Nuclear Inst. and Methods in Physics Research, A 976 (2020) 164269

Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

journal homepage: www.elsevier.com/locate/nima

Global decay chain vertex fitting at Belle II

J.-F. Krohn [a,*], F. Tenchini [b], P. Urquijo [a], F. Abudinén [c], S. Cunliffe [b], T. Ferber [b], M. Gelb [d], J. Gemmler [d], P. Goldenzweig [d], T. Keck [d], I. Komarov [b], T. Kuhr [e], L. Ligioi [c], M. Lubej [f], F. Meier [g,i], F. Metzner [d], C. Pulvermacher [d], M. Ritter [e], U. Tamponi [h], A. Zupanc [f]
Belle II analysis software group

[a] University of Melbourne, Melbourne, Australia
[b] Deutsches Elektronen-Synchrotron, Hamburg, Germany
[c] Max-Planck-Institut für Physik, Munich, Germany
[d] Karlsruhe Institute of Technology, Karlsruhe, Germany
[e] Ludwig Maximilians Universität, Munich, Germany
[f] Jožef Stefan Institute, Ljubljana, Slovenia
[g] University of Sydney, Sydney, Australia
[h] INFN - Sezione di Torino, Torino, Italy
[i] Duke University, Durham, USA

ABSTRACT

In this paper we report the implementation of a global vertex fitting algorithm within the Belle II analysis software environment, which was originally developed for BaBar (Hulsbergen, 2005). We explore the impact of global vertex fitting algorithms for flavour physics analyses with the Belle II detector at the SuperKEKB e^+e^- collider, such as in the reconstruction of final states with neutral particles, and in fits with geometrical constraints from SuperKEKB's nano-beam interaction region. The algorithm is compared to the standard vertex fitting algorithm employed by the Belle experiment. We have developed the fitting framework to utilise the EIGEN library for linear algebra operations, reducing the computation time for vertex fitting operations by an order of magnitude over previous methods. This has a significant impact on physics analysis computing efficiency, where vertex fitting over large combinations of final state particles is one of the most CPU intensive operations at Belle II.

1. Introduction

Particle vertex fitting techniques are widely used in particle and nuclear physics. Beyond the suppression of background, applications range from the improvement of particle momentum resolution (under the assumption they originate from some vertex point), to the determination of the presence of intermediate particles, and the precision determination of decay vertex positions. One can, for example, combine the measurements of two charged pion tracks originating from the decay of a K^0_S to extract the decay vertex position, flight length, and four-vector and their uncertainties. By performing a kinematic fit, one obtains an improvement of the pion track momenta and can use the χ^2 of the fit result to suppress background.

In order to construct more complex decay topologies, one usually combines cascades of these fits starting with long lived stable particles such as electrons, muons, pions and photons, forming intermediate resonances and finally the full decay of interest. For example, in the decay $B^0 \to J/\psi K^0_S$, where $J/\psi \to \mu^+\mu^-$ and $K^0_S \to \pi^+\pi^-$, one would first fit the J/ψ and K^0_S candidates and then use these to construct the B^0 candidates, as depicted in Fig. 1a and 1b. However, this only works well if the final state particles are charged and leave traces in the tracking detectors. Neutral particles cannot be tracked in Belle II; they are only measured by their energy deposition in the calorimeter. Single layer crystal calorimeters do not offer directional information on where the particle associated to an energy deposition originated from. That means that the decay vertex cannot be extracted from a fit that exploits only the calorimeter information. In order to obtain the momentum vector, it could be assumed that the particle originates from the primary interaction point and travels directly into the cluster's centre of gravity. This can introduce a large bias on the momentum direction. Consider, for example, the decay $B^0 \to J/\psi K^0_S$, where the kaon instead decays to neutral final states, $K^0_S \to \pi^0\pi^0$, and the pions decay, $\pi^0 \to \gamma\gamma$, as displayed in Fig. 2. Weakly decaying intermediate particles such as K^0_S mesons have flight lengths of up to tens of centimetres at Belle II. Thus for a neutral particle in such a particle decay chain, the assumption that it originates from the primary e^+e^- collision is not sufficient. Furthermore, decays of particles into an intermediate state and a charged particle, i.e. $D^{*+} \to D^0\pi^+$, form a decay vertex that is only constraint by one measurement and thus susceptible to measurement errors.

The method we present in this paper overcomes these issues by globally fitting the entire decay tree in a single fit, taking into account all

* Corresponding author.
E-mail address: jkrohn@student.unimelb.edu.au (J.-F. Krohn).

https://doi.org/10.1016/j.nima.2020.164269
Received 1 April 2019; Received in revised form 7 June 2020; Accepted 9 June 2020
Available online 12 June 2020

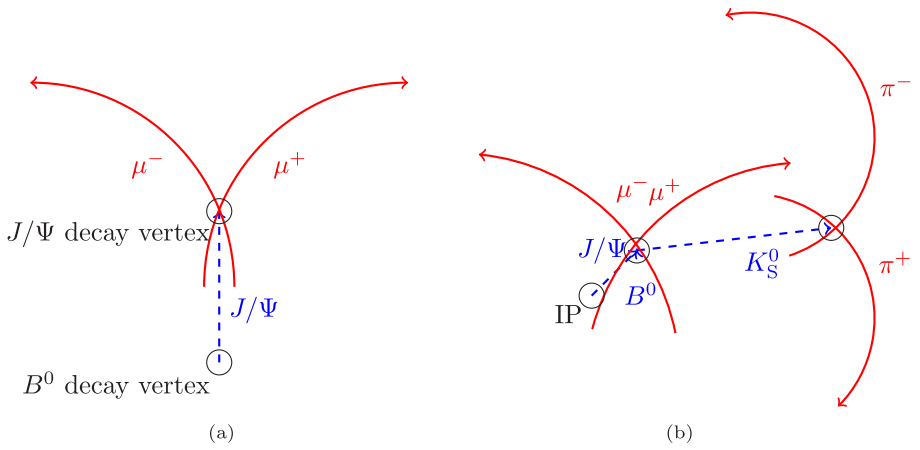

Fig. 1. (a) Depiction of a $J/\psi \to \mu^+\mu^-$ decay. The red lines show the track helix approximations obtained from the tracking detectors, the blue dashed lines show the decaying particle momentum vectors found by the fit. Since the decay length of the J/ψ is too short to be seen in the detector, its decay vertex is taken to be the one of the B^0. (b) Depiction of a $B^0 \to J/\psi(\to \mu^+\mu^-)K^0_S(\to \pi^+\pi^-)$ decay. Measured track helices do not necessarily overlap in three dimensions. The depicted length ratios are not to scale.

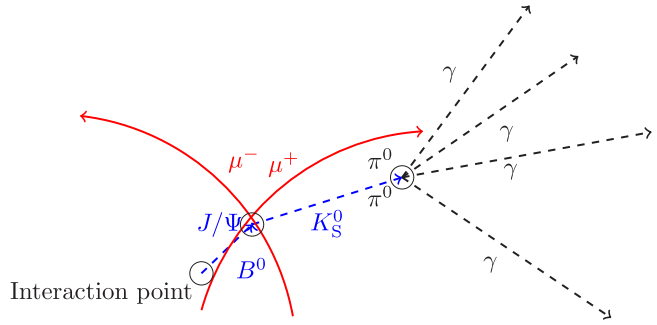

Fig. 2. Depiction of the decay $B^0 \to K^0_S J/\psi$. The red lines show the track helix approximations obtained by the tracking detectors, the blue dashed lines show the composite particle momentum vectors found by the fit. The dashed black lines depict the photon momenta found by the fit. Note that these can only be determined by the fit as the directional information of the calorimeter is not sufficient. The initial guess is that they point from the interaction point towards the calorimeter cluster. The decay lengths of the J/ψ and π^0 are too short to be seen in the detector therefore the vertex positions are taken from the particle above them in the hierarchy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

intermediate particles, extracting all involved particle's four-momenta, vertex positions, flight lengths and their covariance matrices, using a Kalman Filter as described in Ref. [1]. We use the software environment of Belle II and rely on the C++template library EIGEN [2] for matrix operations, which provides fast execution times for the fit algorithm. We furthermore present physics applications of the fitter with Belle II Monte Carlo samples and study performance characteristics.

1.1. The Belle II experiment

The vertex fitting algorithm described in this paper was developed for the analysis software framework of Belle II. The Belle II experiment takes place at the asymmetric e^+e^- collider, SuperKEKB. SuperKEKB provides a beam energy slightly above the mass of the $\Upsilon(4S)$ resonance (10.58 GeV) at an instantaneous luminosity of $8 \cdot 10^{35}$ cm^{-2} s^{-1}. The $\Upsilon(4S)$ resonance decays into pairs of B-mesons just above production threshold, hence this type of experiment is referred to as a B-factory. The asymmetric beam energy gives the B-meson a relativistic boost along a direction close to the detector's axis of symmetry, increasing its flight length in the lab frame, which makes it possible to study the time evolution of B decays — a key observable in the study of CP symmetry violation.

The Belle II detector has a cylindrical structure designed to study the decays of B- and D-mesons, τ-leptons and other processes produced in e^+e^- collisions. Six layers of silicon vertex detectors (2 layers of silicon pixels (PXD), and 4 layers of double sided silicon detectors (SVD)) are located in the central volume of the detector, designed to accurately track the flight paths of charged particles. The following layers are, a central drift chamber (CDC) used to measure track trajectories within a solenoid magnetic field, Cherenkov light based particle identification devices surrounding the CDC in the barrel (TOP) and forward regions (ARICH), followed by the CsI(Tl) electromagnetic calorimeter (ECL). The outermost layers are composed of a magnet solenoid and a K^0_L and muon detector system (KLM), which is also used as the flux return yoke of the magnet. The magnetic field is aligned along the detector's axis of symmetry. For a more detailed description of the detector see Ref. [3].

