

The Neuro-Z-Vertex Trigger of the Belle II Experiment

Sebastian Skambraks^{1,a}, Sara Neuhaus¹, and Christian Kiesling²

¹Technische Universität München, Germany

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

Abstract. A neural network z vertex trigger is planned for the upcoming Belle II detector at the SuperKEKB collider. This neural algorithm is based on a single track 3D parameter estimation using only hit and drift time information from the central drift chamber. Due to the high luminosity ($\mathcal{L} = 8 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$) Belle II will have to face high levels of beam induced background, making a z vertex reconstruction at the first level trigger mandatory. Using the neural z vertex algorithm, the requirements of the standard track trigger can be strongly relaxed. By this, the trigger efficiencies, especially for low multiplicity events, e.g. τ pair production, can be significantly increased.

This contribution presents the foreseen neural network trigger setup and the preceding 2D track finder. Special focus is put on the proposal and evaluation of a possible 3D upgrade of the 2D track finder. Additionally, details are given on a dedicated setup for the upcoming cosmic ray test.

1 Introduction

The Belle II experiment [1] is an upgrade of the successful Belle [2] flavor physics experiment. It is located in Tsukuba, Japan at the upgraded asymmetric $e^+ e^-$ collider SuperKEKB. With beam energies of 4 GeV for e^+ and 7 GeV for e^- the collider runs at the center of mass energy of the $\Upsilon(4S)$ resonance. The design luminosity of SuperKEKB is $\mathcal{L} = 8 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ which is 40 times larger than the world record luminosity reached by its predecessor KEKB. Typical $\Upsilon(4S)$ events will have an average track transverse momentum (p_T) of 500 MeV and an average multiplicity of 11 tracks per event.

Belle II consists of a high precision pixel detector (PXD) as innermost detector, followed by the silicon vertex detector (SVD). Due to the latency and data bandwidth, these high precision measurements are only available for offline reconstruction. The third detector shell is the central drift chamber (CDC), a wire chamber with 56 layers of sense wires. The CDC is the main tracking device of Belle II. With its high speed readout it is perfectly suited for the first trigger level and it provides the input to the z vertex trigger. The outer detector shells comprise calorimeters (ECL), Cherenkov counters (A-RICH), time-of-propagation counters (TOP) and a kaon and muon detector (KLM). More details on the Belle II experiment can be found in [1].

^ae-mail: sebastian.skambraks@tum.de

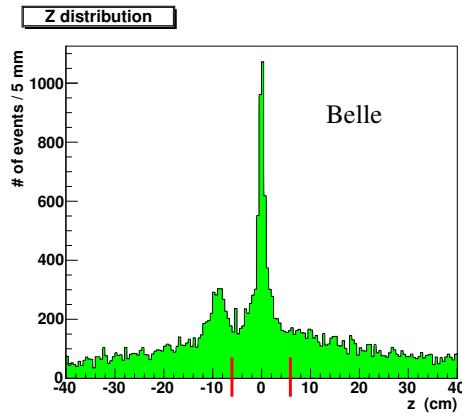


Figure 1. z distribution of random triggered events in Belle [1]. A z vertex trigger with an accuracy of $\sigma_z = 2$ cm would enable a cut at the red marks ($z = \pm 6$ cm) to increase the signal to background ratio.

1.1 Background in Belle II

With the luminosity upgrade, the level of beam background will increase in Belle II. Due to the high beam collimation (nano-beam-scheme [1]), the Touschek effect [3] will become a dangerous new source for background tracks. Other dominant sources of beam background are radiative Bhabha scattering, beam–gas interactions and synchrotron radiation. Typically, background interacts with the beam pipe which results in tracks with displaced z vertices.

Here, the coordinate system to describe the vertex position has its origin at the interaction region, where the e^+ and e^- bunches are brought to collision. The z -axis is oriented along the beam line and the magnetic field points in the z direction. Charged tracks are described as a helix ($p_T, \varphi, \vartheta, z, d$), where the track-vertex is located at z with a displacement d to the z -axis (position of closest approach). p_T is the transverse momentum, φ the azimuthal angle, and ϑ the polar angle of the track momentum at the vertex position. Displacements along the z -axis can be large compared to the radius of the beampipe ($r \approx 1$ cm). Therefore, the displacement d are neglected at the z vertex trigger level.

