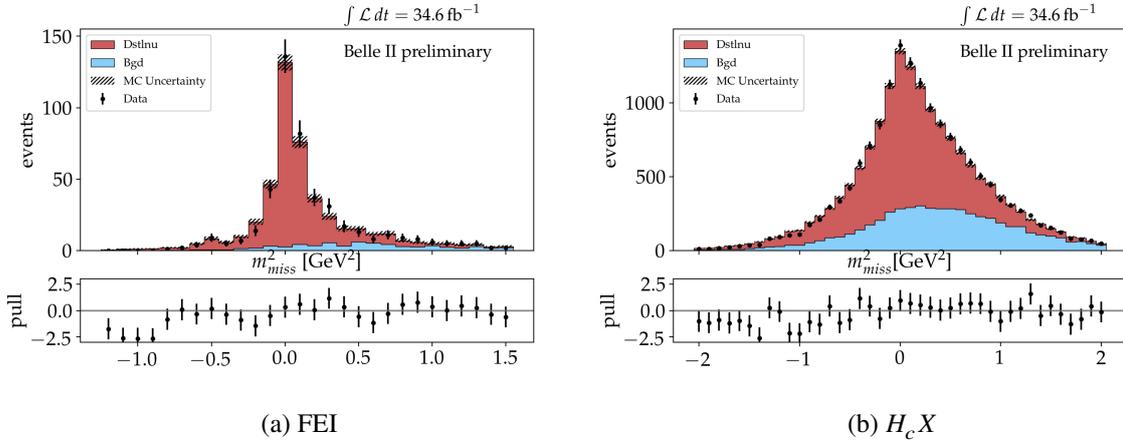
Figure 8.1: Fits to the missing mass squared distributions of signal-side $B^0 \to D^* \ell \nu$ decays in data for the FEI and the $H_c X$ tagging method

the signal and background yields are only considered in the range between $-2.05\,\text{GeV}^2$ and $0.95\,\text{GeV}^2$. This range was chosen in such a way, that the purity was maximised without strongly affecting the efficiency. One obtains a signal yield of $9\,784 \pm 253$ and a background yield of $4\,086 \pm 206$. For the FEI, the fit is performed in the interval of m_{miss}^2 between $-1.25\,\text{GeV}^2$ and $1.55\,\text{GeV}^2$ and the yields are considered between $-1.25\,\text{GeV}^2$ and $0.75\,\text{GeV}^2$. One gets a signal yield of 382 ± 2 and a background yield of 43 ± 1.

This corresponds to 25.6 ± 0.7 times more $D^* \ell \nu$ events tagged with the $H_c X$ tagging than with the FEI. This corresponds to an increase of the performance of the $H_c X$ compared to the FEI in data than in MC. This already indicates, that the calibration factor, computed in the next section, is closer to 1 for the $H_c X$ tagging than for the FEI method.

8.2 Systematic Uncertainties

The statistical error on the signal yield in MC is assumed to be poissonian. As the calibration factor has to be calculated for a specific signal decay, additionally systematic uncertainties on the signal-side must be considered. Errors which are induced by uncertainties on the form of the templates are neglected.

As systematical errors the following corrections are considered: lepton ID corrections, branching fraction corrections, tracking efficiency corrections and slow pion corrections.

In order to introduce these corrections a weight is assigned to each tagged MC event, such that the final corrected signal yield in MC is given by $N_{MC}^{D^* \ell \nu} = \sum_i w_{tot}^i$ where each w_{tot}^i is given by:

$$w_{tot}^i = w_{lID}^i \times w_{\pi_{slow}}^i \times w_{tracks}^i \times w_{BF}^i$$

8.2.1 Lepton ID Corrections

Differences in the shape of the variables used for the lepton identification in MC and data lead to differences in the efficiency of the lepton ID cuts. Consequently correction factors must be introduced to account for this effect.

8.2.2 Branching Fraction Corrections

The branching fractions used in the MC generation for the D^0 decays on the signal-side do not always correspond to the current official measurements given in [2]. Therefore the weights of the events are corrected such that the new values are considered:

$$w_{BF} = \frac{BF_{PDG} \pm \sigma_{BF_{PDG}}}{BF_{gen}}$$

8.2.3 Tracking Efficiency Corrections

The uncertainty on the tracking efficiency of the D^0 decay particles must be accounted for. To this end, to each track an uncertainty of 1 % is attached. Considering these uncertainties to be fully correlated, one gets:

$$w_{track} = 1 \pm n \times 0.01$$

where n is the number of tracks in the D^0 decay on the signal-side.

8.2.4 Slow Pion Corrections

Efficiency problems in the detection of slow neutral pions are known to exist in data compared to MC. As the D^{*+} on the signal-side subsequently decays into a π^+ with relative small momentum, this inefficiency must be considered here. In order to correct for it, correction factors, derived by Chaoyi Lyu, for different momentum ranges of the pion are applied.

8.3 Calibration Factor

Finally the calibration factor between MC and data can be calculated. For the semi-inclusive tagging approach a calibration factor of 1.08 ± 0.04 is computed. The value being compatible with 1 means that the H_cX tagging has a similar performance in data as in Monte Carlo.

For the FEI the computed factor is of 0.71 ± 0.02. The difference between these two values shows that the loss in performance of the BDTs which is seen in the FEI, seems not to strongly affect the H_cX algorithm.

In addition, the calibration factor is calculated for different cuts on the signal probability of H_c and B_{tag}, which is shown in Figure 8.2. For H_cX as well as for the FEI only a relatively little dependence on the chosen signal probability cut is observed. This property is important as this means that the BDTs attribute similar signal probabilities to the particle candidates in MC and in data. Thus, cuts on the signal probability have the same impact in MC as well as in data.
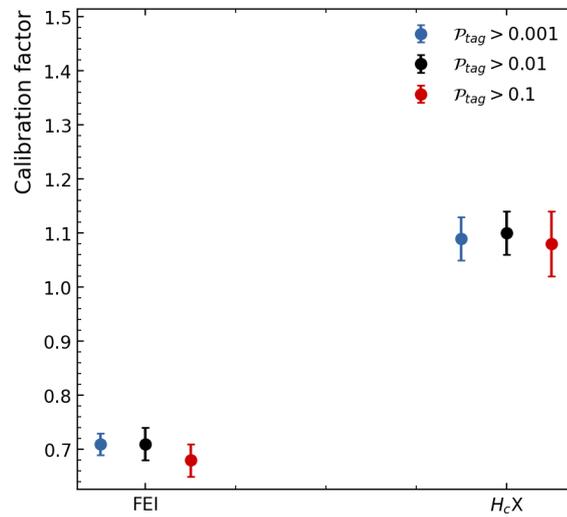
Figure 8.2: Calibration factor of the FEI and the $H_c X$ tagging method for different cuts on the H_c ($H_c X$) classifier output or B classifier output (FEI)

Application of the $H_c X$ Method on Monte Carlo Generated Data with $B \to \mu\nu$ as Signal-Side

After developing the tagging algorithm with the $D^* \ell \nu$ decay, it is tested with $B \to \mu\nu$ as signal-side. Compared to the complex $D^* \ell \nu$ decay with a lot final state particles, $B \to \mu\nu$ has only one single detectable track on the signal-side. These two decays therefore cover very different tagging examples.

