




The Full Event Interpretation

An Exclusive Tagging Algorithm for the Belle II Experiment

T. Keck¹ · F. Abudinén² · Florian U. Bernlochner¹  · R. Cheaib³ · S. Cunliffe⁴ · M. Feindt¹ · T. Ferber⁴ · M. Gelb¹ · J. Gemmler¹ · P. Goldenzweig¹ · M. Heck¹ · S. Hollitt⁵ · J. Kahn⁶ · J.-F. Krohn⁷ · T. Kuhr⁶ · I. Komarov⁴ · L. Ligioi² · M. Lubej⁸ · F. Metzner¹ · M. Prim¹ · C. Pulvermacher¹ · M. Ritter⁶ · J. Schwab¹ · W. Sutcliffe¹ · U. Tamponi⁹ · F. Tenchini⁴ · N. E. Toutounji¹⁰ · P. Urquijo⁷ · D. Weyland¹ · A. Zupanc⁸

Received: 25 July 2018 / Accepted: 16 January 2019 / Published online: 25 February 2019
© The Author(s) 2019

Abstract

The full event interpretation is presented: a new exclusive tagging algorithm used by the high-energy physics experiment Belle II. The experimental setup of Belle II allows the precise measurement of otherwise inaccessible B meson decay modes. The Full Event Interpretation algorithm enables many of these measurements. The algorithm relies on machine learning to automatically identify plausible B meson decay chains based on the data recorded by the detector. Compared to similar algorithms employed by previous experiments, the Full Event Interpretation provides a greater efficiency, yielding a larger effective sample size usable in the measurement.

Keywords Multivariate classification · Full event interpretation · Full reconstruction · Tagging · Belle II · HEP · Machine learning

Introduction

The Belle II experiment is located at the SuperKEKB electron–positron collider in Tsukuba, Japan, and was commissioned in 2018. The experiment is designed to perform a

wide range of high-precision measurements in all fields of heavy flavor physics, in particular it will investigate the decay of B mesons [1]. For this purpose, the experiment is expected to record about 40 billion collision events each containing an $\Upsilon(4S)$ resonance, which at least 96% of the time decays into exactly **two** B mesons (a $B\bar{B}$ pair). Each B meson decays via various intermediate states into a set of final-state particles, which are considered stable in the Belle II detector. In general, charged final-state particles are reconstructed as tracks in the central drift chamber and in the inner silicon-based vertex detectors, whereas neutral final-state particles are reconstructed as energy depositions (called clusters) in the electromagnetic calorimeter. The entire experimental setup of the detector and the collider is described in more detail in Doležal and Uno [1].

The measurement of the branching fraction of rare decays like $B \rightarrow \tau\nu_\tau$, $B \rightarrow K\nu\nu$ or $B \rightarrow l\nu\gamma$, with undetectable neutrinos in their final states, is challenging. However, the second B meson in each event can be used to constrain the allowed decay chains. This general idea is known as **tagging**. Conceptually, each $\Upsilon(4S)$ event is divided into two sides: the signal-side containing the tracks and clusters compatible with the assumed signal B_{sig} decay the physicist is interested in, e.g., a rare decay like $B \rightarrow \tau\nu$; and the tag-side containing

✉ T. Keck
thomas.keck2@kit.edu

✉ Florian U. Bernlochner
florian.bernlochner@kit.edu

¹ Karlsruhe Institute of Technology, Campus Süd, Institut für Experimentelle Teilchenphysik, Wolfgang-Gaede-Str. 1, 76131 Karlsruhe, Germany

² Max-Planck-Institut für Physik, Munich, Germany

³ University of Mississippi, Mississippi, MS, USA

⁴ Deutsches Elektronen-Synchrotron, Hamburg, Germany

⁵ University of Adelaide, Adelaide, Australia

⁶ Ludwig Maximilians Universität, Munich, Germany

⁷ University of Melbourne, Melbourne, Australia

⁸ Jožef Stefan Institute, Ljubljana, Slovenia

⁹ INFN - Sezione di Torino, Turin, Italy

¹⁰ University of Sydney, Sydney, Australia

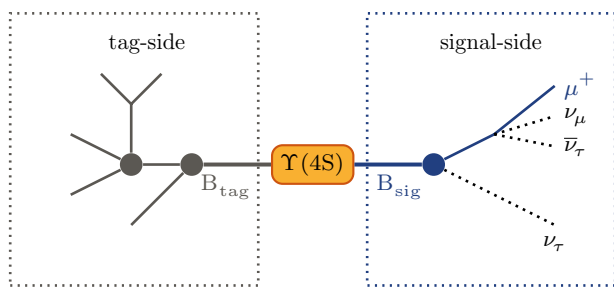


Fig. 1 Schematic overview of a $\Upsilon(4S)$ decay: (Left) a common tag-side decay $B_{\text{tag}}^- \rightarrow D^0 (\rightarrow K_s^0 (\rightarrow \pi^- \pi^+) \pi^- \pi^+) \pi^-$ and (right) a typical signal-side decay $B_{\text{sig}}^+ \rightarrow \tau^+ (\rightarrow \mu^+ \nu_\mu \bar{\nu}_\tau) \nu_\tau$. The two sides overlap spatially in the detector, therefore the assignment of a measured track to one of the sides is not known a priori

the remaining tracks and clusters compatible with an arbitrary B_{tag} meson decay. Figure 1 depicts this situation.

The initial four momentum of the produced $\Upsilon(4S)$ resonance is precisely known and no additional particles are produced in this primary interaction. Therefore, because of the relevant quantum numbers conservation, knowledge about the properties of the tag-side B_{tag} meson allows one to recover information about the signal-side B_{sig} meson which would otherwise be inaccessible. Most importantly, all reconstructed tracks and clusters which are not assigned to the B_{tag} mesons must be compatible with the signal decay of interest.

