BELLE2-REPORT-201X-YYY
August 26, 2016

Belle II Analysis Software

F. Abudinen, T. Keck, L. Li Gioi

Abstract

An overview of the physics analysis software is presented.

1 Physics Analysis Software

Section author(s): L. Li Gioi, P. Goldenzweig, A. Zupanc,

1.1 Introduction

This is the body for section on physics analysis algorithms [1].

1.2 Particle reconstruction

Mass Resolutions.

1.3 Vertex reconstruction

A vertex reconstruction algorithm is a procedure by which the parameters of a decay vertex or interaction vertex are determined from the reconstructed parameters of the outgoing particles. It deals both with finding (pattern recognition) and with fitting (statistical estimation) of the interaction vertices. It usually extracts the vertex position, momentum and invariant mass of the decaying particle. However, also the decay length of an unstable particle inside a decay chain or the decay time difference Δt of the two B mesons from an $\Upsilon(4S)$ decay, are computed with a vertex fit.

1.3.1 Vertex finding algorithms

The Belle II experiment has deployed two implementations of vertex fit: KFit [2], developed for the Belle experiment and RAVE [3], the porting to a standalone package of the CMS vertexing libraries. We use both KFit and RAVE for kinematic fits and RAVE for geometric fits.

Kinematic fits

Kinematic fitting uses the known properties of a specific decay chain for improving the measurements of the process. Lagrangian multipliers are used in order to impose the kinematic constraints to the fit. Considering the measurements $\mathbf{q} = (q_1, ..., q_n)$, with the covariance matrix V and the kinematic constraints $\mathbf{h}(\mathbf{q})$, the function to be minimized in terms of the most suitable vertex is:

$$\chi^2 = (\mathbf{q}-\bar{\mathbf{q}})^T V^{-1} (\mathbf{q}-\bar{\mathbf{q}}) + 2\,\lambda^T (\mathbf{D}(\delta\mathbf{y}) + \mathbf{d}) \quad (1.1)$$

with $h(\bar{\mathbf{q}}) = 0$ and $\mathbf{D} = \partial\mathbf{h}/\partial\mathbf{y}$. Here $\bar{\mathbf{q}}$ represents the improved measurements.

Adaptive Vertex Fit

This set of algorithms [3, 4], included in the RAVE libraries, introduces the concept of soft assignment; a track is associated to a specific vertex with an assignment probability, or weight w_i [4]:

$$w_i(\chi_i^2) = \frac{e^{-\chi_i^2/2T}}{e^{-\chi_i^2/2T} + e^{-\sigma_{cut}^2/2T}} \qquad (1.2)$$

where χ_i^2 is the square of the standardized residual, σ_{cut} is defined as the standardized residual for which $w_i = 0.5$ and the temperature parameter T defines the softness of the weight function.

The fitter is then implemented as an iterated, re-weighted Kalman filter [5]: in every iteration step new track weights are computed and subsequently for the determination of a vertex candidate. This weight can be interpreted as a track-to-vertex assignment probability. Instead of minimizing the least sum of squares, as is expected from a Kalman fitting method, the algorithm minimizes the weighted least sum of squares. In order to avoid falling prematurely into local minima, a deterministic annealing schedule is introduced; in each iteration step the temperature parameter is lowered [4]:

$$T_i = 1 + r \cdot (T_{i-1} - 1) \qquad (1.3)$$

here T_i is the temperature parameter T at iteration i and r denotes the annealing ratio. For convergence, $0 < r < 1$ is needed.

TreeFitter

1.3.2 Primary vertex

Primary vertices are usually identified using kinematic vertex fits. We use, as benchmark for testing the Belle II vertexing performance, the decay vertex of J/ψ coming from the $B^0 \to J/\psi K_S^0$ decay mode. Figure 1.1 shows the fit residuals of the z component of the J/ψ vertex fit. We obtain a resolution of 22 μm and a bias of 2 μm, the latter due to the bias in the helix parameters of the tracks fit.

The same vertex fit performed using Belle Monte Carlo returns a resolution of 43 μm and a bias of 0.2 μm. The value of the resolution, two times of the one expected by Belle II, is compatible with the expected improvement in the impact parameter resolution due the Belle II Pixel Vertex Detector.

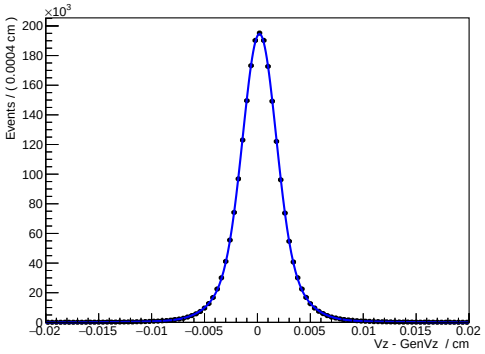

Figure 1.1: $J/\psi \to \mu^+\mu^-$ vertex fit residuals (bias = 2 μm, resolution = 22 μm). The fit is performed using the sum of three Gaussian functions. The values of the shift and resolution are defined as the weighted averages of the mean values and the standard deviations of the three functions.