9.1 Analysis Procedure

The analysis procedure is taken from the $B \to \mu\nu$ analysis performed with Belle [13] . As the $B_{sig} \to \mu\nu$ decay is a two body decay the muon has a monochromatic momentum of $p_\mu = 2.64\,\text{GeV}$ in the B_{sig} rest frame. This property allows for a good discrimination between signal and background events.

In order to boost the muon momentum into the signal B rest frame the tag-side is needed. On the other hand, the very low branching fraction and the subsequently low number of signal events in the data samples makes a high reconstruction efficiency mandatory. Therefore the Belle analysis ([13]) for example was performed with an inclusive tagging methodology.

Figure 9.2 shows the muon momentum distributions from the Belle analysis. It can be seen, that the very broad distribution in the centre-of-mass frame can be transformed to a sharper peak at the expected value by changing to the B_{sig} rest frame.

9.2 MC Samples for Evaluation of the Performance

The MC samples used to evaluate the $H_c X$ algorithm on $B \to \mu\nu$ decay are listed in tabular 9.1. The integrated luminosity of the samples were $200\,\text{fb}^{-1}$.

Additionally to the generic and the continuum events, a signal sample must be used as the $B \to \mu\nu$ events are not contained in the standard MC. For the following studies, the signal events are then scaled to the SM branching fraction.
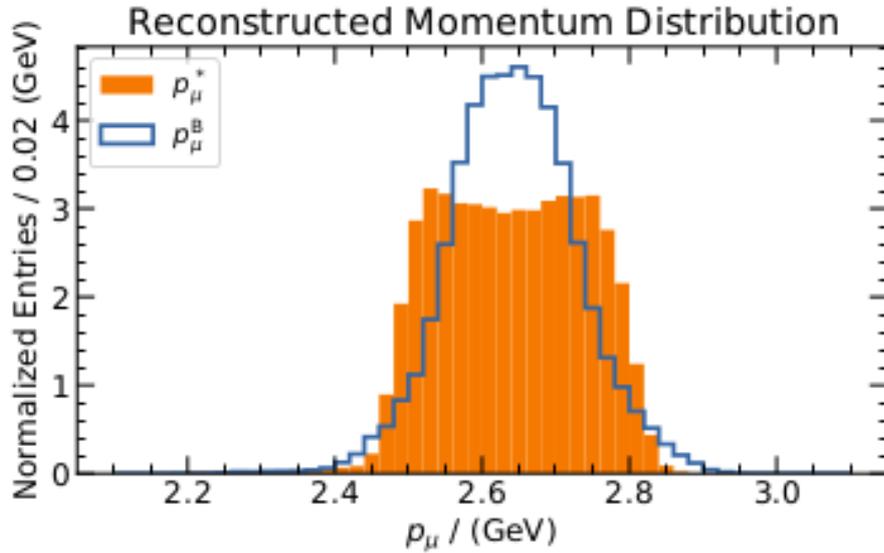
Figure 9.1: Muon momentum distributions in the centre-of-mass frame and in the B_{sig} rest frame performed
with the inclusive tagging in the Belle experiment [13]

Sample	Number of Events ($\times 10^6$)
Signal MC	10
$\Upsilon(4S) \rightarrow B^0 \bar{B}^0$	102
$\Upsilon(4S) \rightarrow B^+ B^-$	108
$c\bar{c}$	265.8
$u\bar{u}$	321.0
$d\bar{d}$	80.2
$s\bar{s}$	76.6

Table 9.1: Summary of simulated samples which are used to evaluate the performance of the $H_c X$ tagging
algorithm

9.3 Signal-Side Reconstruction and Selections

In the $B^+ \rightarrow \mu^+\nu$ decay only the muon can be detected. Therefore the signal-side reconstruction only
consists of one muon. For this muon the momentum in the CM frame must be above 2.2 GeV. In
addition the lepton ID and the muonID must be larger than 0.9 and the electronID smaller than 0.9.

9.4 Offline Cuts

A final selection is again made offline. These cuts slightly differ for $H_c X$ and for the FEI as they
were determined in order to maximise the significance of the signal. For both tagging methods, a full
hadronic decay of the tag-side was required.

9.4.1 $H_c X$

Cuts on the tag-side are made on the beam constrained mass $m_{bc} > 5.265$ GeV, on the energy difference -0.7 GeV $< \Delta E < 0.3$ GeV and on the signal probability of the charmed hadron $\mathcal{P}_{H_c} > 0.001$. On the signal side $\cos(\theta_{BY})$ must be in the interval $(-1.1, 1.1)$ and the CMS momentum of the lepton must be above 2.45 GeV. Additionally the continuum background is suppressed by the cut on the second Fox-Wolfram moment $R2 < 0.25$. Finally completeness constraints can be made by requiring that no additional tracks are in the ROE and the remaining energy must be below 0.9 GeV.

9.4.2 FEI

On the tag-side, cuts are made on $m_{bc} > 5.26$ GeV, on -0.2 GeV $< \Delta E < 0.2$ GeV and on $\mathcal{P}_{B_{tag}} > 0.001$. On the signal side $\cos(\theta_{BY})$ must be in the interval $(-1.5, 1.5)$ and the CMS momentum of the lepton must be above 2.2 GeV. Additionally the cut is made on $R2 < 0.4$. As before, no additional tracks are allowed in the ROE and the remaining energy bust be below 1 GeV.

9.5 Muon Momentum Distribution

Similar to the Belle approach described before, the tag-side is used to boost the momentum of the muon into the B_{sig} rest frame. The distribution of this boosted muon compared to the one in the CM frame is shown in Figure 9.2 for the $H_c X$ tagging and for the FEI. As for the inclusive tagging, the momentum transforms from a relative broad distribution to a peak around the theoretical value $p_\mu = 2.64$ GeV. Therefore, in the following, the muon momentum will only be considered in the B_{tag} rest frame.