Ideally, a full reconstruction of the entire event has to take all reconstructed tracks and clusters into account to attain a correct interpretation of the measured data. The **Full Event Interpretation** (FEI) algorithm presented in this article is a new exclusive tagging algorithm developed for the Belle II experiment, embedded in the Belle II Analysis Software Framework (basf2) [2]. The FEI automatically constructs plausible B_{tag} meson decay chains compatible with the observed tracks and clusters, and calculates for each decay chain the probability of it correctly describing the true process using gradient-boosted decision trees. “Exclusive” refers to the reconstruction of a particle (here the B_{tag}) assuming an explicit decay channel.

Consequently, exclusive tagging reconstructs the B_{tag} independently of the B_{sig} using either **hadronic** or **semileptonic** B meson decay channels. The decay chain of the B_{tag} is explicitly reconstructed and therefore the assignment of tracks and clusters to the tag-side and signal-side is known.

In the case of a measurement of an exclusive branching fraction like $B_{\text{sig}} \rightarrow \tau \nu_\tau$, the entire decay chain of the $\Upsilon(4S)$ is known. As a consequence, all tracks and clusters measured by the detector should be already accounted for. In particular, the requirement of no additional tracks, besides the ones used for the reconstruction of the $\Upsilon(4S)$, is an extremely

powerful and efficient way to remove most reducible¹ backgrounds. This requirement is called the **completeness constraint** throughout this text.

In the case of a measurement of an inclusive branching fraction like $B_{\text{sig}} \rightarrow X_u l \nu$, all remaining tracks and clusters, besides the ones used for the lepton l and the B_{tag} meson, are identified with the X_u system. Hence, the branching fraction can be determined without explicitly assuming a decay chain for the X_u system.

The performance of an exclusive tagging algorithm depends on the **tagging efficiency** (i.e., the fraction of $\Upsilon(4S)$ events which can be tagged), the **tag-side efficiency** (i.e., the fraction of $\Upsilon(4S)$ events with a correct tag) and on the quality of the recovered information, which determines the **tag-side purity** (i.e., the fraction of the tagged $\Upsilon(4S)$ events with a correct tag) of the tagged events.

The exclusive tag typically provides a pure sample (i.e., purities up to 90% are possible). But, this approach suffers from a low tag-side efficiency, just a few percent, since only a tiny fraction of the B decays can be explicitly reconstructed due to the large amount of possible decay channels and their high multiplicity. The imperfect reconstruction efficiency of tracks and clusters further degrades the efficiency.

Both the quality of the recovered information and the systematic uncertainties depend on the decay channel of the B_{tag} , therefore we distinguish further between hadronic and semileptonic exclusive tagging.

Hadronic tagging considers only hadronic B decay chains for the tag-side [3, Section 7.4.1]. Hence, the four momentum of the B_{tag} is well-known and the tagged sample is very pure. A typical hadronic B decay has a branching fraction of $\mathcal{O}(10^{-3})$. As a consequence, hadronic tagging suffers from a low tag-side efficiency and can only be applied to a tiny fraction of the recorded events. Large combinatorics of high-multiplicity decay channels further complicate the reconstruction and require tight selection criteria.

Semileptonic tagging considers only semileptonic $B \rightarrow D l \nu$ and $B \rightarrow D^* l \nu$ decay channels [3, Section 7.4.2]. Due to the presence of a high-momentum lepton, these decay channels can be easily identified and the semileptonic tagging usually yields a higher tag-side efficiency compared to hadronic tagging due to the large semileptonic branching fractions. On the other hand, the semileptonic tag will miss kinematic information due to the neutrino in the final state of the decay. Hence, the sample is not as pure as in the hadronic case.

To conclude, the FEI provides a hadronic and semileptonic tag for B^\pm and B^0 mesons. This enables the measurement of exclusive decays with several neutrinos and

¹ Reducible background has distinct final-state products from the signal.

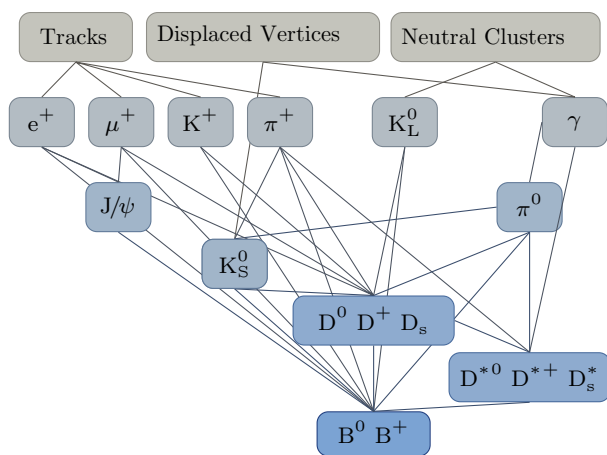


Fig. 2 Schematic overview of the FEI. The algorithm operates on objects identified by the reconstruction software of the Belle II detectors: charged tracks, neutral clusters and displaced vertices. In six distinct stages, these basic objects are interpreted as final-state particles ($e^+, \mu^+, K^+, \pi^+, K_L^0, \gamma$) combined to form intermediate particles ($J/\psi, \pi^0, K_S^0, D, D^*$) and finally form the tag-side B mesons

inclusive decays. In both cases, the FEI provides an explicit tag-side decay chain with an associated probability.

