

University of Mississippi

eGrove

---

Honors Theses

Honors College (Sally McDonnell Barksdale  
Honors College)

---

Spring 5-8-2022

## A Performance Analysis of the Belle II Detector

John Stacy

Follow this and additional works at: [https://egrove.olemiss.edu/hon\\_thesis](https://egrove.olemiss.edu/hon_thesis)



Part of the [Elementary Particles and Fields and String Theory Commons](#)

---

### Recommended Citation

Stacy, John, "A Performance Analysis of the Belle II Detector" (2022). *Honors Theses*. 2598.  
[https://egrove.olemiss.edu/hon\\_thesis/2598](https://egrove.olemiss.edu/hon_thesis/2598)

This Undergraduate Thesis is brought to you for free and open access by the Honors College (Sally McDonnell Barksdale Honors College) at eGrove. It has been accepted for inclusion in Honors Theses by an authorized administrator of eGrove. For more information, please contact [egrove@olemiss.edu](mailto:egrove@olemiss.edu).

A PERFORMANCE ANALYSIS OF THE BELLE II DETECTOR

by  
Wil Stacy

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College

Oxford  
May 2022

Approved by



---

Advisor: Dr. Jake Bennett



---

Reader: Dr. Cecille Labuda



---

Reader: Dr. Alakabha Datta

© 2022  
John Wilson Stacy  
ALL RIGHTS RESERVED

## **Abstract**

The Belle II experiment has recently (2018) started data taking at the SuperKEKB electron-positron collider in Tsukuba, Japan. Detector performance studies are necessary to understand early data and prepare for more complex analyses. This study of the proton detection efficiency of the Belle II detector compares real and simulated data to find discrepancies with the intention to provide useful information for detector and calibration experts to better gauge detector performance. It also attempts to improve the characterization of proton identification efficiency at low momenta, which performs poorly under the current fitting model. This helps analysts exploring final states that include protons to better understand their data.

## Table of Contents

<b>Abstract</b>	<b>3</b>
<b>Table of Contents</b>	<b>4</b>
<b>1: The Belle II Experiment, this Study, and the Science that Makes It All Work</b>	<b>5</b>
1.1: Quarks/Elementary Particles	5
1.2: Baryon Decays	7
1.3: Tracks	8
<b>2: The Belle II Detector</b>	<b>9</b>
2.1: The Central Drift Chamber (CDC)	10
2.2: The Time-of-Propagation (TOP) Counter	13
<b>3: Data/Monte Carlo (MC) Agreement</b>	<b>16</b>
<b>4: Selection Criteria Optimization</b>	<b>21</b>
4.1: Invariant Mass	21
4.2: Selection Criteria	21
4.3: Fits	24
4.4: Figure of Merit (FOM)	24
4.5: Presentation of a Problem	26
4.6: Attempts at Resolution	27
<b>5: Discussion</b>	<b>29</b>
<b>Conclusion</b>	<b>30</b>
<b>References</b>	<b>31</b>

## 1: Introduction

The Belle II experiment takes place at the SuperKEKB accelerator facility in Tsukuba, Japan. Thanks to improvements and innovations, the Belle II experiment plans to gather fifty times as much data as its predecessor, the Belle experiment. The Belle II experiment relies on the SuperKEKB collider, which accelerates beams of electrons and positrons to near the speed of light and then collides them at the interaction region inside the Belle II detector. From this high-energy environment, new particles are created. This study focuses on the proton detection efficiency, the fraction of proton tracks that are correctly identified, of sub-detectors within the Belle II detector. The proton detection efficiency is important because many processes of interest for Belle II include a proton in the final state. If the proton detection efficiency is subpar, any such studies will be adversely affected, potentially even leading to skewed results.