(a) FEI (b) $H_c X$
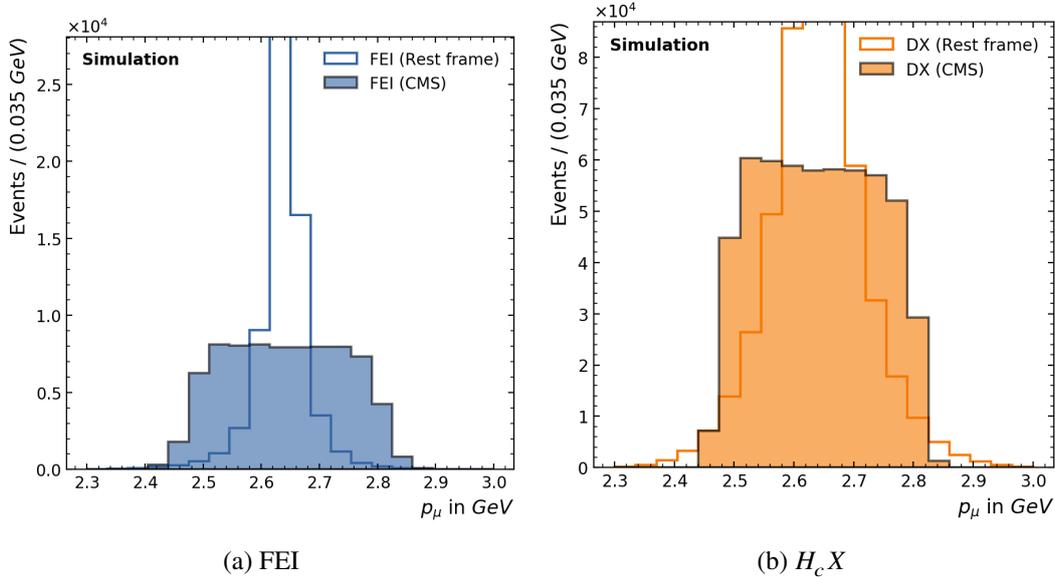
Figure 9.2: Muon momentum distributions in the centre-of-mass frame and in the B_{sig} rest frame performed with the FEI and the $H_c X$ tagging

In Figure 9.3 the muon momentum is shown for different tagging methods. Additionally to the H_cX algorithm and the FEI, first results from Daniel Jacobi with the inclusive tagging are shown.

Compared to the signal yield, a lot of background events, especially continuum events, are falsely tagged by all three tagging approaches. Nevertheless, the sample tagged by the FEI has a much higher purity than the other two tagging methods. If one compares the samples tagged by H_cX and by the inclusive tagging it is to be noted that the H_cX has less background events but that the background is again shifted towards the signal peak, which seems not to be the case for the inclusive tagging.

(a) FEI

(b) H_cX
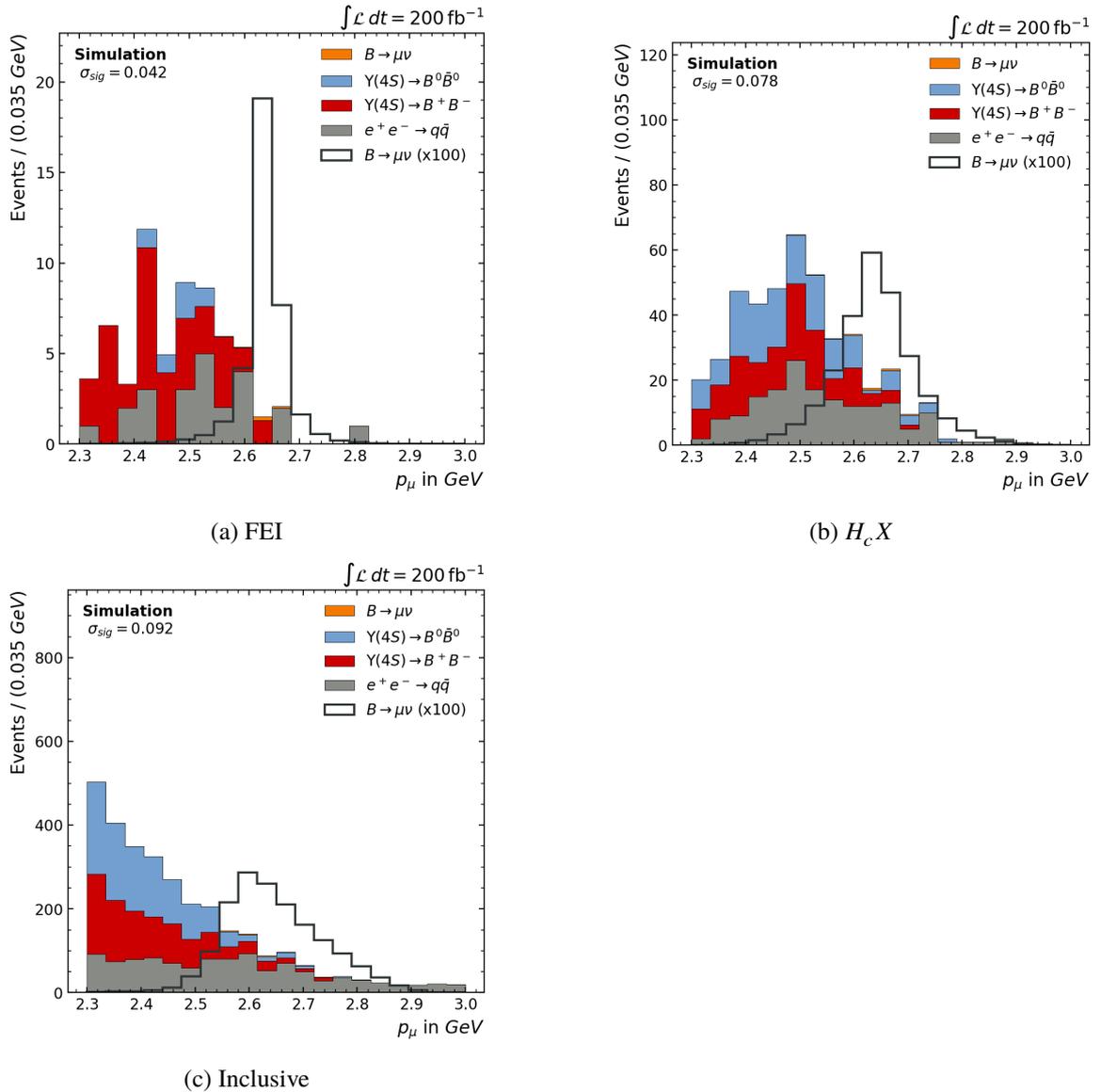
(c) Inclusive

Figure 9.3: Comparison of the p_{mu} distribution of the FEI, the H_cX and the inclusive tagging method. In addition to the number of signal events scaled to the SM branching fraction, the signal yield is represented with a factor of 100 compared to the actual branching ratio