The background situation in the predecessor Belle, which had no z vertex trigger, is shown in Figure 1. Only the peak around $z = 0$ cm corresponds to the interesting physics signal, and a z vertex trigger would allow to suppress the background tails. With the anticipated accuracy of $\sigma_z = 2$ cm, a cut at ± 6 cm (red marks in Figure 1) is possible while maintaining a high efficiency for the physics signal. In Belle II, a z vertex trigger at the first trigger level becomes necessary, to provide a high physics signal rate in spite of the increased background level.

1.2 τ events

Events with a low track multiplicity are most in danger to be mistaken as background. Here a z vertex trigger can significantly improve the efficiency by relaxing the other trigger conditions and otherwise unseen events can be recovered. τ events have a large cross section in Belle II and are a typical candidate for low multiplicity events. A simulation study shows the possible improvement for τ decays

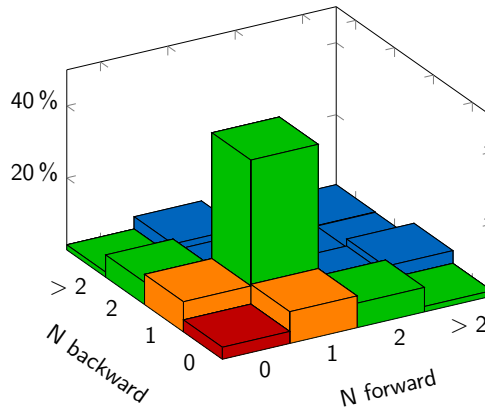


Figure 2. Simulation study of track numbers in hemispheres with τ events as an example for expected low multiplicity events. With a z vertex trigger the otherwise lost green columns can be recovered. This results in a potential efficiency gain by a factor of ≈ 3.9 .

in Belle II for a trigger setup with a perfect z vertex trigger compared to a trigger setup without a z trigger. Without the z trigger at least 3 tracks are required with at least one track in each hemisphere. With a z vertex trigger, only two tracks are required with the additional condition $|z| < 6$ cm. Figure 2 shows the track counting in the forward and backward hemispheres of the Belle II detector in a pure simulation study. Without the z vertex trigger only the events in the blue columns can be triggered. With the z vertex trigger all the green columns can additionally be recovered. This corresponds to a factor of ≈ 3.9 for the total potential efficiency increase.

1.3 CDC first level trigger system

To fulfill the main requirements [1] of the first level trigger (30 kHz output rate, 5 μ s latency, 200 ns event separation) only detector components with a fast readout time can be used. This leaves the CDC as the only tracking and vertexing device at the first trigger level.

A schematic of one half of the CDC is shown in Figure 3. The CDC has 14336 sense wires in 56 layers which are combined into nine superlayers (SL) of wires with the same orientation. There are five axial SLs, oriented parallel to the beam line and the magnetic field (z -axis), and four inclined stereo SLs. The different stereo angles with respect to the z -axis make a 3D reconstruction possible.

The following methods will be integrated in the pipeline of the CDC sub-trigger system. After digitization of the sense wire information of the central drift chamber, the first component is the track segment finder (TSF). It compresses the wire hit information and removes noise by combining the wire layers within the SLs to a total of 2336 track segments (TS). The TSF selects TSs as predefined patterns of wires within a SL (see Figure 3): if at least one wire in four out of the five layers within the TS pattern have a hit, then the TS is hit. The TSF outputs the id of the TS, a priority wire within the TS, and the drift time with respect to the priority wire. Additionally, it provides left/right information based on the wire hit pattern within the TS.

The TSF is followed by the track finder which is based on a Hough transformation [4]. The currently foreseen 2D track finder uses only inputs from the axial SLs. An extended 3D track finder is described in Section 3. The 2D finder provides 2D track parameter estimates (p_T , φ) for all tracks found.

