# Muon and electron identification efficiencies and hadron-lepton mis-identification rates at Belle II for Moriond 2022

Charged PID Group, The Belle II Collaboration

## Abstract

We present a collection of selected results on the performance of Belle II in the identification of electrons and muons. This work is carried out using data collected at the Belle II experiment from 2019 to mid-2021. The dataset corresponds to a total integrated luminosity of $\int L \, \mathrm{dt} = 190 \, \mathrm{fb}^{-1}$ collected at the centre-of-mass energy of the $\Upsilon(4S)$, with an additional sample corresponding to $\int L \, \mathrm{dt} = 18 \, \mathrm{fb}^{-1}$ of off-resonance data collected over the same period.

## 1. DATASET AND DEFINITIONS

The Belle II detector [1] is located around the interaction region of the asymmetric energy SuperKEKB electron-positron collider [2], at the KEK laboratory in Tsukuba, Japan. After a successful commissioning phase in 2018, the experiment has been collecting data at the centre-of-mass (CM) energy of (or nearby) the $\Upsilon(4S)$ resonance since 2019. Electrons and positrons are accelerated at the SuperKEKB collider to energies of $7\,\text{GeV}$ and $4\,\text{GeV}$, respectively, to boost the CM frame relative to the laboratory frame to $\beta\gamma = 0.28$.

In this document we present lepton identification (ID) studies ($\ell^\pm = \{e^\pm, \mu^\pm\}$) performed using the Moriond 2022 on-resonance datasets collected in 2019 (experiments 7, 8 10), 2020 (experiments 12, 14) and the first half of 2021 (experiments 16, 17, 18) amounting to 190 fb$^{-1}$. The $D^{*+}$ channel utilises also an off-resonance dataset amounting to 18 fb$^{-1}$.

The Belle II detector is comprised of several sub-detector components arranged cylindrically around the interaction region. The vertex detector (VXD), the innermost detector element of Belle II, consists of two layers of silicon pixel detectors (PXD) and four layers of double-sided silicon strips detectors (SVD). In the data taking period discussed in this document, only the innermost layer of the PXD was fully installed while, for the second layer, only two sensors were installed. However, the PXD and the SVD were not used for particle identification for the results shown here. The central drift chamber (CDC) is filled with a helium-based gas mixture for tracking charged particles and contributes to their identification via energy loss measurements ($dE/dx$). The time-of-propagation Cherenkov detector (TOP), consisting of 16 bars of fused silica, and the Aerogel Ring Imaging Cherenkov detector (ARICH), are instrumental in the identification of charged hadrons. We note that the TOP detector was excluded for likelihood-based electron identification in this document due to an issue with the description of the probability density function of the electron hypothesis in the software. An electromagnetic calorimeter (ECL) consisting of 8,736 Thallium-doped CsI crystals distributed in a barrel and two endcaps (forward/backward) is used mainly for the identification of electrons/positrons and photons. Finally, the $K_L^0$ and muon detector (KLM) consists of a sandwich-like structure of alternating metal plates and active detector elements based on resistive plate chambers. A superconducting solenoid, situated between the ECL and the KLM, provides a 1.5 T axial magnetic field. A detailed description of the full detector is given in [1]. Information from each particle identification system $D = \{\text{CDC}, \text{TOP}, \text{ARICH}, \text{ECL}, \text{KLM}\}$ is analysed independently to determine the likelihood of each charged particle hypothesis. These likelihoods may then be used to construct a combined global likelihood ratio [3]:

$$\ell\text{ID} = \frac{\mathcal{L}_\ell}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d}, \tag{1}$$

where:

$$\mathcal{L}_i = \prod_d^{d\in D} \mathcal{L}_i^d, \quad i \in \{e, \mu, \pi, K, p, d\} \tag{2}$$

A new method of lepton identification based on the combination of several ECL measurements - such as shower shapes and pulse shape discrimination variables - with the other sub-detector likelihoods in a boosted decision tree [4] (BDT) is also considered. The BDT is trained both in multi-class mode to separate leptons from all other competing charged particle hypotheses, as well as in binary $\ell$-vs.-$\pi$ mode. Only results for the former method

are shown in this document: the identification variables are indicated as $e$BDT and $\mu$BDT, respectively.