9.6 Signal Yield and Significance

In order to compare the performances of the different tagging approaches, the signal yield can be considered. In Figure 9.4(a) the momentum distributions of the signal events is shown for all three tagging methods. Expectedly the number of signal events tagged by the H_cX algorithm is much higher than for the FEI and smaller than for the inclusive tagging.
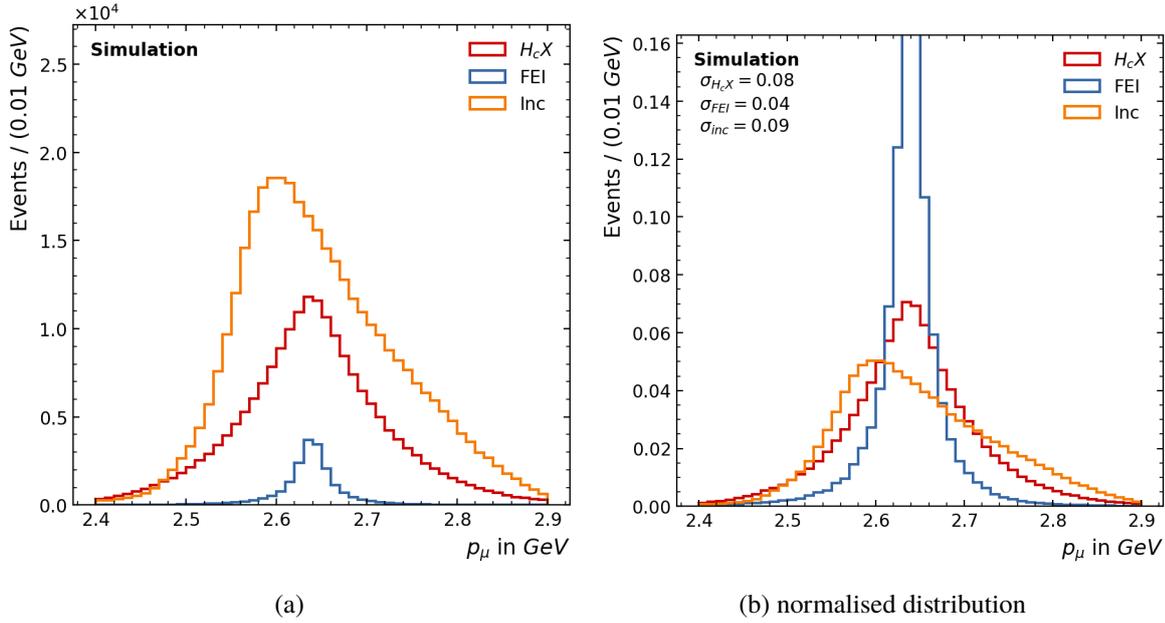
Figure 9.4: Muon momentum distributions in the rest frame of B_{sig}

The normalised muon momentum distribution is shown in figure 9.4(b). It is to be expected, that the resolution depends strongly on the choice of the tag-side reconstruction. In fact, the FEI has a very narrow peak at the expected momentum while the H_cX tagging produces a peak which is a bit broader but still symmetric. The peak of the inclusive tagging is a bit shifted to lower values and not completely symmetric.

The discrimination power between signal and background is indicated with $\frac{s}{\sqrt{b}}$. s and b are the number of signal and background events respectively in the interval where 68 % of the signal events lie. Thus the significance depends on the resolution of the signal distribution and of the amount and the position of the peak of the background events. $\frac{s}{\sqrt{b}}$ is represented in Figure 9.5 as a function of the integrated luminosity of the data sample. The significance of the signal tagged by the inclusive tagging is highly improved compared to the FEI. The H_cX also shows a slight improvement but a relatively small one compared to the much higher efficiency. This is especially due to the fact, that the background tagged by the H_cX algorithm is shifted towards the signal peak.
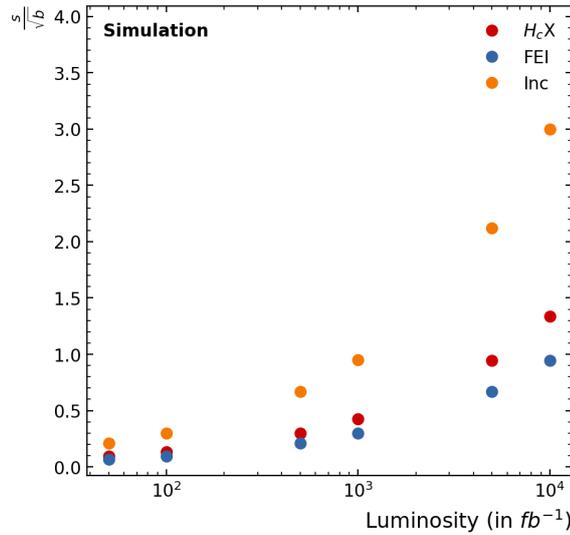
Figure 9.5: Number of signal events over the square root of the number of background events lying in the 68 % signal region for different integrated luminosities

9.7 Continuum Suppression for H_cX

In the $B \to \mu\nu$ signal decay, the continuum events are the main background, especially in the signal region. In order to improve the continuum rejection performed with the cut on the 2nd Fox-Wolfram moment $R2$, a BDT is trained, using further discriminating variables. The feature variables as well as the training of the BDT are described in more details in chapter 10.

Similar to the continuum suppression implemented for the FEI, the continuum suppression for the H_cX tagging is done by using the tag-side B meson and its rest-of-event. As the tag-side ($B_{tag} \to H_cX$) cannot be reconstructed independently of the signal-side, the trained continuum suppression BDT is specific for the $B \to \mu\nu$ decay.

The BDT is trained with the same feature variables and the same hyper- parameters as for the FEI. With an area under the ROC (receiver operating characteristic) curve of 0.98, the BDT has a clear discrimination power between $\Upsilon(4S) \to B\bar{B}$ and continuum events. The performance can be seen in Figure 9.6(a) where a BDT output of 1 means that the event is likely to be a continuum event. When applying the analysis cuts (without the $R2$ cut), the majority of the continuum events is rejected. This can be seen in Figure 9.6(b). The continuum events which are retained are the ones with relatively small BDT output, as the cuts only select events which are similar to the signal events.

Nevertheless, the BDT still allows to reject more continuum events, as can be seen in Figure 9.7. A cut on the BDT output of 0.008 was chosen, by maximising $\frac{s}{\sqrt{b}}$.

