Lepton Identification Performance at Belle II

Miho Wakai

King's College London

September 6th, 2018

Supervisor: Dr. Ilya Komarov

Abstract

The Belle II experiment is located in Japan, using an asymmetric e^+e^- SuperKEKB collider. Originally designed to study precise measurements for CP violation in B mesons, the experiment will also conduct studies on beyond the standard model particles, including the dark sector. This document presents the analysis of lepton identification studies at Belle II, which its data will be used to correctly identify the flavour of leptons. Furthermore, this analysis is intended to contribute to the lepton flavour violating dark matter search at Belle II.

Contents

List of Tables

List of Figures

1 Super KEKB

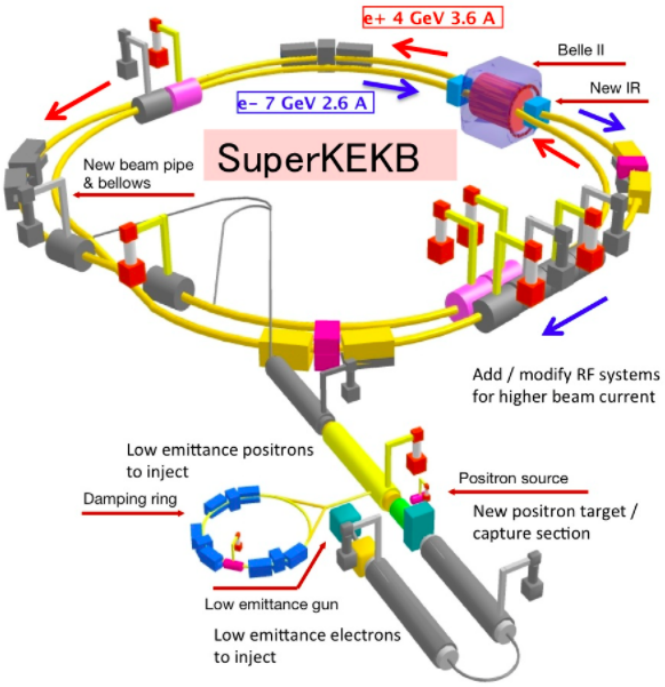

Figure 1: Overview of the SuperKEKB collider [1].

The Super KEKB, shown in Figure 1 is the particle accelerator where the experiment is held. Located in Japan, the collider has a circumference of 3 km, with e^+ and e^- colliding with mostly at a resonance energy of $\Upsilon(4S)$, which is 10.58 GeV [2]. The collider operates with a certain range in center of mass energy to obtain different interactions. Super KEKB is an upgrade from the previous collider KEKB, which the luminosity has been increased by a factor of 40, making it $8 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The KEKB has the current world record for its luminosity, which makes the Super KEKB the next world record holder. The e^+ and e^- have a crossing angle of 83 mrad, and the e^+ has significantly less energy than the e^-. The asymmetry in the energies of the e^+ and e^- are significant as this provides a "boost" to the particles produced. This boosting is significant when searching for CP violation by different particles, especially in B mesons, where the boost causes the B meson pair to travel a distance, allowing a time dependent measure of CP asymmetry [3]. Each beam contains "bunches" of electrons and positrons, for e^- being 6.5×10^{10} and e^+ being 9.0×10^{10}. Each bunch is separated by 4 ns.