1.3.3 B-tag vertex (Δt)

To be sensitive to time-dependent CP violating effects, the vertex resolution must be sufficient to resolve the oscillations of the neutral B mesons in the decay time distribution, performing the measurement of the distance of their decay vertices. The most important contribution to the Δt resolution comes from the tagging B vertex fit. For this fit, in fact, a selection of the tracks originating from the B vertex is required. In the decay tree of a B meson we can divide the tracks in three groups: tracks originating from the B decay, including the ones coming from decay vertices indistinguishable from the B (e.g. μ^+ and μ^- in $B^0 \to [J/\psi \to \mu^+\mu^-]K_S^0$); tracks originating from D mesons and tracks originating from K_S^0 decays. We perform the tag side vertex fit using the RAVE Adaptive Vertex Fit algorithm and giving as input all the tracks with at least one hit on the PXD not used for the fully reconstructed B apart from the ones that originate from K_S^0 decays. In case of a

non converging fit, also tracks with no PXD hits are used.

In order to reduce as much as possible the weight of the tracks originating from D mesons, we constrain the fit to a region defined by an ellipsoid around the boost direction (Figure 1.2), where the B has an higher probability to decay than a possible D meson. In case of non converging fit, a new constraint is defined as a cilinder around the boost direction and the fit is performed again.

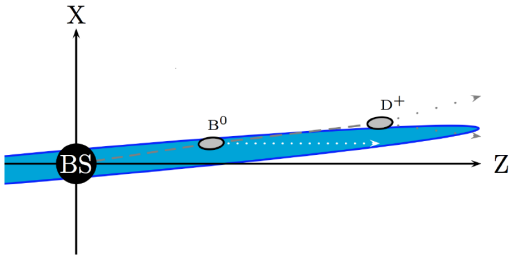

Figure 1.2: Schematic representation of the B-tag vertex fit. The B meson has an higher probability than the D to decay inside the ellipsoid parallel to the boost direction.

We obtain, for the tag side vertex fit of correctly reconstructed B mesons, a bias of $7\,\mu$m and a resolution of $52\,\mu$m, independent of the B decay mode. Figure 1.3 shows the residuals of the B-tag vertex fit of fully reconstructed $B^0 \to [J/\psi \to \mu^+\mu^-][K^0_S \to \pi^+\pi^-]$.

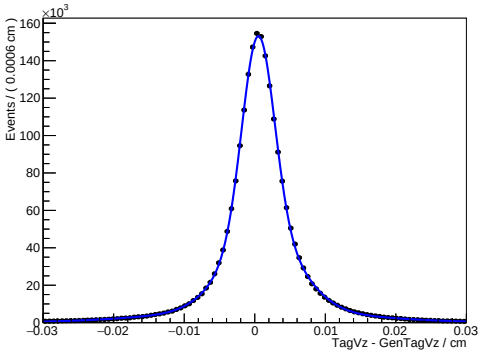

Figure 1.3: Tag side vertex fit residuals of the fully reconstructed $B^0 \to [J/\psi \to \mu^+\mu^-][K^0_S \to \pi^+\pi^-]$ (bias = $7\,\mu$m, resolution = $52\,\mu$m).

The sensitivity to the time-dependent CP violating parameters of equation add ref strongly depends on the detector resolution of the Δt distribution, the last depending (equation add ref) on the resolution of the distance of the decay vertices of the two exclusively produced B mesons. The reduced boost of SuperKEKB produces an average distance between the two B mesons of about $130\,\mu$m, 35% smaller than the $200\,\mu$m of KEKB. This makes it more difficult to resolve the decay vertices of the two B mesons and it is one of the main motivations for the development of the Belle II Pixel Vertex Detector. The new hardware and improvement in the tracks reconstruction and in the vertexing algorithms provide an improvement, compared to Belle [6] of the vertex resolution of both B mesons. This translates to Δt with a resolution of 0.71 ps and a bias of -0.05 ps, which provide a superior separation capability compared to Belle [6]. Figure 1.4 shows the Δt residuals of fully reconstructed $B^0 \to [J/\psi \to \mu^+\mu^-][K^0_S \to \pi^+\pi^-]$.

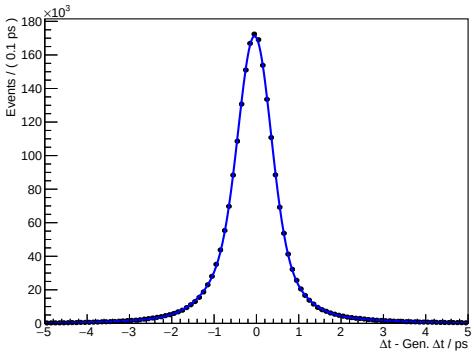

Figure 1.4: Δt residuals of fully reconstructed $B^0 \to [J/\psi \to \mu^+\mu^-][K^0_S \to \pi^+\pi^-]$ (bias = -0.05 ps, resolution = 0.71 ps).