Figs. 3 and 4 show an event display depicting simulated particles traversing the Belle II detector. In the decay $\Upsilon(4S) \to \bar{B}^0 B^0$, one meson decays as $\bar{B}^0 \to D\omega\pi\pi$, and the other as $B^0 \to K_S J/\psi$, with $J/\psi \to \mu^+\mu^-$ and $K^0_S \to \pi^0\pi^0$ with $\pi^0 \to \gamma\gamma$. Fig. 3 depicts the full detector geometry and Fig. 4 shows a close-up of the inner vertex detectors. In this example we show that the decay vertex of the K^0_S can be highly displaced.

2. Extended Kalman filter

Vertex fitting can be formulated as a least squares minimisation problem. The computational challenge in finding a solution lies in matrix inversions, which naively scale as $\mathcal{O}(n^3)$, where n is the dimension of the matrix. In a naive approach, this is equal to the number of parameters extracted in the fit. An extended Kalman Filter is an iterative approach to find the least squares estimator by defining a series of constraints (knowledge of parameters from measurements and symmetries) on a hypothesis (a set of particle parameters in this case). The hypothesis has different states during the stages of the filtering process. The filtering process is an iterative algorithm applying the constraints sequentially and updating the state with respect to the constraints. The sequence is repeated until convergence is reached or divergence is observed. In the case of convergence, the last state describes the best hypothesis for the parameters that can be found. We use an Extended Kalman Filter in the gain matrix formulation [1,4,5] in our algorithm. The state vector $\boldsymbol{x}$ holds the particle parameters to be optimised by the fit. The parameters depend on the type of particle. The most general parametrisation takes the form

$$\boldsymbol{x} = \{x_1, y_1, z_1, \theta_1, p_{x1}, p_{y1}, p_{z1}, E_1, \ldots, x_n, y_n, z_n, \theta_n, p_{xn}, p_{yn}, p_{zn}, E_n\}, \quad (1)$$

where vertex coordinates of the ith particle are denoted as $\{x_i, y_i, z_i\}$, its decay length is denoted as θ_i and the four momentum is $\{p_{x,i}, p_{y,i}, p_{z,i},$

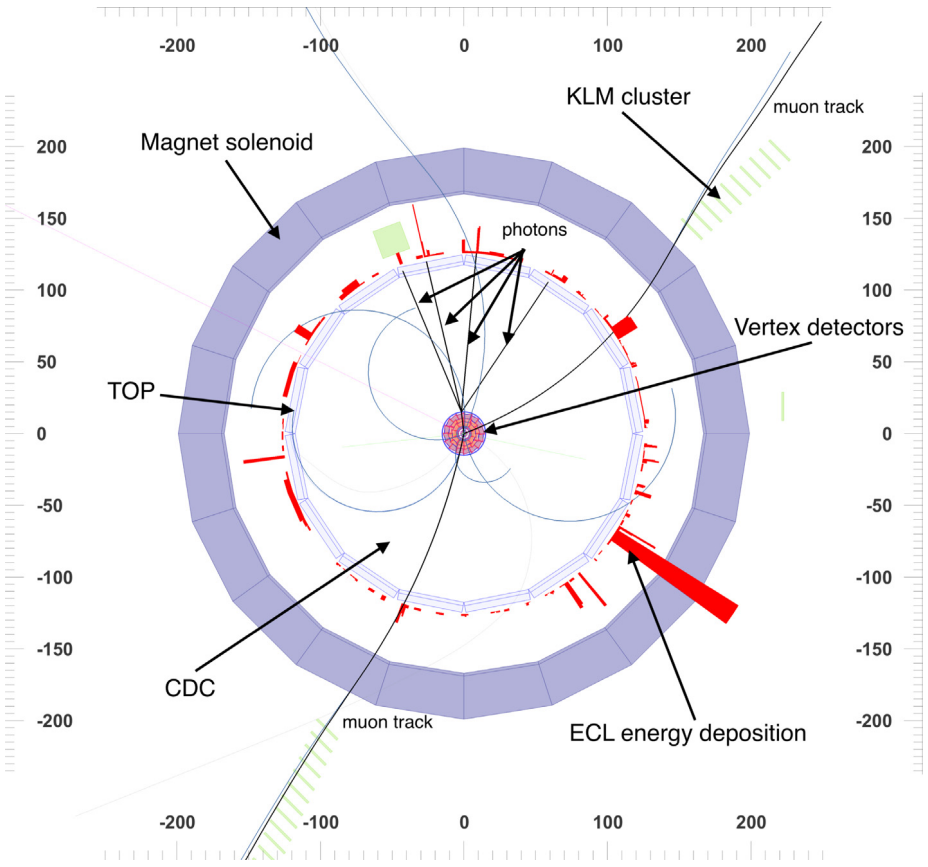

Fig. 3. Event display projected onto the x–y plane depicting the process of $\bar{e}^+e^- \to Y(4S) \to \bar{B}^0 B^0$, with $\bar{B}^0 \to D\omega\pi\pi$ and $B^0 \to K_S J/\psi$, where $J/\psi \to \mu^+\mu^-$ and $K_S^0 \to \pi^0\pi^0$ with $\pi^0 \to \gamma\gamma$. All particles of the signal B^0 decay are indicated by black lines. The rest of the event correspond to the decay of the $\bar{B}^0$. All distances are measured in centimetres. For a close-up of the vertex detector see Fig. 4.

$E_i\}$, and n is the number of particles in the fitted topology. This implies that $n \approx 8 \times$ number of particles. However, for final state particles the decay vertex is not of interest, so that only the momenta have to be parametrised. The problem is split into k constraints given by measurements and other knowledge of parameters (see Section 3 for the definitions of the measurement $\boldsymbol{m}$ and hypothesis $\boldsymbol{h}$ for each of the different constraints). The minimised χ^2 can be expressed as a weighted sum of k sets of equations, which are the constraints. For a given iteration, α, of the Kalman Filter, one can write

$$\chi^2_\alpha = \sum_k \chi^2_k, \tag{2}$$

and

$$\begin{aligned}\chi^2_k = &(\boldsymbol{x}_k - \boldsymbol{x}_{k-1})^T \boldsymbol{C}^{-1}_{k-1}(\boldsymbol{x}_k - \boldsymbol{x}_{k-1}) + \\ &(\boldsymbol{m}_k - \boldsymbol{h}_k(\boldsymbol{x}_k))^T \boldsymbol{V}^{-1}_k (\boldsymbol{m}_k - \boldsymbol{h}_k(\boldsymbol{x}_k)),\end{aligned} \tag{3}$$

using the measurement covariance matrix $\boldsymbol{V}^{-1}_k$ transported to the $\boldsymbol{m}-\boldsymbol{h}$ system as a weight, and the covariance matrix $\boldsymbol{C}^{-1}_{k-1}$ of the hypothesis, such that the hypothesis of each constraint k depends on the outcome of the previous constraint. Minimising Eq. (3) yields a rule to find a new state $\boldsymbol{x}^\alpha_k$ for the current iteration and constraint, that is

$$\boldsymbol{x}^\alpha_k = \boldsymbol{x}^\alpha_{k-1} - \boldsymbol{K}^\alpha_k \boldsymbol{r}^\alpha_k, \tag{4}$$

where $\boldsymbol{K}$ is a gain matrix, which will be defined below. However, we first define the residual $\boldsymbol{r}$ as the distance between measurement and hypothesis

$$\boldsymbol{r}^\alpha_k = \boldsymbol{m}_k - \boldsymbol{h}_k(\boldsymbol{x}^\alpha_{k-1}) \,. \tag{5}$$

The hypothesis $\boldsymbol{h}_k(\boldsymbol{x}^\alpha)$ can be linearised around a reference state of the previous iteration $\boldsymbol{x}^{\alpha-1}$ using the Jacobian

$$\boldsymbol{H}^{\alpha-1}_k := \frac{\partial \boldsymbol{r}_k(\boldsymbol{x}^{\alpha-1})}{\partial \boldsymbol{x}^{\alpha-1}} = -\frac{\partial \boldsymbol{h}_k(\boldsymbol{x}^{\alpha-1})}{\partial \boldsymbol{x}^{\alpha-1}}, \tag{6}$$

such that

$$\boldsymbol{h}_k(\boldsymbol{x}^\alpha_{k-1}) = \boldsymbol{h}_k(\boldsymbol{x}^{\alpha-1}) - \boldsymbol{H}^{\alpha-1}_k \cdot (\boldsymbol{x}^\alpha_{k-1} - \boldsymbol{x}^{\alpha-1}). \tag{7}$$

Note that $\boldsymbol{x}^{\alpha-1}$ is always taken after the last constraint in iteration $\alpha-1$ was filtered and the minus sign is coming from generalising $\boldsymbol{H}$ as the derivative of the residual. Eq. (5) thus becomes

$$\boldsymbol{r}^\alpha_k = \boldsymbol{m}_k - \boldsymbol{h}_k(\boldsymbol{x}^{\alpha-1}) + \boldsymbol{H}^{\alpha-1}_k \cdot (\boldsymbol{x}^\alpha_{k-1} - \boldsymbol{x}^{\alpha-1}). \tag{8}$$

The gain matrix is calculated for every constraint in every iteration and is defined as

$$\boldsymbol{K}_k = \boldsymbol{C}_{k-1}(\boldsymbol{H}^{\alpha-1}_k)^T \boldsymbol{R}^{-1}_k. \tag{9}$$

The only matrix to be inverted is of the dimension of the constraint and not the dimension of the state-space. $\boldsymbol{R}_k$ is defined below. The current state's covariance matrix is obtained via propagation of uncertainties

$$\boldsymbol{C}_k = \boldsymbol{C}_{k-1} - \boldsymbol{C}_{k-1}\boldsymbol{H}^{\alpha-1}_k \boldsymbol{R}^{-1}_k \boldsymbol{H}^{\alpha-1}_k \boldsymbol{C}^T_{k-1} \tag{10}$$

There are other formulations of Eq. (10) that are numerically more stable as discussed in [1]. The covariance matrix of the residual system can be found to be

$$\boldsymbol{R}_k = \boldsymbol{V}_k + \boldsymbol{H}^{\alpha-1}_k \boldsymbol{C}_{k-1}(\boldsymbol{H}^{\alpha-1}_k)^T. \tag{11}$$

For constraints where no measurement is available, for example the kinematic or geometric constraint, it can be shown that Eq. (11) is valid if $\boldsymbol{V}_k = 0$ [1].