### 1.1: Quarks/Elementary Particles

Subatomic particles, despite having been discovered over 120 years ago, are still a recent discovery in the scope of physics. This is because physics has been a field of study since Classical Greece. In the scope of *high energy* physics, however, knowledge of protons, neutrons, and electrons is considered relatively old. The more recent discovery, in the mid to late twentieth century, was that many particles are actually composite states called hadrons, which are made up of fundamental particles called quarks. Protons and neutrons were previously thought to be elementary particles, meaning they have no smaller constituent parts, but the discovery of quarks

stripped them of that distinction. Now, quarks are considered to be elementary particles along with electrons and other leptons. Additionally, each quark and lepton has an anti-matter partner, which have the same mass but opposite charge. Quarks come in different generations, where higher generations have more mass and decay to the smaller mass, lower generation quarks. Those different generations can be observed in Figure 1-1, which summarizes the Standard Model of particle physics. This is the basis for understanding baryon decays, which are the focus of this project and are described below.

## 1.2: Baryon Decays

To create new particles, the Belle II experiment accelerates a beam of electrons and a beam of positrons, antimatter partners of electrons, to almost the speed of light, then forces the two beams to collide at a specific point known as the interaction region (IR) of the detector. When these beams collide, the immense energy present (approximately 10.58 GeV) from both the mass and kinetic energy of the beam particles creates an environment in which particles from higher generations may be created. These particles will quickly decay down to “final state” particles consisting primarily of first generation quarks. For example, the  $\Lambda^0$  baryon, which consists of an up quark, a down quark, and a strange quark, decays into a proton (with 2 up quarks and a down quark) and a negative pion (with a down quark and an up antiquark). Since the  $\Lambda^0$  has such a reliable decay pattern and is abundantly produced, it can be used to study the detector performance related to proton and pion identification. Figure 1-2 shows an example decay process, starting with the electron-positron collision, that includes a proton in the final state.