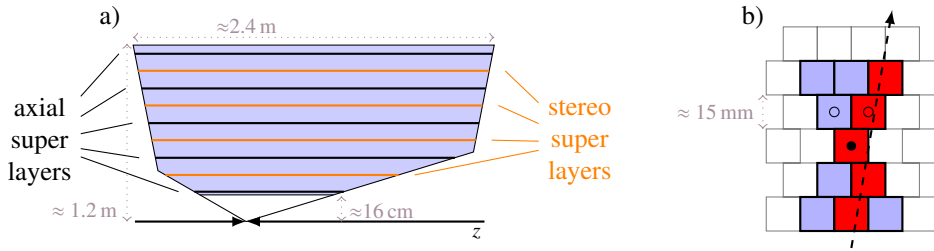


Figure 3. a) The superlayer (SL) structure of the Belle II central drift chamber (CDC). SLs with axial orientation (parallel to the z -axis) alternate with inclined stereo SLs. b) The shape of one track segment (TS), combining 6 wire layers within one SL, is shown as a colored region. The wire cells hit by the track are colored red. A track leaving a hit in at least four out of the five wire layers within the track segment shape is recognized as a TS hit. The TS hit consists of the TS id, the drift time at the priority wire and a left/right information with respect to the priority wire. The central wire has the first priority; if it is missing, one of the two wires above it is used.

All TSs and the 2D track estimates are then used as input to the 3D track reconstruction (ϑ, z). This includes the neural network, which is stable even in the presence of noise and nonlinearities occurring under realistic conditions, for example a nonlinear drift time to distance relation.

Finally, the results of the CDC sub-trigger system are fed into the global decision logic of the first level trigger.

2 Neural network for track parameter estimation

The general ideas for the neural network track parameter estimation in the Belle II experiment were developed in [5–8]. More details on the setup of the single networks, the input transformation and the specialization of networks to phase space sectors can be found there.

2.1 Network setup and training

Neural networks are powerful machine learning tools, that can be used for general function approximation [9]. In the feed-forward setup used here [8], the network has the structure of a directed, acyclic graph, where each node computes the weighted sum over its inputs and evaluates this with an activation function (hyperbolic tangent). This structure is well suited for an implementation in parallel hardware and has a deterministic runtime. Here, the neural network is used for single track z vertex estimation (see Figure 4). A preprocessing step is required to select the associated hits for each track and transform the raw hit information to the input values for the network.

The neural network receives one TS per SL per track as input. One TS is represented by two values for the position and one value for the drift time. To match the possible values of the hyperbolic tangent, all input and output values are scaled to the interval $[-1, 1]$. The floating point output value is then directly interpreted as the scaled z vertex value of the single track. The training is carried out with the iRPROP⁻ [10] backpropagation algorithm on a simulated dataset. As soon as the detector data is available, a retraining with real tracks is planned, where the precise offline reconstruction values are used as training targets.

The required preprocessing for the neural network comprises the association of TS hits to tracks and the representation of the hits in appropriate coordinates. The TS selection is currently based

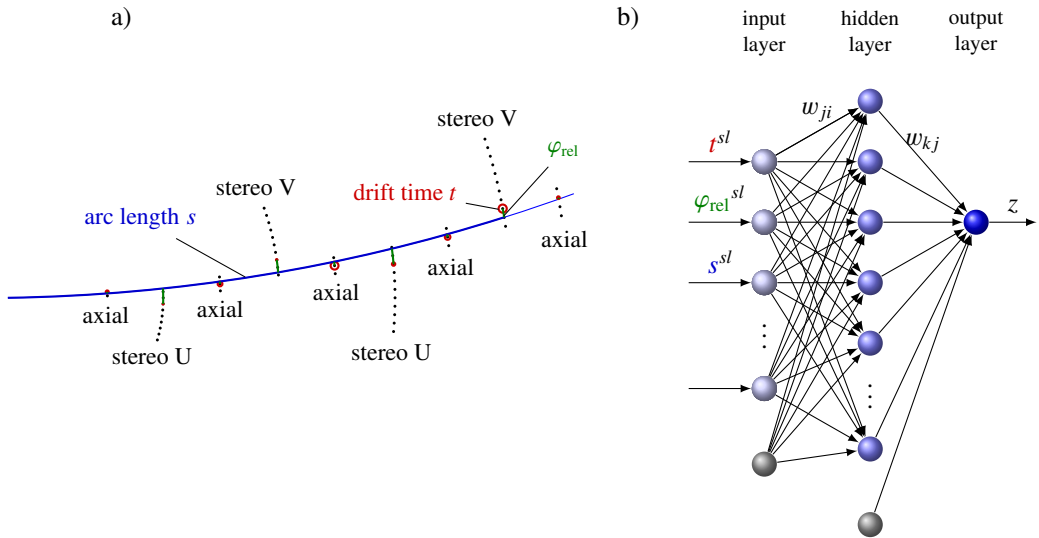


Figure 4. a) Input selection for the neural networks. Hits in small intervals around the found 2D tracks are selected and transformed to the input for the neural network. b) Structure of a neural network receiving 3 input values per superlayer (SL).