Linearising the hypothesis around a reference state, see Eq. (7), is a major improvement to the stability of the fit. The authors of Ref. [1] pointed us towards this improvement and more details can be found in Ref. [6]. In this formulation of the LSE problem, the update of the covariance matrix, Eq. (10), is the computational bottleneck, as it is a

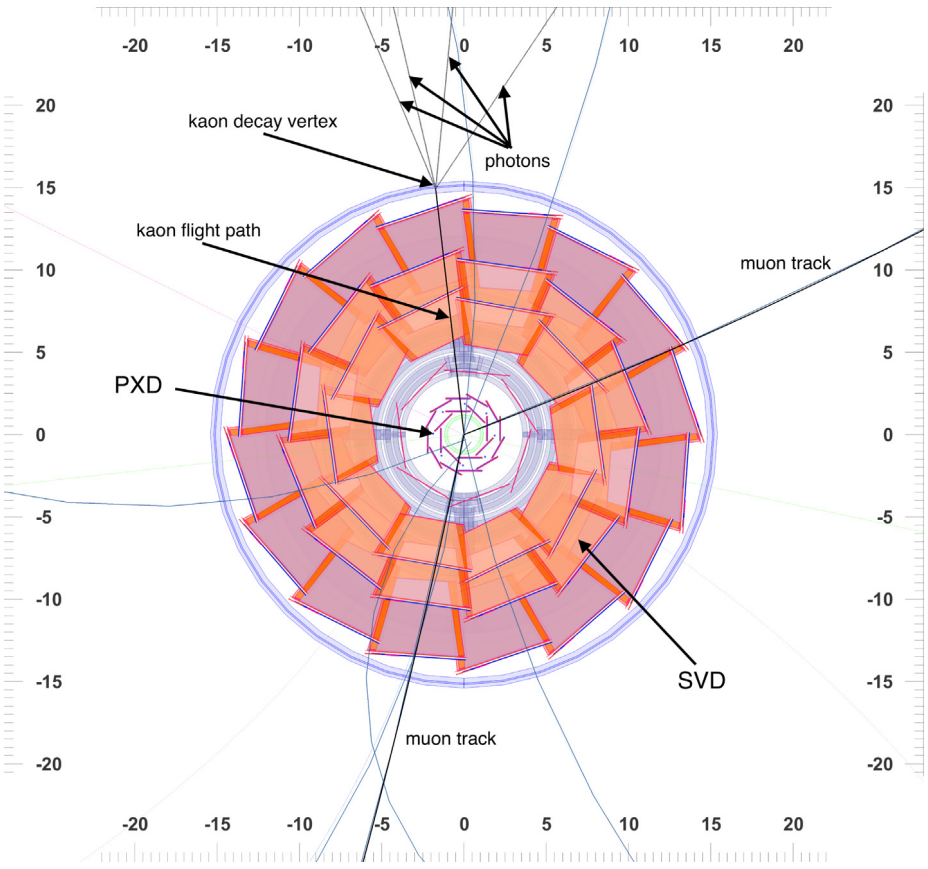

Fig. 4. Close-up of Fig. 3. In this event, the K_S^0 travels until the edge of the vertex detectors before decaying.

dense $m \times m$ matrix with the dimension of the state vector m and must be calculated and filled k times for each iteration. We define a fit as converged when the change of the χ^2 is less than 2% and the resulting covariance matrix is invertible, which is tested by checking its rank. A Kalman Filter is sensitive to the initial state vector and its associated covariance matrix chosen. We choose values of the measurement in the respective constraint multiplied by a factor 1000 as the initial value of the hypothesised covariance matrix. For the initial vertex positions we use the points of closest approach of track-helices or those of the "parent particles", if they are available, zero otherwise.

3. Parametrising and constraining the decay chain

To parametrise the decay chain, we use a set of parameters that describe the properties of the particles. We perform a number of reductions on these properties to reduce the dimensionality of the problem. For final state particles, we only save the momenta, as they do not have decay vertices. For their production vertices we use the decay vertices of their respective composite particles, referred to as the parents. The energy of each final state particle is calculated using the momenta and its nominal mass hypotheses, taking the mass values provided by the particle data group (PDG) [7]. Intermediate particles are classified in two categories: particles that decay dominantly via the strong force, and those that decay via the weak force. Strongly decaying particles with a lifetime of less than 10^{-14} s, which corresponds to boosted flight lengths of less than 1 μm, are treated as if they instantly decay (the detector has a vertex position resolution in the x-y plane for charged particles of 20–30 μm). The hypothesised quantities are their energy and momenta, while the production and decay vertices coincide and are taken from their parent particle's decay vertex. For weakly decaying particles, we additionally measure a decay vertex and a flight length, defined as the distance between the production and decay vertices in three dimensions.

3.1. Parametrising the constraints

Constraints are defined by Eq. (7), note that Lagrange multipliers could be eliminated, which allows for this convenient definition of constraints [1]. The resulting Jacobians take the form of $m \times n$ matrices where m is the dimension of the state vector and n is the dimension of the respective constraint. Thus, only few of its elements are non-zero. For example, for a three dimensional point constraint k, the hypothesis of particle number j with $\boldsymbol{h}_j = \{x_j, y_j, z_j\}$ and $\boldsymbol{x}$ as in Eq. (1), only the jth diagonal block is non-zero

$$-\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}} = \boldsymbol{H} = \begin{pmatrix} 0 & -\mathbb{1}_3 & 0 \end{pmatrix}. \tag{12}$$

The blocks filled with zero correspond to the parameters of particle $\boldsymbol{x}_i \neq \boldsymbol{x}_j$. We will omit the columns filled with zeros throughout this section, for brevity.

In the following we list the definitions of constraints that have been implemented in the Belle II software, based on the specific geometry of Belle II.

3.1.1. Reconstructed track

A track can be parametrised with a five parameter helix. In Belle II it was chosen to use a perigee-parametrised helix, such that the helix is defined at the perigee, the point of closest approach of the helix to the origin of the coordinate system. The corresponding transformations to transport a helix to that point are discussed in Ref. [9]. A description of the parameters can be found in Table 1, and a depiction of the helix is in Fig. 5. We parametrise tracks such that we can express the model's

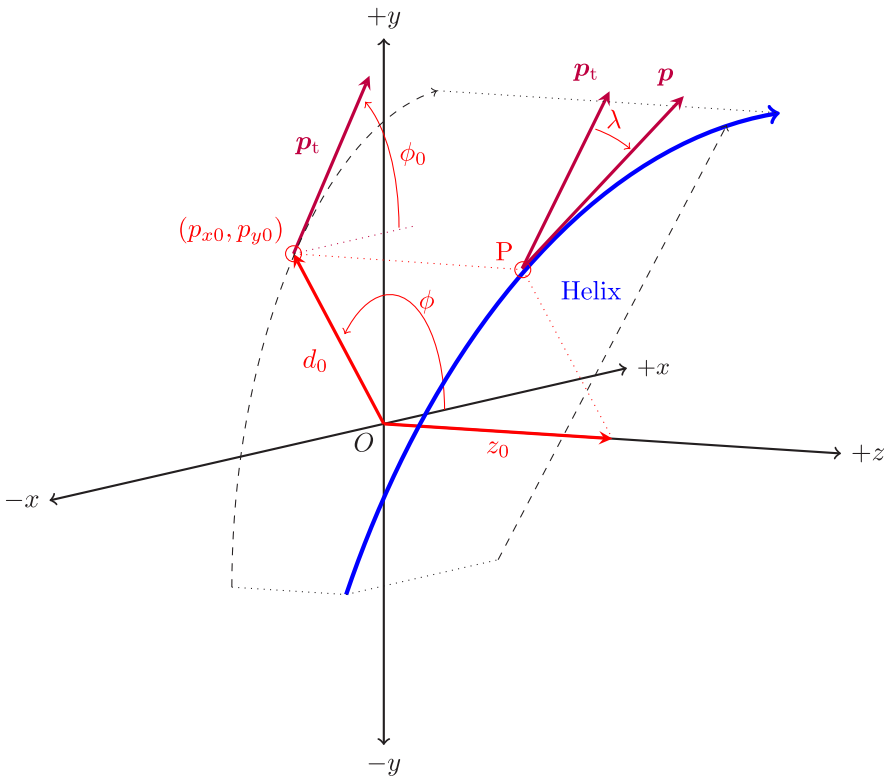

Fig. 5. The perigee parametrisation of the track helix, depiction adapted from [8]. A description of the parameters can be found in Table 1.