# Standard Model of Elementary Particles

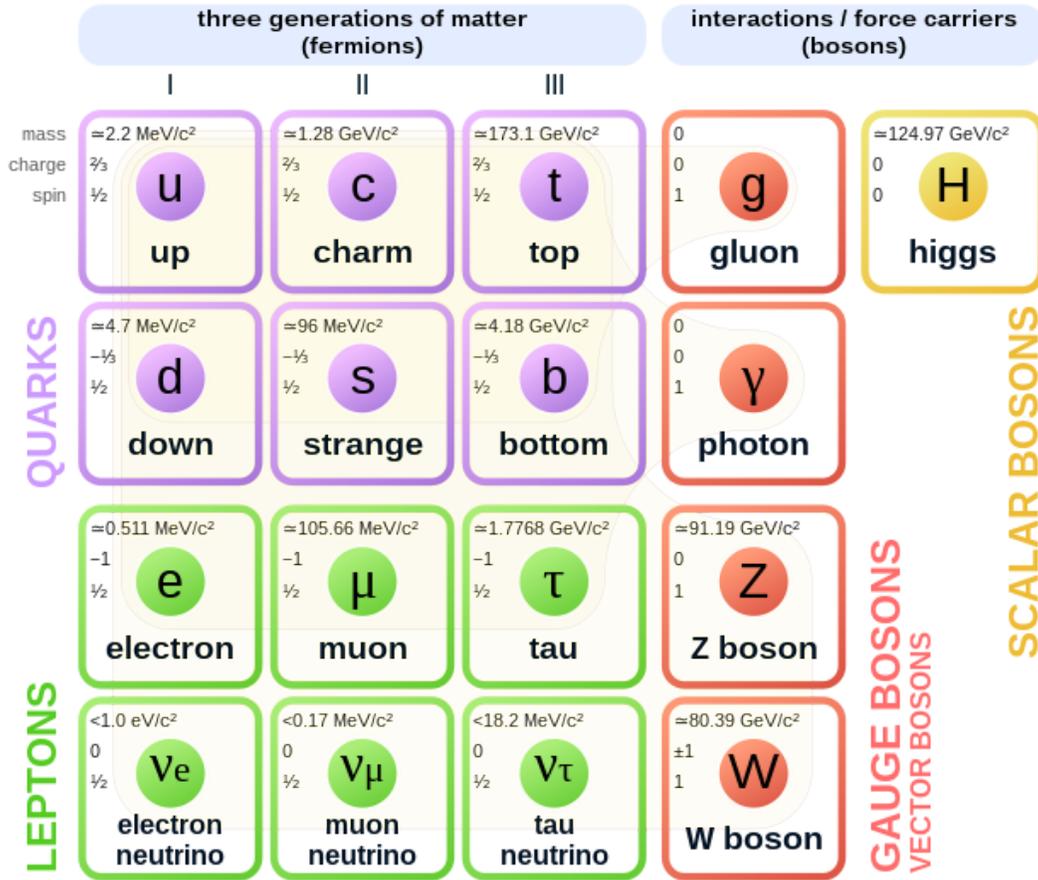


Fig. 1-1: Standard Model of Elementary Particles

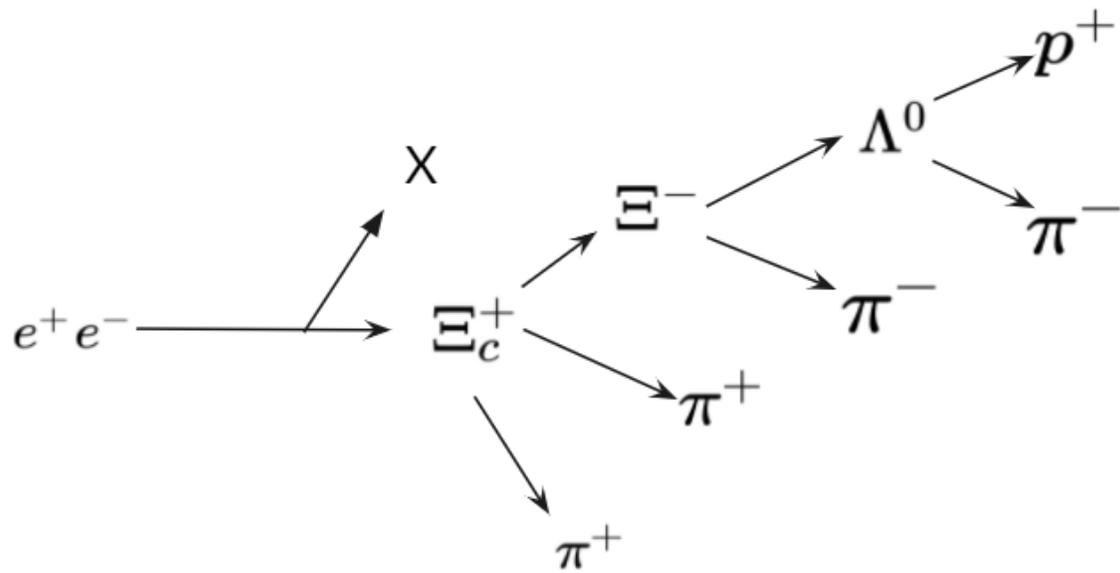


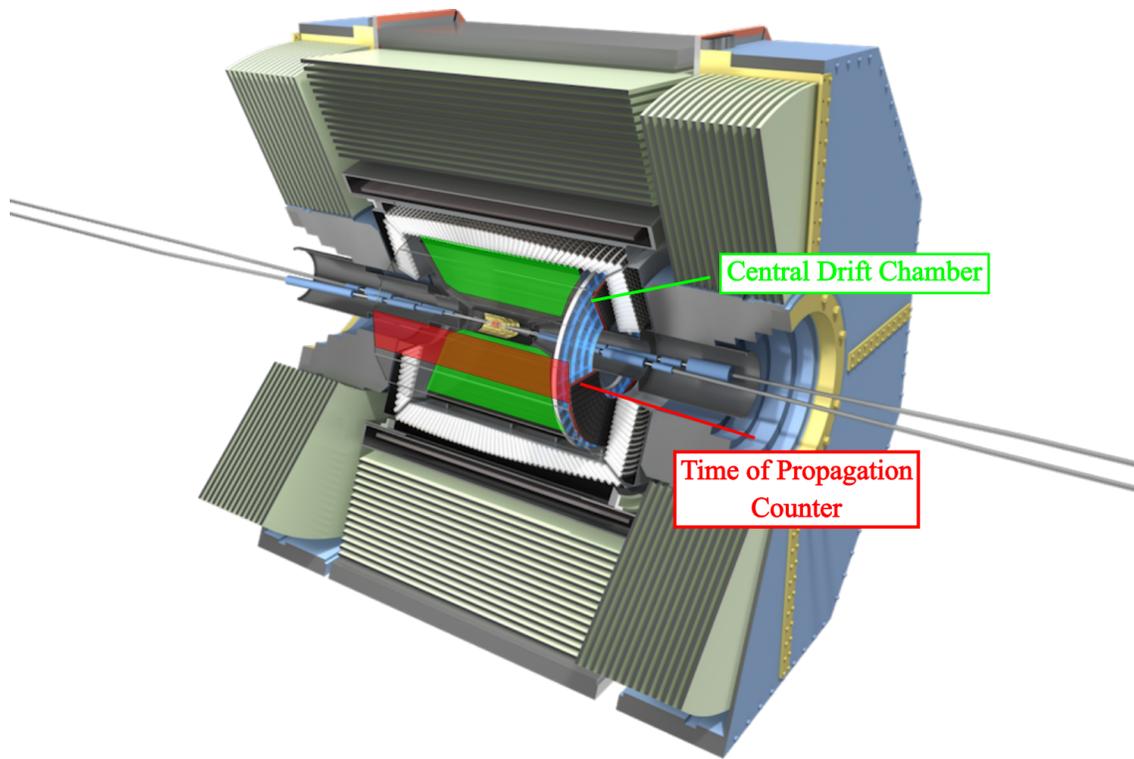
Fig 1-2: Feynman Diagram of the decays from collision to a proton and a pion

### 1.3: Tracks

As will be discussed below, particle detection is performed by reading out electronic signals rather than spatial recognition. In other words, there are no videos of particles traveling through the detector or photographs with time stamps. Instead, a process called “track reconstruction” is used to identify the path a particle took through the detector. For this reason, the recreated path of a particle through the detector is called a track.

## **2: The Belle II Detector**

The Belle II Detector, which is shown in Figure 2-1, has several subcomponents. While most of them have specific jobs that only they can accomplish, some have purposes that overlap to create a system of redundancy, thereby improving the accuracy of its measurements. One such example is particle identification. The Time of Propagation (TOP) counter, Aerogel Ring Imaging Cherenkov Detector (ARICH), and Central Drift Chamber (CDC) all play a part in particle identification. The TOP counter is the most powerful when it comes to distinguishing between kaons and pions. However, when a particle fails to reach the TOP counter or the ARICH, the CDC provides the primary information that can be used for particle identification for charged tracks. The analysis presented here focuses on the CDC, but includes some studies of the TOP performance, as well.