Methods

The FEI algorithm follows a hierarchical approach with six stages, visualized in Fig. 2. Final-state particle candidates are constructed using the reconstructed tracks and clusters, and combined to intermediate particles until the final B candidates are formed. The probability of each candidate to be correct is estimated by a multivariate classifier. A multivariate classifier maps a set of input features (e.g., the four momentum or the vertex position) to a real-valued output, which can be interpreted as a probability estimate. The multivariate classifiers are constructed by optimizing a loss function (e.g., the misclassification rate) on Monte Carlo simulated $\Upsilon(4S)$ events and are described later in detail.

All steps in the algorithm are configurable. Therefore, the decay channels used, the cuts employed, the choice of the input features, and hyper-parameters of the multivariate classifiers depend on the configuration. A more detailed description of the algorithm and the default configuration can be found in Keck [4] and in the following we give a brief overview over the key aspects of the algorithm.

Combination of Candidates

Charged final-state particle candidates are created from tracks assuming different particle hypotheses. Neutral final-state particle candidates are created from clusters and displaced vertices constructed by oppositely charged tracks.

Each candidate can be correct (signal) or wrong (background). For instance, a track used to create a π^+ candidate can originate from a pion traversing the detector (signal), from a kaon traversing the detector (background) or originates from a random combination of hits from beam background (also background).

All candidates available at this stage are combined to intermediate particle candidates in the subsequent stages, until candidates for the desired B mesons are created. Each intermediate particle has multiple possible decay channels, which can be used to create valid candidates. For instance, a B^- candidate can be created by combining a D^0 and a π^- candidate, or by combining a D^0 , a π^- and a π^0 candidate. The D^0 candidate could be created from a K^- and a π^+ , or from a K_S^0 and a π^0 .

The FEI reconstructs more than 100 explicit decay channels, leading to $\mathcal{O}(10000)$ distinct decay chains.

Multivariate Classification

The FEI employs multivariate classifiers to estimate the probability of each candidate to be correct, which can be used to discriminate correctly identified candidates from background. For each final-state particle and for each decay channel of an intermediate particle, a multivariate classifier is trained which estimates the signal probability that the candidate is correct. To use all available information at each stage, a network of multivariate classifiers is built, following the hierarchical structure in Fig. 2.

For instance, the classifier for the decay of $B^- \rightarrow D^0 \pi^-$ would use the signal probability of the D^0 and π^- candidates, as input features to estimate the signal probability of the B^- candidate created by combining the aforementioned D^0 and π^- candidates.

Additional input features of the classifiers are the kinematic and vertex fit information of the candidate and its daughters. The multivariate classifiers used by the FEI are trained on Monte Carlo simulated events. The training is fully automatized and distributed using a map reduce approach [5]. Monte Carlo simulated data used to train the FEI is partitioned. At each reconstruction stage, the partitioned data is distributed to nodes where the reconstruction is performed and training datasets are produced (the mapping stage). The reduction stage consists of merging the training datasets and training multivariate classifiers with these training datasets.

The available information flows from the data provided by the detector through the intermediate candidates into the final B meson candidates, yielding a single number which can be used to distinguish correctly from incorrectly identified B_{tag} mesons. The process is visualized in Fig. 2. This allows one to tune the trade-off between tag-side efficiency and tag-side purity of the algorithm by requiring a minimal

signal probability. By contrast, most exclusive measurements by Belle, which used the previous `FR` algorithm, chose a working point near the maximum tag-side efficiency as described in “[Previous work](#)” section.

Combinatorics

It is not feasible to consider all possible B meson candidates created by all possible combinations. The amount of possible combinations scales with the factorial in the number of tracks and clusters. This problem is known as **combinatorics** in high-energy physics. Furthermore, it is not worthwhile to consider all possible B meson candidates, because all of them are wrong except for two in the best-case scenario.

The `FEI` uses two sets of the so-called **cuts**. A cut is a criterion that a candidate has to fulfill to be considered further. For instance, one could demand that the beam-constrained mass of the B meson candidate is near the nominal mass 5.28 GeV of a B meson particle, or that a μ^+ candidate has a high muon particle identification likelihood, which combines sub-detector information to identify muons.

Directly after the creation of the candidate (either from a track/cluster, or by combining other candidates), but before the application of the multivariate classifier, the `FEI` uses loose and fast **pre-cuts** to remove wrongly identified candidates (background), without losing signal. The main purpose of these cuts is to save computing time and to reduce the memory consumption. These **pre-cuts** are applied separately for each decay channel.

At first, a very loose fixed cut is applied on a quantity which is fast to calculate, e.g., the energy for photons, the invariant mass for D mesons, the energy released in the decay for D^* mesons, or the beam-constrained mass for hadronic B mesons. Second, the remaining candidates are ranked according to a quantity, which is fast to calculate (usually the same quantity as above is used here). Only the n (usually between 10 and 20) best candidates in each decay channel are further considered, the others are discarded. This best candidate selection ensures that each decay channel and each event receives roughly the same amount of computing time.

Next, the computationally expensive parts of the reconstruction are performed on each candidate: the matching of the reconstructed candidates to the generated particles (in case of simulated events), the vertex fitting, and the multivariate classification.

After the multivariate classifiers have estimated the signal probability of each candidate, the candidates of different decay channels can be compared. Here the `FEI` uses tighter **post-cuts** to aggressively remove incorrectly reconstructed candidates using all the available information. The main purpose of these cuts is to restrict the number of candidates per particle to a manageable number.

At first, there is a loose fixed cut on the signal probability, to remove unreasonable candidates. Second, the remaining candidates are ranked according to their signal probability. Only the m (usually between 10 and 20) best candidates of the particle (i.e., over all decay channels) are further considered, the others are discarded. This best candidate selection ensures that the amount of candidates produced in the next stage is reasonably low and can be handled by the computing system.