2 Belle II

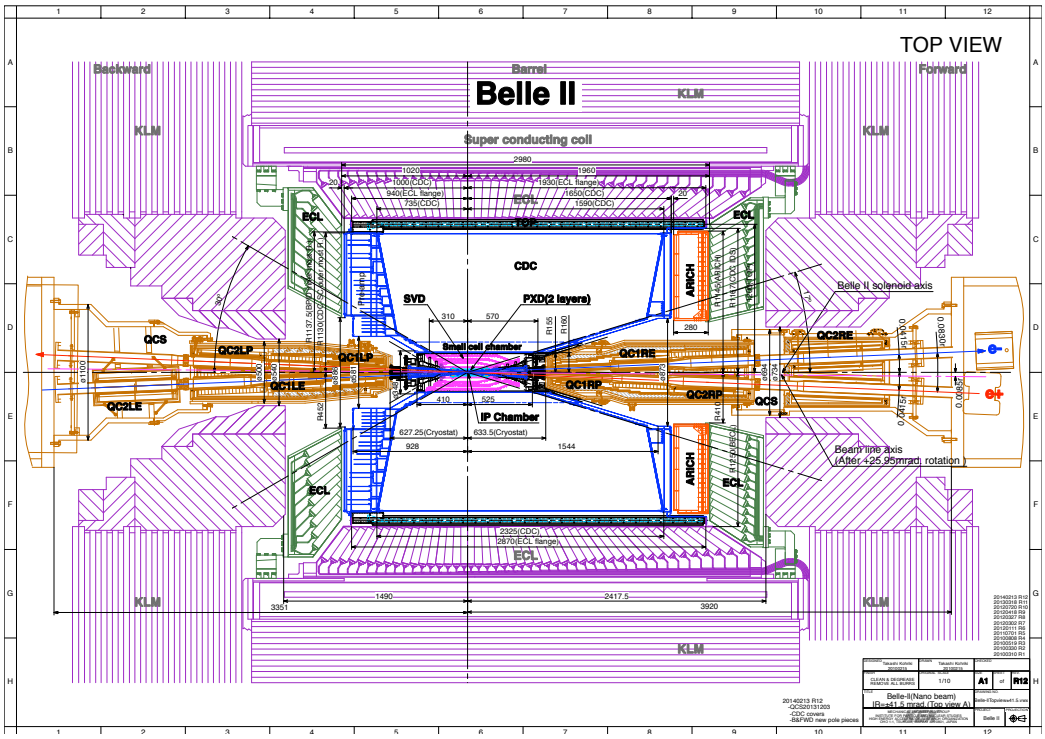

Figure 2: Overview of the Belle II Detector [4].

The Belle II detector has a width of around 7.5 m and a height of 7 m shown in Figure 2. The main parts of the detector relevant for this study are the central drift chamber (CDC), electromagnetic calorimeter (ECL) and the KLong Muon Detector (KLM). The CDC allows precise measurement of the momentum of charged particles. The main purpose of the ECL is to measure energy of e^- and γ that are completely absorbed by the detector. Photons and charged particles in the ECL hit the CsI(Tl) crystals and create showers of particles. There are photodetectors attached to the crystals in order to detect the scintillation light produced from the deposited energy, which is converted to an electrical signal and is used for analysis [5].

3 Lepton Flavour Violating Dark Matter (LFVDM)

Recently, an excess of γ-rays from the Galatic Center was observed by an independent analysis of the Fermi Large Area Telescope (LAT) data and directly by the Fermi-LAT collaboration [6]. An explanation for this can be made using light flavour violating bosons, which this mediator will interact with the Standard Model through chiral lepton flavour violating couplings. At Belle II, we aim to search for this lepton flavour violating mediator which will then produce two dark matter particles, represented as a Feynman diagram in Figure 3. Assuming the Z' particle to be invisible, when reconstructing the $e^+\mu^-$ we will see a peak for a missing momentum which was carried by the Z'. We are currently in the stage of preparing for analysis through understanding the tracking of charged particles, trigger efficiency, the background events, and lepton identification.

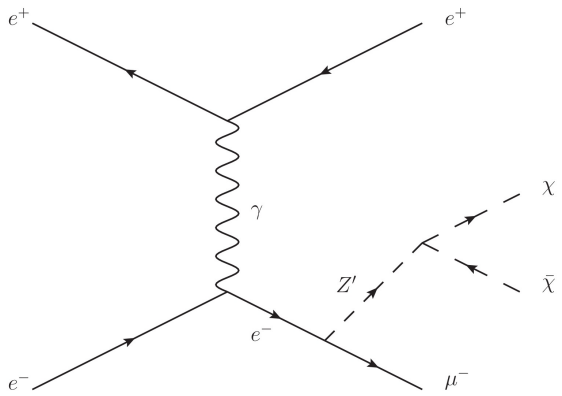

Figure 3: Feynman diagram for lepton flavour violating dark matter. Courtesy of Mr. Giacomo De Pietro.

4 Project

My contribution was to conduct analysis on lepton identification performance (LID). Lepton identification is used to differentiate between different flavour leptons. The main variable which the identification currently relies on is the E/p variable, which is the ratio of the energy deposited in the ECL to the momentum measured by the CDC. As electrons are light and fully absorbed by the ECL, the ratio of these two numbers must be close to 1, while muons are highly penetrative and does not deposit much energy in the ECL, hence the ratio would be smaller than electrons.

In the future, lepton identification in Belle II will be done through the variables depending on the combined likelihood of the particle to be of a certain flavour from the subdetetectors. Reliable combination of the likelihoods is not yet implemented in the Belle II software, but in this work we examine the likelihood obtained from the CDC and ECL detetctors as such variables. Comparing various LID variables, we choose the best variable for the lepton identification that can be used for early data analysis, such as searches for LFV Z'. We compare performance of selection on that variable in data and simulations and provide corrections necessary for quantitative predictions based on simulations. The performance study presented here became a base for the calibration framework, allowing

Belle II analysis on LIDs. This framework can be used by physicists to perform analysis using arbitrary particle identification variables (PID) as a function of all possible kinematic variables.

5 Analysis

Analysis was conducted using the Belle II analysis software framework, as well as python, especially the pandas package [7]. This was all documented using Jupyter Notebook.

5.1 Event Selection

Both data and Monte-Carlo simulated samples were used for analysis. For data, we compared the effects of different software versions for the LID performances (Production 4, Production 5). Six different event types for MC were explored, listed below.