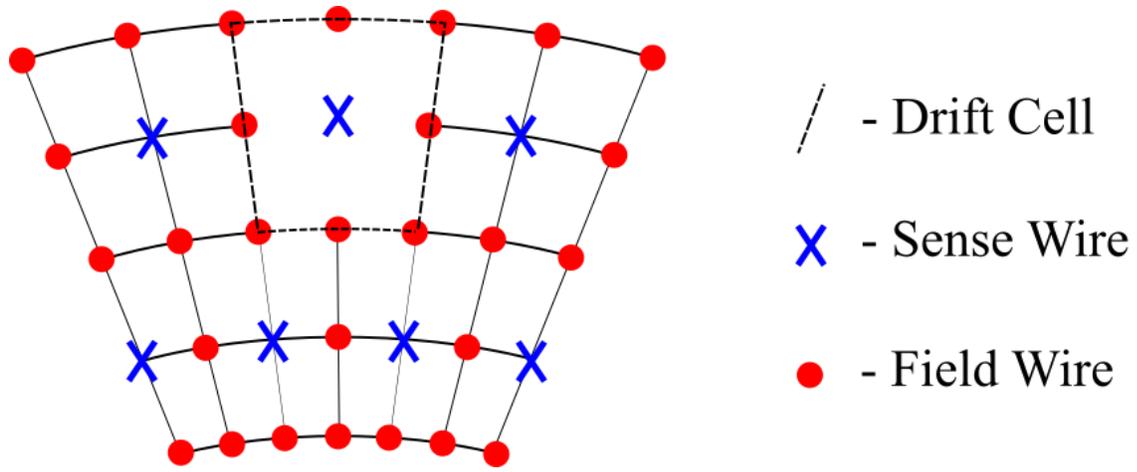


*Fig. 2-1: Cutaway diagram of the Belle II detector. The Central Drift Chamber is highlighted in green, while one of the 16 Time of Propagation Counter radiators surrounding the CDC is highlighted in red.*

## 2.1: The Central Drift Chamber (CDC)

The central drift chamber is a 2.2-meter wide drum filled with wires, a gas mixture of 50% helium - 50% ethane, and a uniform magnetic field. The CDC serves 3 purposes: reconstruction of the trajectory of charged particles moving through the detector (also referred to as “tracks”), provision of a trigger for data collection if sufficient high-quality charged tracks are present, and particle identification using ionization energy loss measurements [1].

As mentioned in Section 1.2, baryons decay into elementary particles which then exit the interaction region. On their way through the CDC, these charged particles ionize helium-ethane gas molecules, resulting in positively charged gas molecules and free electrons. Once ionized, these gas molecules and free electrons are attracted to field and sense wires, respectively. Field wires are wires that have been grounded; they are used to shape the electric field of the chamber. Sense wires are held at a high voltage from an external source. The gas molecules drift toward the field wires and the free electrons move toward the sense wires due to the resulting electric field. Sense and field wires are configured such that they lie approximately collinear with the beam direction. Each sense wire lies at the center of what is known as a drift cell. Ionization created in the drift cell bordered by 8 field wires will drift to the sense wire at its center. As the wires are in a drum and groups of layers have similar numbers of wires, they do not actually form a square, but an annulus sector instead, as shown in Figure 2-2.



*Fig. 2-2: Configuration of wires in the CDC*

As free electrons drift toward the sense wire, they ionize more gas molecules, which frees additional electrons and so on, such that by the time the electrons reach the sense wire, a Townsend avalanche - a chain reaction of free electrons - is created. This avalanche reduces the voltage of the sense wire, which acts as the trigger to inform the detector that a charged particle passed a wire at a specific time. The wire locations are known up to a resolution of  $100\mu\text{m}$  [1]. With trigger times from many wires and their known locations, the Belle II detector is able to recreate the path taken by the particle. Since the magnetic field inside the detector is well measured, the radius of curvature of the track allows for a measurement of the particle momentum according to the Lorentz force.

As a charged particle collides with the gas molecules, the law of conservation of energy dictates that it loses some amount of kinetic energy to the resultant free electron. This is known as the ionization energy loss, or  $dE/dx$ . The ionization energy loss is dependent only on the velocity of the particle, as shown in the right-hand side of Figure 2-3. When plotted against the velocity ( $\beta\gamma$ ) of the particle, all types of particles follow the same distribution for  $dE/dx$ . In the left-hand side of Figure 2-3, however, it is clear that when plotted against the particle's momentum, the  $dE/dx$  distribution produces distinct bands for each type of particle. This distinction can be used to identify the most likely mass of a particle for a given measurement of  $dE/dx$  and momentum.