solely on the 2D track parameter estimates provided by the Hough finder. Small intervals around the 2D track estimates are defined for each SL in a pre-training run (see Figure 4). Due to the different stereo angles, the intervals of the stereo SLs are asymmetric and larger than the intervals of the axial SLs. If more than one hit is contained within the interval, the hit with the shortest drift time is used. After the hit selection, the hit positions are represented in a unique way by s and φ_{rel} , using the 2D information: the parameter s is the arc length along the 2D track to the hit, and φ_{rel} is the distance in φ of the 2D track to the hit (see Figure 4). The third component is the drift time t , where the left/right information of the TS is used as the sign. More details on the input representation can be found in [8].

To address asymmetries in ϑ , p_T and charge, the possibility for sectorization of the phase space is introduced. Sectorization in p_T and charge is possible with the present 2D finder alone, whereas ϑ sectorization is only possible if an additional ϑ estimate is introduced. One option for ϑ estimation is the 3D track finder described in Section 3. In [7] it was shown that many “expert” neural networks, specialized to small sectors of the phase space, help to improve the z estimation accuracy. Because sectorization costs memory resources and introduces a sensitive dependency on the sector selection accuracy, a small total number of sectors is desirable.

Figure 5 shows the z estimation capabilities of a basic setup with only two neural networks for different charges with a 90% RMS as resolution measure. For the realistic events with noise, the $O(2\text{ cm})$ resolution is achieved in the high p_T region, but further improvements are necessary in the low p_T region.

2.2 Cosmic ray setup

The first practical test of the z vertex trigger is the upcoming CDC cosmic ray test. Only a part of the CDC is used (two sectors with opposite φ), and there is no magnetic field. Cosmic tracks passing

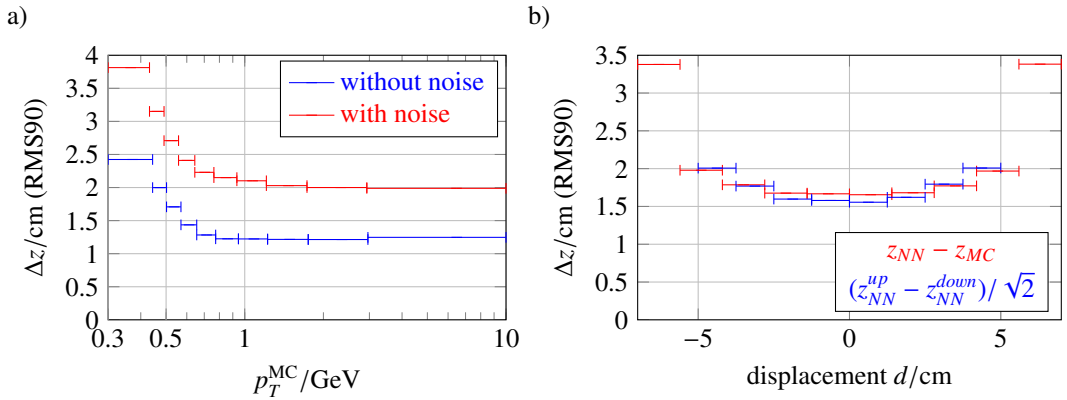


Figure 5. a) z vertex resolution using two neural networks for different charges on ideal noiseless events and realistic events with noise. b) z resolution in the cosmic ray setup. The red line shows single track z vertices vs. simulated values. The blue line shows a comparison of the apparent two track z vertices. The displacement d is the distance of closest approach to the z axis.

close to the z -axis will usually leave hits in both the upper and the lower half of the CDC and should be recognized as two tracks by the 2D finder. Due to the larger distance of closest approach to the z -axis (d) and the 2D finder constraint for tracks to pass through the z axis, displaced tracks will be recognized as curved tracks. Dedicated neural networks are trained with the altered event topology and detector setup. As reference value, the tracks from the upper CDC half are compared to the tracks in the lower CDC half. Figure 5 shows the z vertex resolution of simulated cosmic ray tracks vs. the displacement to the z axis. It shows that the neural network can easily adapt to the altered conditions.