The improvement of the purity of the tagged sample can be seen in Figure 9.8. While the signal yield is only very slightly affected, the cut on the BDT output rejects a lot more continuum events than the simple $R2$ cut.

The improvement of the continuum suppression leads to a much higher signal significance for the measurement of the $B \to \mu\nu$ decay. In addition to results with the FEI, the inclusive tagging and the H_cX tagging with a simple $R2$ cut, Figure 9.9 shows $\frac{s}{\sqrt{b}}$ for the H_cX tagging with the applied

(a) without cuts   (b) with cuts
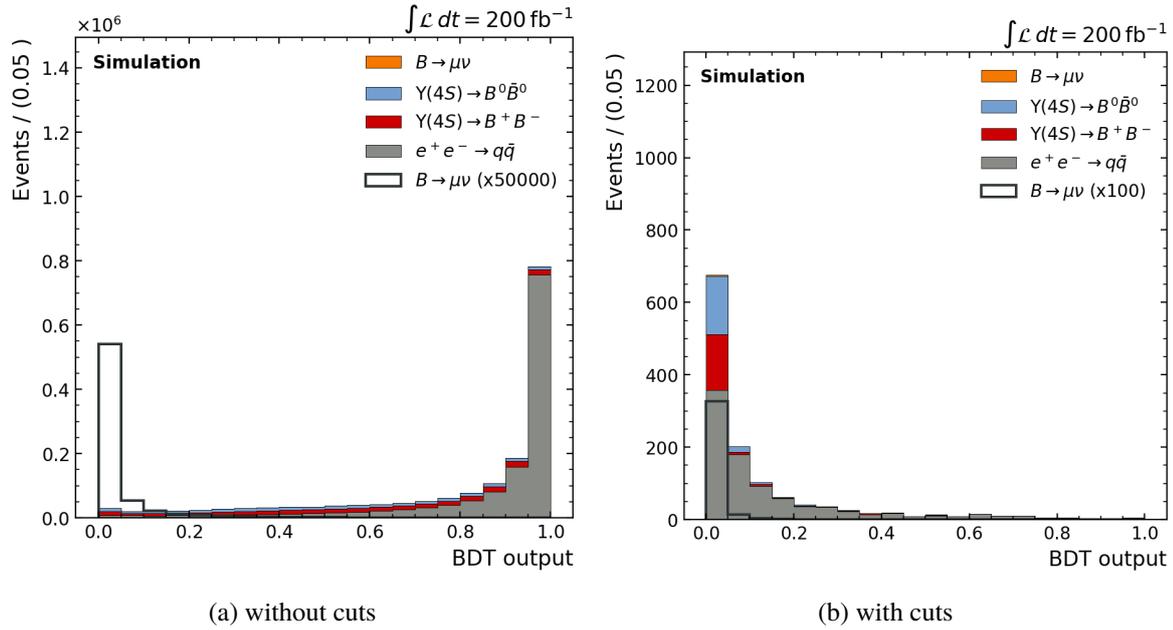
Figure 9.6: Output of the continuum suppression BDT with and without applied cuts

continuum suppression BDT. It can be seen, that the significance increases further more, leading to a quite important improvement compared to the FEI. While for the inclusive tagging a continuum suppression with a BDT is already applied, for the FEI the rejection was done with the cut on $R2$. Thus it must be considered, that a continuum suppression would surely also improve the results of the FEI.
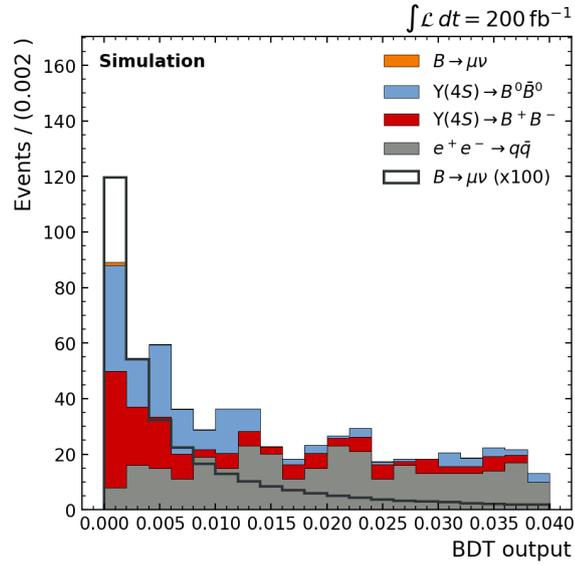
Figure 9.7: Output of the continuum suppression BDT with applied cuts