Performance

Applying the `FEI` to $\mathcal{O}(1$ billion) events is a CPU-intensive task. An optimized runtime and a small memory footprint are key for a practical application and to save computing resources. The `FEI` spends most CPU time on vertex fitting (38%), particle combination (27%), and classifier inference (15%). All three tasks have been carefully optimized.

The `FEI` uses only a fast and simple unconstrained vertex fit during the reconstruction, and feeds the calculated information into its multivariate classifiers. The user can refit the whole decay chain of the final B candidates, including mass and/or interaction point profile constraints if desired. A dedicated fitter (called `FastFit`) based on a Kalman Filter [6] was implemented for the `FEI`, which requires drastically less computing time than the default implementation used by Belle II and yields very similar results. Due to this fitter, an overall speedup of the `FEI` of 2.74 was observed. The `FastFit` code is licensed under GPLv3 and available on GitHub [7].

As explained in “[Combinatorics](#)” section, the number of candidates which have to be processed scales as the factorial of the multiplicity of the channel. In previous approaches, the runtime and the maximum memory consumption was dominated by a few high-multiplicity events and tight cuts had to be applied to high-multiplicity channels. By contrast, the `FEI` addresses the combinatorics problem by performing best candidate selections during the reconstruction of the decay chain instead of fixed cuts. As a consequence, for each event and each decay channel, the `FEI` processes the same number of candidates in vertex fitting and classifier inference, i.e., consumes similar amounts of CPU time. Moreover, the maximum memory consumption is limited due to the fixed number of best candidates per event, which is a key requirement for using the computing infrastructure.

Finally, the `FEI` uses `FastBDT` [8], a gradient-boosted decision tree (BDT) implementation, as its default multivariate classification algorithm. The algorithm was originally designed for the `FEI` to speed up the training and application phase. Compared to other popular BDT implementations such as those provided by `TMVA` [9], `SKLearn` [10] and `XGBoost` [11], it originally improved the execution time by more than one order of magnitude, both in training

and application. In addition, an improved classification quality was observed. Most of the time when using `FastBDT` is spent during the extraction of the necessary features, therefore no further significant speedups can be achieved by employing a different method.

Automatic Reporting

The FEI includes an automatic reporting system called Full Event Interpretation Report (FEIR).

The FEIR contains efficiencies and purities for all particles and decay channels at different points during the reconstruction. Individual reports containing control plots for each multivariate classifier and input variables are also automatically created. Control plots include receiver operating characteristics (ROC) curves, which show the tag-side efficiency against purity. Additionally, for each classifier, the purity is plotted as a function of classifier output, to check for a linear relationship as this confirms the classifier output can be treated as a probability. This built-in monitoring capability upgrades the FEI from a black-box to a white-box algorithm, which the user can understand and inspect on all levels of reconstruction.

Previous Work

Previous experiments have already developed and successfully employed tagging algorithms. To compare the algorithms, the maximal achievable tag-side efficiency is of particular interest, because it is directly related to the signal selection efficiency of the measurement. On the other hand, the achievable tag-side purity is only of limited use, because the achievable final purity of the final selection used for the measurement is dominated by the completeness constraint. Hence, most of the incorrect tags can be easily discarded and the final purity depends strongly on the considered signal decay channel. Moreover, signal-side independent ROC curves are not available for most of the previously employed algorithms. The area under the ROC curve allows one to compare the performance of the tagging algorithms.

The BaBar experiment [12] used the `Semi-Exclusive B reconstruction (SER)` algorithm for hadronic tagging [3, Section 7.4.1.1]. The algorithm used exclusive D and D^* mesons candidates as a seed, and combined those with up to five charmless hadrons to form a B_{tag} without assuming an exclusive B decay mode. The tag-side efficiency and tag-side purity of each B decay chain was extracted by fitting the beam-constrained mass [3, Section 7.1.1.2] spectrum of the constructed B_{tag} meson candidates. The beam-constrained mass is defined as $M_{\text{bc}} = \sqrt{E_{\text{beam}}^2/c^4 - p_B^2/c^2}$ where p_B denotes the three momentum of the reconstructed

B meson candidate and E_{beam} denotes half of the centre-of-mass energy of the colliding electron–positron pair. The maximum hadronic tag-side efficiency achieved by this algorithm was 0.2% for $B^0\bar{B}^0$ and 0.4% for B^+B^- , with a tag-side purity around 30%. The tag-side purity could be further increased by rejecting B meson candidates from low-purity decay chains. The semileptonic tag was usually constructed by combining an exclusive D or D^* meson with a lepton. The maximum semileptonic tag-side efficiency was typically 0.3% for $B^0\bar{B}^0$ and 0.6% for B^+B^- with an unknown tag-side purity.

The Belle experiment [13] used the so-called `Full Reconstruction (FR)` algorithm [14] for hadronic tagging [3, Section 7.4.1.2]. The FR introduced an hierarchical approach, which is still used by its successor and is presented in this article (see “`Methods`” section). The tag-side efficiency and tag-side purity was extracted by fitting the beam-constrained mass spectrum of the constructed B_{tag} meson candidates. The maximum hadronic tag-side efficiency achieved by this algorithm was 0.18% for $B^0\bar{B}^0$ and 0.28% for B^+B^- , with a tag-side purity around 10%. Multivariate classifiers [15] were used to estimate the signal probability of each candidate. The tag-side purity could be further increased by requiring a minimal signal probability. Variants of the FR were used for semileptonic tagging (see [16, 17]). The maximum semileptonic tag-side efficiency was 0.31% for $B^0\bar{B}^0$ and 0.34% for B^+B^- , with a typical tag-side purity of 5%.