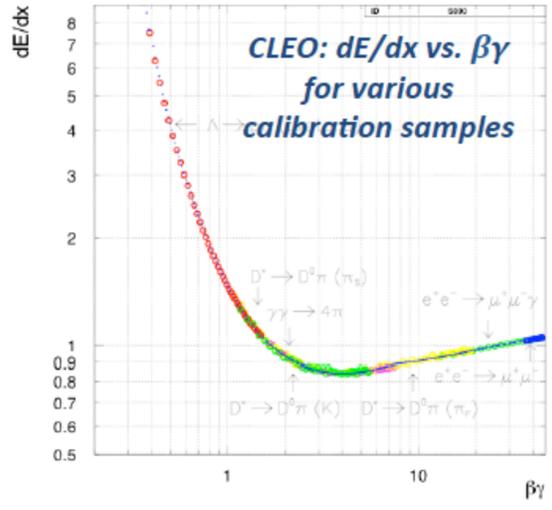
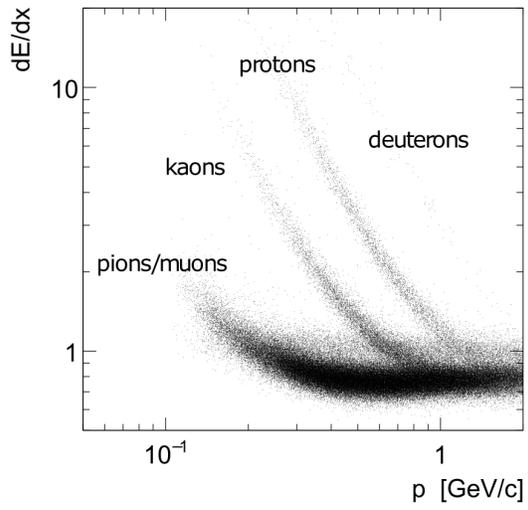
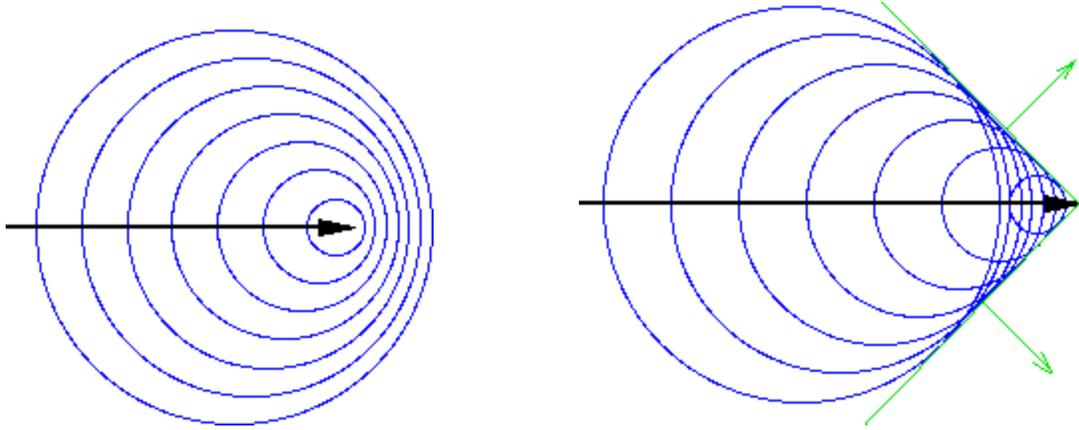


Fig 2-3:  $dE/dx$  plotted against the momentum and velocity ( $\beta\gamma$ ) of the particle

## 2.2: The Time-of-Propagation (TOP) Counter

The TOP counter consists of 16 radiators positioned radially around the CDC. Each radiator is a quartz prism approximately 1200 millimeters long, 400 millimeters wide, and 20 millimeters thick with photon detectors at one end [1]. The TOP counter is responsible for a significant portion of the particle identification in the Belle II experiment. As noted above, the CDC supplements the TOP counter in the case of low-momentum particles that fail to reach the TOP prisms and in the backward direction where neither the TOP nor the ARICH are active. The TOP counter utilizes a phenomenon known as Cherenkov radiation to identify particles.