1.3.4 Fit of the Decay Chain

1.4 Continuum Suppression

1.4.1 Event topology

1.4.2 Performance

1.5 Flavor Tagger

The task of the flavor tagger is to determine the flavor of the accompanying B^0 meson at the time of its decay. Flavor tagging is required for measurements of the time-dependent CP-Violation and B meson mixing where usually one of the neutral B mesons is fully reconstructed. Disregarding initial state radiation, B mesons pairs are produced in isolation at the $\Upsilon(4S)$ resonance. Therefore, the tracks together with the ECL and KLM clusters which remain from the fully reconstructed side can be assumed to a good approximation to belong to the decay of B_{tag}.
The B Mesons are relatively exhibit a vast plethora of possible decay channels. A large fraction of decay channels provide unambiguous flavor signatures through a flavor-specific final state. Nevertheless, the wide range of possible decay channels makes it unfeasible to fully reconstruct a sufficiently large number of flavor specific B_{tag} decays. In lieu of a full reconstruction, the flavor tagger applies inclusive techniques to maximally exploit the information provided by different flavor specific signatures inherent to flavor specific decays.

1.5.1 Definitions

In general, the efficiency ε of a flavor tagging algorithm is defined as the fraction of tagged events over the total number of events, i.e. the fraction of events to which a flavor tag can be assigned. The fraction of wrong identifications over the number of tagged events is denoted by w. Thus, the number of tagged B and $\overline{B}$ events is given by

$$\begin{aligned} N_B^{\text{tag}} &= \varepsilon(1-w)N_B + \varepsilon w N_{\overline{B}^0} \\ N_{\overline{B}^0}^{\text{tag}} &= \varepsilon(1-w)N_{\overline{B}^0} + \varepsilon w N_B. \end{aligned} \tag{1.4}$$

The observed time dependent probability density add ref becomes then

$$\begin{aligned} &\mathcal{P}^{\text{obs}}(\Delta t, q, \varepsilon, w) \\ &\quad = \varepsilon\left[(1-w)\mathcal{P}(\Delta t, q) + w\mathcal{P}(\Delta t, -q)\right] \\ &\quad = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \cdot \varepsilon\Big[1 + q\cdot(1-2w)\cdot \\ &\qquad (\mathcal{A}_{\text{CP}}\cos(\Delta m \Delta t) + \mathcal{S}_{\text{CP}}\sin(\Delta m \Delta t))\Big] \end{aligned} \tag{1.5}$$

and consequently the CP-violating asymmetry add ref

$$\begin{aligned} &a_{\text{CP}}^{\text{obs}}(\Delta t, w) \\ &= \frac{\mathcal{P}^{\text{obs}}(\Delta t, q, \varepsilon, w) - \mathcal{P}^{\text{obs}}(\Delta t, -q, \varepsilon, w)}{\mathcal{P}^{\text{obs}}(\Delta t, q, \varepsilon, w) + \mathcal{P}^{\text{obs}}(\Delta t, -q, \varepsilon, w)} \\ &= (1-2w)\left[\mathcal{A}_{\text{CP}}\cos(\Delta m\Delta t) + \mathcal{S}_{\text{CP}}\sin(\Delta m \Delta t)\right] \\ &= (1-2w)\cdot a_{\text{CP}}(\Delta t) \end{aligned} \tag{1.6}$$

which are diluted by the so-called dilution factor $r = (1-2w)$. Since the amplitude of the observed CP asymmetries is proportional to the dilution $(1-2w)$, the wrong tag fraction has a direct impact on the measurement of the CP violating parameters $\mathcal{A}_{\text{CP}}$ and $\mathcal{S}_{\text{CP}}$. In order to reduce systematic uncertainties, a precised measurement of w is required.
Due to Poisson statistics the statistical uncertainty is inversely proportional to $\sqrt{\varepsilon}\cdot r$. Therefore, the effective tagging efficiency is defined as

$$\varepsilon_{\text{eff}} = \varepsilon\cdot(1-2w)^2 = \varepsilon\cdot r^2. \tag{1.7}$$

A maximization of the effective efficiency results then in a minimization of the statistical uncertainty.
Up to now it has been ignored that w and ε could be slightly different for $q = +1(-1)$ as result of a not charge-symmetric detector performance. In consequence one redefines

$$\varepsilon = \frac{\varepsilon_B + \varepsilon_{\overline{B}^0}}{2} \tag{1.8}$$

$$w = \frac{w_B + w_{\overline{B}^0}}{2} \tag{1.9}$$

and introduces the differences

$$\Delta\varepsilon = \varepsilon_B - \varepsilon_{\overline{B}^0} \quad (1.10)$$

$$\Delta w = w_B - w_{\overline{B}^0}, \quad (1.11)$$

where the subindex correspond to the true flavor, i.e. w_B is for example the fraction of B wrongly classified as $\overline{B}^0$.

1.5.2 Tagging Categories

The flavor tagger relies on flavor specific decay modes whose branching fraction is relatively high ($\gtrsim 2\%$). Since the wrong tag fraction w and the efficiency ε of flavor specific signatures are different for different decay modes, the flavor specific signatures have been grouped into individual categories. The considered decay modes are characterized by one or two flavor specific final state tracks. This tracks are denominated as targets since their charges are the flavor specific signatures which tag the flavor of B_{tag}. The current flavor tagger is based on 13 categories. An overview is presented in table 1.1.