Table 1
Definitions of the Belle II helix parametrisation and their dependencies.

d_0	The distance of closest approach to the z axis (POCA) signed with the z component of the angular momentum w.r.t. to the origin.
ϕ_0	Azimuthal angle of the momentum at the POCA.
ω	Scaled inverse of the track momentum.
z_0	The pivotal point is the perigee that is the POCA.
$\tan\lambda$	The angle of the momentum at the POCA with respect to the x–y plane.
$\phi = \mathrm{atan2}(p_y, p_x)$	Angle of the x–y plane to the helix.
$\Delta_\parallel = -x\cdot\cos\phi - y\cdot\sin\phi$ $\Delta_\perp = -y\cdot\cos\phi + x\cdot\sin\phi$ $A = 2\cdot\Delta_\perp + \omega\cdot(\Delta_\perp^2 + \Delta_\parallel^2)$ $U = \sqrt{1+\omega\cdot A}$	Quantities used to transport the coordinate system, such that $\{x, y, z\}$ points to the perigee.
λ	Angle between the x–y plane and the helix.
$l = \mathrm{atan2}(\omega\cdot\Delta_\parallel, 1+\omega\cdot\Delta_\perp)$	Arc length from the perigee to a point.
$p_x,\ p_y,\ p_z$	Momenta along the x, y, z axes.
q	Charge of the particle.
B_z	Magnetic field strength in z-direction.
$a = B_z/c$	Magnetic field strength in the z direction divided by the speed of light.
$p_t = \sqrt{p_x^2 + p_y^2}$	Transverse momentum.
$p_{t0} = \sqrt{p_{x0}^2 + p_{y0}^2}$	Pseudo transverse momentum.
$p_{x0} = p_x - a\cdot q\cdot y$	Pseudo momentum along the x direction.
$p_{y0} = p_y + a\cdot q\cdot x$	Pseudo momentum along the y direction.
$r^2 = x^2 + y^2$	Radius squared.
$\beta = 1 + \frac{p_{t0}}{p_t}$	Quantity used for reading convenience.

dependence on Cartesian parameters as

$$\boldsymbol{h}_{\mathrm{track}}(\boldsymbol{x}) = \begin{pmatrix} d_0 \\ \phi_0 \\ \omega \\ z_0 \\ \tan\lambda \end{pmatrix} = \begin{pmatrix} A(1+U)^{-1} \\ \mathrm{atan2}(p_y, p_x) - \mathrm{atan2}(\omega\cdot\Delta_\parallel, 1+\omega\cdot\Delta_\perp) \\ a\cdot q/p_t \\ z + l\cdot\tan\lambda \\ p_z/p_t \end{pmatrix}. \tag{13}$$

Where atan2 refers to the ϕ domain corrected inverse tangent function. We use the same parametrisation for the hypothesis and the measurement. We label the measurement quantities with the index m. The residuals of iteration α then become

$$\boldsymbol{r}^{\alpha}_{track}(\boldsymbol{x}) = \begin{pmatrix} d_{0,m} - d_0 \\ \phi_{0,m} - \phi_0 \\ \omega_m - \omega \\ z_{0,m} - z_0 \\ \tan\lambda_m - \tan\lambda \end{pmatrix} + \boldsymbol{H}^{\alpha-1}\cdot(\boldsymbol{x}^{\alpha}_{k-1} - \boldsymbol{x}^{\alpha-1}) \,. \tag{14}$$

We define the Jacobian block $\boldsymbol{A} := -\partial\boldsymbol{h}/\partial\boldsymbol{x}$ as the derivatives with respect to the vertex position, and $\boldsymbol{B} := -\partial\boldsymbol{h}/\partial\boldsymbol{p}$ as the derivatives with respect to momentum. The positions of these blocks in the Jacobian depend on the topology fitted and the particle represented by the track. We choose to order the state vector hierarchically. This means that the decay vertex parameters come before its momentum, followed by the child particle's parameters. The full Jacobian $\boldsymbol{H}$ then takes the following form

$$\boldsymbol{H} = \begin{pmatrix} \dots & \dots & \dots \\ \dots & \dots & \dots \\ \dots \quad \boldsymbol{A} & \dots & \boldsymbol{B} \quad \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{pmatrix}. \tag{15}$$

For the non-zero elements of the Jacobian blocks, denoted by $-\partial d_0/\partial x = \boldsymbol{A}_{d_0,x}$, we derive the spatial components as

$$\begin{aligned} \boldsymbol{A}_{d_0,x} &= -\frac{p_{y0}}{p_{t0}}, & \boldsymbol{A}_{d_0,y} &= \frac{p_{x0}}{p_{t0}}, \\ \boldsymbol{A}_{\phi_0,x} &= -\frac{a\cdot q\cdot p_{x0}}{p_{t0}^2}, & \boldsymbol{A}_{\phi_0,y} &= -\frac{a\cdot q\cdot p_{y0}}{p_{t0}^2}, \\ \boldsymbol{A}_{z_0,x} &= \frac{p_x\cdot p_{x0}}{p_{t0}^2}, & \boldsymbol{A}_{z_0,y} &= \frac{p_x\cdot p_{y0}}{p_{t0}^2}, & \boldsymbol{A}_{z_0,z} &= -1, \end{aligned} \tag{16}$$

and for the momenta

$$\begin{aligned}
\boldsymbol{B}_{d_0,p_x} &= \frac{y((aq)^2 r + 2aqp_y x + 2p_y^2\beta)}{p_{t0}p_t^2\beta^2} \\
&\quad + \frac{p_x(2p_y x\beta + aq(y^2(-2+\beta) + x^2\beta))}{p_{t0}p_t^2\beta^2}, \\
\boldsymbol{B}_{d_0,p_y} &= -\frac{2p_x^2 x\beta + 2p_x y(p_y - aqx + p_y p_{t0}/p_t)}{p_{t0}p_t^2\beta^2} \\
&\quad - \frac{aq(aqrx - p_y(x^2(-2+\beta) + y^2\beta))}{p_{t0}p_t^2\beta^2},
\end{aligned} \tag{17}$$

and

$$\begin{aligned}
\boldsymbol{B}_{\omega,p_x} &= \frac{aqp_x}{p_t^3}, \ \boldsymbol{B}_{\omega,p_y} = \frac{aqp_y}{p_t^3}, \\
\boldsymbol{B}_{z_0,p_x} &= -\frac{p_z(p_x^2 x - p_y(aqr + p_y x) + 2p_x p_y y)}{p_{t0}^2 p_t^2}, \\
\boldsymbol{B}_{z_0,p_y} &= -\frac{p_z(p_x(aqr + 2p_y x) - p_x^2 y + p_y^2 y)}{p_{t0}^2 p_t^2}, \\
\boldsymbol{B}_{z_0,p_y} &= -(aq)^{-1}\mathrm{atan2}\left(aq(p_y y - p_x x),\ p_x^2 + p_y p_{y0} - aqp_x y\right), \\
\boldsymbol{B}_{\tan\lambda,p_x} &= \frac{p_z p_x}{p_t^3}, \ \boldsymbol{B}_{\tan\lambda,p_y} = \frac{p_z p_y}{p_t^3}, \ \boldsymbol{B}_{\tan\lambda,p_z} = -p_t^{-1}.
\end{aligned} \tag{18}$$

3.1.2. Reconstructed photon

For photons we measure the position of the calorimeter cluster and its energy and can infer the vertex parameters. The geometry, depicted in Fig. 6, gives

$$0 = \boldsymbol{u}_{\text{parent}} + \boldsymbol{\delta} - \boldsymbol{m}, \tag{19}$$

substituting $\boldsymbol{\delta} = \tau \cdot \boldsymbol{p}$ and inserting the energy relation, we get

$$\boldsymbol{h}_{\text{photon}}(\boldsymbol{x}) = \begin{pmatrix} u_x + \tau \cdot p_x \\ u_y + \tau \cdot p_y \\ u_z + \tau \cdot p_z \\ \sqrt{p_x^2 + p_y^2 + p_z^2} \end{pmatrix} \text{ and } \boldsymbol{m}_{\text{photon}}(\boldsymbol{x}) = \begin{pmatrix} m_x \\ m_y \\ m_z \\ E_m \end{pmatrix}, \tag{20}$$

where $\{u_x, u_y, u_z\}$ are the production vertex coordinates, $\{p_x, p_y, p_z\}$ are the parameters of the momentum vector pointing from the production vertex to the calorimeter cluster, $\{m_x, m_y, m_z, E_m\}$ are the position and measured energy of the corresponding ECL cluster. The parameter τ is a scalar with the units of time, it can be eliminated when writing down the residual to reduce the dimensionality of the equation system and avoid a trivial local minimum of $\boldsymbol{r}_\gamma$ at $\tau = 0$ when taking $\{u_x, u_y, u_z\} = 0$ as the starting point of the first iteration. Since the geometry of the detector is cylindrical, we cannot simply eliminate any of the dimensions as this could introduce a pole in the residual equations. Therefore we sort the momenta and eliminate the dimension with the highest momentum such that we form a 3-dimensional equation system

$$\boldsymbol{r}'^{\alpha}_{\text{photon}}(\boldsymbol{x}) = \begin{pmatrix} (m_i - u_i) - (m_k - u_k)\frac{p_i}{p_k} \\ (m_j - u_j) - (m_k - u_k)\frac{p_j}{p_k} \\ E_m - \sqrt{p_i^2 + p_j^2 + p_k^2} \end{pmatrix} + \boldsymbol{H}^{\alpha-1} \cdot (\boldsymbol{x}^{\alpha}_{k-1} - \boldsymbol{x}^{\alpha-1}), \tag{21}$$

where the indices i, j, k indicate the dimensions by order of increasing momentum $p_k \geq p_i \geq p_j$. We define $\boldsymbol{A}_{i,u_k} := -\partial h'_i/\partial u_k$ and $\boldsymbol{B}_{i,p_k} := -\partial h'_i/\partial p_k$ with the hypothesis of the reduced system r'. Thus, the non-zero entries are

$$\begin{aligned}
\boldsymbol{A}_{0,u_k} &= \frac{p_i}{p_k}, & \boldsymbol{A}_{0,u_i} &= -1, \\
\boldsymbol{A}_{1,u_k} &= \frac{p_j}{p_k}, & \boldsymbol{A}_{1,u_j} &= -1, \\
\boldsymbol{B}_{0,p_k} &= p_k^{-2}, & \boldsymbol{B}_{0,p_i} &= \frac{u_k - m_k}{p_k}, \\
\boldsymbol{B}_{1,p_k} &= p_k^{-2}, & \boldsymbol{B}_{1,p_j} &= \frac{u_k - m_k}{p_k}, \\
\boldsymbol{B}_{2,p_k} &= -\frac{p_k}{|\boldsymbol{p}|}, & \boldsymbol{B}_{2,p_i} &= -\frac{p_i}{|\boldsymbol{p}|}, & \boldsymbol{B}_{2,p_j} &= -\frac{p_j}{|\boldsymbol{p}|}.
\end{aligned} \tag{22}$$

The full Jacobian then takes the form

$$\boldsymbol{H} = \begin{pmatrix} \ldots & & \ldots & & \ldots \\ \ldots & \boldsymbol{A} & \ldots & \boldsymbol{B} & \ldots \\ \ldots & & \ldots & & \ldots \end{pmatrix}. \tag{23}$$

We must transform the covariance matrix of the measurement into the reduced system. For that, we use

$$\boldsymbol{V}' = \boldsymbol{F}\boldsymbol{V}\boldsymbol{F}^T, \tag{24}$$

with the transport matrix $\boldsymbol{F} = \partial r'/\partial m$, which depends on the sorting of the momenta such that the non-zero entries are

$$\begin{aligned}
\boldsymbol{F}_{0,m_k} &= -\frac{p_i}{p_k}, & \boldsymbol{F}_{0,m_i} &= 1, \\
\boldsymbol{F}_{1,m_k} &= -\frac{p_j}{p_k}, & \boldsymbol{F}_{1,m_j} &= 1, \\
\boldsymbol{F}_{2,E_k} &= 1.
\end{aligned} \tag{25}$$

We do not parametrise this constraint in p_t in order to keep the derivatives in Eq. (22) as computationally simple as possible.