Compared to the previously employed algorithms, the FEI provides a greater tagging and tag-side efficiency, with a equal or better tag-side purity. The improvements with respect to the FR can be attributed equally to the additional

Table 1 Summary of the maximum tag-side efficiency of the Full Event Interpretation and for the previously used exclusive tagging algorithms

	B^\pm (%)	B^0 (%)
Hadronic		
FEI with FR channels	0.53	0.33
FEI	0.76	0.46
FR	0.28	0.18
SER	0.4	0.2
Semileptonic		
FEI	1.80	2.04
FR	0.31	0.34
SER	0.3	0.6

For the FEI simulated data from the last official Monte Carlo campaign of the Belle experiment were used. The maximum tag-side efficiency on recorded data is lower (see “`Hadronic tag`” section). The numbers for the older algorithms (see “`Previous work`” section), are not directly comparable due to different selection criteria, like best candidate selections and selections to suppress non- $\Upsilon(4S)$ events

decay channels and the new candidate selection criteria. The reported maximum tag-side efficiencies for the previously used exclusive tagging algorithms are summarized in “Results” section, Table 1. The stated efficiencies are not directly comparable due to different selection criteria, e.g., a threshold on the beam-constrained mass or the deviation of the nominal energy from the reconstructed energy $\Delta E = E_{\text{beam}} - E_B$ with E_B denoting the energy of the B candidate, best candidate selections, or cuts on the event shape used to suppress background from non- $\Upsilon(4S)$ events.

Results

The FEI algorithm was developed for the Belle II experiment. To quantify the improvements with respect to the previously used FR algorithm, the FEI is applied to data recorded by the Belle experiment. Simulated events and recorded data from the Belle experiment are converted into the new Belle II data format [4, Chapter 2]. This conversion tool was used to validate the entire Belle II analysis software and will be described in a separate publication [18]. The remainder of this article focuses on the results obtained for the hadronic tag on data recorded by the Belle experiment. The results for the semileptonic tag and for Belle II are based on simulated events and are only summarized briefly. A detailed validation of the entire algorithm can be found in Keck [4, Chapter 4].

Hadronic Tag

The performance of the hadronic tag provided by the FEI using simulated and recorded Belle events is studied and compared to the previously used FR algorithm.

At first, the considered decay channels of the FEI are restricted to the set of hadronic decay channels used by the FR. The performance of the FEI to the FR are compared using the same hardware and the same simulated charged (neutral) $B\bar{B}$ Belle events. The FEI required 33% less computing time and achieved a maximum tag-side efficiency of 0.53% (0.33%) on simulated events, which is significantly higher than the previously reported tag-side efficiencies (see “Previous work” section). The increase in the maximum tag-side efficiency is due to the improved candidate selection criteria, in particular the best candidate selections.

Second, all decay channels of the FEI are used, including the 38 additional hadronic decay channels. The performance of the FEI to the FR using the same hardware and the same simulated charged (neutral) Belle events are then compared. The FEI required 48% more computing time and achieved a maximum tag-side efficiency of 0.76% (0.46%) on simulated events. The further increase in the maximum tag-side efficiency is due to the additional decay channels.

As mentioned before, the maximum tag-side efficiency is an important performance indicator for exclusive measurements, which can employ the completeness constraint to achieve a high final purity. The achieved maximum tag-side efficiencies are summarized in Table 1.

To validate the results for the hadronic tag obtained from the simulation study, we conducted exclusive measurements of ten different semileptonic B decay channels using the full $\Upsilon(4S)$ dataset recorded by Belle. The branching fractions of the considered semileptonic decay channels are well-known from independent untagged measurements. The branching fraction of those well-known decay channels is measured using the hadronic tag, taking into account all known disagreements between simulation and data, e.g., in the particle identification performance and the track reconstruction efficiency. We assume that the remaining disagreement between simulation and data is caused by the tag-side. Therefore, the ratio ϵ of the measured and the expected branching fraction is proportional to the ratio of the tag-side efficiency on recorded data and simulated events. Our assumption is supported by the compatibility of the extracted ratios within their uncertainties. Figure 3 summarizes the results for the ten decay channels. The ratios averaged over all control channels for the charged and neutral B_{tag} mesons are

$$\begin{aligned}\epsilon_{\text{charged}} &= 0.74_{-0.013}^{+0.014} \pm 0.050 \\ \epsilon_{\text{neutral}} &= 0.86_{-0.050}^{+0.045} \pm 0.054,\end{aligned}$$

where the first uncertainty is statistical and the second systematic. The systematic uncertainties arises from the signal-side, e.g., through uncertainties on the particle identification performance or the track reconstruction efficiency.

A detailed description of the control measurements, including results for each tag and control channel, can be found in Schwab [19]. A similar study was conducted in the past for the FR by Sibidanov et al. [20], yielding a similar overall ratio of $\epsilon_{\text{comb.}} = 0.75 \pm 0.03$. The rather large discrepancy between simulated events and recorded data is caused by the uncertainty on the branching fractions and decay models of the simulated B decay channels used for the tag-side and the large number of multivariate classifiers involved in the process.

The uncertainty on the tag-side efficiency of the FEI is one of the most important systematic uncertainties in the measurement of branching fractions of rare decays. The tag-side efficiency can be corrected using the extracted ratios. It is possible to apply this correction as a function of the tag-side decay channel and signal probability. A measurement which uses the ratios to correct the tag-side efficiency is performed relative to the considered calibration decay channels. The systematic uncertainty of the correction is given by the uncertainty of the ratios.