(a) $R2$ cut
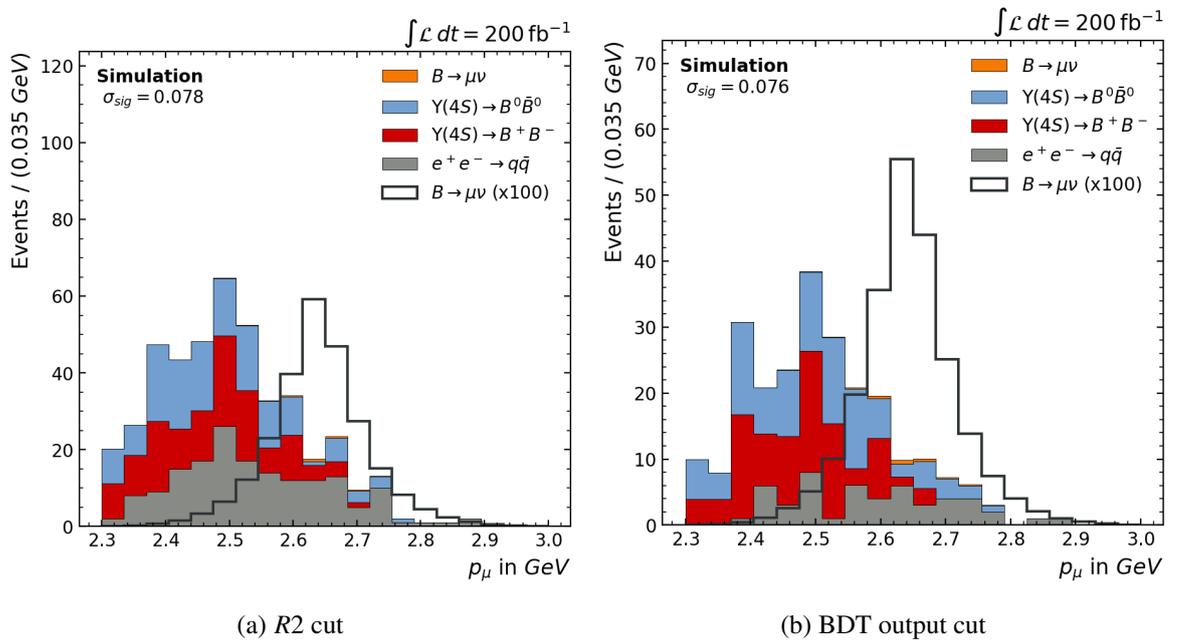
(b) BDT output cut

Figure 9.8: Comparison of the p_{mu} distribution of the H_cX tagging with an $R2$ cut and with a BDT output cut
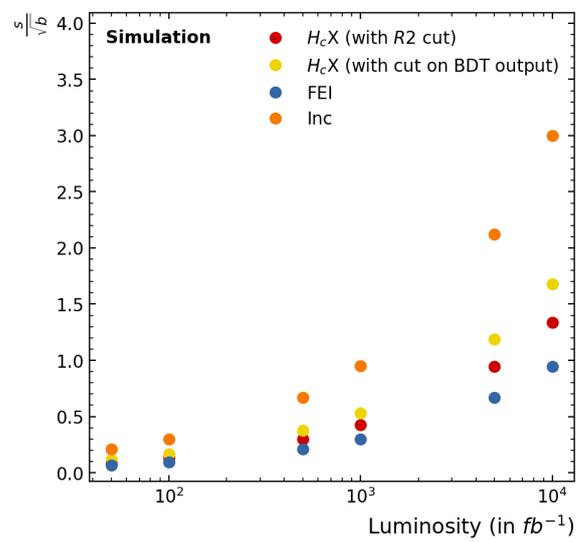
Figure 9.9: $\frac{s}{\sqrt{b}}$ for different integrated luminosities: in addition to the H_cX tagging with the cut on $R2$, the H_cX tagging with a cut on the BDT output is represented

Implementation of a Continuum Suppression in the FEI

As previously seen with the $B \to \mu\nu$ signal decay, continuum events can be an important background, limiting the measurements of some decays. Therefore a continuum suppression with a BDT is implemented into the FEI.

10.1 Event shape of $B\bar{B}$ and $q\bar{q}$ events

The continuum suppression discriminates between $B\bar{B}$ and $q\bar{q}$ events by exploiting differences in their event shape. In the case where two B mesons are produced ($e^+e^- \to \Upsilon(4s) \to B\bar{B}$) the two mesons are almost at rest in the centre of mass frame of the collision, as the $\Upsilon(4S)$ mass is only slightly above the B pair threshold. This leads to the fact, that the decay products of the B mesons are emitted isotropically in the rest frame. In comparison, quarks which are directly created in the e^+e^- collision ($e^+e^- \to q\bar{q}$) have a high initial momentum. In the CM frame the particles produced by quark hadronisation are emitted back-to-back [8]. These two different event topologies, shown in Figure 10.1, allow one to define variables, which discriminate well between $\Upsilon(4S)$ and continuum events.
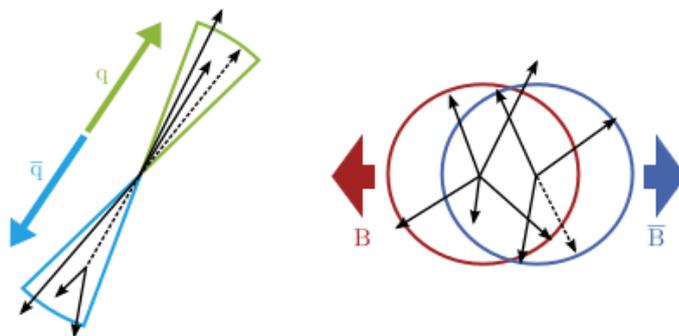
Figure 10.1: Event shape of a continuum and a $B\bar{B}$ event [14]

10.2 Variables

Before defining the discriminating variables, a quantity called thrust axis must be introduced, which is central to the calculation of many of these variables. The thrust axis is defined as the unit vector along which the projection of the momenta of the considered particles is maximal. More details on the variables can be found in [11].

- R2: second reduced Fox Wolfram moment, which corresponds to the ratio of the 2nd to the 0th order Fox Wolfram moment.

  In case of N particles with momenta $\boldsymbol{p}_i$, the lth order Fox-Wolfram moment H_l is defined as:

$$H_l = \sum_{i,j}^{N} |\boldsymbol{p}_i||\boldsymbol{p}_j|P_l(\cos(\theta_{ij})) \text{ and } R_l = \frac{H_l}{H_0} \qquad (10.1)$$

  with $\theta_{i,j}$ the angle between $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$, and P_l the lth order Legendre polynomial. [11]

- cosTBTO: cosine of the angle between the thrust axis of the B and the thrust axis of the particles in the ROE

- cosTBz: cosine of the angle between the thrust axis of the B and the z-axis

- thrustBm: magnitude of the B thrust axis

- thrustOm: magnitude of the ROE thrust axis

- CleoConeCS: these 9 variables were first introduced in the CLEO experiment. They correspond to the scalar momentum flow around the thrust axis which fall into concentric cones in angular intervals of 10°. The cones in direction of the thrust axis are merged with the corresponding ones in opposite direction.

- KSFW Variables: the 18 Kakuno-Super-Fox-Wolfram variables also describe the sphericity of the event. It consists of m_{miss}^2, E_t and 16 KSFW moments (H_{xl}^{so} and H_l^{oo} with $l = 0, 1, 2, 3, 4$).

10.3 Method

To suppress the continuum background, physics analyses often simply cut on $R2$ ($R2 < 0.3$). As it can be seen in Figure 10.2 this cut already eliminates the majority of the continuum events. In order to improve the separation, a BDT can be used including the discriminating power of the variables described above. Some of the discriminating variables are shown in Figure 10.2. The other variables can be found in the appendix A.