- ee$\rightarrow ee$: Bhabha event
- ee$\rightarrow eeee$: 2γ event
- ee$\rightarrow \mu\mu$: Two muon event
- ee$\rightarrow ee\mu\mu$: 2γ with muon pair production
- ee$\rightarrow \pi\pi$: Two pion event
- ee$\rightarrow \tau\tau$: Two tau event

MC Event	Cross Section [nb]
ee$\rightarrow ee$	300
ee$\rightarrow eeee$	39.7
ee$\rightarrow \mu\mu$	1.148
ee$\rightarrow ee\mu\mu$	18.9
ee$\rightarrow \pi\pi$	0.102
ee$\rightarrow \tau\tau$	0.919

Table 1: Cross section for MC events [8].

The data and MC files were all produced by a steering file which contains initial cuts. Tracks used for event reconstruction indicates that the event must be produced within 2 cm along the beam axis and within 0.5 cm of the radial distance from the interaction point.

To select events important for this analysis, we applied the cuts on the individual and combined kinematics of the two tracks. For performance studies, our interest was in the 2γ events, which we wanted a clean sample of. The previous Belle I analysis [9] listed below was implemented to select these events.

- At least one of the two tracks satisfies $|\overrightarrow{p}| \geq 0.5\,\mathrm{GeV/c}$
- $|\overrightarrow{p}^{*}_{\mathrm{T}}| \geq 0.25\,\mathrm{GeV/c}$
- $|\overrightarrow{p}^{*}_{\mathrm{T(tot)}}| = |\overrightarrow{p}^{*}_{\mathrm{T+}} + \overrightarrow{p}^{*}_{\mathrm{T\text{-}}}| \geq 0.2\,\mathrm{GeV/c}$

- $|\overrightarrow{p}^{*}_{\mathrm{z(tot)}}| = |\overrightarrow{p}^{*}_{\mathrm{z+}} + \overrightarrow{p}^{*}_{\mathrm{z\text{-}}}| \leq 2.5\,\mathrm{GeV/c}$
- $|\overrightarrow{p}^{*}_{\mathrm{(tot)}}| = |\overrightarrow{p}^{*}_{+} + \overrightarrow{p}^{*}_{-}| \leq 5.0\,\mathrm{GeV/c}$
- $0.6\,\mathrm{GeV} \leq E^{\mathrm{cluster}}_{\mathrm{tot}} \leq 6.0\,\mathrm{GeV}$

Three tighter cuts were also added which further reduces background.

- $0.6\,\mathrm{GeV} \leq E^{\mathrm{cluster}}_{\mathrm{tot}} \leq 4.0\,\mathrm{GeV}$
- $|\overrightarrow{p}^{*}_{\mathrm{T(tot)}}| = |\overrightarrow{p}^{*}_{\mathrm{T+}} + \overrightarrow{p}^{*}_{\mathrm{T\text{-}}}| \leq 1.5\,\mathrm{GeV/c}$
- Number of clean tracks = 2

In addition, we have replaced $|\overrightarrow{p}^{*}_{\mathrm{T(tot)}}|$ as the total visible momentum of event in CMS, which is the absolute value of the sum of the individual CMS momentum for both particles.

The highest discriminating power belongs to the cut on $\mathrm{E}^{\mathrm{cluster}}_{\mathrm{tot}}$, which the distribution for the MC events are shown in Figure 4 as a variable called visible energy of events. This cut heavily reduces the background from pion events.

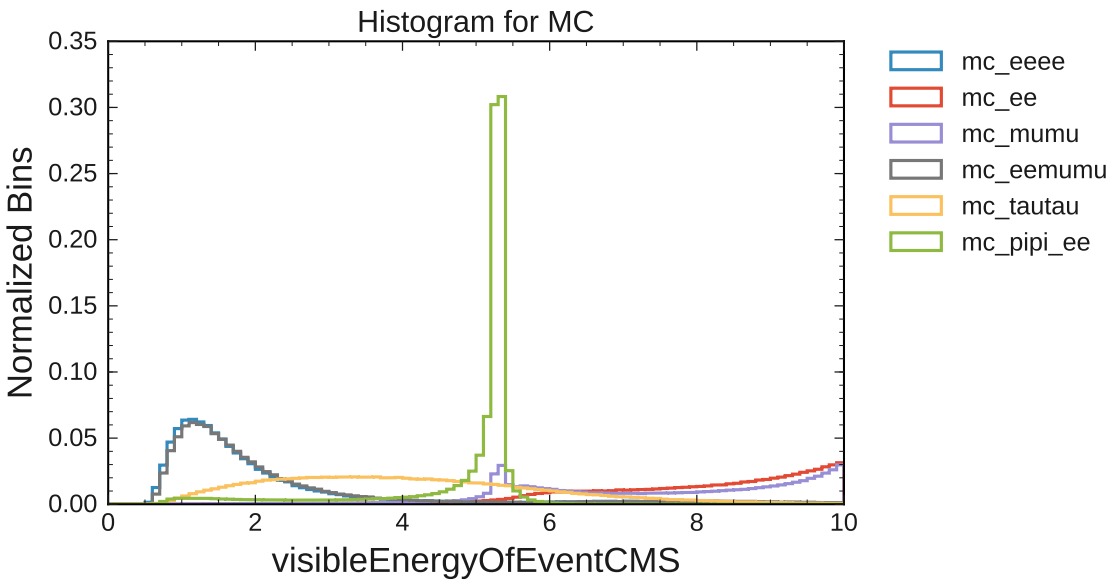

Figure 4: Normalized histogram for visible energy of events in MC when no initial cuts are applied.

