Journal of Physics: Conference Series **3206** (2026) 012082 doi:10.1088/1742-6596/3206/1/012082

Calibrated particle identification for Belle II

M. Hohmann[†], M. Milesi[†], P. Urquijo[†], C. Hainje[‡], J. Strube[‡]

[†]School of Physics, The University of Melbourne, Parkville, Victoria, 3010, Australia
[‡]Pacific Northwest National Laboratory, Richland, WA, USA

E-mail: marcel.hohmann@belle2.org

Abstract. We present several efforts aimed at improving charged particle identification at the Belle II experiment. We define an ablation test to quantify and evaluate the impact of each sub-detector on the global particle identification performance. We demonstrate that the performance of the identification scheme can be improved via a simple calibration of the sub-detector likelihoods through a set of per hypothesis, per sub-detector weights. Finally, we present preliminary results on an improved definition of the likelihoods contributed by the electromagnetic calorimeter. A set of multiclass boosted decision trees is trained to exploit the shape of energy depositions of different particle species. In simulated $B\overline{B}$ events, the pion-to-electron and muon fake rates are reduced by 55% and 31% respectively at low-medium momentum.

1. Introduction

The Belle II experiment [1] at the SuperKEKB [2] electron-positron collider in Tsukuba, Japan is a new-generation B-factory experiment with an extensive physics program [3]. Among the main goals of the experiment are precision tests of the standard model and searches for beyond-standard model phenomena via measurements of CP asymmetries and rare decays of B-mesons and tau leptons. Belle II has collected an integrated luminosity of 428 fb^{-1} of data and intends to collect 50 ab^{-1} over the next decade, predominantly at the $\Upsilon(4S)$ resonance. This will exceed the world-leading data sample of its predecessor, the Belle experiment, by a factor of 50. The continued success of the Belle II physics program will depend crucially on particle identification performance.

2. Particle identification at Belle II

Belle II is a hermetic detector with almost 4π coverage consisting of several sub-detectors. From the beam pipe out, these are: two layers of pixelated silicon sensors (PXD)[1] and four layers of silicon strip sensors (SVD) which together form the vertexing system; a central drift chamber (CDC) for tracking; a time of propagation system in the barrel (TOP) and aerogel ring-imaging Cherenkov (ARICH) detector in the forward end-cap, both designed for hadron identification; a CsI(TI) electromagnetic calorimeter (ECL), primarily intended to measure the energy of electrons and photons; a set of scintillating strips and resistive plate chambers placed beyond the superconducting coil which generates a 1.5T magnetic field, to identify K_L^0 mesons and muons (KLM).

[1] Only one segment of the second layer was installed for all data-taking periods to date.

All sub-detectors apart from the PXD contribute to the global charged particle identification scheme. Their measurements are used to define likelihoods for each of the six charged particle species that are long-lived enough to be detected: electrons, muons, pions, kaons, protons and deuterons, and their respective anti-particles. In each sub-detector, the likelihoods are defined as a function of the probability density function (PDF) parameters given a set of observables, where the PDFs are either analytic or predicted from simulated data. The observables are: the particle energy loss by ionisation (dE/dx) in the SVD and CDC, velocity-dependent optical responses in the TOP and ARICH, the ratio of the energy deposited in the ECL to the momentum of the particle as measured in the tracking system (E/p), and measurements of the scattering and penetration ranges in the KLM. Under the assumption that the measurements of the observables are independent the likelihoods from the sub-detectors are combined into a global likelihood for each charged stable hypothesis (i) as:

$$\log \mathcal{L}_i = \sum_{\text{det}}^{\{\text{SVD, CDC, } \dots\}} \log \mathcal{L}_i^{\text{det}}. \tag{1}$$

Finally, a set of global likelihood ratios (P_i) are defined to serve as particle identification variables upon which to impose selections in analyses:

$$P_i = \frac{\mathcal{L}_i}{\sum_j^{\{e,\mu,\pi,K,p,d\}} \mathcal{L}_j} \tag{2}$$