3 3D track finding

A track finding algorithm is crucial to enable the application of the neural networks to single tracks. It uses the CDC hits as input to find the initial track estimates and relate them to the hits. The correct hit selection is essential for the z estimation accuracy, but difficult in the presence of noise. The track estimates are used to calculate the neural network inputs and to enable phase space sectors with “expert” neural networks. The 2D finder in the foreseen setup [1, 11] only uses the axial hits as input and finds 2D track estimates (p_T, φ). In order to allow the z vertex estimation with the neural network, the stereo hits subsequently need to be related to the 2D estimates.

The recently developed 3D track finder is able to use stereo hits and axial hits simultaneously in order to find the 3D track estimates (p_T, φ, ϑ) and their hit relations. The early use of the additional stereo hit information helps to improve the track finding efficiency, the correctness of the stereo hit selection and the accuracy of the track parameter estimates. The sectors around the track estimates used to select the neural network input (one hit per SL per track, see Subsection 2.1) become significantly smaller for the stereo hits. Additionally, a phase space sectorization in ϑ becomes possible. The method is related to the phase space sectorization developed in [5, 6].

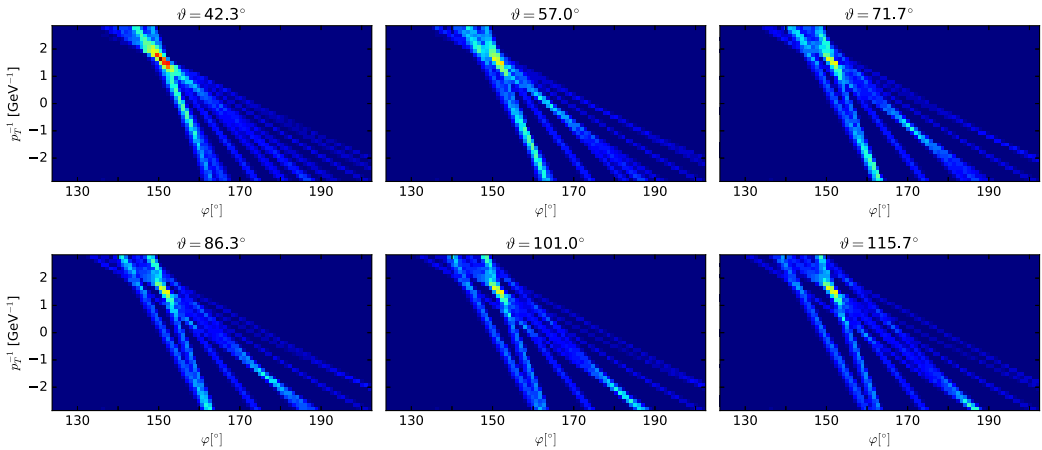


Figure 6. 3D track parameter space for a simulated single muon track with track parameters $p_T^{-1} = 1.62 \text{ GeV}^{-1}$, $\varphi = 151^\circ$, $\vartheta = 38^\circ$ after summing up the hits. For 6 different ϑ bins a 2D track parameter plane in p_T^{-1}, φ is shown. Only close to the correct ϑ , the lines of the stereo hits intersect with the crossing point of the axial hits.

3.1 Track finding setup

The basic concept of the proposed 3D finder is the Bayesian principle of conditional probabilities. A model of the joint probability space of hits and tracks can be formulated as

$$P(T|H) = \frac{P(H|T) \times P(T)}{P(H)}$$

where $P(T|H)$ denotes the probability that a set T of tracks has caused the observed set H of hits. In the proposed approach, the right hand side of the equation is approximated by a multi-dimensional histogram of dimension $\dim(H) + \dim(T)$ that can be trained using simulated events. With this general formulation, independent of the hit and track model, the concrete hit and track parametrization can easily be adapted to varying conditions.