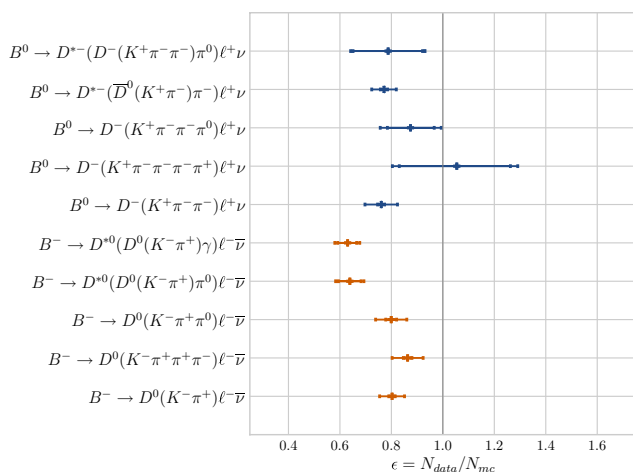


Fig. 3 The ratios calculated by measuring ten semileptonic decay channels on converted Belle data using the hadronic tag. The procedure is described in Schwab [19]

To compare the hadronic tag provided by the FEI and the FR in a well-defined manner, which is independent of the signal-side, both algorithms are applied to the same set of ten million events. These events are randomly sampled from the full $\Upsilon(4S)$ dataset of 772 million events recorded by the Belle experiment. After the tag-side reconstruction, only B meson candidates are kept, which fulfill cuts on the beam-constrained mass of $M_{bc} > 5.24$ GeV and on the deviation of the reconstructed energy from the nominal energy of -0.15 GeV $< \Delta E < 0.1$ GeV calculated on the candidate. In addition, a best candidate selection is performed, taking the B meson candidate with the highest signal probability in each event.

The same cuts on the beam-constrained mass $M_{bc} > 5.24$ GeV and the deviation of the reconstructed energy from the nominal energy -0.15 GeV $< \Delta E < 0.1$ GeV were applied and only the best (i.e., the highest signal probability) B meson candidate in each event was used.

From this dataset, we determined the tag-side efficiency and tag-side purity for different cuts on the signal probability. We followed the procedure established in previous publications [3, Chapter 7.1]. For different cuts on the signal probability, extended unbinned maximum likelihood fits of the beam-constrained mass spectrum are performed. The signal peak consisting of correct B_{tag} mesons is modeled with a Crystal Ball function [21], whereas the background is described using an ARGUS function [22]. The Gaussian mean of the Crystal Ball function was fixed to the B meson mass and its power law exponent was fixed to $m = 4$ based on the expected shape obtained from Monte Carlo simulations. The location and the width of the ARGUS were fixed using the known kinematic end point of the spectrum. All other parameters: the normalization of both functions, the width of the Crystal Ball, and the remaining

shape parameters of both functions were adjusted by the fit. The tag-side efficiency and tag-side purity are determined in a window of 5.27 GeV $< M_{bc} < 5.29$ GeV using the fitted yields of the signal and background component.

In addition, we checked for a potential peaking combinatorial background component, which would bias the results. This test was done using ten million events recorded 60 MeV below the $\Upsilon(4S)$ resonance. This dataset does not contain B mesons, hence no signal is expected. The fitted signal yields were compatible with zero.

The resulting ROC curves are shown in Figs. 4 and 5 for charged and neutral B_{tag} mesons, respectively. The FEI exhibits a larger overall tag-side efficiency compared to the FR. We observe a slightly better performance for the FR than reported in Feindt et al. [14]. Both algorithms perform equally well when requiring a high tag-side purity. We suspect this is because there are only a finite number of cleanly identifiable B_{tag} meson candidates and both algorithms identify them with similar performance. The results for tag-side purities above 70% cannot be extracted reliably and depend strongly on the chosen signal or background fit model. For practical applications, the low tag-side purity regions is of particular interest for exclusive measurements. The beam-constrained mass distributions corresponding to the low-purity region with about 15% tag-side purity and the high-purity region with approximately 80% tag-side purity are shown in Figs. 6 and 7, respectively, for the charged B_{tag} .

The maximum tag-side efficiency on recorded data is not determinable by this method, as the fits are restricted to the best B_{tag} candidates. However, a significant contribution to the improvement of the FEI compared to the FR is the increased number of provided candidates per event. A physics measurement will benefit from these additional tag-side candidates by first combining them with potential signal-side candidates, applying the completeness constraint (i.e.,

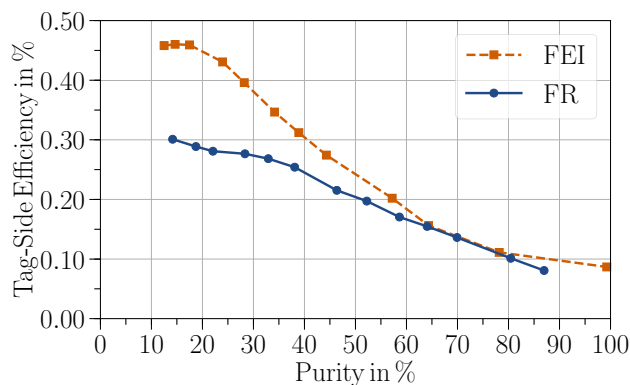


Fig. 4 Receiver operating characteristic of charged B_{tag} mesons extracted from a fit of the beam-constrained mass on converted Belle data. The FEI outperforms the FR algorithms performance at low and high purity

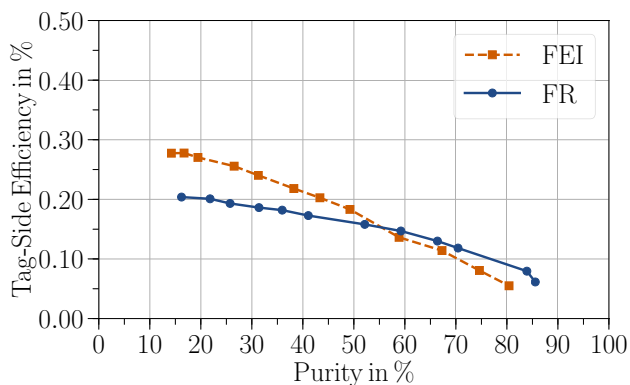


Fig. 5 Receiver operating characteristic of neutral B_{tag} mesons extracted from a fit of the beam-constrained mass on converted Belle data. The FEI outperforms the FR algorithms performance at low and intermediate purity. At high purity, the tag-side efficiency cannot be extracted reliably