3. Evaluating impact of detectors on PID

To evaluate the impact of each sub-detector on the global charged particle identification performance an *ablation* test is performed. We first define a separation score metric s_d as:

$$s_d = 1 - \mathcal{A}_{o,d}, \tag{3}$$

where d can be all or a subset of sub-detectors, and $\mathcal{A}_{o,d}$ is the unity-normalised area of the overlap of histograms of simulated pure single particle samples of the signal and background species in the likelihood ratio for the signal hypothesis calculated from the subset of sub-detectors d. By construction, a value of $s_d = 1$ indicates perfect separation and $s_d = 0$ indicates no separation. To perform the ablation test we consider

$$\Delta s_{\not d} = s_{\not d} - s, \tag{4}$$

where s is the separation score considering all sub-detectors and $s_{\not d}$ the score considering all sub-detectors bar $\not d$. An example of the test for electron-pion separation is given in Fig. 1. We note that electron-pion separation is driven by the SVD and CDC at low momenta and the calorimeter at higher momenta. The TOP is observed to worsen the separation performance across the momentum spectrum. For the TOP this is associated with a known issue in the modelling of δ-ray electron emissions in the analytic evaluations of the electron hypothesis PDF in the samples considered for this work. The performance degradation caused by the contributions of the ARICH at low momentum and the SVD at high momentum are still under investigation.

4. Calibrating likelihoods

Under the assumption that that the sub-detector likelihoods are independent and accurate the likelihood ratio of Eqn. 2 is the theoretically optimal quantity [4] for identifying charged particle species. It has however been observed that several of the sub-detectors can provide over- or under-confident likelihoods degrading the separation performance. As improving or fixing the

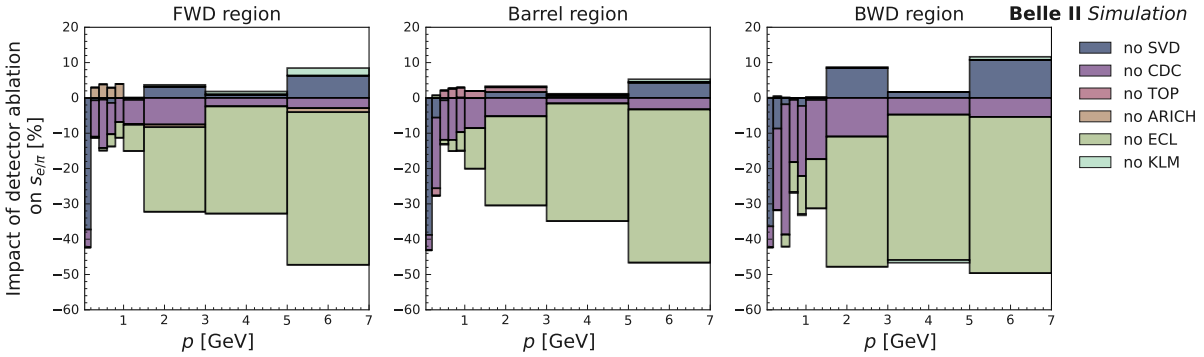

Figure 1. The impact on the $e^{\pm}$ vs $\pi^{\pm}$ separation by removing a sub-detector from the combined likelihood as a function of the track momenta in three detector regions (forward endcap, barrel, backward endcap). Positive scores indicate removing the detector improves the particle identification performance.

likelihood calculation of each sub-detector can be done every new major software release and data re-processing only, we introduce a set of 36 calibration weights, w_i^{det}, for each particle hypothesis for each sub-detector. The calibration weights enter the likelihood calculation by modifying Eqn. 1:

$$\log \mathcal{L}_i = \sum_{\text{det}}^{\{\text{SVD, CDC, } \dots\}} w_i^{\text{det}} \log \mathcal{L}_i^{\text{det}}.. \tag{5}$$

The weights are extracted from single particle Monte Carlo simulations. Events of each particle species are generated with flat (p, θ) distributions and the weights optimized by minimising the loss function:

$$L(\hat{y}, y) = \text{CE}(\hat{y}, y) + \beta \text{BCE}(\hat{y}_\pi, y_\pi), \tag{6}$$

where $\hat{y}, y$ are the predicted and true particle species for each track, and CE and BCE the cross-entropy and binary cross-entropy respectively. The subscript π indicates that only true pions are considered. The hyper-parameter β allows for tuning the relative importance of pions, the dominant particle species produced at Belle II. The parameter should be carefully tuned for good performance across all physics analyses. For the purpose of this work, however, it has been set to $\beta = 0.1$ which gives reasonable performance. Fig. 2 shows the optimized weights and a comparison of pion fake rate at fixed electron, muon and kaon performance for the standard and calibrated likelihood ratios. Significant improvements in pion rejection of up to factors of seven and three are observed at constant electron and kaon efficiencies respectively. For muon identification the calibrated likelihood ratios are observed to provide worse separation at high momentum, performance is expected to be recoverable with an improved tuning of the β parameter.

