Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers

Marco Milesi[1,*], *Justin* Tan[1,**], and *Phillip* Urquijo[1,***]

[1]The University of Melbourne, VIC, Australia

Abstract. We present a major overhaul to lepton identification for the Belle II experiment, based on a novel multi-variate classification algorithm. Boosted decision trees are trained combining measurements from the electromagnetic calorimeter (ECL) and the tracking system. The chosen observables are sensitive to the different physics that governs interactions of hadrons, electrons and muons with the calorimeter crystals. Dedicated classifiers are used in various detector regions and lepton momentum ranges. The tree output is eventually combined with classifiers that rely upon independent measurements from other sub-detectors. Using simulation, the performance of the new algorithm is compared against the method used for analysis of the 2018 Belle II data, namely a likelihood discriminator based on the ratio of energy measured in the ECL over the momentum measured by the trackers. In the low momentum region, we largely improve the lepton-pion separation power, decreasing misidentification probability by a factor of 10 for electrons, and 2 for muons at fixed identification efficiency.

1 Introduction

The Belle II experiment [1] is a *B*-factory at the SuperKEKB [2] asymmetric e^+e^- collider in Tsukuba, Japan. It is designed to cover a broad research programme in the flavour physics sector, including searches for new physics in rare *B* meson decays and high precision measurements of Standard Model parameters. Of particular interest is the study of semi-tauonic *B* decays to test lepton flavour universality, in both exclusive modes like $B \to D^*\tau\nu$ and inclusive $B \to X\tau\nu$. The analysis of these decays relies on the capability of correctly separating low-momentum lepton candidates (e, μ) in the decay of the τ from hadronic backgrounds. At the mean momentum of $\lesssim 600 \, \text{MeV}/c$, muons do not reach the dedicated muon detector. Furthermore, at low momenta electrons suffer significant energy losses due to bremsstrahlung, making them more easily mimicked by hadrons.

Owing to a smaller beam profile and higher current, the design SuperKEKB luminosity is $8 \times 10^{35} \, \text{cm}^{-2}\text{s}^{-1}$, about 40 times larger than the one of its predecessor, KEKB. The resulting harsher beam background conditions, as well as the smaller centre-of-mass boost in the laboratory frame due to the reduced beam energy asymmetry, require enhancements in the Belle II algorithms for both decay vertex reconstruction and particle identification.

---

*e-mail: marco.milesi@unimelb.edu.au
**e-mail: justin.tan@student.unimelb.edu.au
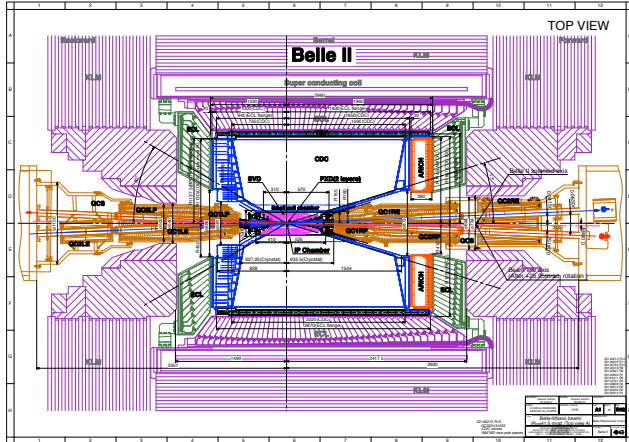***e-mail: phillip.urquijo@unimelb.edu.au

Figure 1. Cut view of the Belle II detector in the $r - z$ plane [3].

The method hereby presented is developed to improve the identification of low-momentum leptons, by combining several measurements in the electromagnetic calorimeter with information from the other sub-detectors in a multivariate classifier.