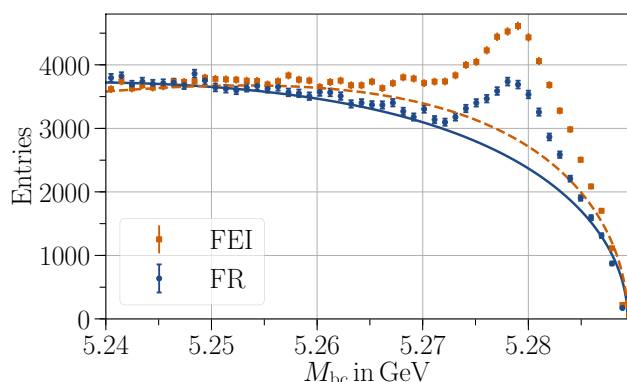


Fig. 6 Beam-constrained mass distribution of charged B_{tag} mesons in the low tag-side purity region on converted Belle data

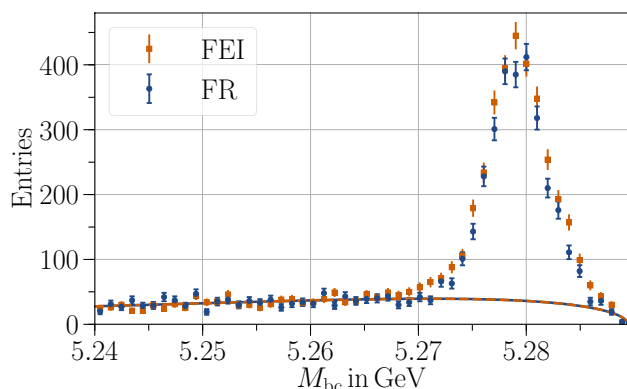


Fig. 7 Beam-constrained mass distribution of charged B_{tag} mesons in the high tag-side purity region on converted Belle data

requiring no additional tracks in the event), and performing the best B_{tag} candidate selection as the final step of the selection procedure. This procedure was successfully used by several measurements to validate the expected improvements on recorded data: [4, 19, 23].

Semileptonic Tag

The performance of the semileptonic tag provided by the FEI is studied using simulated Belle events. The maximum tag-side efficiencies are summarized in Table 1. Receiver operating characteristics extracted from simulated events can be found in Keck [4]. The results obtained from simulated events, and the fact that the hadronic and semileptonic tag only share five out of six reconstruction stages, indicate a significant increase in the maximum tag-side efficiency. The semileptonic tag was successfully used by Keck [4] to determine the branching fraction of $B \rightarrow \tau \nu_\tau$ on the full $\Upsilon(4S)$ dataset recorded by the Belle experiment, with a smaller relative statistical uncertainty than obtained previously. However, no studies with well-known calibration channels as described in Kronenbitter [24] and no signal-side independent determination of the ROCs as described in Kirchgessner [16], are available yet.

Outlook for Belle II

As the Belle II reconstruction software is still being optimized and no large recorded experimental data set was available at the time of writing, hence the final tag-side efficiency cannot be determined reliably for Belle II at this point. Preliminary results can be found in [4] which indicate a worse overall performance. This is likely due to the increased beam background caused by the higher luminosity of the collider, which does lead to additional tracks and neutral energy depositions. This additional detector activity is not yet fully rejected by the Belle II reconstruction algorithms [4] and future improvements are likely possible.

Discussion

The multivariate classifiers used by the FEI are trained on Monte Carlo simulated events. Depending on the training procedure and the type of events provided to the training, the multivariate classifiers of the FEI are optimized for different objectives.

In this article, we presented a so-called **generic** adaption of the FEI. The generic refers to that the FEI was trained independently of any specific signal-side using 180 million simulated $\Upsilon(4S)$ events. This setup optimizes the tag-side efficiency of a “generic” $\Upsilon(4S)$.

Other versions of the FEI exist which optimize the tag-side efficiency of specific signal events like $B \rightarrow \tau \nu_\tau$. The so-called **specific** FEI is trained on the remaining tracks and clusters after a potential signal B meson was already identified. The training uses simulated $\Upsilon(4S)$ events and simulated signal events. As a consequence, the classifiers can be specifically trained to identify correctly reconstructed B_{tag} mesons for signal events and can focus on reducing non-trivial background which is not discarded by the completeness constraint. The specific FEI was first introduced as a proof of concept by Keck [25] and used in Metzner [23].

Roughly half of the improvements with respect to the previous algorithm can be attributed to the additionally considered decay channels. Future extensions are currently investigated which use semileptonic D meson decays, baryonic decays and decays including K_L^0 particles.

It should also be noted that the FEI algorithm can be applied, with little modification, to the $\Upsilon(4S)$ resonance. This resonance decays into a pair of $B^{(*)}B^{(*)}$ and $B_s^{0(*)}B_s^{0(*)}$ mesons. The powerful completeness constraint can still be applied in this situation.

Conclusion

The Full Event Interpretation is a new exclusive tagging algorithm developed for the Belle II experiment that will be used to measure a wide range of decays with a minimum of detectable information. The algorithm exploits the unique setup of B factories and significantly improves the tag-side efficiency compared to its predecessor algorithms.