5. Improving ECL likelihoods

In addition to calibrating the likelihoods, we are also improving the particle identification performance by maximising the information extracted from each sub-detector. We expand on the scheme of Ref.[5], training multiclass boosted decision trees (BDT) as implemented in Ref. [6] to exploit information related to the shape of the energy deposition within the calorimeter. The BDTs consist of 500 trees of depth three, and are trained in 18 regions corresponding to charge-dependent training for low ($p < 0.6$ GeV), medium ($0.6 \leq p < 1.0$ GeV) and high ($p \geq 1.0$ GeV)

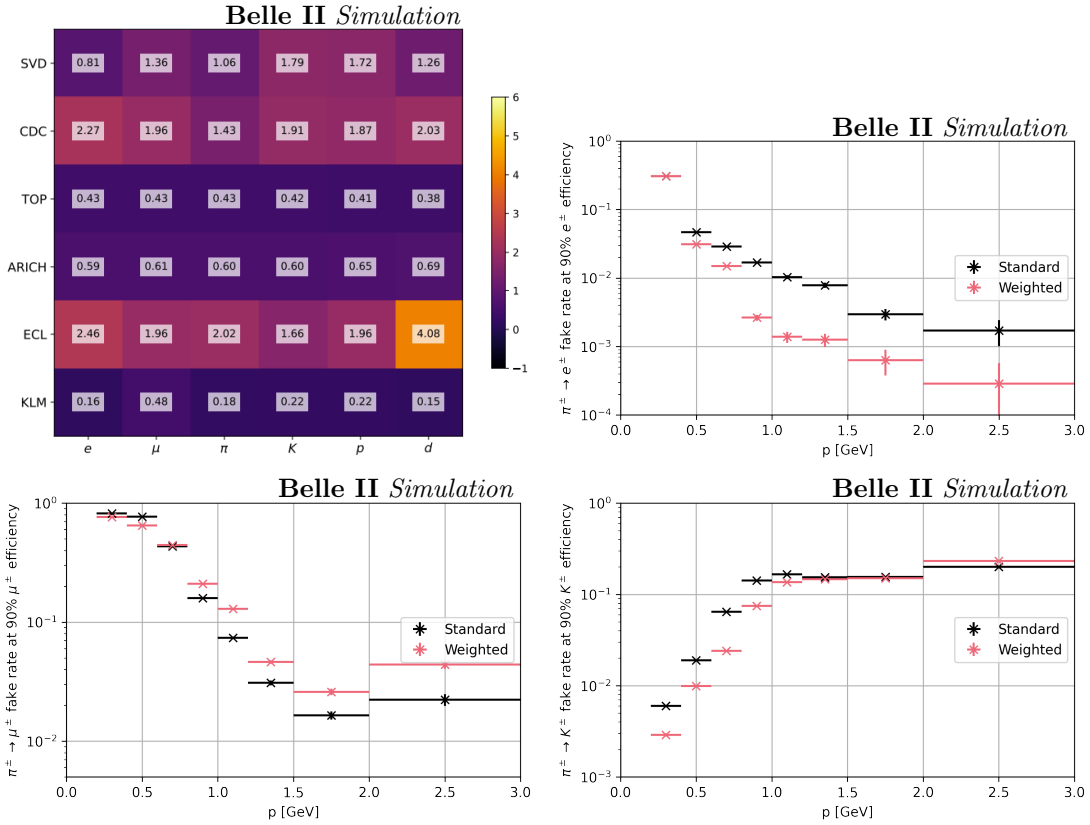

Figure 2. Matrix of calibrated weights *(upper left)*. Comparison of pion rejection as a function of track momentum for electron identification *(upper right)*, muon identification *(lower left)* and kaon identification *(lower right)*, for the standard *(black)* and weighted *(red)* particle identification variables on simulated $B\overline{B}$ samples. In each momentum bin the threshold on the particle identification variables is selected to give a 90% efficiency in the signal species.