5.2 Tag and Probe Method

The tag and probe method was used to define the performance of LID. For this method, we need to have a clean sample of tracks with known flavour selected without any initial cuts on PID variables. Using the selection on kinematics described in the previous section, we get a clean sample of 2γ events. To differentiate $\gamma \rightarrow$ee and $\gamma \rightarrow \mu\mu$, we apply a further cut on one of the tracks requiring it to be of the given flavour (μ and e). This is the "tag" track. Since the flavour of the leftover track ("probe" track) is defined by the flavour of the tag track, we can study the efficiency of the LID cut on the probe track knowing both number of tracks before and after the selection. The cuts applied for tagging and probing for e^- and μ^- are listed below.

- Tag on e^+: $|\mathrm{E/p} - 1| < 0.2$
- Probe on e^-: $0.8 \leq \mathrm{E/p} \leq 1.2$
- Probe on μ^-: $0.15 \leq \mathrm{E} \leq 0.4$ and $\mathrm{E/p} < 0.4$
- Probe on μ^+: $0.15 \leq \mathrm{E} \leq 0.4$ and $\mathrm{E/p} < 0.4$

5.3 Distribution of Data and Monte Carlo

Distribution of simulated and recorded events in E/p and E variables allows us to understand the composition of the data.

Figure 5 shows the data and MC without any cuts applied. The particle of interest, for this case being the electron, is plotted for variables E/p against p. Each of the MC events are weighted according to their cross section, listed in Table 1. The numbers shown in all the MC events are the expected percentage of events to appear in data. This seems to mostly agree, as the distribution of data and bhabha events labeled as mc_ee are showing similar events. As this is a logarithmic scaled plot, it was interesting to see the two curves on both data and MC. The smaller curve at around 4 GeV/c is most likely to be misidentified positrons as electrons, as this is the beam energy of the positrons. Furthermore, there is a horizontal line in the 2 γ event labelled as mc_eemumu, which is displaying an electron. Therefore this suggests that there are events where when reconstructing the event, an electron was taken as a daughter particle instead of a muon.

Figure 6 shows the data and MC with the additional cuts applied explained in the previous section. Many of the misidentified electrons were eliminated by this cut in mc_eemumu, showing a cleaner sample of the muons. Although this cut sufficiently eliminates most background events, many of the Bhabha events (mc_ee) still passes these cuts. Hence a further cut was necessary to reduce this background.

Figures 7 and 8 represents the tag and probe method used for analysis. It is clear from the distributions that through this method we can extract a clean sample of the event of interest.