The tag-side efficiency for hadronically tagged B mesons was validated and calibrated using Belle data. Furthermore, the hadronic and the semileptonic tag provided by FEI have already been used in several validation measurements [4, 19, 26] using the full $\Upsilon(4S)$ dataset recorded by the Belle experiment. Similar studies and measurements for Belle II are anticipated as soon as the experiment records a sufficient amount of collision events.

There are several ways that the FEI algorithm could be further refined and applied to so far unexplored applications. These will provide an exciting and fruitful area of future research.

Acknowledgements We thank the KEKB accelerator group, the Belle collaboration, and the Belle II collaboration for the provided data and infrastructure. This research was supported by: the Federal Ministry of Education and Research of Germany (BMBF), the German Research Foundation (DFG), the Doctoral School “Karlsruhe School of Elementary and Astroparticle Physics: Science and Technology” funded by the German Research Foundation (DFG), and the DFG-funded Research Training Group “GRK 1694: Elementary Particle Physics at Highest

Energy and highest Precision”. F.B. and W.S. are supported by DFG Emmy-Noether Grant No. BE 6075/1-1

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Abe T et al (2010) Belle II technical design report. Technical report, KEK-REPORT-2010-1. [arXiv:1011.0352](https://arxiv.org/abs/1011.0352)
2. Moll Andreas (2011) The software framework of the Belle II experiment. J Phys Conf Ser. <https://doi.org/10.1088/1742-6596/331/3/032024>
3. Bevan AJ (2014) The physics of the B factories. Eur Phys J C 74:3026. <https://doi.org/10.1140/epjc/s10052-014-3026-9>
4. Keck T (2017) Machine learning algorithms for the Belle II experiment and their validation on Belle data. PhD thesis, KIT. <https://doi.org/10.5445/IR/1000078149>
5. Dean J, Ghemawat S (2004) MapReduce: simplified data processing on large clusters. In: OSDI'04: sixth symposium on operating system design and implementation, San Francisco, CA, pp 137–150
6. Fruhwirth R (1987) Application of Kalman filtering to track and vertex fitting. Instrum Methods Nucl. [https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4)
7. <https://github.com/thomaskeck/FastFit>. Accessed 02 Oct 2017
8. Keck T (2017) FastBDT: a speed-optimized multivariate classification algorithm for the Belle II experiment. Comput Softw Big Sci 1(1):9. <https://doi.org/10.1007/s41781-017-0002-8>
9. Hocker A et al (2007) TMVA—toolkit for multivariate data analysis. PoS ACAT:040
10. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
11. Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. CoRR. [arXiv:abs/1603.02754](https://arxiv.org/abs/1603.02754)
12. Aubert B et al (2013) The BABAR detector: upgrades, operation and performance. Nucl Instrum Methods A. <https://doi.org/10.1016/j.nima.2013.05.107>
13. Abashian A et al (2002) The Belle detector. Nucl Instrum Methods A479:117–232. [https://doi.org/10.1016/S0168-9002\(01\)02013-7](https://doi.org/10.1016/S0168-9002(01)02013-7)
14. Feindt M, Keller F, Kreps M, Kuhr T, Neubauer S, Zander D, Zupanc A (2011) A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories. Nucl Instrum Methods A654:432–440. <https://doi.org/10.1016/j.nima.2011.06.008>
15. Feindt M, Kerzel U (2006) The NeuroBayes neural network package. Nucl Instrum Methods 559(1):190–194. <https://doi.org/10.1016/j.nima.2005.11.166>
16. Kirchgessner K (2012) Semileptonic tag side reconstruction. Master's thesis, KIT. <http://ekp-invenio.physik.uni-karlsruhe.de/record/48181>
17. Kronenbitter B (2015) Measurement of the branching fraction of $B^+ \rightarrow \tau^+ \nu_\tau$ decays with the semileptonic tagging method. Phys Rev D 92(5):051102. <https://doi.org/10.1103/PhysRevD.92.051102>

18. Gelb M et al (2018) B2BII: data conversion from Belle to Belle II. *Comput Softw Big Sci* 2(1):9. <https://doi.org/10.1007/s41781-018-0016-x>
19. Schwab J (2017) Calibration of the full event interpretation for the Belle and the Belle II experiment. Master's thesis, KIT. <https://ekp-invenio.physik.uni-karlsruhe.de/record/48931>
20. Sibidanov A (2013) Study of exclusive $B \rightarrow X_u \ell \nu$ decays and extraction of $|V_{ub}|$ using full reconstruction tagging at the Belle experiment. *Phys Rev D* 88(3):032005. <https://doi.org/10.1103/PhysRevD.88.032005>
21. Skwarnicki T (1986) PhD thesis, Institute for Nuclear Physics, Krakow
22. Albrecht H (1990) Search for Hadronic $b \rightarrow u$ decays. *Phys Lett B* 241:278–282. [https://doi.org/10.1016/0370-2693\(90\)91293-K](https://doi.org/10.1016/0370-2693(90)91293-K)
23. Metzner F (2016) Analysis of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decays with the Belle II analysis software framework. Master's thesis, KIT. <https://ekp-invenio.physik.uni-karlsruhe.de/record/48845>
24. Kronenbitter B (2014) Measurement of the branching fraction of $B^+ \rightarrow \tau^+ \nu_\tau$ decays at the Belle experiment. PhD thesis, KIT. <https://ekp-invenio.physik.uni-karlsruhe.de/record/48604>
25. Keck T (2014) The full event interpretation for Belle II. Master's thesis, KIT. <https://ekp-invenio.physik.uni-karlsruhe.de/record/48602>
26. Gelb M et al (2018) Search for the rare decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ with improved hadronic tagging. *Phys Rev D* 98(11):112016. <https://doi.org/10.1103/PhysRevD.98.112016>