Categories	Targets
Electron	e^-
Int. Electron	e^+
Muon	μ^-
Int. Muon	μ^+
KinLepton	l^-
Int. KinLepton	l^+
Kaon	K^-
KaonPion	K^-, π^+
SlowPion	π^+
MaximumP*	l^-, π^-
FSC	l^-, π^+
FastPion	π^-
Lambda	Λ
Total= 13	

$\overline{B}^0 \to D^{*+}\ \overline{\nu}_\ell\ \ell^-$
$\quad\hookrightarrow D^0\ \pi^+$
$\quad\quad\hookrightarrow X\ K^-$

$\overline{B}^0 \to D^+\ \pi^-$
$\quad\hookrightarrow K^0\ \nu_\ell\ \ell^+$

$\overline{B}^0 \to \Lambda_c^+\ X^-$
$\quad\hookrightarrow \Lambda\ \pi^+$
$\quad\quad\hookrightarrow p\ \pi^-$

Table 1.1: Individual categories and their targets with some characteristic examples of the considered decay modes. Int. stays here for Intermediate.

In order to extract the flavor specific signatures, the targets have to be identified among the tracks which remain from the fully reconstructed B_{sig}. For this task, discriminating variables are calculated. In some calculations also the clusters which remain from the full reconstruction are taken into account. The discriminating variables are employed not only to identify the target tracks but also to determine the categories that provide the correct flavor tag for a specific event.
In the following the different flavor signatures and the discriminating variables are discussed in detail. An overview of the different discriminants for each category is presented in table 1.2.

Leptons

Kaons

Pions

High momentum particles

Lambda baryons

Correlation between kaons and slow pions

Correlation between fast and slow particles

1.5.3 Workflow and Algorithms

1.5.4 Performance

1.6 Full Event Interpretation

1.6.1 Physics Motivation

Measurements of decays including neutrinos, in particular rare decays, suffer from missing kinematic information. The FEI recovers this information partially and infers strong constraints on the signal candidates by automatically reconstructing the rest of the event in thousands of exclusive decay channels. The Full Event Interpretation (FEI) is an essential component in a wide range of important analyses, including: the measurement of the CKM element —V_{ub}— through the semileptonic decay $b \to u\nu$; the search for a charged-Higgs effect in $B \to D\tau\nu$; and the precise measurement of the