1.1 The Belle II detector layout

The structure of the Belle II detector is displayed in Figure 1. Two layers of pixelated silicon sensors (PXD) together with 4 layers of silicon strip sensors (SVD) are located closest to the beam pipe for reconstructing decay vertices. A central drift chamber (CDC) then fills the larger outer radius of the tracking volume. A time of propagation system in the barrel region (TOP), in combination with a ring-imaging Cherenkov detector in the forward endcap (ARICH) are specifically designed for hadron identification. A CsI(Tl) laterally-segmented electromagnetic calorimeter (ECL), with a longitudinal size of 16.2 X_0 in units of radiation length, is used to measure the energy of photons and electrons. Finally, the scintillator strips and resistive plate chambers of the the outermost KLM detector serve for K_L^0 meson and muon identification.

2 Charged particle identification at Belle II

Charged stable particle identification (ID) at Belle II is made by a combination of measurements, namely:

- Particle energy loss by ionisation (dE/dx) in the SVD and CDC.
- Measurements of the velocity-dependent optical response in the TOP and ARICH.
- Measurements of the energy deposition pattern in the CsI(Tl) scintillation crystals of the ECL.
- Measurements in the KLM of the different penetration range and scattering of muons and hadrons.

Herein, "stable" refers to charged particles that live long enough to travel across entire subsections of the detector: e, μ, π, K, p, d and their respective antiparticles. In this study, we focus solely on electrons, muons and charged pions.
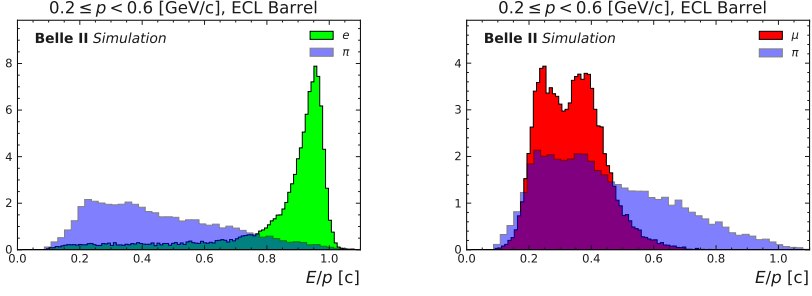
Figure 2. Distribution of E/p for simulated single particle candidates: $e^{\pm}$ (green), $\mu^{\pm}$ (red) and $\pi^{\pm}$ (blue) for $0.2 \leq p < 0.6\,\text{GeV/c}$ in the ECL barrel region (defined by the polar angle range $0.56 \leq \theta < 2.24$ rad). The bimodal distribution in the muon case is a result of the design of Belle II ECL clustering algorithm, which favours formation of radially-symmetric clusters around the crystal with highest energy (the *seed* crystal). These energy deposition patterns are typical of photons and electrons, less so of minimally-ionising particles such as muons, which therefore often spread their energy over more than one cluster.

In each sub-detector, a likelihood $\mathcal{L}_i^{\text{det}}$ is defined for each charged stable particle hypothesis i as a function of the probability density function (PDF) parameters for a given set of observables. The PDF parameters are either predicted from simulated data, or determined analytically. Assuming the sub-detectors' measurements of each of the identifying observables are independent, a global likelihood for hypothesis i is defined as:

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

Given all possible, mutually exclusive outcomes of identification, $\{A_j\} = \{e, \mu, \pi \dots\}$, and a set of measurements $\mathbf{x}$ for a reconstructed particle candidate, the likelihood ratio is defined as a proxy for identification of the candidate as particle of type i:

$$P(\mathbf{x})_i = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j}. \tag{2}$$

A cut on the $P(\mathbf{x})_i$ distribution is thus used as a particle identification criterion.

2.1 Standard likelihood-based identification in the ECL.

The standard Belle II particle ID algorithm in the ECL defines a univariate likelihood as a function of E/p, the ratio of energy measured in the calorimeter over the momentum measured by the trackers. This variable is generally very powerful in discriminating electrons against hadrons, such as π's: for the former, it is expected to peak sharply around unity given that they are almost always stopped by the ECL. The joint probability density functions of E/p for various particle species are determined by fits to templates from simulation.