Similar to a Hough transformation [4], the track finding problem is transformed to a peak finding problem in the track parameter space. Peaks can be interpreted as local maxima of the probability distribution, i.e. the tracks most probable to have caused the observed hits. Instead of analytically calculating the single hit representations in track parameter space, they are obtained from a histogram using simulated tracks. A weight $w(t)$ for a specific track t in the track parameter space ($t \in T$), given a set H , is calculated as the sum over all single hit contributions $P(t|h)$ for this track t :

$$w(t) = \sum_{h \in H} P(t|h) \quad (1)$$

In the special case of using only axial hits and ignoring ϑ in the track parametrization, the results are equivalent to the analytically calculated Hough transformation implemented by the current 2D finder. For a specific value of ϑ , a 2D track parameter plane can be constructed from the stereo hits and used in the same way as the 2D plane from the axial hits.

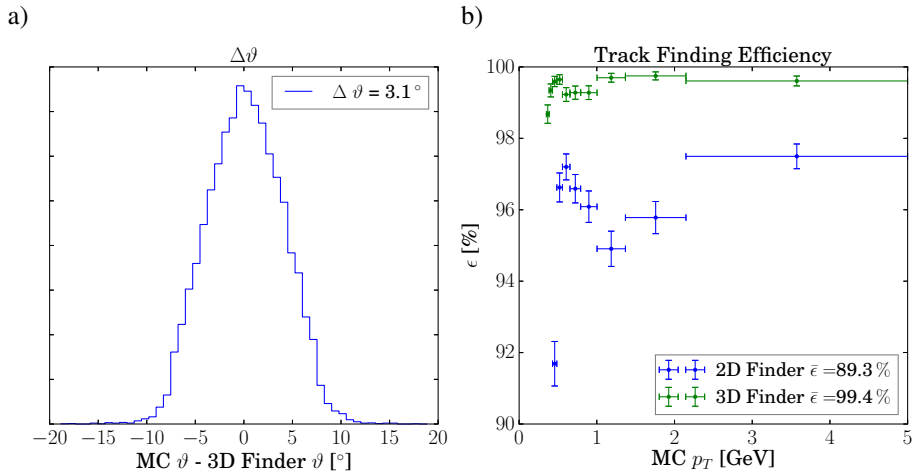


Figure 7. Results of the Bayesian 3D track finder. a) Accuracy of the ϑ estimation. b) Track finding efficiency compared to a 2D finder.

3.2 Training of the track finder

In the proposed 3D finder the distribution $P(t|h)$ is approximated by a 5D histogram representing a 3D track space (p_T , φ , ϑ) and a 2D hit space (TS-id and priority). With 2336 TS and 3 options for the priority, the hit space is already discrete. On the track space the following binning is applied: 384 bins in φ , 40 bins in p_T and 6 bins in ϑ . Drift times and the left/right information are currently unused, but they could easily be included in a future extension of this model. The training is carried out by filling the histogram, i.e. incrementing each (track, hit) pair for a sufficiently large set of training events. To finalize the training, the filled histogram is normalized such that each track is equally probable. During the normalization, the bit width of the histogram is limited to 3 bits.

3.3 Application of the track finder

A track finding and track parameter estimation can be initiated by looking up the 3D track spaces corresponding to each hit in an event and then summing them up (Equation (1)). An example of the summed 3D track space is shown in Figure 6. In the 6 ϑ bins a p_T^{-1} , φ plane is shown, containing one line for each SL. The lines of the five SL have the same intersection point in all ϑ bins, whereas the four stereo lines have the same intersection point only in the correct ϑ bin.

The next step is peak finding, which can be done by a clustering algorithm. The current setup requires cluster members to have a weight of at least 90% of the peak weight. Once clusters in track space are found, the contributing hits can be selected. Only hits with a high weight contribution to a cluster are related to the found track. Finally, an estimate of the track parameters for a found cluster is calculated with its related hits as the weighted mean of the cluster in track space. By using the weighted mean, with a 3 bit weight contribution of each hit to each cluster cell, the track parameter resolution can become smaller than the bin width.