3.1.3. Reconstructed K_L^0

We treat K_L^0 in the same way as photons, except that we use the nominal mass provided by the PDG in the energy calculation. The KLM detector is used for the cluster position measurement instead of the calorimeter. It cannot provide a precise energy measurement. Instead, it extrapolates the energy deposited by a particle as $E = c \cdot n$, where n is the number of hit cells in the cluster and c is a constant with the units GeV. This approach makes the energy measurement for K_L^0 much less resolved than for photons.

3.1.4. Kinematic constraint

The kinematic constraint enforces four-momentum conservation, meaning it fits the four-momentum of the parent as the sum of the child momenta

$$\boldsymbol{r}^{\alpha}(\boldsymbol{x}) = \boldsymbol{p}_{\text{particle}} - \sum_i \boldsymbol{p}_{i,\text{child}} + \boldsymbol{H}^{\alpha-1} \cdot (\boldsymbol{x}^{\alpha}_{k-1} - \boldsymbol{x}^{\alpha-1}). \tag{26}$$

The Jacobian for a particle with *two* children, which in this example are taken to be stable particles, can be defined as

$$\boldsymbol{H} = \begin{pmatrix} \ldots & & \ldots & & \ldots & & \ldots \\ \ldots & & \ldots & & \ldots & & \ldots \\ \ldots & \boldsymbol{A} & \ldots & \boldsymbol{B}_1 & \ldots & \boldsymbol{B}_2 & \ldots \\ \ldots & & \ldots & & \ldots & & \ldots \end{pmatrix}, \tag{27}$$

with the blocks $\boldsymbol{A}, \boldsymbol{B}$ as

$$\boldsymbol{A} = -\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}_{\text{particle}}} = \mathbb{1}_4, \tag{28}$$

and

$$\boldsymbol{B}_i = -\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}_{\text{child},i}} = -1 \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ p_{x,i}/E_i & p_{y,i}/E_i & p_{z,i}/E_i & 0 \end{pmatrix}, \tag{29}$$

Note that the energy row of $\boldsymbol{B}_i$ depends on how the particle is parametrised. Composite particles, for example, are parametrised with an energy variable in the state vector, resulting in $\boldsymbol{B} = -\mathbb{1}_4$, while for stable particles Eq. (29) is used.

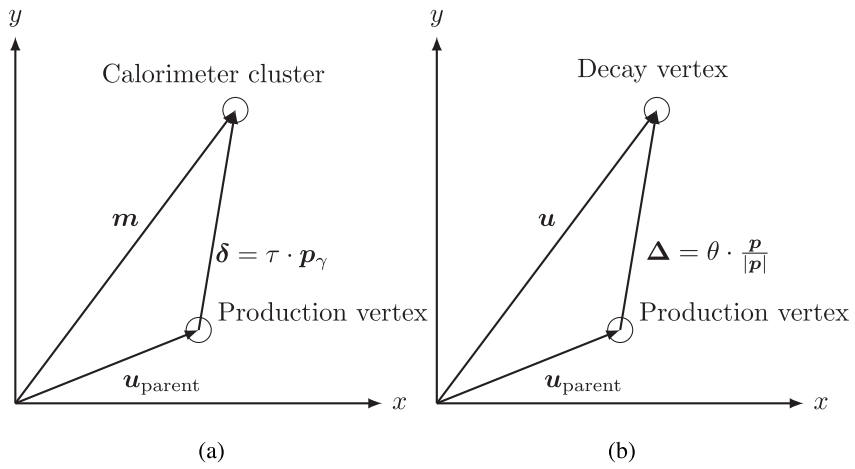

Fig. 6. (a) The photon constraint, Eq. (21), reduced to two dimensions for simplicity. The vector $\boldsymbol{\delta}$ is defined as pointing from the photon's production vertex to the measured calorimeter cluster, indicated with the photons parent's coordinate vector $\boldsymbol{u}$ and measurement vector $\boldsymbol{m}$. (b) Geometric constraint, Eq. (31). The vector $\boldsymbol{\Delta}$ is defined pointing from the particles decay and production vertex, indicated with the particle's and its parent's coordinate vector $\boldsymbol{u}$.

3.1.5. Geometric constraint

The geometric constraint fits the decay length parameter θ for composite particles, see Fig. 6. Accounting for the geometry we have

$$0 = \boldsymbol{u}_{\text{parent}} + \boldsymbol{\Delta} - \boldsymbol{u}\,. \tag{30}$$

Instead of directly extracting a flight vector $\boldsymbol{\Delta}$, we use the unit vector of the momentum as it is well constrained by the previously filtered kinematic constraints, substituting $\boldsymbol{\Delta} = \theta \cdot \boldsymbol{p}/|\boldsymbol{p}|$, allows for a more accurate estimation of θ, the decay length parameter in our model. In contrast with [1], we choose decay length rather than decay time, because it makes the fit more linear. We define the residual as

$$\boldsymbol{r}^{\alpha}(\boldsymbol{x}) = \boldsymbol{u}_{\text{parent}} + \theta \cdot \frac{\boldsymbol{p}}{|\boldsymbol{p}|} - \boldsymbol{u} + \boldsymbol{H}^{\alpha-1} \cdot (\boldsymbol{x}^{\alpha}_{k-1} - \boldsymbol{x}^{\alpha-1})\,. \tag{31}$$

using

$$\boldsymbol{A} = -\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{u}_{\text{parent}}} = \mathbb{1}_3,\ \boldsymbol{B} = -\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{u}} = -\mathbb{1}_3,\ \boldsymbol{C} = -\frac{\partial \boldsymbol{h}}{\partial \theta} = \frac{1}{|\boldsymbol{p}|}\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix}, \tag{32}$$

and

$$\boldsymbol{D} = -\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{p}} = \frac{\theta}{|\boldsymbol{p}|^3}\begin{pmatrix} (p_y^2 + p_z^2) & -p_x p_y & -p_x p_z \\ -p_y p_x & (p_x^2 + p_z^2) & -p_y p_z \\ -p_z p_x & -p_z p_y & (p_x^2 + p_y^2) \end{pmatrix}, \tag{33}$$

such that

$$\boldsymbol{H} = \begin{pmatrix} \ldots & & \ldots & & \ldots & & \ldots & & \ldots \\ \ldots & \boldsymbol{A} & \ldots & \boldsymbol{B} & \ldots & \boldsymbol{C} & \ldots & \boldsymbol{D} & \ldots \\ \ldots & & \ldots & & \ldots & & \ldots & & \ldots \end{pmatrix}. \tag{34}$$

3.1.6. Mass constraint

The mass constraint requires a particle four-vector to be consistent with its nominal mass. We treat the particle as a measurement with infinite precision and use the mass value provided by PDG such that

$$r^{\alpha}(\boldsymbol{x}) = m^2_{\text{PDG}} - E^2 + |\boldsymbol{p}|^2 + \boldsymbol{H}^{\alpha-1} \cdot (\boldsymbol{x}^{\alpha}_{k-1} - \boldsymbol{x}^{\alpha-1}), \tag{35}$$

with

$$\boldsymbol{H} = \begin{pmatrix} \ldots & 2p_x & 2p_y & 2p_z & -2E & \ldots \end{pmatrix}. \tag{36}$$

3.1.7. Beam spot constraint

B-mesons have a short lifetime and decay very close the beam spot, therefore it is useful to constrain the reconstructed particles to a volume within that region. The constraint is implemented by considering the initial e^+e^- collision as an abstract parent particle with the parameters and uncertainties of the beam spot, thus constraining the *production* vertex position of its children to an area within the beam spot's three dimensional uncertainty region using

$$\boldsymbol{r}^{\alpha}(\boldsymbol{x}) = \boldsymbol{s} - \boldsymbol{h}_{\text{production}} + \boldsymbol{H}^{\alpha-1} \cdot (\boldsymbol{x}^{\alpha}_{k-1} - \boldsymbol{x}^{\alpha-1}), \tag{37}$$

with the Jacobian

$$\boldsymbol{H} = \begin{pmatrix} \ldots & & \ldots \\ \ldots & -\mathbb{1}_3 & \ldots \\ \ldots & & \ldots \end{pmatrix}. \tag{38}$$

Here $\boldsymbol{s}$ denotes the beam spot position vector and $\boldsymbol{h}_{\text{production}}$ is the fitted production vertex of the particle. The beam spot is typically determined by averaging the x, y and z positions of the e^+e^- collisions over many interactions. SuperKEKB's nano-beam scheme will give x, y and z collision vertex distributions of $\sigma^*_x = 11$ μm, $\sigma^*_y = 0.062$ μm, and $\sigma^*_z = 500$ μm respectively. This will imply powerful *production* vertex constraints for physics analyses.

