Background Suppression with the Belle II Neural Network Trigger

Sebastian Skambraks, Sara Neuhaus, Christian Kiesling
Max Planck Institute for Physics, Munich, Germany

E-mail: sskambra@mpp.mpg.de

Abstract. Neural networks are going to be used in the pipelined first level trigger of the upgraded flavor physics experiment Belle II at the high luminosity B factory SuperKEKB in Tsukuba, Japan. An instantaneous luminosity of $\mathcal{L} = 8 \times 10^{35}\text{cm}^{-2}\text{s}^{-1}$ is anticipated, 40 times larger than the world record reached with the predecessor KEKB. Background tracks, with vertices displaced along the beamline (z-axis), are expected to be severely increased due to the high luminosity. Using input from the central drift chamber, the main tracking device of Belle II, the online neural network trigger provides 3D track reconstruction within the fixed latency of the first level trigger. In particular, the robust estimation of the z-vertices allows a significantly improved suppression of the machine background. Based on a Monte Carlo background simulation, the high event rate faced by the first level trigger is analyzed and the benefits of the neural network trigger are evaluated.

1. Introduction

Belle II [1] at SuperKEKB [2] is a high luminosity upgrade of Belle [3], which was successful in confirming the CKM mechanism as the main source of CP violation in the standard model. In asymmetric e^+e^- collisions at the $\Upsilon(4S)$ resonance (e^+ with 4 GeV and e^- with 7 GeV), Belle II will be used to search for new physics at the precision frontier. A key feature that will enable the high instantaneous luminosity of $\mathcal{L} = 8 \times 10^{35}\text{cm}^{-2}\text{s}^{-1}$ is the nano-beam scheme: at the Interaction Region (IR), the vertical beam size will be squeezed to only $\mathcal{O}(60\,\text{nm})$. Besides $\Upsilon(4S)$ events, with a large average track multiplicity (≈ 10), Belle II has access to important low multiplicity physics decay channels like τ pair production, initial state radiation, dark photons searches, etc. [4].

The downside of the luminosity upgrade is a severely increased background rate. This background manifests itself in noise and tracks. All events suffer from background noise (single hits), which increases the occupancy in all detectors and thus requires stable algorithms. Background tracks, found at the first trigger level, are a dangerous threat to low multiplicity physics events. Fortunately, many background tracks have their vertices displaced in the longitudinal coordinate z. The expected z vertex distribution of background trigger tracks is shown in figure 1 and will be discussed in section 3.

A 3D track reconstruction at the first trigger level will enable a significant reduction of the background track rate. The neural network trigger [5] is a noise robust solution to this 3D tracking problem that will be implemented on FPGA hardware [6]. After reviewing the neural network trigger and its integration into the first level trigger in section 2, the simulation of the expected backgrounds and its possible suppression are discussed in section 3.