The continuum rejection starts by reconstructing the tag-side B meson with the FEI. This B meson and the corresponding rest of event are used to compute the aforementioned variables. A pre-trained BDT is then applied, which attributes to each event a BDT output, which can interpreted as the probability that the event is a continuum event. The discriminating power of the BDT is illustrated in
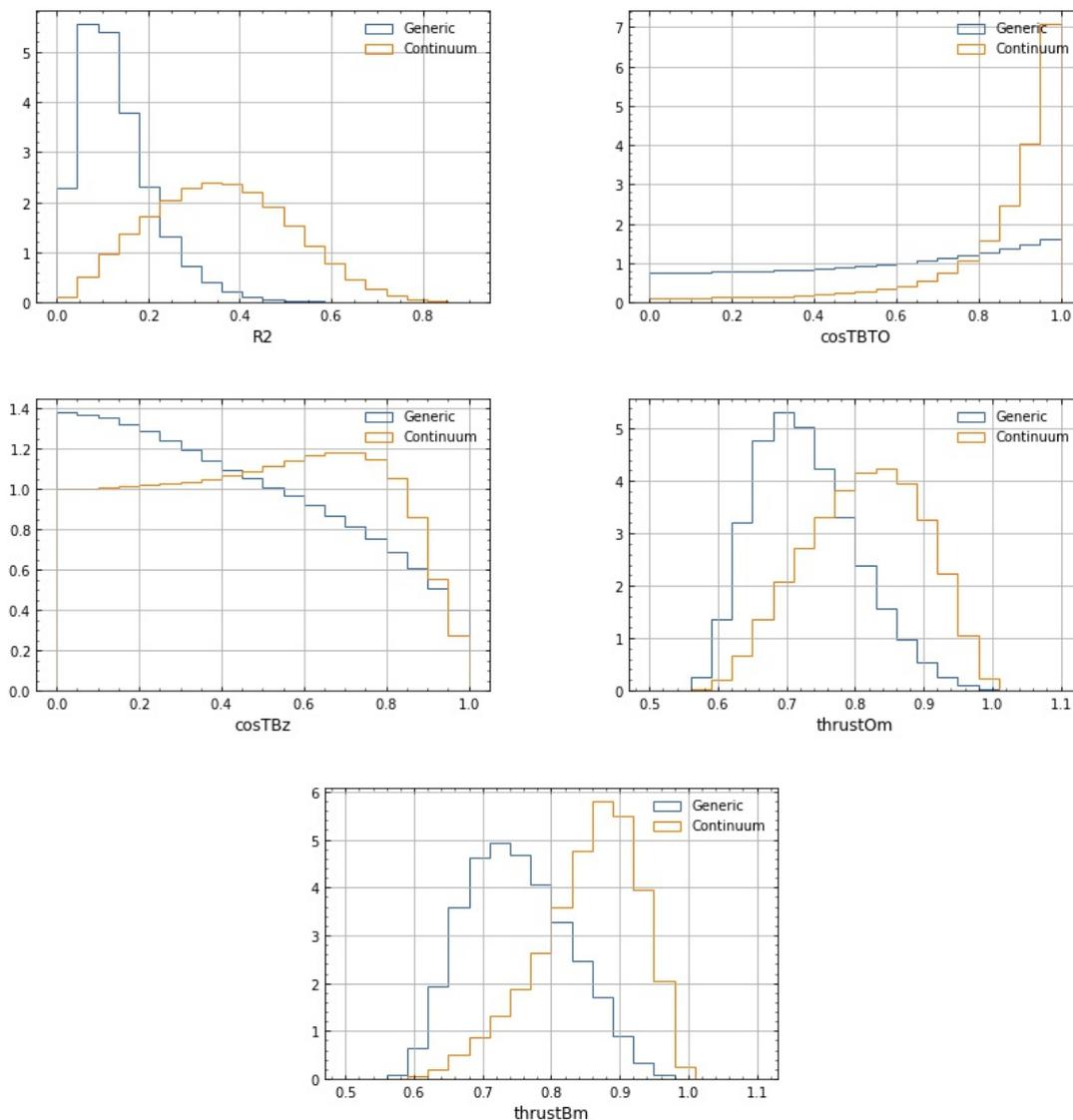
Figure 10.2: Example of the most discriminating variables

figure 10.3 as ROC curve. Additionally the separation power for different $R2$ cuts is depicted, which shows that the BDT brings a small improvement especially for smaller false positive rates.

An artificial neural network (ANN) was also trained by William Sutcliffe with Tensorflow using the same input variables. As the area under curve of the ROC curve of the BDT (0.92) and of the ANN (0.918) do not differ significantly, a simple BDT is used in the following.
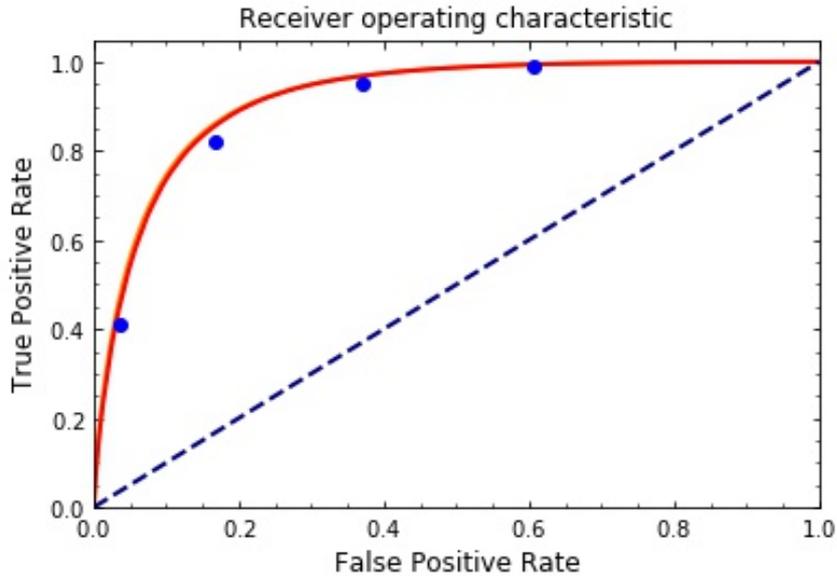
Figure 10.3: Performance of the BDT (full line) and for different $R2$ cuts (single points)

10.4 Integration in the FEI

In order to simplify its application, the continuum suppression is integrated into the FEI module. In addition to the reconstruction of the tag-side and the determination of a signal probability for each particle candidate, the FEI should provide a probability for each event to be a continuum event.

As the continuum suppression is based on the reconstructed B_{tag} the new module is added right at the end of the FEI module. Similarly to the other FEI steps, a pre-reconstruction, a training and a post-reconstruction module are implemented.

10.4.1 Pre-Reconstruction

In the pre-reconstruction the events are prepared, such that the BDT can be trained or applied. This step takes as input the reconstructed B meson candidates from the FEI. For each candidate its rest of event is built. Additionally the "buildContinuumSuppression" module is applied in order to calculate the feature variables needed for the BDT.

10.4.2 Training

This step can be enabled if the BDT will be trained. It prepares the events to save them into a file, which can subsequently be used for training offline. A best candidate selection of the reconstructed B candidates is performed, to ensure that each event cannot occur twice in the sample. The discriminating variables mentioned above are then written to a root file. Additionally the target of the BDT, containing the information if the event is a continuum event or not, is saved.

10.4.3 Post-Reconstruction

The post-reconstruction represents the last step, where for each event a continuum probability is computed. In the case where a trained continuum suppression BDT can be found, the BDT is applied to each B meson candidate.