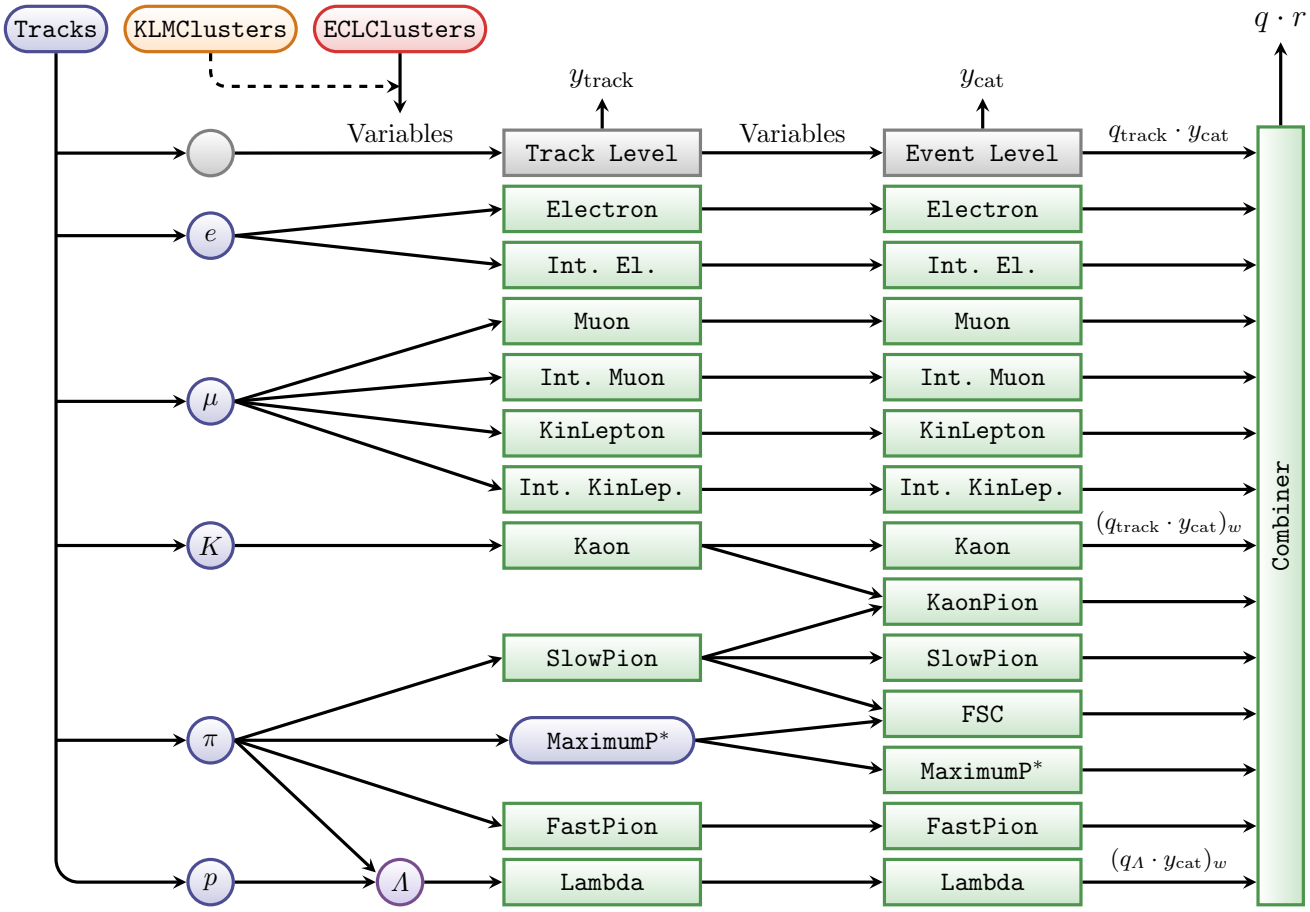

Figure 1.5: Workflow of the Flavor Tagger: The grey path above represent only the steps of the procedure. Reconstructed tracks are available for five different mass hypothesis. Each green box corresponds to a multivariate method: y_{track} and y_{cat} correspond to the output of track and event levels respectively. The variables used as input for each method are shown in table 1.2. The variables $(q \cdot y_{\text{cat}})_w$ stay here for the weighted sum over the three tracks with highest y_{cat}.

Categories	Discriminating input variables
Electron	$p^*,\ p_\mathrm{t}^*,\ p,\ p_\mathrm{t},\ \mathcal{L}_e,\ M_\mathrm{recoil},\ p_\mathrm{miss}^*,\ \cos\theta_\mathrm{miss}^*, E_{90}^W,\ \chi^2$
Intermediate Electron	$p^*,\ p_\mathrm{t}^*,\ p,\ p_\mathrm{t},\ \mathcal{L}_e,\ M_\mathrm{recoil},\ p_\mathrm{miss}^*,\ \cos\theta_\mathrm{miss}^*, E_{90}^W,\ \chi^2$
Muon	$p^*,\ p_\mathrm{t}^*,\ p,\ p_\mathrm{t},\ \mathcal{L}_\mu,\ M_\mathrm{recoil},\ p_\mathrm{miss}^*,\ \cos\theta_\mathrm{miss}^*, E_{90}^W,\ \chi^2$
Intermediate Muon	$p^*,\ p_\mathrm{t}^*,\ p,\ p_\mathrm{t},\ \mathcal{L}_\mu,\ M_\mathrm{recoil},\ p_\mathrm{miss}^*,\ \cos\theta_\mathrm{miss}^*, E_{90}^W,\ \chi^2$
KinLepton	$p^*,\ p_\mathrm{t}^*,\ p,\ p_\mathrm{t},\ \mathcal{L}_\mu,\ \mathcal{L}_e,\ M_\mathrm{recoil},\ p_\mathrm{miss}^*,\ \cos\theta_\mathrm{miss}^*, E_{90}^W,\ \chi^2$
Intermediate KinLepton	$p^*,\ p_\mathrm{t}^*,\ p,\ p_\mathrm{t},\ \mathcal{L}_\mu,\ \mathcal{L}_e,\ M_\mathrm{recoil},\ p_\mathrm{miss}^*,\ \cos\theta_\mathrm{miss}^*, E_{90}^W,\ \chi^2$
Kaon	$p^*,\ p_\mathrm{t}^*,\ p_\mathrm{t},\ \mathcal{L}_K,\ \cos\theta,\ n_{K_S^0},\ \sum p_\mathrm{t},\ \lvert\mathbf{x}\rvert,\ \chi^2$
KaonPion	$y_\mathrm{Kaon},\ y_\mathrm{SlowPion},\ \cos\theta_{K,\pi},\ q_K \cdot q_\pi,\ \mathcal{L}_K$
SlowPion	$p^*,\ p_\mathrm{t}^*,\ p,\ p_\mathrm{t},\ \mathcal{L}_\pi,\ \mathcal{L}_K,\ \mathcal{L}_e,\ \cos\theta,\ \cos\theta_\mathrm{Thrust},\ \chi^2$
MaximumP*	$p^*,\ p_\mathrm{t}^*,\ p,\ p_\mathrm{t},\ \cos\theta_\mathrm{Thrust},\ d_0$
FSC	$p_\mathrm{Slow}^*,\ p_\mathrm{Fast}^*,\ \mathcal{L}_K,\ \cos\theta_\mathrm{ThrustSlow},\ \cos\theta_\mathrm{ThrustFast},\ \cos\theta_\mathrm{SlowFast},\ q_\mathrm{Slow} \cdot q_\mathrm{Fast}$
FastPion	$p^*,\ p_\mathrm{t}^*,\ p,\ p_\mathrm{t},\ \mathcal{L}_\pi,\ \mathcal{L}_K,\ \mathcal{L}_e,\ \cos\theta,\ \cos\theta_\mathrm{Thrust},\ \chi^2$
Lambda	$p_\Lambda^*,\ p_\Lambda,\ p_\mathrm{proton}^*,\ p_\mathrm{proton},\ q_\Lambda,\ M_\Lambda,\ n_{K_S^0},\ \cos\theta_{\mathbf{x}_\Lambda,\mathbf{p}_\Lambda},\ \lvert\mathbf{x}_\Lambda\rvert,\ \sigma_\Lambda^{zz},\ \chi^2_\Lambda$

Table 1.2: Discriminating input variables for each category.

branching fraction of $B \to \tau\nu$, which is sensitive to new physics effects.