We report the lepton identification performance of electron-hadron, and muon-hadron separation ($h^{\pm} = \{\pi^{\pm}, K^{\pm}\}$) using a complementary set of decay channels. Efficiency and mis-identification rate are defined as the probability for an electron (muon) track to be correctly identified as such, and the probability for a hadron track to be wrongly identified as a lepton track. Electron and muon identification efficiencies are studied using $e^+e^- \to \ell^+\ell^-(\gamma)$, $e^+e^- \to e^+e^-\ell^+\ell^-$, and $J/\psi \to \ell^+\ell^-$, while pion mis-identification rates are studied using $K_S^0 \to \pi^+\pi^-$ and $e^+e^- \to \tau^{\pm}(1P)\tau^{\mp}(3P)$. The $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ channel is used to determine kaon mis-identification rates.

Performance is evaluated in the polar angle acceptance regions corresponding to the electromagnetic calorimeter (ECL) for electrons (0.22 to 2.71 rad), and to the $K_L^0$-muon detector (KLM) for muons (0.40 to 2.60 rad). Combined, the set of probe channels covers a lab-frame momentum range of 0.4 GeV/$c$ to 7.0 GeV/$c$ for electrons and of 0.4 GeV/$c$ to 6.5 GeV/$c$ for muons. Results are also binned with respect to the track lab-frame polar angle and measured track charge. However, all results shown in this document refer to the central "barrel" region of the detector: $0.56 \leq \theta < 2.23$ rad for electrons, $0.82 \leq \theta < 2.22$ rad for muons.

Lepton identification performance is studied for different selection working points tuned in simulation to achieve uniform 60%, 70%, 80%, 90%, 95% identification efficiency across track momentum, charge and polar angle-binned phase space. For the likelihood-based ID only, we also study the performance for fixed cuts on the $\ell$ID variable of greater than 0.5, 0.9 and 0.95.

For brevity, the plots presented here show results for a selection of uniform 80% efficiency, and $\ell$ID $> 0.9$. The description of the analysis of each channel is provided elsewhere [5]. Here we show the combination of all results assessing the lepton identification performance of Belle II in terms of efficiency and fake rates obtained as a function of particle momentum.

### 1.1. Results overview

We have evaluated efficiency and mis-identification rates as a function of track polar angle $\theta$, lab-frame momentum $p$ and charge. For conciseness, the figures only show results for negatively charged tracks. We overlay efficiencies and mis-identification rates for all channels in the polar angle region corresponding to the barrel for electron identification in Fig. 1, for muon identification in Fig. 2.

In Fig. 3 we also overlay the efficiency and mis-identification rates for the hadronic channels only - $J/\psi$, $K_S^0$, $D^*$ - for $e$ID $> 0.9$ in the ECL barrel region and $\mu$ID $> 0.9$ in the KLM barrel region. When integrating over the track momentum and charge, the average lepton identification efficiency measured in the $J/\psi$ sample in the barrel region is:
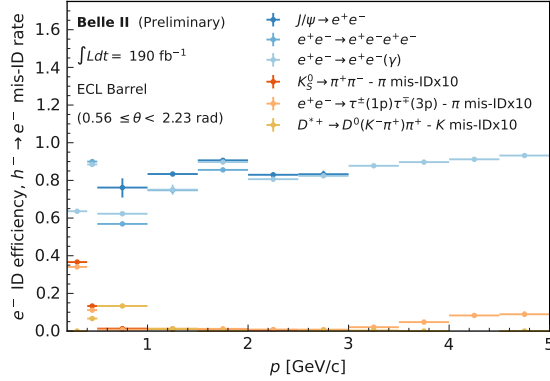
$$\epsilon(e^{\pm}) = \left(8.60^{+0.03}_{-0.03} \text{ (stat.)}^{+0.02}_{-0.01} \text{ (syst.)}\right) \times 10^{-1} \tag{3}$$

$$\epsilon(\mu^{\pm}) = \left(8.85^{+0.03}_{-0.02} \text{ (stat.)}^{+0.01}_{-0.01} \text{ (syst.)}\right) \times 10^{-1}, \tag{4}$$