3.1.8. Custom origin constraint

Similar to the beam spot constraint, one can construct another geometric constraint by defining a custom vertex position and an associated uncertainty to be the origin of the decay chain. This can be very useful when the decay chain contains particles that cannot be detected, for example, neutrinos or long lived dark matter particles. In these decay chains it is not useful to fit the full chain, as the missing particle's four momentum would lead to an incorrect assumption for the kinematic constraint of the parent particle. However, knowing that the particle originates from a B-meson decay, one can define a geometric constraint corresponding to the volume where B-mesons decay on average. A beam energy constraint is not necessarily needed, as one can mass constrain the $\Upsilon(4S)$ particle, since all four-vectors in $e^+e^- \to \Upsilon(4S)$ are well known.

4. Applications of the fitter in Belle II

In this section we present use cases of the algorithm within the Belle II experiment. Of special interest are decay chains containing one and two π^0-mesons, as well as D^{*+}-mesons. There are numerous channels where the phase space is large enough to add a π^0 to a vertex with two charged tracks, which makes this a very common structure in decay trees and therefore an interesting target to fit in a wide spectrum of analyses. Many decay channels of the B-meson contain D^{*+}-mesons, hence improving the reconstruction of these is a major goal. We use Monte Carlo samples from the Belle II experiment and evaluate background rejection and signal resolution improvements, as

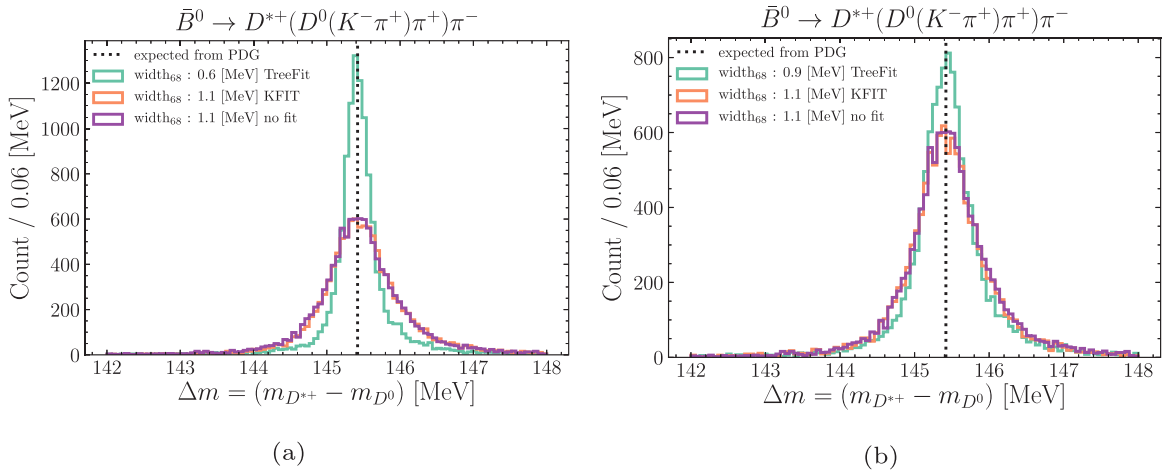

Fig. 7. Resolution comparison of the variable Δm for TreeFit (a) with and (b) without geometrically constraint D^0-meson and B^0-meson to KFIT.

well as the extraction of various other parameters and their resolutions. Background can arise from non-signal processes, and from incorrect combinations of particles used in building the tree. The selection criteria used for all analyses in this section are listed in Table 2. We use implicit charge conjugation, meaning that charged particles appearing in the decay chains imply the inclusion of opposite-sign charges. We abbreviate decays $x \to yz$ as $x(yz)$. Because some distributions are not Gaussian we use the inter quantile width containing 68% of the distribution as a measure to compare them. It is defined as $\text{width}_{68} = q(84\%) - q(16\%)$.

4.1. Comparing the global fit to staged fits — the effect of the geometric constraint

We compare the described method to the standard staged fit method, KFIT [10], that was employed by the Belle experiment. KFIT performs stage wise fits, fitting the children particles of each vertex in a separate fit, so that only the measurements of the children particles are accounted for. This is done in a global fit to the 4-vectors and vertex position in cartesian space. In the case of the decay $D^{*+} \to D^0\pi^+$ in the decay chain $\bar{B}^0 \to D^{*+}(D^0(K^-\pi^+)\pi^+)\pi^-$, the first decay vertex is fitted using $D^0 \to K^-\pi^+$ the resulting D^0-meson is then fitted together with a pion track to form a D^{*+}-meson. This means that the decay vertex of the D^{*+}-meson is only constrained by the measurement of the π^+-track and the direction of the D^0-meson. The π^+-tracks of D^{*+}-meson decays have in the environment of the Belle II experiment usually low transversal momenta and thus are more likely to have large measurement uncertainties. In the global fit the final state particles are constrained via their respective constraints, tracks in this case. The four-vectors of intermediate particles are built using kinematic constraints. These, together with the decay vertex of the D^0-meson and the decay vertex of the D^{*+} are optimised at the same time, via a geometric constraint. As a result the resolution on the correlated quantity $\Delta m = m(D^{*+}) - m(D^0)$ can be improved by about 30% to a width of 0.57 MeV, depicted in Fig. 7a. If this constraint is lifted the improvement is diminished, see Fig. 7b. However, the individual resolutions on the masses of the D^{*+}-meson and D^0-meson are not improved. The improvement can be observed in semi-leptonic decay chains such as $\bar{B}^0 \to D^{*+}(D^0(K^-\pi^+)\pi^+)\mu^-\bar{\nu}_\mu$ displayed in Fig. 8. However, in this case the resulting kinematic parameters of the B^0-meson will be strongly biased due to the missing four-vector of the neutrino.

4.2. Fitting decay chains containing neutral particles

A high rate mode with a single π^0, is $\bar{B}^0 \to D^{*+}(D^0(K^-\pi^+\pi^0(\gamma\gamma))\pi^+)\pi^-$. This channel is of interest as it is the highest branching ratio D^0

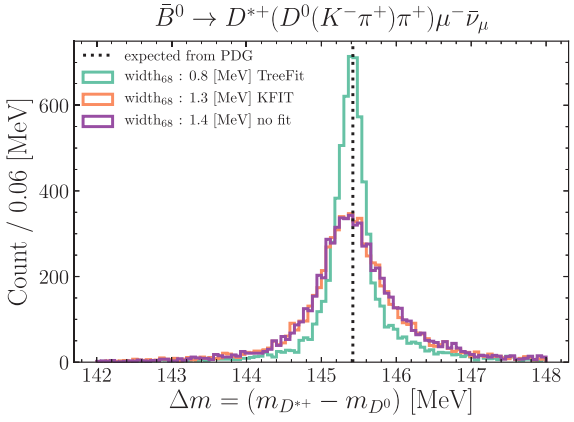

Fig. 8. Comparison of the Δm-resolution in a semi-leptonic decay chain for different fitters. The resolution is worse than in a kinematically fully constrained decay chain, Fig. 7a, but the relative improvement compared to KFIT is similar.

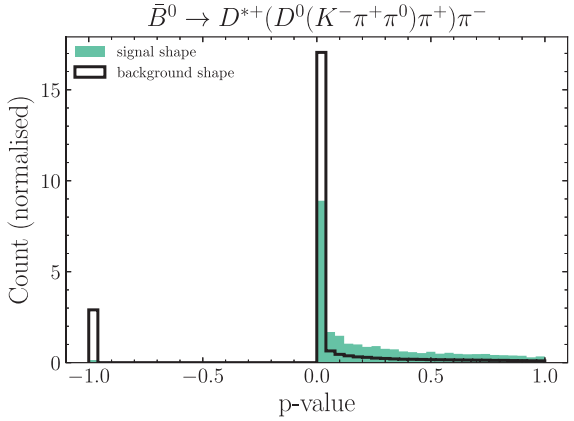

Fig. 9. P-value distributions of fits to $\bar{B}^0 \to D^{*+}(D^0(K^-\pi^+\pi^0)\pi^+)\pi^-$. Failed fits were assigned a p-value of -1. The shape of the distribution is non-flat due to the use of a mass constraint on the π^0. The preselection criteria are listed in Table 2.

decay mode. We use the fitter in this example to suppress background. If we reject combinations for which the fit failed – i.e. those with a p-value ≤ 0 – we are able to reject about 15% of the background while rejecting only about 1% signal, as displayed in Fig. 9. If a fit fails, a p-value of -1 is assigned for display purposes only. Note that since some

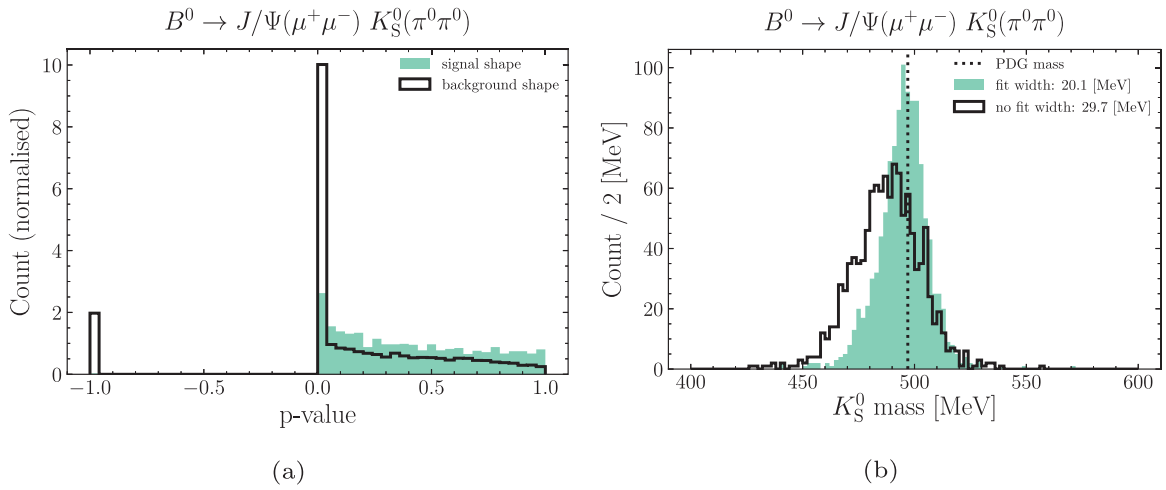

Fig. 10. (a) P-value of the fits to $B^0 \to J/\psi K^0_S$. (b) Fitted mass of the K^0_S (green) and the mass before the fit (black). The mass distribution is centred around the true value after performing the fits with a π^0 mass constraint. Qualitatively this is the same result as obtained in Ref. [1]. The bias was removed and the width reduced by 9.6 MeV.