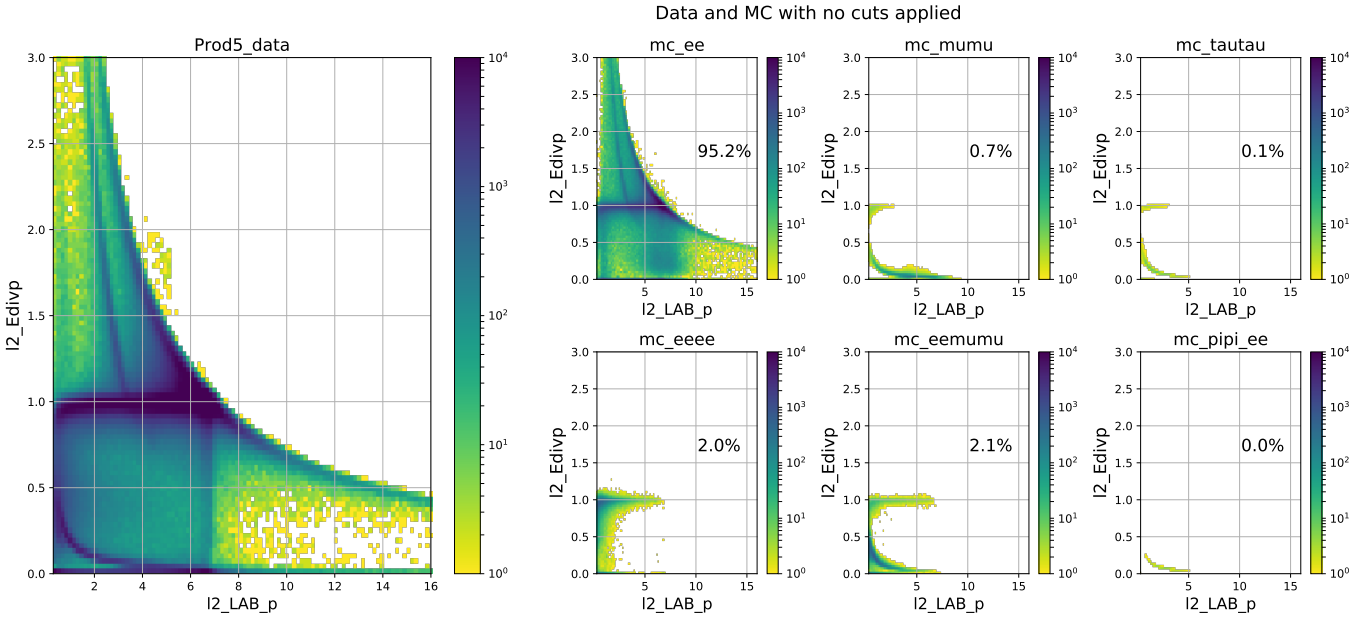

Figure 5: E/p against p for data and MC.

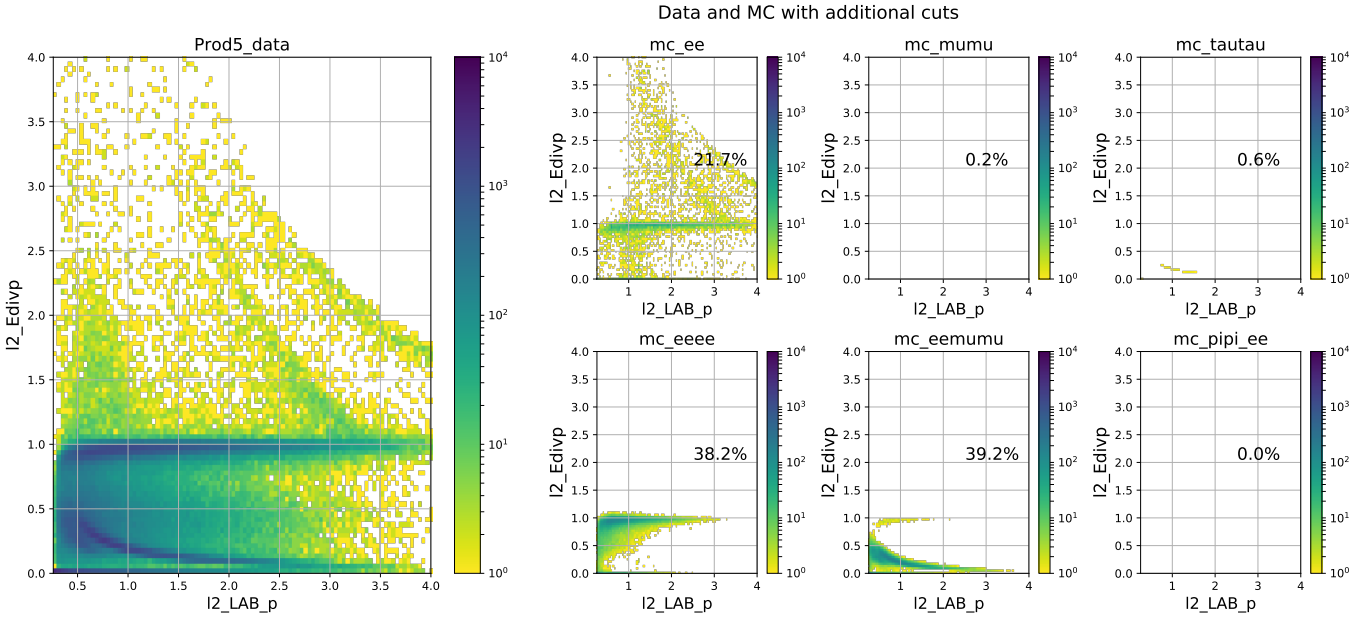

Figure 6: E/p against p with for data and MC with additional cuts.