As an analysis technique unique to B factories, the Full Event Interpretation will play an important role in the measurement of rare decays. This technique reconstructs one of the B mesons and infers strong constraints for the remaining B meson in the event using the precisely known initial state of the $\Upsilon(4S)$. The actual analysis is then performed on the second B meson. The two mesons are called tag-side B tag and signal-side B_{sig} , respectively. In effect the FEI allows one to reconstruct the initial $\Upsilon(4S)$ resonance, and thereby recovering the kinematic and flavour information of B_{sig} . Furthermore, the background can be drastically reduced by discarding $\Upsilon(4S)$ candidates with remaining tracks or energy clusters in the rest of event.

Belle already employed a similar technique called Full Reconstruction (FR) with great success. As a further development the Full Event Interpretation is more inclusive, provides more automation and analysis-specific optimisations. Both techniques heavily rely on multivariate classifiers (MVC). MVCs have to be trained on a Monte Carlo (MC) simulated data sample. However, the analysis-specific signal-side selection strongly influences the background distributions on the tag-side. Yet this influence had to be neglected by the FR, because the training of the MVCs was done independently from the signal-side analysis. In contrast, the FEI can be trained for each analysis separately and can thereby take the signal-side selection into account. The analysis-specific training is possible due to the deployment of speed-optimized training algorithms, full automation and the extensive usage of parallelization on all levels. The total training duration for a typical analysis is in the order of days instead of weeks. In consequence, it is also feasible to retrain the FEI if better MC data or optimized MVCs become available.

1.6.2 Hadronic, Semileptonic and Inclusive Tagging

As previously described, the FEI automatically reconstructs one out of the two B mesons in an $\Upsilon(4S)$ decay to recover information about the remaining B meson. In fact, there is an entire class of analysis methods, so-called tagging-methods, based on this concept. In the past there were three distinct tagging-methods: hadronic, semileptonic and

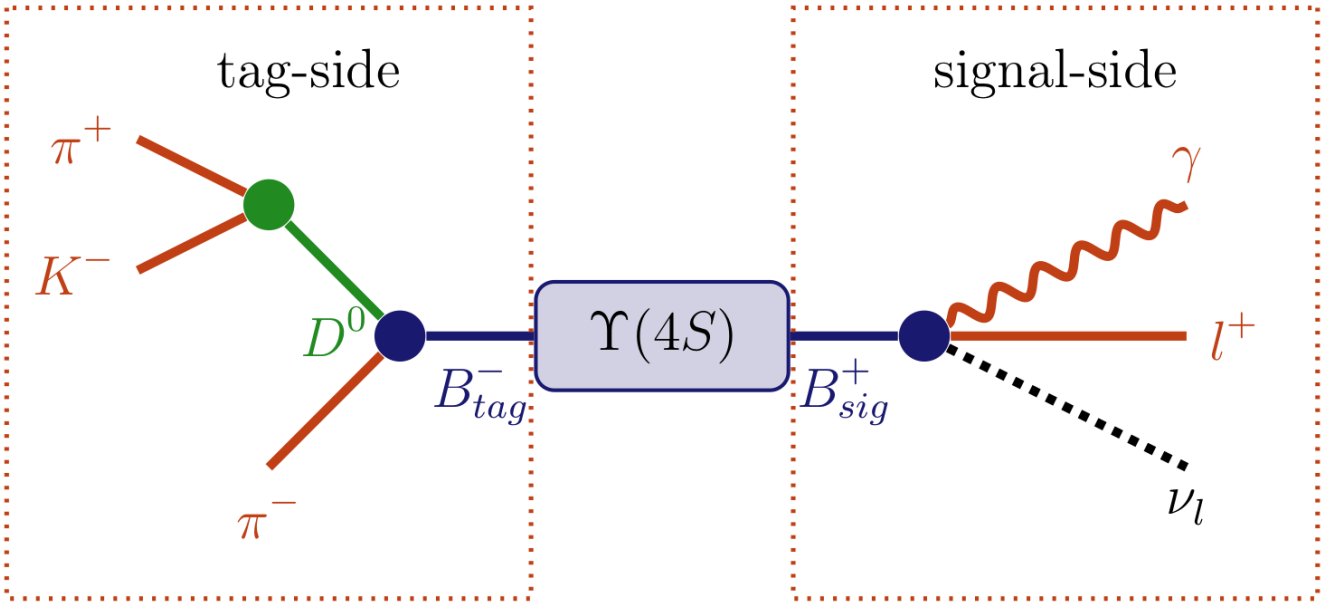

Figure 1.6: Example of a signal decay tagged by the FEI.

inclusive tagging.

- Hadronic tagging solely uses hadronic decay channels for the reconstruction. Hence, the kinematics of the reconstructed candidates are well known and the tagged sample is very pure. Then again, hadronic tagging is only possible for a tiny fraction of the dataset on the order of a few per mille.

- Semileptonic tagging uses semileptonic B decays. Due to the high branching fraction of semileptonic decays this approach usually has a higher tagging efficiency. On the other hand, the semileptonic reconstruction suffers from missing 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.

- Inclusive tagging combines the four-momenta of all particles in the rest of the event of the signal-side B candidate. The achieved tagging efficiency is usually one order of magnitude above the hadronic and semileptonic tagging. Yet the decay topology is not explicitly reconstructed and cannot be used to discard wrong candidates. In consequence, the methods suffers from a high background and the tagged sample is very impure.