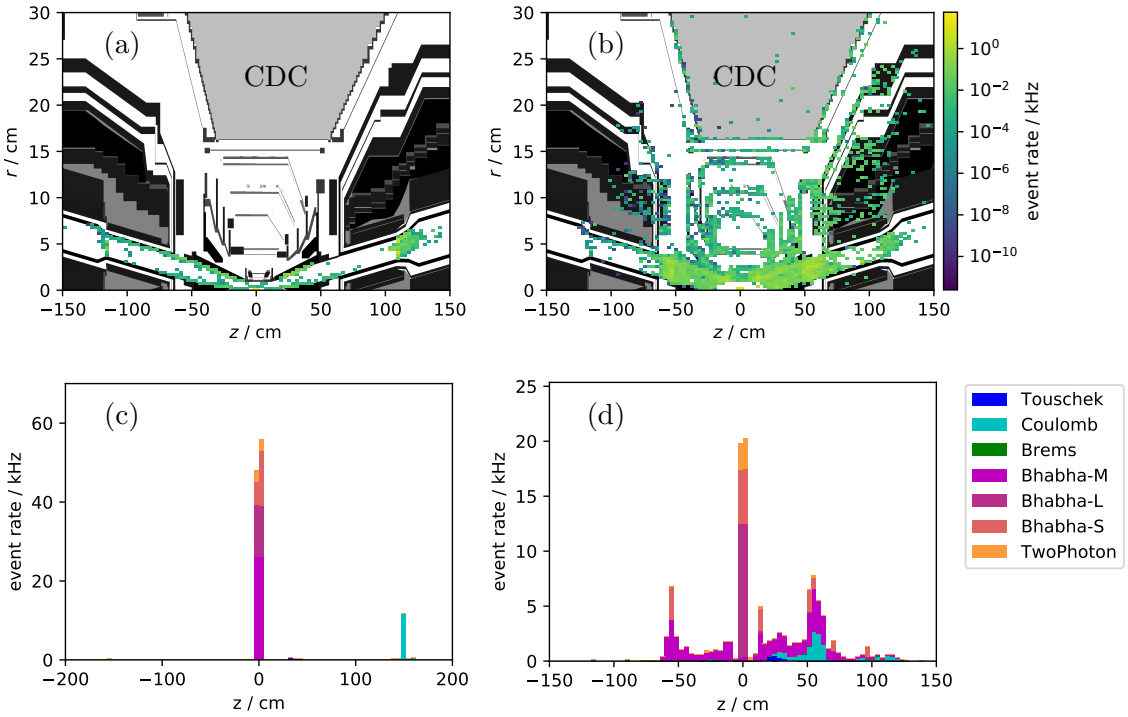

Figure 1. Rate of background events with at least one 2D trigger track. (a) & (b) Longitudinal (z) and radial (r) components of the track vertices on top of the material (detectors, beam-pipe, etc.). (c) & (d) z vertex distribution for different background types. (a) & (c) Generated particles before the simulation of the material interaction with Geant4. (b) & (d) Vertices of the simulated particles with a match to the tracks found by the 2D trigger.

2. Neural Network Trigger

At the first trigger level of Belle II, events are processed online using only information from detectors with a fast readout time. The main sub-triggers use the Central Drift Chamber (CDC) and the electromagnetic calorimeter (ECL). The neural network trigger is integrated in the CDC subtrigger [7]. Detectors with a longer readout time, especially the precise silicon detectors (silicon vertex detector - SVD & pixel detector - PXD), are only read out if a trigger signal is received. All the sub-trigger systems and the combination of their results in the global decision logic (GDL) are operated as deadtime-free pipelines using FPGAs. Required is a fixed latency of under 5 µs while providing a trigger rate of 30 kHz [1].

2.1. CDC Trigger

The CDC is a drift chamber consisting of 56 layers of sense wires, with high voltage field wires in between. The measurement of the drift time provides a distance information of a charged track to a sense wire. The 56 layers are grouped to 9 Super Layers (SL) of wires with the same orientation: 5 axial SL with wires parallel to the z axis (direction of the 1.5 T magnetic field) and 4 stereo SL inclined with a stereo angle w.r.t. the z axis that enables 3D track reconstruction [1].

In the CDC subtrigger, the first component is the Track Segment Finder (TSF): it is a noise suppression and data reduction element that combines all layers within a SL to Track Segments (TS) [8]. A TS is active if at least four out of five layers within the TS pattern have a hit. An active TS contains one priority wire with its drift time and indicates whether the track

passed left or right of the wire. The TSF is followed by a 2D track finder based on a Hough transformation [9] using only the axial TS positions (no drift times) as input [5]. For each found track two parameters are provided: p_T and ϕ, the transverse momentum of the track and the azimuthal angle. Parallel to the 2D finder, the event time is determined to enable the use of the drift times. Next, each found track is extended with 3D information: z and θ, the longitudinal vertex position and the polar angle of the track. This 3D reconstruction is carried out by two alternative approaches, the neural network and a least square fit. Finally, in the GDL, the CDC results are combined with the other subtriggers and a final trigger decision is made.

2.2. Neural Network

The essential part in the neural network trigger is a multilayer perceptron with one hidden layer using the hyperbolic tangent as activation function. The neural network is applied on single tracks and trained to estimate z and θ, scaled to the codomain of the activation function. This scaling is realized by expecting tracks to have their z vertex within 1 m around the IR (i.e. $z \in [-50, 50]$ cm). The neural network trigger receives p_T and ϕ of the tracks found, the event time t_0 and all the TS (axial and stereo). At first, a preprocessing step associates each 2D track with the corresponding TS and thus enables the neural network to process single tracks. The input to a network is one TS per SL per track, e.g. 9 TS for a track with a TS in every SL. In order to increase the efficiency, a dedicated pretrained network is selected, based on the actual number of TS per track. To form the input, each TS is represented with three values: two for the position of the TS (ϕ_{rel}, α) and the drift time t. The position is calculated relative to the 2D track, ϕ_{rel} is the distance of the 2D track to the TS and α is the arc length along the 2D track to the TS divided by the radius of the 2D track. Altogether, a network for a track with 9 TS has 27 inputs and 2 outputs. The size of the hidden layer is set to 127 nodes. A detailed description of the current neural network structure and its inputs is provided in [5].