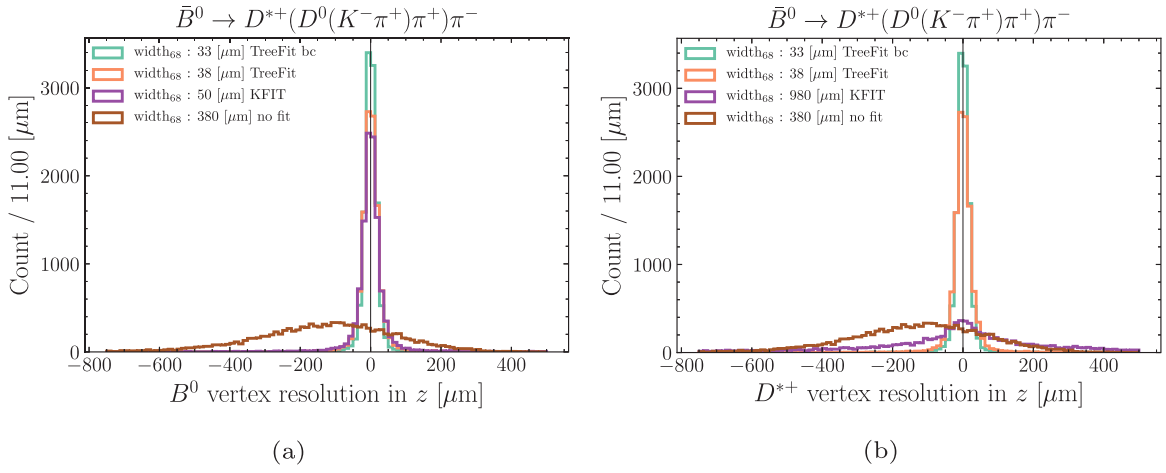

Fig. 11. (a) Resolution of the z coordinate of B-mesons with and without applying a beam constraint. (b) Resolution of the z coordinate of D^{*+}-mesons with and without applying a beam constraint. The fits using a beam constraint are indicated as TreeFit bc. Over- and underflow bins are not depicted.

of the constraints are nonlinear and the uncertainties are not exactly Gaussian in some cases, p-values can have non-flat shapes.

4.3. Fitting decay vertices of neutral particles

We will next study the decay chain $B^0 \to K^0_S(\pi^0(\gamma\gamma)\pi^0(\gamma\gamma))J/\psi(\mu^+\mu^-)$. By performing the fit a large amount of the background can be removed by requiring that it passes the fit, see Fig. 10a. The only well defined vertex in this chain is given by the $J/\psi \to \mu^+\mu^-$ decay. The other vertices are very uncertain due to the absence of charged tracks. The best assumption that can be made for the production vertex position of the four photons is that they originate from the interaction point, if they are fitted in a single stage fit oblivious of the J/ψ. Performing a fit with mass constrained π^0-mesons improves the extracted mass of the K^0_S, so that after the fit, it is centred around the true value, as depicted in Fig. 10b. It is then possible to further reject background outside the nominal mass window and improve the signal purity when analysing this channel.

4.4. Using a beam spot constraint to improve the decay vertex resolution of b-mesons

A beam constraint enforces the production vertex of the B^0-meson to lie within the beam spots volume, see Section 3.1.7. It can be used to improve the decay vertex resolution of B-mesons and all children particles in the decay chain, depicted in Figs. 11a, 11b and 12. The resolution for B-mesons was improved by 15% to a value of about 33 µm in the example of the decay chain $\bar{B}^0 \to D^{*+}(D^0(K^-\pi^+)\pi^+)\pi^-$. In the no fit case 0.0.0 is taken to be the vertex and no error is assigned. In terms of the defined resolution this can be better than the fit using KFIT in the D^{*+} case, because the pion involved in the decay has a low momentum, typically of around 150 MeV. These have measurement with a very large uncertainties and a fit involving only the D^0 and the pion then is dominated by this uncertainty. The very long tails for KFIT are due to very forward or very backward tracks. When fitting the entire decay chain with TreeFit, the tracks with better measurements help to constrain the vertex via geometric constraints. The efficiency of both algorithms is comparable.

4.5. Extracting the decay length of D^0-mesons from D^{*+} decays using a geometric constraint

The geometric constraint, see 3.1.5, constrains production and decay vertices of long lived particles in the decay chain. This allows for the extraction of flight lengths and thus lifetimes of intermediate particles such as D^0-mesons. We perform this study on $\bar{B}^0 \to D^{*+}(D^0(K^-\pi^+)\pi^+)\pi^-$ decays and extract the decay length of the D^0-meson. The results are depicted in Fig. 13. Since D^{*+}-mesons decay

Table 2
Selection criteria used in the analyses in this section. All particles use the same criteria if appearing in the decay chains. The invariant mass is obtained by summing the particle's daughter four-momenta before performing the fit. The particle ID for K^+/π^+ is defined as the likelihood ratio $\mathcal{LR}(K^+(\pi^+)) = \mathcal{L}(K^+(\pi^+))/(\mathcal{L}(\pi^+) + \mathcal{L}(K^+))$ and the beam energy constrained mass m_{bc} [11].

Particle	Preselection applied
γ	$E_\gamma > 0.075$ GeV
π^0	$0.145\ \mathrm{GeV} > m(\gamma\gamma) > 0.125\ \mathrm{GeV}$
π^+	$\mathcal{LR}(\pi^+) > 0.5$
K^+	$\mathcal{LR}(K^+) > 0.5$
D^0	$1.9\ \mathrm{GeV} > m(K^-\pi^+\pi^0) > 1.7\ \mathrm{GeV}$
D^*	$\Delta m = m(D^*) - m(D^0) < 0.155$ GeV
B^0	$m_{\mathrm{bc}} = \sqrt{E^2_{\mathrm{beam}}/4 - p^2} > 5.27$ GeV

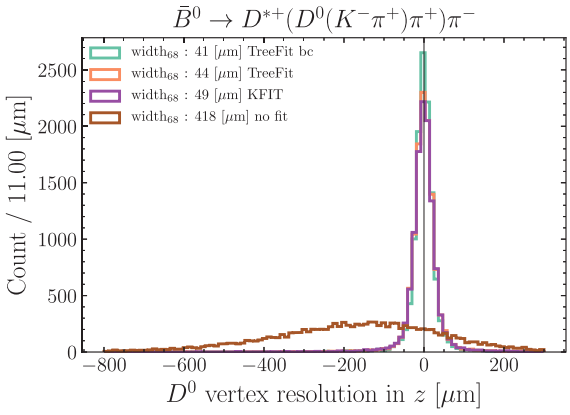

Fig. 12. Resolution of the z coordinate of D-mesons with and without applying a beam constraint. The fits using a beam constraint are indicated as TreeFit bc. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

almost instantly, we will use the three dimensional distance between the B^0 and the D^0 decay vertices. To improve the resolution on the B^0 vertex we apply a beam spot constraint.

5. Performance

Vertex fitting is computationally the most expensive operation during physics analysis. This is due to the large number of possible combinations that arise when reconstructing complex decay topologies. Physics analyses in the Belle II experiment are performed on a global grid of computing clusters [12] occupying thousands of CPUs simultaneously. Reducing the execution time of these analyses is a major effort of the collaboration. To achieve this we use the EIGEN library for matrix operations. EIGEN ensures memory locality is respected and operations are vectorised when possible. Modern CPUs are much faster in performing operations than retrieving memory. Allocating memory used locally in time and space allows modern CPUs to guess and prefetch the memory that is used next, such that no clock cycles are wasted [13]. The algorithm based on EIGEN was profiled using the Valgrind [14] module Callgrind. We find that one iteration of the algorithm spends 69% of the time on the implementation of Eq. (10), 27% on Eq. (9) and 4% on other operations, when fitting the decay chain $B^+ \to D^0(K^-\pi^+)\pi^+$. The performance using EIGEN was compared to a previous implementation using the CLHEP library [15], a C++library widely used for algebraic operations in particle physics. By using EIGEN we could speed up the execution time by a factor two in simple cases and more than an order of magnitude in more complex topologies, see Fig. 14. These numbers were obtained using EIGEN-3.3.4 and CLHEP-2.2.0.4 and compiled with clang-4.0 using the $-O3$ flag for full optimisation including vectorisation.