The FEI combines the first two tagging-methods: hadronic and semileptonic tagging, into a single algorithm. Simultaneously it increases the tagging efficiency by reconstructing more decay channels in total. The long-term goal is to unify all three methods in the FEI.

1.6.3 Hierarchical Approach

The basic idea of the Full Event Interpretation is to reconstruct the particles and train the MVCs in a hierarchical approach. At first the final-state particle candidates are selected and corresponding classification methods are trained using the detector information. Building on this, intermediate particle candidates are reconstructed and a multivariate classifier is trained for each employed decay channel. The MVC combines all information about a candidate into a single value the signal-probability. In consequence, candidates from different decay channels can be treated equally in the following reconstruction steps.For instance, the

FEI currently reconstructs 15 decay channels of the D^0 . Afterwards, the generated D^0 candidates are used to reconstruct D^{*0} in 2 decay channels. All information about the specific D0 decay channel of the candidate is encoded in its signal-probability, which is available to the D^{*0} classifiers. In effect, the hierarchical approach reconstructs 2 * 15 = 30 exclusive decay channels and provides a signal-probability for each candidate, which makes use of all available information. Finally, the B candidates are reconstructed and the corresponding classifiers are trained.

1.6.4 Training modes

The FEI has to be trained on Monte Carlo data and is applied subsequently on real data after an analysis-specific signal-side selection. There are three different types of events one has to consider in the training and application of the FEI:

- double-generic events - $e^+e^- \to \Upsilon(4S) \to BB$ for charged and neutral B pairs, where both B mesons decay generically;
- continuum events $e^+e^- \to c\bar{c}, ss\bar{s}, d\bar{d}, u\bar{u}$;
- and signal events $e^+e^- \to \Upsilon(4S) \to BB$, where one B decays generically and the other decays in an analysis-specific signal-channel like $B^+ \to \tau^+\nu$.

The final classifier output for the B tag mesons has to identify correctly reconstructed B tag mesons in the signal events of the analysis and reject background B tag mesons from double-generic, continuum and signal events efficiently. To accomplish a high efficiency for correctly reconstructed B tag in signal events a training on pure signal Monte Carlo after the signal-side selection would be appropriate, but in this scenario background components from double-generic and continuum events would not be considered in the training and therefore could not be rejected efficiently. On the other hand, a training on double-generic and continuum Monte Carlo after signal-side selection suffers from low statistics especially for correctly reconstructed B tag mesons, because the constraint that the reconstructed candidate has to use all remaining tracks is very strict. Moreover, it is not clear if D mesons from continuum background should be considered as signal in the corresponding trainings.

The background components are factorized into background from $\Upsilon(4S)$ decays and from continuum events. It is assumed that the continuum events can be suppressed efficiently with the ContinuumSuppression module, therefore no Monte Carlo data for continuum events is used in the training of the FEI. Further studies have to be performed to test this assumption.

The FR of Belle was trained on double-generic and continuum Monte Carlo without considering the signal-side selection. In consequence, the background distributions were fundamentally different in training and application. For example, most of the CPU time in the training was used for events with more than 12 tracks, yet these events never led to a valid B tag meson in an analysis with only one track on the signal-side like $B \to \tau\nu$. Therefore the FEI employs two different training modes:

- **generic-mode**; the training is done on double-generic Monte Carlo without signal-side selection, which corresponds to the FR of Belle. Hence, the training is independent of the signal-side and is only trained once for all analyse. The method is optimized to reconstruct tag-side of generic MC. e.g. in an inclusive analysis like $B \to X_{cc}K$-
- **specific-mode**; the training is optimized for the signal-side selection and trained on double-generic and signal Monte Carlo, in order to get enough signal statistics despite the no-remaining-tracks constraint. In this mode the FEI is trained on the RestOfEvent after the signal-side selection, therefore the training depends on the signal-side and one has to train it for every analysis separately. The method is optimized to reconstruct the tag-side of signal MC. This mode can be used in searches for $B^+ \to \tau^+\nu$, $B^+ \to l\nu\gamma$, $B \to \nu\nu(\gamma)$, $B \to K^*\nu\nu$, $B \to D^*\tau\nu$, ... Another advantage is that global constraints on the beam-