For low momentum e's, however, stronger bending of the trajectory in the detector's solenoidal magnetic field leads to longer paths through material before the ECL, increasing energy losses from bremsstrahlung. This effect reduces E/p separation power. The rate of hadronic inelastic interactions is also higher at lower momenta, resulting in a broader E/p shape which in turn also strongly limits $\pi - \mu$ separation. This situation is represented in Figure 2, which shows the E/p distribution for simulated e, μ, π candidates with $0.2 \leq p < 0.6$ GeV/c in the central ECL barrel region.

3 Boosted decision trees for lepton identification

In the Belle II ECL software, several observables are defined to describe lateral shower shape development - such as Zernike moments [4] - and the extrapolated track penetration depth into the ECL [5], ΔL. Lateral shower development is expected to differ between EM-interacting particles (e), pure minimally-ionising particles (μ) and hadrons (π), which are minimally ionising but can also undergo hadronic split-offs due to strong inelastic interactions with the nuclei of the ECL material. Furthermore, quantities such as ΔL provide information about the longitudinal development of the shower, which is also sensitive to the underlying physics that governs interactions of different charged particles with matter.

Access to such diverse information in the ECL leads to a multivariate approach for particle ID. Since observables are in general highly correlated, machine learning algorithms such as boosted decision trees (BDTs) provide a better handle at exploiting non-trivial dependencies across inputs to improve classification performance. In order to exploit the particle ID capability of Belle II in its entirety, we further combine ECL-based inputs with the high-level likelihoods for the e, μ and π hypotheses from the other sub-detectors.

In this study, we use gradient boosting decision trees as implemented in the TMVA [6] package for binary classification: e vs. π and μ vs. π.

The method has been fully integrated into the Belle II analysis software [7], making it readily available to all users of the collaboration. Independent training configurations suitable for different data taking periods and detector conditions are stored in a central Conditions Database [8] as serialised ROOT [9] objects.

3.1 Inputs and categories, train and test datasets

The complete list of input variables is summarised in Table 1. The ECL observables' shapes generally depend on a particle's momentum, as well as on geometrical effects related to the calorimeter structure. Furthermore, the likelihoods of the other sub-detectors are often defined only in specific subsets of the full detector acceptance. Therefore, a categorisation is performed by training independent classifiers in 9 different subsets ("categories") defined by track momentum p (three bins of low, medium and high momentum) and ECL cluster polar angle $\theta_{cluster}$ (three bins of ECL forward, backward endcaps and barrel region), as outlined in Table 2.

The training dataset consists of 10^6 simulated particles for each species ($e^\pm$, $\mu^\pm$, $\pi^\pm$), generated with momenta, polar and azimuthal angles distributed uniformly in the $0.2 \leq p < 5$ GeV/c, $0 \leq \theta < \pi$ rad and $0 \leq \phi < 2\pi$ rad ranges, respectively. In the simulation stage, effects of beam-induced backgrounds sampled from pseudo-random triggered data are taken into account.

For testing, a set of 2×10^5 events of the $B^0\bar{B}^0$ process are used. Charged particles are then inclusively reconstructed from tracks with $p > 0.2$ GeV/c, transverse ($|dr| < 2.0$ cm) and longitudinal ($|dz| < 5.0$ cm) impact parameters, that are geometrically matched to correctly-labelled candidates at generation level. Only tracks that have a matching calorimeter cluster are retained. For electrons, reconstructed energies and momenta are corrected for bremsstrahlung losses.