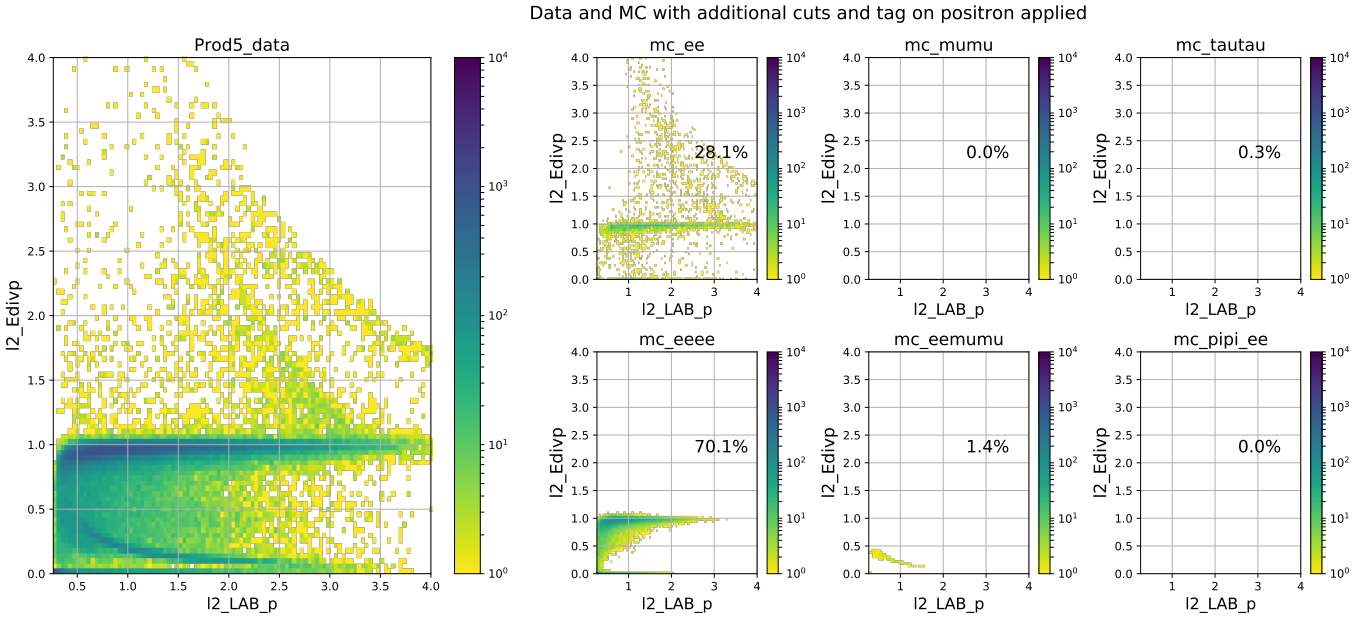

Figure 7: E/p against p for data and MC with additional cuts and a further e^+ probe cut.

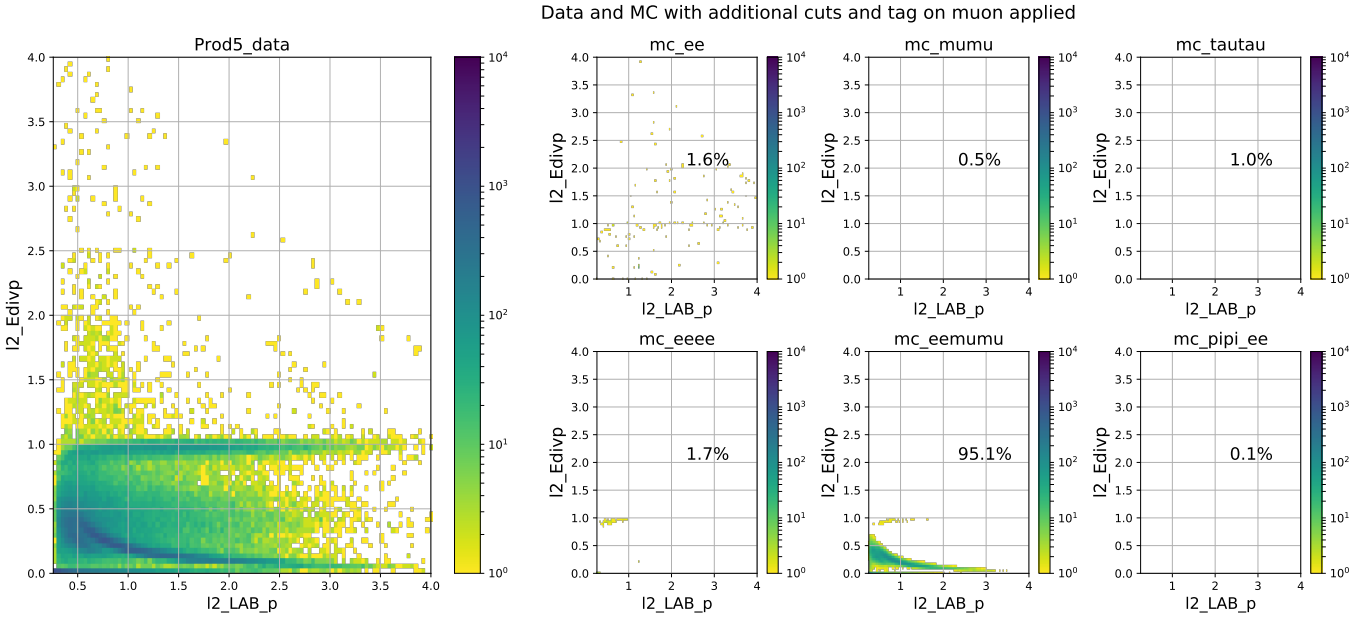

Figure 8: E/p against p for data and MC with additional cuts and a further μ^+ probe cut.

5.4 Distribution of Lepton Identification Performance

The Belle II detector has three main regions, defined as forward, barrel, and backward. By taking a clean sample of the 2γ events, we compared the distributions for the LID variables of the probe particle when it is in different detector regions. Furthermore, we compared the efficiency scan for these regions, as well as the efficiency ratios between data and MC. The results can be seen in Figure 9, where the variable of cluster E, which is the total energy deposited in the calorimeter, divided by lab frame momentum (E/p) is used. We conclude that the quality depends on the different detector regions and therefore the E/p variable cannot be used in all regions.