with a pion mis-identification rate measured in the $K_S^0$ sample of:
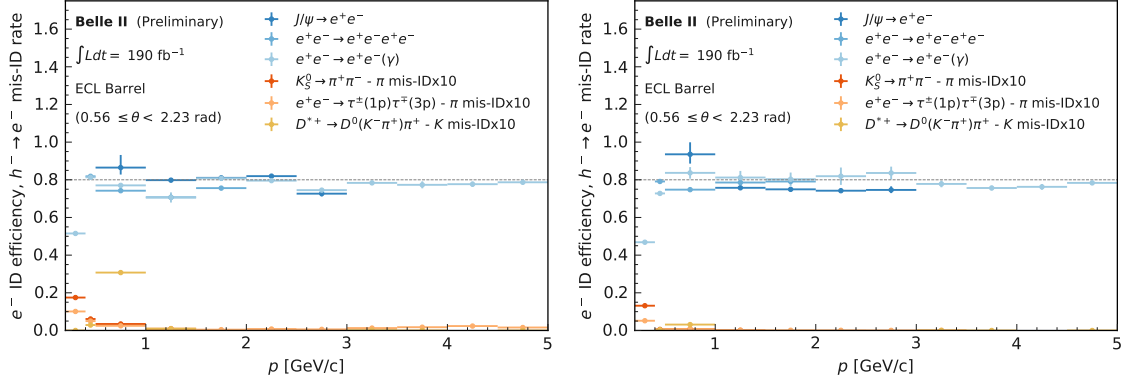
(a) *e*ID, uniform 80% efficiency selection.



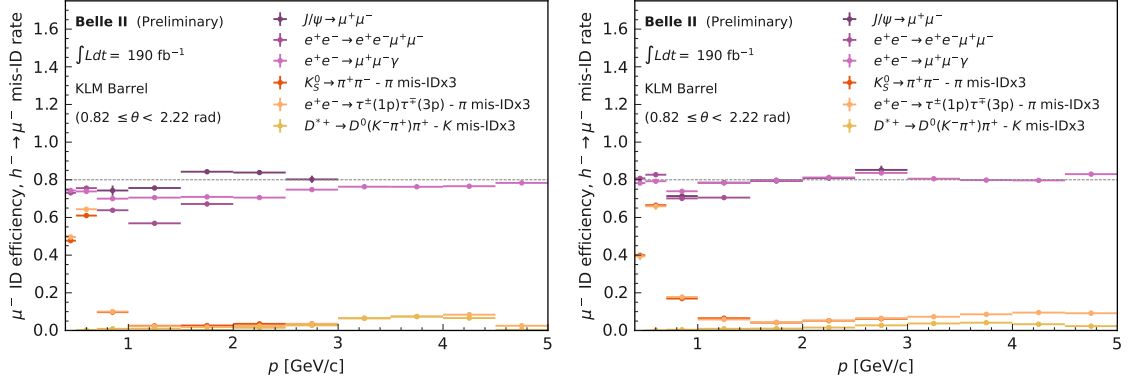(b) *e*BDT, uniform 80% efficiency selection.



(c) *e*ID > 0.9 selection.

FIG. 1: Electron identification and hadron-electron mis-identification rates from all the channels overlaid, ECL barrel region ($0.56 \leq \theta < 2.23$ rad). Top row: uniform 80% efficiency selection across $(p, \theta)$ bins (tuned in simulation), comparing likelihood-based ID (left) with BDT-based ID (right). Bottom row: *e*ID > 0.9 selection, likelihood-based ID only. Note that the mis-identification rate is multiplied by a factor of 10 for illustration purposes. The momentum range is clipped at 5 GeV/$c$.

$$\text{mis} - \text{ID}(\pi^{\pm} \to e^{\pm}) = \left(4.11^{+0.20}_{-0.20} \text{ (stat.)}^{+<0.01}_{-<0.01} \text{ (syst.)}\right) \times 10^{-3} \qquad (5)$$

$$\text{mis} - \text{ID}(\pi^{\pm} \to \mu^{\pm}) = \left(7.33^{+0.02}_{-0.01} \text{ (stat.)}^{+0.01}_{-0.01} \text{ (syst.)}\right) \times 10^{-2}. \qquad (6)$$

A combination of the ratios of efficiencies and mis-identification rates between data and MC in each bin from all channels is done following the procedure outlined in Ref. [6], approximating the likelihoods of each measurement as Gaussian functions and under the following assumptions: the measurements are statistically independent; if more than two measurements can be combined in a specific bin, then the combination is done associatively; systematic errors sources are independent in a given bin. The first condition is satisfied because each measurement applies its selection criteria, and these are different and independent of the others; consequently the phase-space regions selected are different. The independence of the systematic source is guaranteed by the fact that they are specific to the analysis methodology followed in each measurement. If any measurement in a bin is not consistent

4

(a) $\mu$ID, uniform 80% efficiency selection.