4 Performance in official Belle II simulated samples

Figure 3 shows the pion-lepton mis-identification probability (indicated as *fake rate*) as a function of track momentum for likelihood-based (E/p-only in the ECL) and BDT-based PID, in the barrel region. Efficiency for selecting leptons is arbitrarily fixed to 95% in each

Variable	Range	Description
E/p [c]	–	Ratio of cluster energy over track momentum.
$E_{cluster}$ [GeV]	–	Cluster energy.
E_1/E_9	–	Ratio of the energy of the seed crystal over the energy sum of the 9 surrounding crystals.
E_9/E_{21}	–	Ratio of the energy sum of 9 crystals surrounding the seed over the energy sum of the 25 surrounding crystals (minus 4 corners).
$\|Z_{40}\|$	–	Zernike moment $n = 4$, $m = 0$, calculated in a plane orthogonal to the EM shower direction.
$\|Z_{51}\|$	–	Zernike moment $n = 5$, $m = 1$, calculated in a plane orthogonal to the EM shower direction.
Z_{MVA}	–	Score of BDT trained on 11 Zernike moments.
ΔL [mm]	–	Projection on the extrapolated track direction of the distance between the track entry point in the ECL and the cluster centroid.
$\Delta \log \mathcal{L}(\ell/\pi)_{CDC}$	–	Log-likelihood difference between $\ell - \pi$ hypothesis in the CDC.
$\Delta \log \mathcal{L}(\ell/\pi)_{TOP}$	ECL barrel	Log-likelihood difference between $\ell - \pi$ hypothesis in the TOP.
$\Delta \log \mathcal{L}(\ell/\pi)_{ARICH}$	ECL FWD endcap	Log-likelihood difference between $\ell - \pi$ hypothesis in the ARICH.
$\Delta \log \mathcal{L}(\mu/\pi)_{KLM}$	$p > 0.6$ GeV/c	Log-likelihood difference between $\mu - \pi$ hypothesis in the KLM.

Table 1. Description of the input variables for the e/π, μ/π BDTs. The "Range" field indicates whether a variable is defined only in a particular subset of the selected phase space region.

Polar angle $\theta_{cluster}$ [rad]	Track momentum p [GeV/c]
0.21 - 0.56 (ECL FWD endcap)	0.2 - 0.6
0.56 - 2.24 (ECL barrel)	0.6 - 1.0
2.24 - 2.70 (ECL BWD endcap)	> 1.0

Table 2. The polar angle and momentum ranges that define the 9 independent categories where BDTs are trained.

of the three momentum categories for illustrative purposes. In the most interesting low p category, the BDT achieves a reduction in the $\pi \to e$ fake rate by about one order of magnitude with respect to the standard Belle II lepton identification algorithm. In the muon case, the pion fake rate at low momentum is reduced by about a factor two. At higher momenta, the improvement of the BDT is less accentuated, as in the electron case E/p recovers very strong discrimination power against hadrons, and in the muon case the KLM alone achieves by far the best performance for $\pi - \mu$ separation.

5 Summary and outlook

We showed that the combination of several calorimetric measurements together with particle likelihoods from other sub-detectors in a boosted decision tree indicates very promising
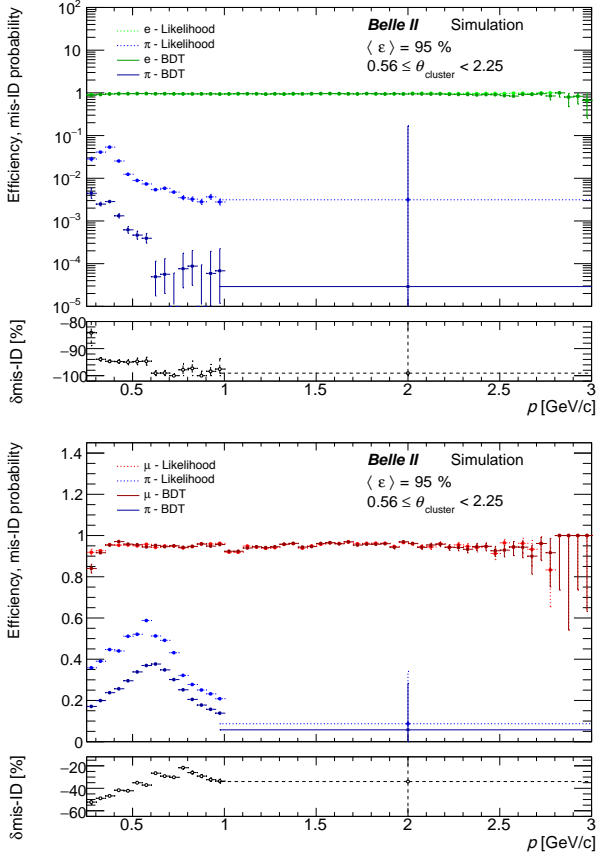
Figure 3. Lepton identification efficiency and $\pi \to e$ (top), $\pi \to \mu$ (bottom) mis-identification probability f as a function of p for likelihood-based and BDT-based PID, in the ECL barrel region. The cut on the classifier is arbitrarily chosen to result in a flat 95% average efficiency for correctly identifying e and μ in each of the three momentum categories. The quantity shown in the bottom pad - $\delta_{\mathrm{mis-ID}} = f_{BDT}/f_{\mathcal{L}} - 1$ - represents the relative difference in mis-identification probability of the BDT method with respect to the likelihood method.