The networks are trained with simulated single muon tracks with overlaid background noise. The track parameters are uniformly distributed in z, $\cos(\theta)$, ϕ and p_T^{-1} and the vertices are located on the z axis. The training is carried out using the iRPROP$^-$ [10] backpropagation algorithm within the library FANN [11]. The size of the training data set was chosen to be 10× the number of weights in the network.

Figure 2(a) shows the neural network reconstructed z vertices of the background tracks. The true vertex positions in figure 1(d) are smeared, with a smaller total rate (at least 3 stereo TS are required). The limited z range is clearly visible and tracks with vertices outside of the z

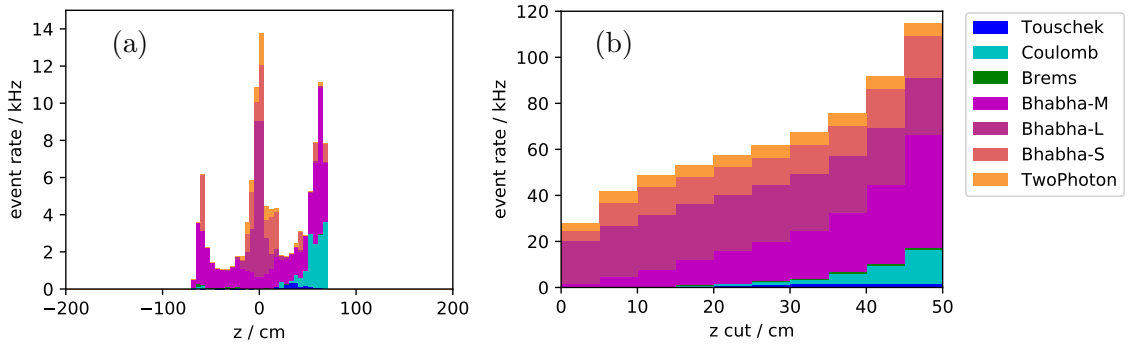

Figure 2. (a) Neural network estimates of the z vertices. (b) Integrated background rate for cuts based on the neural network z estimates.

range are nicely predicted at the boundaries. Figure 2(b) is an integrated version of 2(a) for cuts in z (z_{cut}) varied in 5 cm steps. It shows the remaining background event rate, if only tracks are triggered where the estimated z value z_{NN} satisfies: $-z_{\text{cut}} < z_{\text{NN}} < z_{\text{cut}}$.

3. Background

Two main classes of background can be distinguished: luminosity and machine background. Luminosity backgrounds are e^+e^- events ($z = 0\,\text{cm}$) from uninteresting QED processes. The main luminosity backgrounds are 2γ events ($e^+e^- \rightarrow e^+e^-e^+e^-$) and (radiative) Bhabha ($e^+e^- \rightarrow e^+e^-\gamma$). Machine background is caused by particle loss from a single beam, i.e. when particles hit the beam wall at any position along the beamline ($z \neq 0\,\text{cm}$). The main sources of machine background [12] are the Touschek effect [13] (intra beam scattering) and scattering of beam particles on residual gas nuclei (elastic/Coulomb and inelastic/Brems). Touschek background will increase due to the nano-beam scheme, beam-gas scattering due to the bad vacuum conditions coming along with the small beam-pipe radius. The machine background rate in the IR is reduced by collimators [14]. Via material scattering, the vertex positions of the background trigger tracks may deviate from the primary background vertices, i.e. luminosity background can also produce tracks with $z \neq 0\,\text{cm}$. The background types considered are listed in table 1.

Table 1. Background types considered in the simulation with their physics generators and their expected rate of events containing at least one 2D trigger track.

background	process	generator	2D trigger rate
TwoPhoton	$e^+e^- \rightarrow e^+e^-e^+e^-$	AAFH [15]	6 kHz
Bhabha-S		BBBrem [16]	20 kHz
Bhabha-M	$e^+e^- \rightarrow e^+e^-\gamma$	BHWide [17]	52 kHz
Bhabha-L		BHWide	26 kHz
Touschek	intra bunch	SAD [18]	2 kHz
Coulomb	elastic beam-gas	SAD	15 kHz
Brems	inelastic beam-gas	SAD	1 kHz

3.1. Background Simulation

In machine background processes, beam particles experience a small deviation in momentum or energy and have to be tracked along the collider ring for several turns until they eventually get lost. In the simulation, a sample of deviated beam particles, weighted by the physics probability of the deviation, is tracked with SAD. Here all generated particles (weighted) are simulated $10\times$ with Geant4, i.e. low rated events are included. The rates are used directly in the evaluation.