(b) $\mu$BDT, uniform 80% efficiency selection.
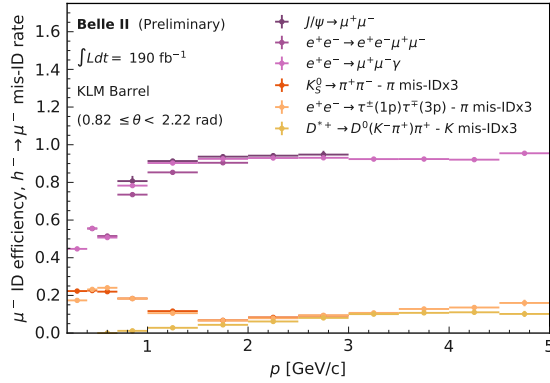


(c) $\mu$ID > 0.9 selection.

FIG. 2: Muon identification and hadron-muon mis-identification rates from all the channels overlaid, KLM barrel region ($0.82 \leq \theta < 2.22$ rad). Top row: uniform 80% efficiency selection across $(p, \theta)$ bins (tuned in simulation), comparing likelihood-based ID (left) with BDT-based ID (right). Bottom row: $\mu$ID > 0.9 selection, likelihood-based ID only. Note that the mis-identification rate is multiplied by a factor of 3 for illustration purposes. The momentum range is clipped at 5 GeV/$c$.

within $3\sigma$ with the combination, we assign an extra systematic uncertainty as the "distance" between the central value of the combination to the minimum (maximum) central values among individual methods in each bin. We consider this discrepancy between channels to be caused predominantly by different levels of activity in the detector from nearby tracks and clusters around particle candidates selected in hadronic events - such as in $J/\psi$, $K_s^0$, $D^*$ decays - as opposed to candidates from low multiplicity events. These differences are of $\mathcal{O}$(few %) in some of the bins, and are currently being thoroughly investigated. Results for the $e$ID > 0.9, $\mu$ID > 0.9 efficiency ratios between data and simulation as a function of momentum in the barrel region, together with the combination of individual channels, are shown in Fig.4.
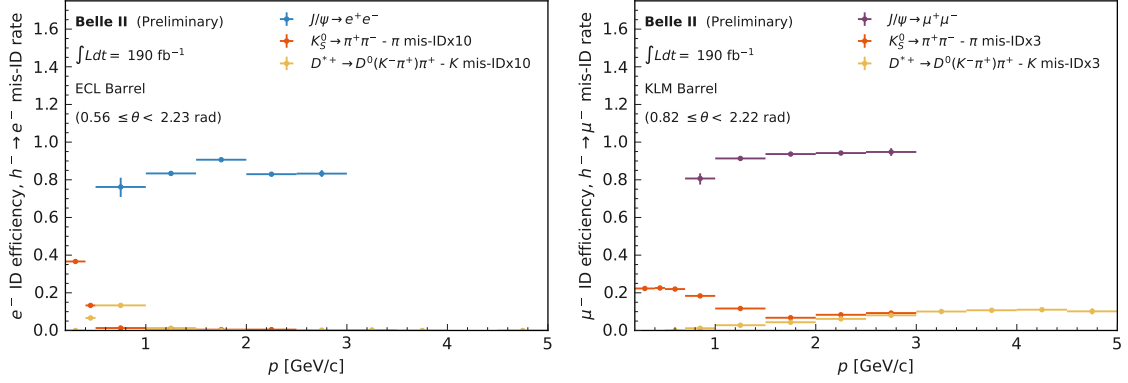
FIG. 3: $J/\psi$-based lepton ID efficiency, $K_S^0$-based pion fake rate and $D^*$-based kaon fake rate overlay as a function of track momentum. Left: $e$ID > 0.9 in the ECL barrel region ($0.56 \leq \theta < 2.23$ rad). Note that the hadron mis-identification rate has been inflated by a factor 10 for illustration purposes. Right: $\mu$ID > 0.9 in the KLM barrel region ($0.82 \leq \theta < 2.13$ rad). Note that the hadron mis-identification rate has been inflated by a factor 3 for illustration purposes.
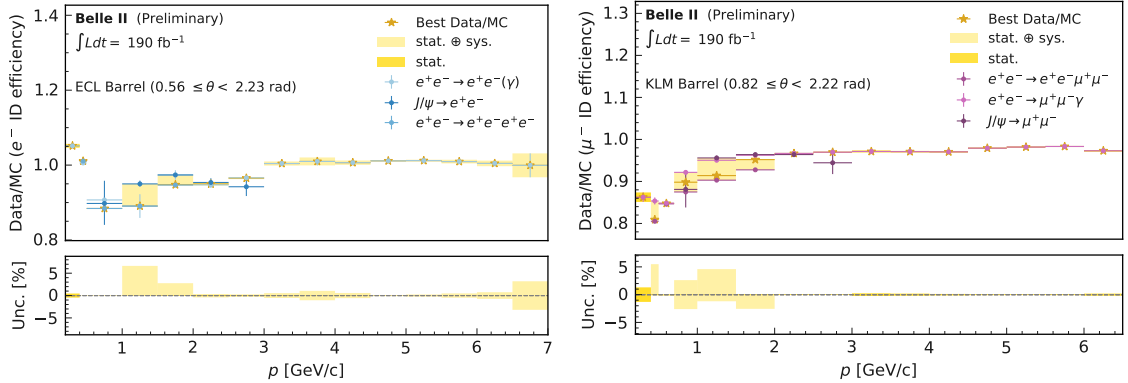