10.5 Training Process

The training of the FEI has to be performed for every new MC campaign. The implementation of the continuum suppression is implemented such that its training is done separately after the training of the FEI. The whole training process is managed by a signal script. The infrastructure managing the training process is adopted from the FEI training.

The script calls a steering files which runs the FEI with enabled training of the continuum suppression. The tag-side reconstruction is done for $\Upsilon(4S) \rightarrow B\bar{B}$ and continuum MC samples. After the B_{tag} lists are created, the pre-reconstruction and the training modules of the continuum suppression are called. With the obtained training files, the BDT can be trained offline. To construct and train the BDT the FastBDT package (described in [15]) is used. The same hyper-parameter as for the FEI BDTs are chosen (given in table 10.1).

Parameter	Value
nTrees	400
nCutLevels	10
nLevels	3
shrinkage	0.1
randRatio	0.5

Table 10.1: Hyper-Parameters used for the BDT of the continuum suppression

Once the training is completed, the post-reconstruction is performed, where the BDT can be applied to the events. The final step is then to upload the BDT of the continuum suppression such that the training process has only to be done once.

Conclusion

11.1 Development of the Semi-Inclusive Tagging Algorithm

The semi-inclusive tagging approach is a compromise between the exclusive and the inclusive tagging by partially reconstructing the tag-side. The novel method aims to keep a relatively high purity while also retaining a high efficiency. The method was tested on signal decays with very different topologies with $B \to D^* \ell \nu$ being a decay with many final state particles and $B \to \mu \nu$ containing only one detectable track.

For the signal decay $B \to D^* \ell \nu$, the tagging efficiency can be greatly improved with the semi-inclusive tagging approach. The number of tagged signal events increases by a factor of 17.3 ± 0.4 in MC and of 25.6 ± 0.7 in data compared to the Full Event Interpretation. The purity of the sample only slightly decreases. Nevertheless the new approach comes with some drawbacks. First the signal distribution in the relevant fitting variable m^2_{miss} is broadened. Second, the background is shaped such that it peaks at the same value as the signal. This means a loss in the discrimination power between signal and background.

In the $B \to \mu \nu$ signal decay, the same behaviour can be seen. The shift of the background events towards the signal peak strongly affects the significance of the observed signal peak. Thus, the $\frac{s}{\sqrt{b}}$ ratio is only slightly improved by the $H_c X$ method compared to the FEI, despite the high increase of efficiency. The significance can though be further increased by a continuum suppression with a pre-trained BDT, which rejects most of the continuum events.

Before applying the novel approach in physics analyses, the sculpting of the background must be further investigated. One approach to solve this issue would be to improve the choice whether a particle is added to the X system or not. Attempts with graph neural networks are currently investigated by Axel Heimeroth.

11.2 Integration of a Continuum Suppression into the FEI

Continuum events are one of the main backgrounds in many physics analyses, which makes its suppression essential. As the event shape of the $e^+ e^- \to q\bar{q}$ events differ strongly from the one of $\Upsilon(4S) \to B\bar{B}$ events, several variables can be computed in order to discriminate between $B\bar{B}$ events and continuum. A BDT trained on these features has a good discrimination power, with a performance of 0.92.

The continuum suppression as well as the training of the BDT was implemented into the FEI as an additional step. Optionally, the FEI adds a probability to each event, that this event is a continuum event. In physics analyses, a cut can then be made on this probability, depending on the chosen working point.

Bibliography

[1] T. Keck et al.,
The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment,
(2018), arXiv: 1807.08680 [hep-ex] (cit. on pp. 1, 13).

[2] P. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01 (cit. on pp. 3, 15, 17, 29).

[3] *Particle content of the Standard Model*,
https://en.wikipedia.org/wiki/Standard_Model#/media/File:
Standard_Model_of_Elementary_Particles.svg, [Online; August 2021] (cit. on p. 3).

[4] Y. Amhis et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*,
The European Physical Journal C **81** (2021), ISSN: 1434-6052,
URL: http://dx.doi.org/10.1140/epjc/s10052-020-8156-7 (cit. on pp. 4, 5).

[5] *Super KEKB and Belle II*,
https://www.belle2.org/project/super_kekb_and_belle_ii,
[Online; August 2021] (cit. on p. 7).

[6] *Belle II Luminosity*,
https://confluence.desy.de/display/BI/Belle+II+Luminosity,
[Online; August 2021] (cit. on pp. 8, 9).

[7] T. Abe et al., *Belle II Technical Design Report*, 2010,
arXiv: 1011.0352 [physics.ins-det] (cit. on pp. 8, 10).

[8] A. J. Bevan et al., *The Physics of the B Factories*, The European Physical Journal C **74** (2014),
ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-014-3026-9
(cit. on pp. 11, 41).

[9] T. Keck,
Machine learning algorithms for the Belle II experiment and their validation on Belle data,
PhD thesis: Karlsruhe Institute of Technology (KIT), 2017 (cit. on pp. 12, 13).

[10] W. Sutcliffe and F. Bernlochner,
A calibration of the Belle II hadronic tag-side reconstruction algorithm with $B \to X\ell\nu$ decays,
(2020) (cit. on pp. 13, 27).

[11] E. Kou et al., *The Belle II Physics Book*,
Progress of Theoretical and Experimental Physics **2019** (2019), ISSN: 2050-3911,
URL: http://dx.doi.org/10.1093/ptep/ptz106 (cit. on pp. 15, 22, 23, 42).

[12] *Belle II: World record in the accelerator*,
https://particle-physics.desy.de/events__news/news/news_archive_2020/
belle_ii_world_record_in_the_accelerator/index_eng.html,
[Online; August 2021] (cit. on p. 15).

[13]   M. T. Prim et al., *Search for B + ßμ + νμ and B + ßμ + N with inclusive tagging*,
       Physical Review D **101** (2020), ISSN: 2470-0029,
       URL: http://dx.doi.org/10.1103/PhysRevD.101.032007 (cit. on pp. 16, 17, 31, 32).

[14]   D. Weyland,
       Continuum Suppression with Deep Learning techniques for the Belle II Experiment,
       MA thesis: Karlsruhe Institute of Technology (KIT), 2017 (cit. on pp. 17, 41).

[15]   T. Keck, *FastBDT: A speed-optimized and cache-friendly implementation of stochastic
       gradient-boosted decision trees for multivariate classification*, 2016,
       arXiv: 1609.06119 [cs.LG] (cit. on p. 45).

30th August 2021 11:19

Continuum Suppression

A.1 Feature Variables for the Continuum Suppression

List of Figures

List of Tables