In luminosity backgrounds, the number of generated (unweighted) events corresponds to a given time equivalent. The radiative Bhabha events are divided into three parts (S/M/L), because the cross section strongly depends on the polar scattering angle of one outgoing lepton with the z axis ($\theta_{e^\pm}$) and the whole phase space should be represented. In Bhabha-S (small), at least one scattering angle is required to be small $\theta_{e^\pm} < 0.5°$. Bhabha-L (large) contains the events, where for both scattering angles holds $\theta_{e^\pm} > 1°$ and at least one angle has $\theta_{e^\pm} > 10°$. Bhabha-M (medium) covers the remaining phase space.

3.2. Material Scattering

The products of a scattering process with the material might be in the acceptance region of the CDC trigger. Therefore, generated background particles not in the acceptance produce detectable secondaries and tracks from the IR have secondaries with vertices at $z \neq 0\,\text{cm}$. Figure 1 shows the vertices of the initial background particles and the vertices of the particles that are actually found by the trigger, i.e. after material scattering. The particles in figure 1(a) which are generated on the beam-pipe can be identified with machine background in figure 1(c). Figure 1(b) shows the vertices after material scattering at various positions and figure 1(d) shows that both luminosity and machine background undergo scattering. Figure 1(d) is the expected z vertex distribution of events with at least one 2D trigger track.

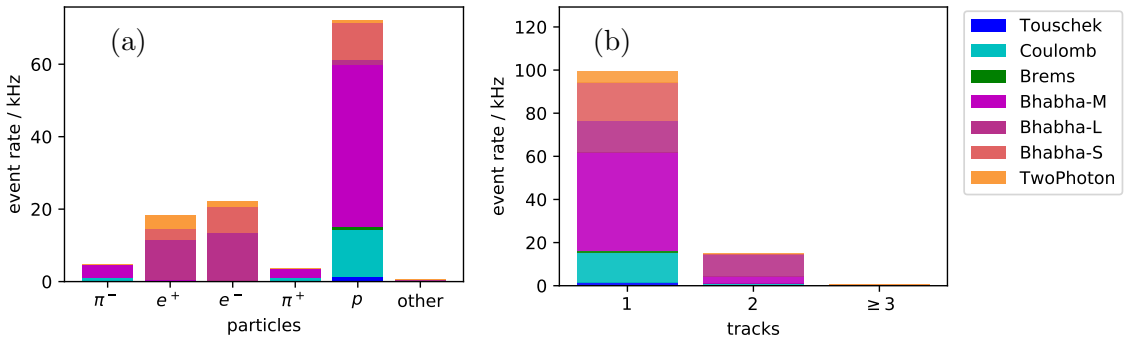

Figure 3. Properties of the background events. (a) Particle spectrum of the simulated Monte Carlo particles matched to the 2D trigger tracks. (b) Trigger track multiplicity in the background events.

Figure 3 shows the spectrum of the particle types causing trigger tracks and the track multiplicity. The spectrum is dominated by protons (p), produced by nuclear spallation of the material. The final states of inelastic scattering processes (p, π^+, π^-) are produced by all background types, while e^+ and e^- tracks without scattering are observed only from the luminosity backgrounds.

3.3. Background Suppression

The goal of the z vertex reconstruction is to reject background tracks with z vertices outside of the IR. From the perspective of the neural network trigger alone, the background sample splits into two parts: the reducible ($z \neq 0\,\text{cm}$) and the irreducible ($z = 0\,\text{cm}$) background. Substantially, the irreducible backgrounds are e^+e^- tracks from the IR caused by the luminosity backgrounds. Countermeasures are installed in the GDL of the trigger, where the results of all subtriggers are available (Bhabha vetos). Basically, ECL and CDC information is combined in order to identify electron and positron tracks.

Figure 4(a) shows the Monte Carlo z vertices of the reducible background events, here defined as $|z| \geq 1\,\text{cm}$, with an event rate of 81.4 kHz. The largest contribution to the reducible background are scattered Bhabha-M events. In Figure 4(b), the integrated rates for different z_{cut} values show that a huge suppression of the background rate is enabled by the neural network z vertex trigger. This allows to open up the other trigger conditions and thus to enhance the sensitivity for new physics. In order to improve the efficiency for low multiplicity physics, even a single track trigger can be considered. Given the maximum trigger rate of 30 kHz, a multiplicity dependent z_{cut} and pre-scaling may be used.