FIG. 4: The efficiency ratio between data and MC for individual channels and their combination as a function of track momentum. Left: $e$ID > 0.9 in the ECL barrel region ($0.56 \leq \theta < 2.23$ rad). Right: $\mu$ID > 0.9 in the KLM barrel region ($0.82 \leq \theta < 2.13$ rad). The yellow starred markers represent the best available measurement in each bin, which can be either the result of the combination of multiple channels, or an individual channel if that is the only available. The light yellow shaded band represents the total statistical plus systematic uncertainty on the best available result, the darker yellow band is the statistical uncertainty only.

### 1.2. Conclusion

We have presented the status of lepton identification efficiencies and hadron mis-identification rates at Belle II based on a likelihood classifier method as well as a new BDT-based method. A broad set of calibration channels in different event topologies have been analysed. In the benchmark barrel region of the detector, we measure lepton identification efficiencies with $J/\psi \to \ell\ell$ to be $\epsilon(e^\pm) = \left(8.60^{+0.03}_{-0.03}\,(\text{stat.})^{+0.02}_{-0.01}\,(\text{syst.})\right) \times 10^{-1}$ for electrons with $e$ID > 0.9, $\epsilon(\mu^\pm) = \left(8.85^{+0.03}_{-0.02}\,(\text{stat.})^{+0.01}_{-0.01}\,(\text{syst.})\right) \times 10^{-1}$ for muons

with $\mu$ID $>0.9$, corresponding to pion mis-identification rates with $K_S^0 \rightarrow \pi^+\pi^-$ of $\mathrm{mis-ID}(\pi^\pm \rightarrow e^\pm) = \left(4.11^{+0.20}_{-0.20}\ (\text{stat.})^{+<0.01}_{-<0.01}\ (\text{syst.})\right) \times 10^{-3}$ for electron candidates, and $\mathrm{mis-ID}(\pi^\pm \rightarrow \mu^\pm) = \left(7.33^{+0.02}_{-0.01}\ (\text{stat.})^{+0.01}_{-0.01}\ (\text{syst.})\right) \times 10^{-2}$ for muon candidates. As for the lepton identification efficiency, the MC simulation is found to be fairly consistent with data, with the largest differences - within 20% - appearing mostly at very low momenta; differences between calibration channels are accounted for as systematic uncertainty, and are of order few %. Further studies to reduce these inconsistencies between individual calibration channels are underway.

[1] T. Abe (Belle II Collaboration) (2010), arXiv:1011.0352.

[2] Y. Ohnishi, T. Abe, T. Adachi, K. Akai, Y. Arimoto, K. Ebihara, K. Egawa, J. Flanagan, H. Fukuma, Y. Funakoshi, et al., Progress of Theoretical and Experimental Physics **2013** (2013), ISSN 2050-3911, URL `https://doi.org/10.1093/ptep/pts083`.

[3] E. Kou, P. Urquijo, et al. (Belle-II), PTEP **2019**, 123C01 (2019), [Erratum: PTEP 2020, 029201 (2020)], arXiv:1808.10567.

[4] M. Milesi, J. Tan, and P. Urquijo, EPJ Web Conf. **245**, 06023 (2020), URL `https://doi.org/10.1051/epjconf/202024506023`.

[5] LeptonID group and Belle II Collaboration, BELLE2-CONF-PH-2021-002 (2021).

[6] R. Barlow, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology* (2004), physics/0406120.