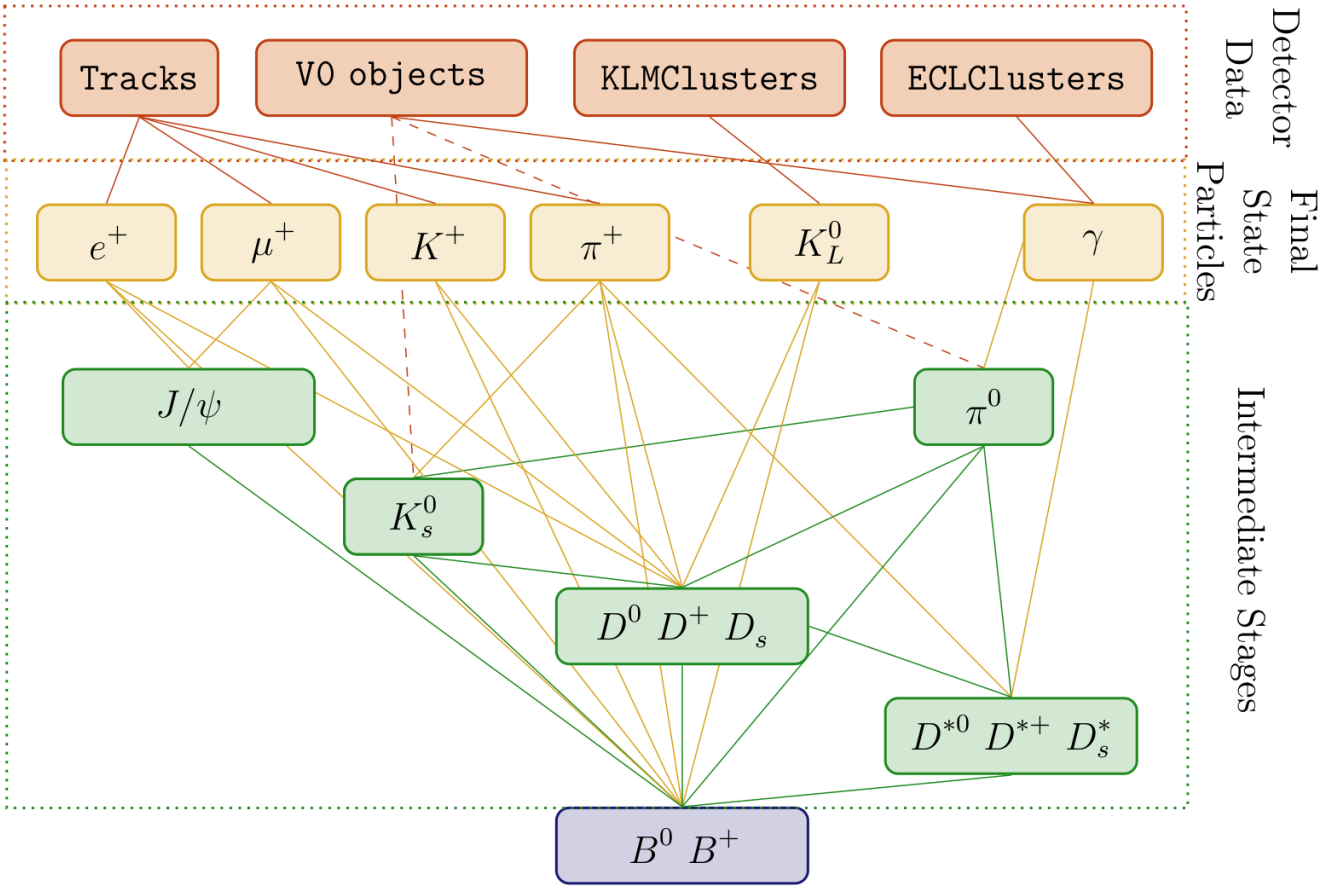

Figure 1.7: Hierarchy of the FEI

constrained mass and delta E can be enforced at the beginning of the training.

TODO Current Plots of FEI on Belle 1 and Belle 2

1.6.5 Calibration

[P. Urquijo] An important systematic error in analyses using tag methods is the efficiency calibration. Several techniques for calibration have been used in Belle, and are described in turn.

- $B \to D^{(*)}\ell\nu$ calibration. Events are double tagged, where the signal side is reconstructed in a known semileptonic decay mode, in bins of the tag quality variables. This has been used in $B \to X_u\ell\nu$ analyses. The systematic errors were approximately 4.5%, shared between statistical (1.5%), reconstruction (2.7%), and branching fraction uncertainties (3%). The detection uncertainties are mostly based on data driven techniques, while the branching fractions are more difficult to improve in the future.

- $B \to X\ell\nu$ calibration. Events are also double tagged, however the signal side selected only via the presence of a charged lepton originating from a semileptonic B decay. This has been used in precision exclusive $B \to D^{(*)}\ell\nu$ decay analyses. The technique is systematics limited but higher precision than the $B \to D^{(*)}\ell\nu$ calibration approach. The uncertainty can be controlled via the choice of tighter tag side criteria at a cost of statistical power.

- Control mode calibration. An analysis sideband region is chosen that is enhanced in a well known decay mode, and calibrated accordingly. This technique has been used by rare decay analyses that are

Bibliography

[1] T. Abe, Belle II Collaboration, *Belle II Technical Design Report*, `arXiv:1011.0352 [physics.ins-det]`.

[2] J. Tanaka, *Kinematic fitting*, Belle Note 194 .

[3] F. M. B. P. W. Waltenberger, W. Mitaroff and H. V. Riedel, *The RAVE/VERTIGO vertex reconstruction toolkit and framework*, Journal of Physics: Conference Series **119** (2008) 032037.

[4] R. F. W. Waltenberger and P. Vanlaer, *Adaptive Vertex Fitting*, CERN-CMS-NOTE-2008-033 .

[5] R. Fröhwirth, *Application of Kalman Filtering to Track and Vertex fitting*, Nuclear Instruments and Methods in Physics Research A **262** (1987) 444–450.

[6] T. Kawasaki, *Simulation Study for Measurement of CP violation in $B \to$ charmonium + Ks at KEK B-factory*, Ph.D. Thesis .