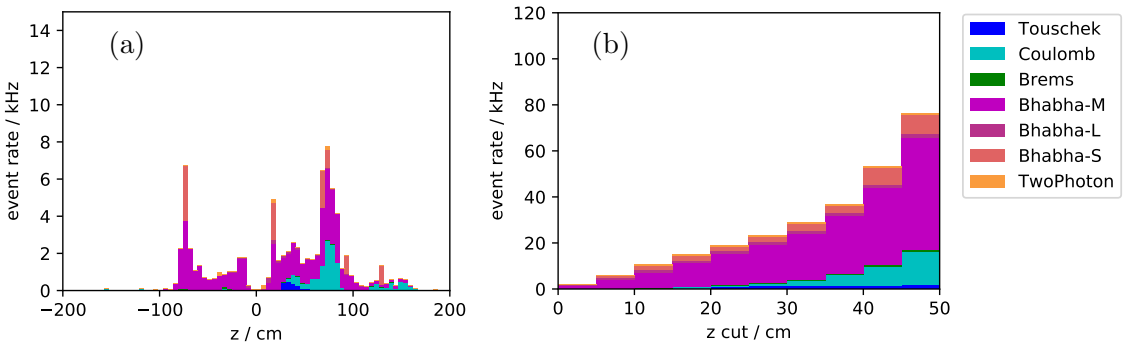

Figure 4. Suppression of the reducible background with vertices $|z| \geq 1\,\text{cm}$. (a) Monte Carlo z distribution of the simulated particles matched to 2D trigger tracks. (b) Integrated background event rate of the reducible background at different cut values (z_{cut}) on the neural network z estimates.

4. Conclusion

A neural network tracking algorithm, robust against noise, has been developed for the Belle II first level trigger. It allows a significant suppression of the background rate and enables even a single track trigger. Luminosity and machine background tracks, with a total event rate of $\approx 120\,\text{kHz}$, undergo material scattering. About 80 kHz is reducible by a z vertex trigger, the remaining $\approx 40\,\text{kHz}$ from QED events (at $z = 0\,\text{cm}$) have to be suppressed by a special veto in combination with the calorimeter. The exact determination of z_{cut} will depend on the efficiency of the other trigger logics in the GDL and has to be the subject of a combined study.

References

[1] Abe T *et al.* (Belle-II) 2010 Belle II Technical Design Report (*Preprint* `1011.0352`)
[2] Hashimoto S *et al.* 2004 LoI for KEK Super B Factory, Part I: Physics
[3] Abashian A *et al.* 2002 *Nucl. Instrum. Methods A* **479** 117–232 ISSN 0168-9002
[4] Bevan A J *et al.* 2014 *The European Physical Journal C* **74** 3026 ISSN 1434-6052
[5] Neuhaus S, Skambraks S and Kiesling C 2017 *EPJ Web Conf.* **150** 00009
[6] Baehr S, Skambraks S, Neuhaus S, Kiesling C and Becker J 2017 *Journal of Instrumentation* **12** C03065
[7] Skambraks S *et al.* 2015 *IEEE Transactions on Nuclear Science* **62** 1732–1740 ISSN 0018-9499
[8] Iwasaki Y, Cheon B, Won E, Gao X, Macchiarulo L, Nishimura K and Varner G 2011 *IEEE Transactions on Nuclear Science* **58** 1807–1815 ISSN 0018-9499
[9] Hough P V C 1959 *Conf. Proc.* **C590914** 554–558
[10] Igel C and Hüsken M 2000 *Proc. 2nd Int. Symp. Neural Computation* pp 115–121
[11] Nissen S "FANN" [software], version 2.2.0, Available from https://github.com/libfann/fann
[12] Nakayama H, Iwasaki M, Kanazawa K, Nakano H, Ohnishi Y, Tanaka S and Tsuboyama T 2012 *Proc. of IPAC* pp 1825–1827
[13] Bernardini C, Corazza G F, Di Giugno G, Ghigo G, Haissinski J, Marin P, Querzoli R and Touschek B 1963 *Phys. Rev. Lett.* **10**(9) 407–409
[14] Nakayama H, Suetsugu Y, Kanazawa K, Ohnishi Y, Funakoshi Y, Ohmi K and Zou D a 2012 *Proc. of IPAC* pp 1104–1106
[15] Berends F, Daverveldt P and Kleiss R 1985 *Nuclear Physics B* **253** 441 – 463 ISSN 0550-3213
[16] Kleiss R and Burkhardt H 1994 *Comput. Phys. Commun.* **81** 372–380 (*Preprint* `hep-ph/9401333`)
[17] Jadach S, Placzek W and Ward B 1997 *Physics Letters B* **390** 298–308 ISSN 0370-2693
[18] Hirata K 1988 *Proceedings of the Second Advanced ICFA Beam Dynamics Workshop* CERN 88-04