Cherenkov radiation occurs when a particle exceeds the propagation speed of light in a medium ( $c/n$ , where  $n$  is the index of refraction of the medium). When this happens, the particle experiences the Cherenkov effect in which it radiates photons at a constant angle dependent on the speed of the particle and the refractive index of the material. A good analogy is the sonic boom experienced by a jet breaking the sound barrier. Figure 2-4 (left) shows a particle traveling slower than the speed of light in the medium. The wave fronts, represented by the blue circles, do not collide, and Cherenkov radiation does not occur. Figure 2-4 (right), on the other hand, shows a particle that is moving faster than the speed of light in the medium, resulting in an overlap of light wavefronts. This results in a cone of light propagating outward, as shown by the green arrows in the diagram. The high energy of these photons results in a high frequency, meaning the Cherenkov radiation tends to be in the blue part of the EM spectrum. For the case of the TOP detector, the refractive material is the quartz radiator, which has a refractive index between 1.43-1.58 depending on the wavelength of light [2].



*Fig 2-4: Charged particles traveling slower (left) and faster (right) than the speed of light in the medium, the latter of which results in the phenomenon known as Cherenkov radiation*

Source: [<https://cbooth.staff.shef.ac.uk/phy6040det/coherent.html>]

The Cherenkov photons produced in the TOP detector reflect internally through the quartz radiator until colliding with a photon detector at one end. The location of impact and the time of propagation (hence the name of the detector) are used to identify the particle passing through the quartz. Figure 2-5 illustrates this process using a kaon track. The dashed lines represent the theoretical path of a photon emitted by a pion from the same incident angle. Notice that the often-lower-momentum pion has a wider Cherenkov angle, measured as the angle away from the direction of the particle's momentum. This is because the Cherenkov angle is solely dependent on the speed of the particle and the refractive index of the medium [3].

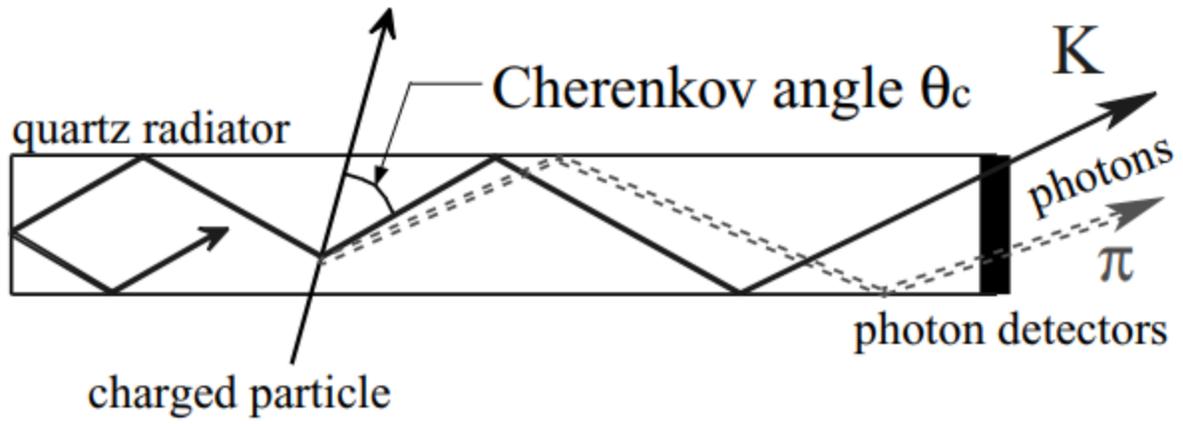
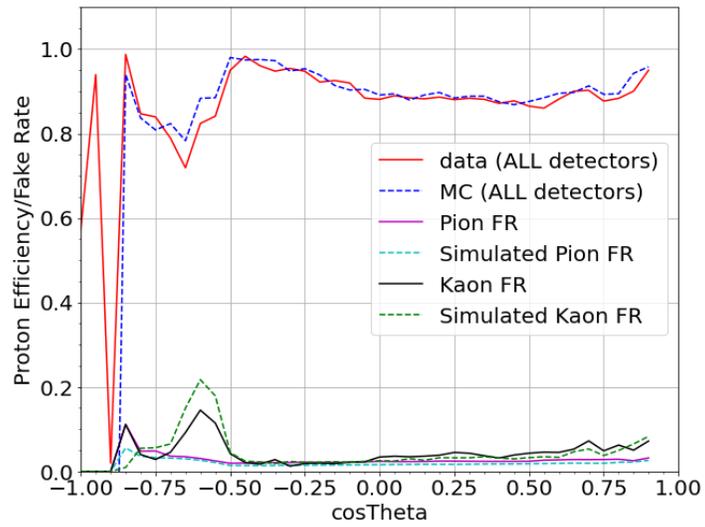