In order to assess the scaling with increasing matrix sizes, we compiled a stand-alone programme to directly compare the libraries. We investigate n times m matrices, M, and symmetrical matrices, S. For the latter we used symmetric matrix types in both libraries and. Optimised matrix operation implementations provided by the libraries are used where possible. We fix one matrix dimension to the largest occurring number in the algorithm, which is $m = 5$ for track constraints. Then we increase the other dimension n, which is equivalent to adding more particles to the decay chain. The computation of Eq. (10) is repeated 500 times and the median is taken as a data point. The process is repeated 100 times, each time with randomised initial matrices and the width of the distribution is taken as the uncertainty. We observe much stronger exponential scaling of the operations involving the CLHEP library, depicted in Fig. 15. Since the version of CLHEP used in the algorithm is outdated, we additionally compare to a more recent version of the library. We do not find a change in performance for the operations investigated. We used the compiler flags $-O3$ and $-march = native$ when building our programme to enable all possible optimisations the hardware allows. We find that omitting these flags drastically worsens the performance of EIGEN, while CLHEP has to be compiled prior to usage and therefore is oblivious to these options. We measured the cache-miss rate to be close to zero, which persists if we specify an upper limit for the size of the matrices in EIGEN. From this

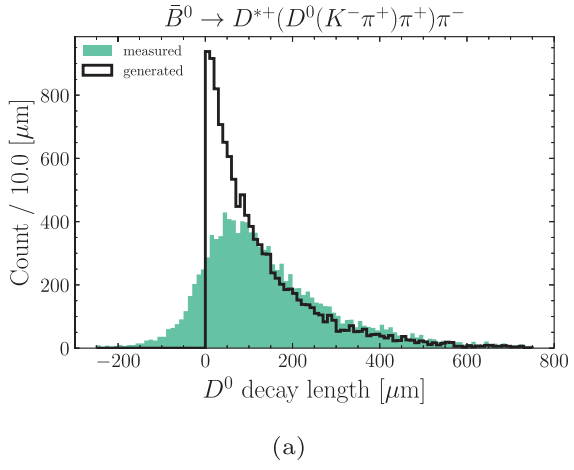

Fig. 13. (a) The extracted decay length (green) and the generated decay length (black). The negative tail of the extracted decay length is due to the detectors resolution function. (b) Pull (measured – –generated/uncertainty) of the decay length distribution. The shape of a Gaussian with a mean of zero and a RMS of one is plotted for reference. The shape of the pull matches with a Gaussian, apart from a small mismatch in the central region. This mismatch can be attributed to an underestimation of the tracking parameter uncertainties, where a similar mismatch was observed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

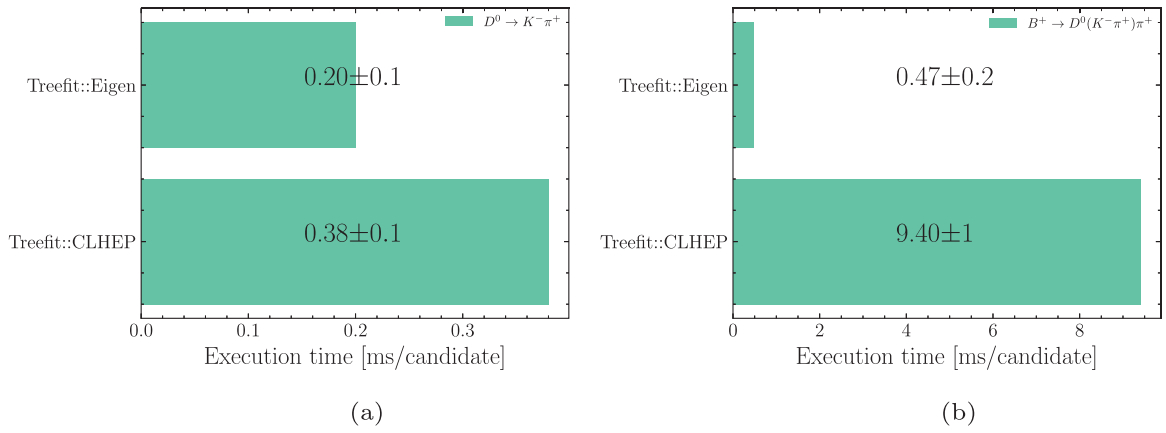

Fig. 14. (a) Average execution time per candidate fitting two tracks. (b) Average execution time per candidate fitting a topology with an intermediate decay. The errors are estimations from repeating the study 10 times.

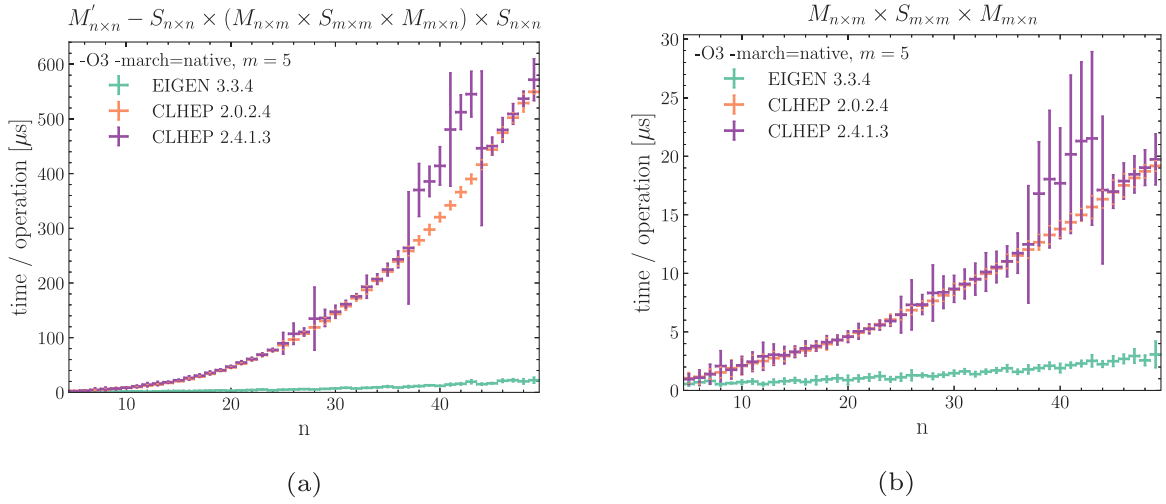

Fig. 15. (a) Execution time scaling of Eq. (10). For increasing matrix sizes n. (b) Execution time scaling for the inner matrix product of Eq. (10). The notation S indicates a symmetrical matrix, M a non symmetrical, the dimension m was fixed to $m = 5$, which is equivalent to Eq. (10) being calculated for a track constraint.

we conclude that our matrices are mostly heap allocated and the main performance gain is driven by vectorisation.

6. Conclusion

We presented an improved implementation of a global vertex fitting tool, based on Ref. [1], tailored for the environment of the Belle II experiment. It can be used for various purposes, such as the extraction of particle production and decay vertices, decay lengths, particle four-momenta and rejection of backgrounds, as well as the extraction of the respective uncertainties. The global fitting technique is particularly powerful in fitting and reducing background in modes that contain neutral particles and can significantly improve the resolution on Δm an important variable for the reconstruction of D^*-mesons, which will be very important for the Belle II physics analysis programme. We were able to reduce the execution time of the algorithm by an order of magnitude, by restricting the maximum possible size of dynamic matrices and enabling vectorised matrix operations.

CRediT authorship contribution statement

F. Tenchini: Software, Writing - review & editing. **P. Urquijo:** Supervision, Conceptualization, Writing - review & editing. **F. Abudinén:** Software, Writing - review & editing. **S. Cunliffe:** Software, Writing - review & editing. **T. Ferber:** Software, Writing - review & editing. **M. Gelb:** Software, Writing - review & editing. **J. Gemmler:** Software, Writing - review & editing. **P. Goldenzweig:** Software, Writing - review & editing. **T. Keck:** Software, Writing - review & editing. **I. Komarov:** Software, Writing - review & editing. **T. Kuhr:** Software, Writing - review & editing. **L. Ligioi:** Software, Writing - review & editing. **M. Lubej:** Software, Writing - review & editing. **F. Meier:** Software, Writing - review & editing. **F. Metzner:** Software, Writing - review & editing. **C. Pulvermacher:** Software, Writing - review & editing. **M. Ritter:** Software, Writing - review & editing. **U. Tamponi:** Software, Writing - review & editing. **A. Zupanc:** Software, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was made possible by the Belle II collaboration and funding from ARC (Australia), ARRS (Slovenia), BMBF (Germany), EXC153 (Germany), HGF (Germany), INFN (Italy), MSMT (Czech Republic), NSERC (Canada) and U.S. DOE. We would like to thank Wouter Hulsbergen for his input.

References

[1] W.D. Hulsbergen, Decay chain fitting with a Kalman filter, Nucl. Instrum. Methods A552 (2005) 566–575, http://dx.doi.org/10.1016/j.nima.2005.06.078, arXiv:physics/0503191.

[2] G. Guennebaud, B. Jacob, et al., Eigen v3, URL http://eigen.tuxfamily.org.

[3] The Belle II collaboration, Belle II Technical Design Report, 2010, arXiv:1011.0352.

[4] R.E. Kalman, A new approach to linear filtering and prediction problems, Trans. ASME–J. Basic Eng. 82 (Series D) (1960) 35–45.

[5] R. Frühwirth, Application of kalman filtering to track and vertex fitting, Nucl. Instrum. Methods Phys. Res. A 262 (HEPHY-PUB-503) (1987) 444, 19 p. URL https://cds.cern.ch/record/178627.

[6] R. Frühwirth, R.K. Bock, Data analysis techniques for high-energy physics experiments, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. 11 (2000) 252.

[7] M.e.a. Tanabashi, Particle Data Group Collaboration Collaboration, Review of particle physics, Phys. Rev. D 98 (2018) 030001, http://dx.doi.org/10.1103/PhysRevD.98.030001, URL https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[8] N. Dawe, GitHub repository, URL https://github.com/ndawe/tikz-track.

[9] V. Karimaki, Effective circle fitting for particle trajectories, Nucl. Instrum. Methods A305 (1991) 187–191, http://dx.doi.org/10.1016/0168-9002(91)90533-V.

[10] P. Avery, Applied Fitting Theory I, IV, VI, 1999.

[11] E. Kou, et al., The belle II physics book, 2018, URL arXiv:1808.10567.

[12] H. Takanori, and, Computing at the Belle II experiment, J. Phys. Conf. Ser. 664 (1) (2015) 012002, http://dx.doi.org/10.1088/1742-6596/664/1/012002.

[13] M. Kowarschik, C. Weiß, An overview of cache optimization techniques and cache-aware numerical algorithms, in: Algorithms for Memory Hierarchies — Advanced Lectures, in: Lecture Notes in Computer Science, vol. 2625, Springer, 2003, pp. 213–232.

[14] N. Nethercote, J. Seward, Valgrind: A framework for heavyweight dynamic binary instrumentation, SIGPLAN Not. 42 (6) (2007) 89–100, http://dx.doi.org/10.1145/1273442.1250746, URL http://doi.acm.org/10.1145/1273442.1250746.

[15] L. Lonnblad, CLHEP: A project for designing a c++ class library for high-energy physics, Comput. Phys. Comm. 84 (1994) 307–316, http://dx.doi.org/10.1016/0010-4655(94)90217-8.