Using the same approach, we repeated the same exercise and created distributions for different PID variables, namely electron and muon likelihoods defined by the ECL subdetector in the different subdetector regions. By making a comparison with the E/p distributions, it was clear that the particle identification based on likelihoods are less reliable, and still needs improvement to tag the leptons correctly. It is interesting to note that the ECL electron ID variable (ECL_PID) shown in Figure10 should heavily rely on the E/p value obtained in its detector, therefore this has raised some questions regarding the reliability of the information being passed within the detector, where further research is necessary.

5.5 Correction Ratio

During analysis, the efficiency for LID cuts applied on data and MC behaved differently, which needs correction. This correction ratio showing the different behaviour in data and MC may also depend on the phase space of the track. Therefore we created a mapping of these correction ratios shown in Figure 11 in the most sensitive region which is the barrel. The corrections were calculated in bins of cos θ and momentum. This mapping procedure is listed below.

1. Choose cut on the probe particle (e^-: $0.8 \leq \mathrm{E/p} \leq 1.2$)
2. Obtain the efficiency of the cut for both data and MC using (1)
3. Obtain the efficiency ratio of data and MC with uncertainties
4. Map these ratios among two different variables (Ex. cos θ of the angle between the momentum of the track and beam direction against lab frame momentum of particle)

$$\epsilon = \frac{N_{\text{tag \& probe}}}{N_{\text{tag}}} \tag{1}$$

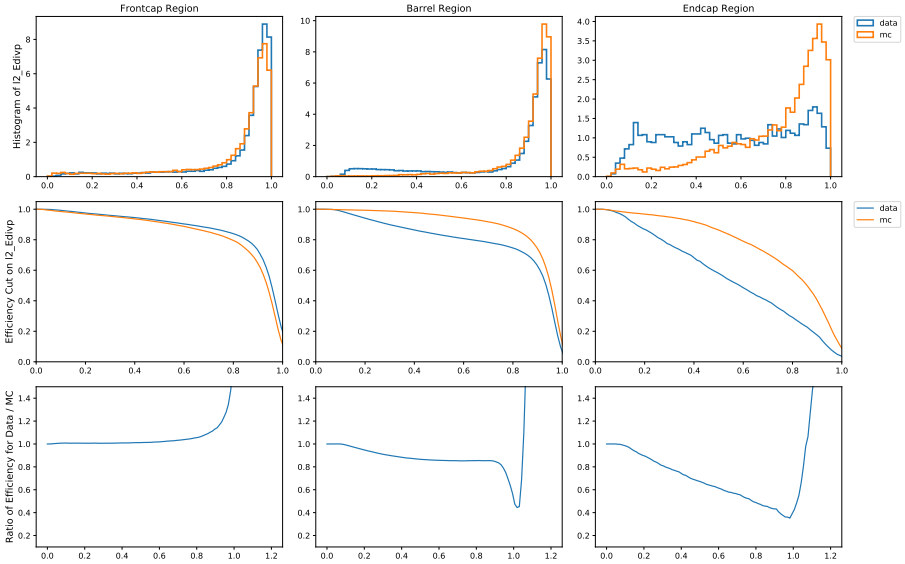

Figure 9: Distributions of particle identification variable (E/p) for data and MC. (Top row: Distribution of PID variable for data and MC. Second Row: Efficiency of cuts applied on the PID variable. Third row: Ratio of the efficiency for E/p produced for the probe particle being in different detector regions.)

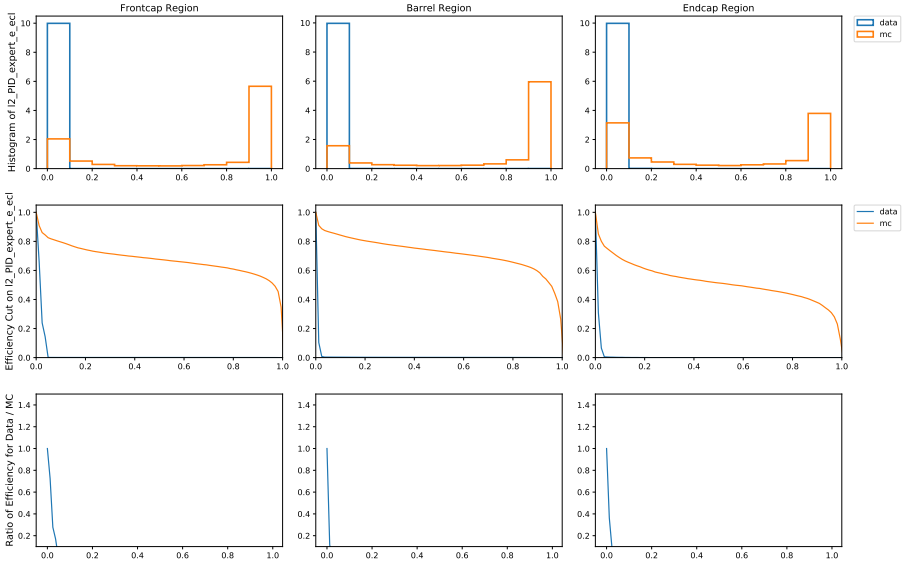

Figure 10: Distributions of particle identification variable (ECL_PID) for data and MC. (Top row: Distribution of PID variable for data and MC. Second Row: Efficiency of cuts applied on the PID variable. Third row: Ratio of the efficiency for E/p produced for the probe particle being in different detector regions.)