momentum particles as seen in the lab frame, with matching clusters in the end-cap, barrel, or backward end-cap of the calorimeter. In addition to E/p, we include as inputs to the BDTs 24 variables describing the lateral shower shape [7] and extrapolated track penetration depth in the ECL [8], and 50 variables comprising of five variables describing the per-crystal energy and position relative to the center of the calorimeter cluster for the ten most energetic crystals associated with the cluster. The training is performed on simulated single particle samples. As the BDT outputs are observed to be well-calibrated, the classification scores are interpreted as probabilities of identification and integrated directly into the global likelihood scheme. Fig. 3 compares the calorimeter-only performance of the BDT-based particle identification with the standard method which considers only E/p. In simulated $B\overline{B}$ samples that reflect the typical event activity at Belle II, we observe reductions in pion-to-lepton fake rates of 55% and 31% for electron and muon identification respectively at low-medium momentum. Further optimisation and the integration of a convolutional neural network that exploits the regular grid structure of the barrel section of the ECL [9] is ongoing.

6. Summary and outlook

We presented several efforts aimed at characterising and improving charged particle identification at Belle II. The impact of individual sub-detectors in the likelihood-based global identification

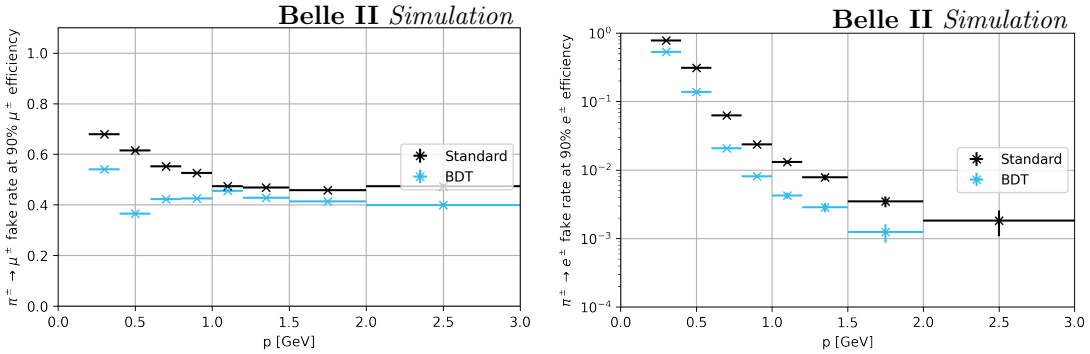

Figure 3. Comparison of pion-to-lepton fake rate at fixed electron *(left)* and muon *(right)* efficiencies in ECL only performance for the E/p univariate based likelihoods (*black*) and the likelihoods extracted from BDTs *(blue)* on simulated $B\overline{B}$ samples.

model is quantified via an ablation test, that accounts for possible correlations between the likelihoods of different sub-detectors. Each likelihood in the combination is calibrated via weights defined for each particle hypothesis and sub-detector which are found to significantly improve electron-pion and kaon-pion separation, at the cost of worsening muon separation at high momentum, which may be recoverable via hyper-parameter tuning. Finally, we presented preliminary results for an improved definition of the ECL likelihood, which exploits a combination of shapes of the energy depositions and crystal-level quantities in a set of boosted decision trees. Using ECL-only information, we found the pion-to-electron and muon fake rates are reduced by 55% and 31% respectively at low-medium momentum. Further work is ongoing to integrate these improvements into the core Belle II software [10].

References

[1] T. Abe et al. 2010. *Preprint arXiv:1011.0352*
[2] K. Akai, K. Furukawa, and H. Koiso 2018. *Nucl. Instrum. Meth. A* **907** 188–99
[3] E. Kou et al. 2019. *PTEP* **2019.12** 123C01
[4] N. Jerzy and P. E. Sharpe 1933. *Philos. Trans. Royal Soc. A* 289
[5] M. Milesi, J. Tan, and P. Urquijo 2020. *EPJ Web Conf.* **245** 06023
[6] A. Hoecker et al. 2009. *Preprint arXiv:physics/0703039*
[7] F. Zernike 1934. *Physica* **1.7** 689–704
[8] D. N. Brown, J. Ilic, and G. B. Mohanty 2008. *Nucl. Instrum. Meth.* **A592** 254–60
[9] A. N. Charan 2023. *J. Phys. Conf. Ser.* **2438** 012111
[10] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun 2018. *Comput. Softw. Big Sci.* **3.1**