improvements in the Belle II lepton identification performance, especially in the critical low momentum region. The method is fully integrated in the Belle II analysis software framework, and will be employed for the analysis of the early Belle II collision dataset in 2020. Although only tested so far for binary classification, the method can be simply extended to multi-class particle identification. Furthermore, additional discriminating variables recently introduced in Belle II, such as a novel ECL pulse shape discrimination-based classifier [10], will be included to further improve the algorithm perfomance.

We noticed that, despite categorical training, a residual p dependence of the classifer response persists, especially towards the lower end of the momentum spectrum. Achieving a uniform identification efficiency with respect to momentum is desirable, since it can often reduce systematic uncertainties affecting a measurement that is sensitive to the lepton p distribution [11]. To achieve this goal, development of a neural network classifier with mutual information regularisation to de-correlate the classifier score from the particle momentum

is underway. This method has been originally implemented to avoid background sculpting effects in searches of rare loop-induced decays involving $b \rightarrow s\gamma$ and $b \rightarrow s\ell^-\ell^+$ transitions [12].

References

[1] T. Abe *et al.*, *Belle II Technical Design Report*, 1011.0352 (2010).

[2] K. Akai, K. Furukawa, H. Koiso, *SuperKEKB Collider*, Nucl. Instrum. Meth., **A907**, 188-199 (2018).

[3] E. Kou, P. Urquijo, *The Belle II Physics Book*, Progress of Th. and Exp. Phys. (PTEP) **12** (2019).

[4] V. Lakshminarayanan, A. Fleck, *Zernike polynomials: a guide*, Journal of Modern Optics 58 **7**, 545561 (2011).

[5] D.N. Brown, J. Ilic, G.B. Mohanty, *Extracting longitudinal shower development information from crystal calorimetry plus tracking*, 0711.1593 (2007).

[6] A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, CERN-OPEN-2007-007 (2007).

[7] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, N. Braun, *The Belle II Core Software*, Comput. Softw. Big Sci. **3**, 1 (2019).

[8] L. Wood, T. Elsethagen, M. Schram, E. Stephan, *Conditions Database for the Belle II Experiment*, Journal of Physics: Conference Series, **898** (2017).

[9] R. Brun, F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth., **A389**, 81-86 (1997).

[10] S. Longo, J. Roney, *Hadronic vs. electromagnetic pulse shape discrimination in CsI(Tl) for high energy physics experiments*, Journal of Instr. 3, **13**, (2018).

[11] A. Rogozhnikov, A. Bukva, V.V. Gligorov, A. Ustyuzhanin, M. Williams, *New approaches for boosting to uniformity*, Journal of Instr. 3. **10**, (2015).

[12] J. Tan, P. Urquijo, *Invariance through Mutual Information Regularization*, Presentation at the 3rd IML Machine Learning Workshop, CERN (2019).