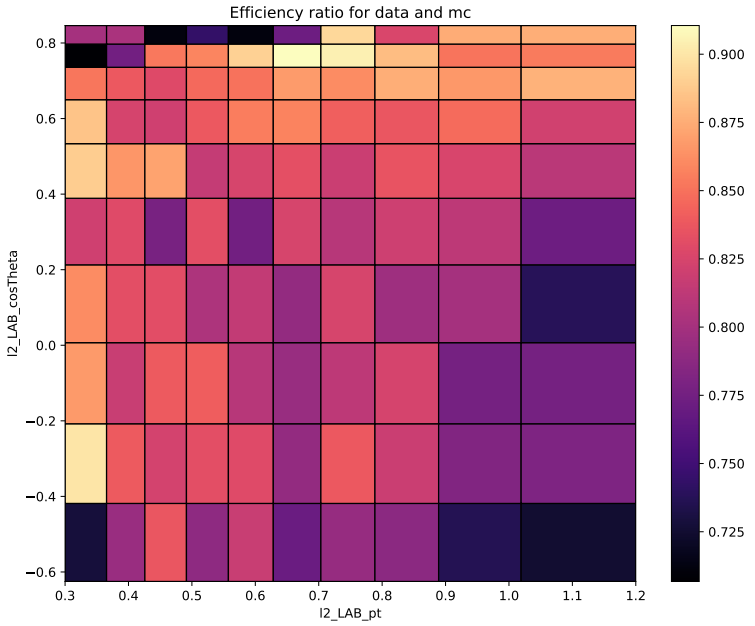

Figure 11: Correction ratio presented as a color map between data and 2γ (mc_eeee) event.

In Figure11, the lab cos theta range is the exact angular region for the barrel detector, where most good tracks lie. The average value for this entire map was 0.8655 ± 0.0028.

6 Conclusion

This report summarizes the study of lepton identification performance in the Belle II detector using the first data collected in spring and summer of 2018. Efficiency of the electron and muon selection was studied using the 2γ sample that was selected with purity of $70.1\%\pm0.4\%$ for $\gamma \rightarrow$ ee and $95.1\% \pm 0.2\%$ for the $\gamma \rightarrow \mu\mu$ channel. We show that currently the most robust way to define the flavour of the lepton in the Belle II experiment is to cut on the energy deposit in the electromagnetic calorimeter and its ratio to the absolute value of the tracks' momentum.

By conducting analysis, we have also created a framework for further analysis for leptons. The identification methods listed here will be implemented for the search of lepton flavour violating dark matter particles.

Acknowledgements

I would first like to express my deep gratitude to DESY for organizing the summer student programme. I had an enjoyable summer with lovely people that I hope to see again. Thank you to the entire Belle II team at DESY who were always so kind and welcoming. I am thankful for the tremendous support I received. I would also like to thank Dr. Simon Wehle for teaching me so much about both physics and computing, and always giving me a helping hand. Lastly, I would like to thank Dr. Ilya Komarov for being such a wonderful supervisor. It was very motivating to see his dedication towards physics, and his encouragement was very much appreciated. I am truly grateful for his patience and kindness that I received.

References

[1] Thomas Browder. Super KEKB and Belle II. `https://www.belle2.org/project/super_kekb_and_belle_ii`.

[2] Matthew J. Dolan, Torben Ferber, Christopher Hearty, Felix Kahlhoefer, and Kai Schmidt-Hoberg. Revised constraints and Belle II sensitivity for visible and invisible axion-like particles. *JHEP*, 12:094, 2017.

[3] Alfio Lazzaro. CP Violation in B Mesons. In *Proceedings, 45th International Winter Meeting on Nuclear Physics (Bormio 2007): Bormio, Italy, January 15-20, 2007*.

[4] Martin Ritter. Belle II Confluence Detector WebHome. `https://confluence.desy.de/display/BI/Detector+WebHome`.

[5] V. Aulchenko et al. Electromagnetic calorimeter for Belle II. *J. Phys. Conf. Ser.*, 587(1):012045, 2015.

[6] Iftah Galon, Anna Kwa, and Philip Tanedo. Lepton-Flavor Violating Mediators. *JHEP*, 03:064, 2017.

[7] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.

[8] E. Kou et al. The Belle II Physics book. 2018.

[9] L Hinz. Lepton ID efficiency correction and systematic error. *Belle Note*, 954, 2006.