The achievable ϑ resolution (RMS 90%) and the track finding efficiency ϵ , with the proposed 3D track finder compared to a 2D finder, are shown in Figure 7. For better illustration only the

region with $\epsilon > 90\%$ is shown, whereas, for the 2D finder, the two points at the lowest p_T values are cut (values: $\epsilon = 48.5\%$ and 78.4%). The experiment was carried out with 10000 simulated two-muon events, with a uniform distribution in the track parameters p_T^{-1} , φ , ϑ and ranges $p_T \in [0.35, 5]$, $\varphi \in [0, 360]^\circ$, $\vartheta \in [35, 123]^\circ$. The good resolution in ϑ of $\Delta\vartheta \approx 3^\circ$ clearly enables the possibility to use 3D input information for the neural network. The efficiency plot illustrates the significant improvement of a 3D finder compared to a 2D finder. The low efficiency of the 2D finder is due to the requirement that hits are present in all five axial SLs. In particular, the dip around 1 GeV in Figure 7 is caused by an inefficiency of the TSF for certain crossing angles. By including the stereo hits, the 3D finder can also identify tracks with missing TS hits, reducing the effect of the TSF inefficiency and at the same time increasing the acceptance in p_T and ϑ .

Compared to the 2D track finder, the 3D track finder requires more hardware resources. It has to receive approximately twice the number of TS as input, the phase space has to be 6 times larger for the ϑ bins, and it requires 3 bits for each bin in $P(\vartheta|h)$. The possibilities for a hardware implementation will be evaluated by future studies. If the track finding in 3D is too exhaustive for the present hardware, the method is alternatively suited for the ϑ estimation only. The p_T , φ part of the 3D finder is analog to the present 2D finder, such that the ϑ estimation can be executed as a separate step after the 2D finder.

4 Conclusion

A z vertex trigger is essential for the Belle II experiment in order to cope with the huge beam background levels coming along with the luminosity upgrade. Specifically, the efficiency for low multiplicity decay channels can be significantly enhanced, e.g. by a factor of ≈ 4 for τ -lepton events. With its high accuracy and stability in the presence of noise, the presented neural network algorithm is well suited for the single track z vertex estimation. However, its accuracy is sensitive to the quality of the preprocessing. Hence, further optimization will mostly focus on the improvement of the track finder and the hit selection. The neural network setup is prepared for the first cosmic ray run, where the correctness of the hardware implementation will be evaluated.

A 3D track finder motivated by the Bayesian principle of conditional probabilities has been presented in a machine learning setup. Simulation studies have demonstrated a high accuracy in 3D track parameter estimation and hit selection. Compared to a 2D track finder, a significant improvement in the track finding efficiency has been achieved. Future studies will evaluate the hardware implementation of the proposed 3D track finder.

References

- [1] T. Abe et al., *Belle II Technical Design Report* (2010), <http://arxiv.org/abs/1011.0352>
- [2] A. Abashian et al., *Nucl. Inst. & Meth. A* **479**, 117 (2002)
- [3] C. Bernardini et al., *Phys. Rev. Lett.* **10**, 407 (1963)
- [4] P.V.C. Hough, *Machine Analysis Of Bubble Chamber Pictures*, in *HEACC, Proc. 2nd Int. Conf. on High-Energy Accelerators and Instrumentation* (1959), Vol. C590914, pp. 554–558, http://inspirehep.net/record/919922/files/HEACC59_598-602.pdf
- [5] S. Skambraks, Diplomarbeit, Institute for Informatics, Ludwig-Maximilians-Universität, München (2013), <https://publications.mppmu.mpg.de/2013/MPP-2013-339/FullText.pdf>
- [6] S. Skambraks, M.S. thesis, Faculty of Physics, Ludwig-Maximilians-Universität, München (2013), <https://publications.mppmu.mpg.de/2013/MPP-2013-377/FullText.pdf>
- [7] S. Skambraks et al., *IEEE Trans. Nucl. Sci.*, **62**, 1732 (2015)

- [8] S. Neuhaus et al., *J. Phys. Conf. Ser.* **608**, 012052 (2015)
- [9] R. Hecht-Nielsen, *Theory of the backpropagation neural network*, in *IJCNN, IEEE Int. Joint Conf. Neural Networks* (1989), Vol. 1, pp. 593–605
- [10] C. Igel, M. Hüsken, *Improving the Rprop Learning Algorithm*, in *Proc. 2nd Int. Symp. Neural Computation* (2000), pp. 115–121
- [11] Y. Iwasaki et al., *IEEE Trans. Nucl. Sci.* **58**, 1807 (2011)