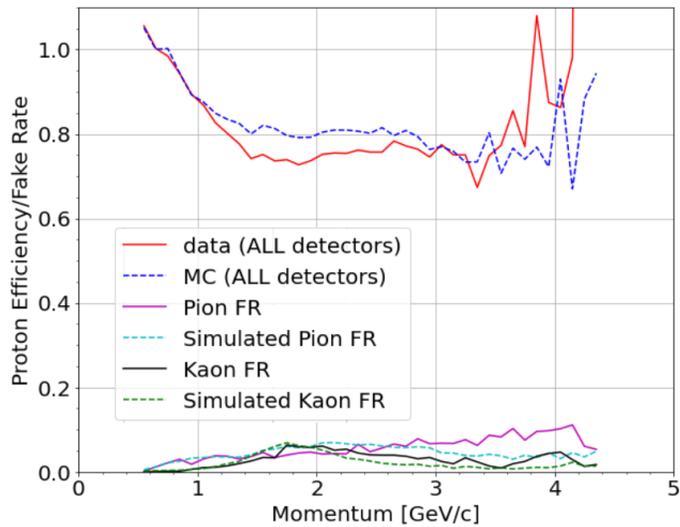
Fig 2-5: Diagram of TOP counter intercepting Cherenkov photons for particle identification

### **3: Data/Monte Carlo (MC) Agreement**

Early studies for this analysis focused on identifying discrepancies when comparing real data and simulated data (Monte Carlo). The goal was to relay these discrepancies to the calibration experts on the Belle II team to get Monte Carlo data to agree better with the results derived from real-life runs of the detector. As shown in Figure 3-1, the red line representing real data and the dashed blue line representing Monte Carlo (MC) data are remarkably similar until approaching a  $\cos(\theta)$  of -1, where  $\theta$  is the polar angle of the track momentum relative to the beam line. This erratic pattern arises due to the fact that the TOP fails to measure tracks going in the backward direction, leaving all particle identification to the CDC. In Figure 3-2, which shares its legend with Figure 3-1, there is more divergence between the MC data and the real life data, while the pattern becomes erratic towards higher momenta due to low statistics. That is, there are not many events with individual particles retaining that much momentum.



*Fig 3-1: Proton Identification Efficiency as a function of the particle's cosine of the angle, along with fake rates for other particles*



*Fig 3-2: Proton Identification Efficiency as a function of the particle's momentum, along with fake rates for other particles*

The 2D histograms pictured in Figure 3-3 help to visualize the proton identification weak spots of the detector as well as the Monte Carlo's approximation of those weak spots. The third histogram shows where the Monte Carlo fails to align with the real data. If the two aligned perfectly, the third histogram would be a uniform greenish-blue, as denoted in its respective legend. At around  $\cos(\theta) = -1$  and 3.5 GeV/c, though, a bright yellow bin stands out, flagging an issue in which the Monte Carlo's simulated data is significantly lower than the real life data. Conversely, the very dark blue bin at around  $\cos(\theta) = -0.2$  and in the higher momentum region tells us that the Monte Carlo data is significantly higher than it should be.

Figure 3-4 follows the same pattern as Figure 3-3, but instead of proton identification efficiency, we measure the kaon fake rate, which indicates how often a kaon was improperly identified as a proton. As before, a histogram of the pure greenish-blue color (found at 0.00 on figure 3-3c) would be ideal, as this would mean that our MC data aligned perfectly with our real data, but once again there are distinct bins in which the datasets disagree.

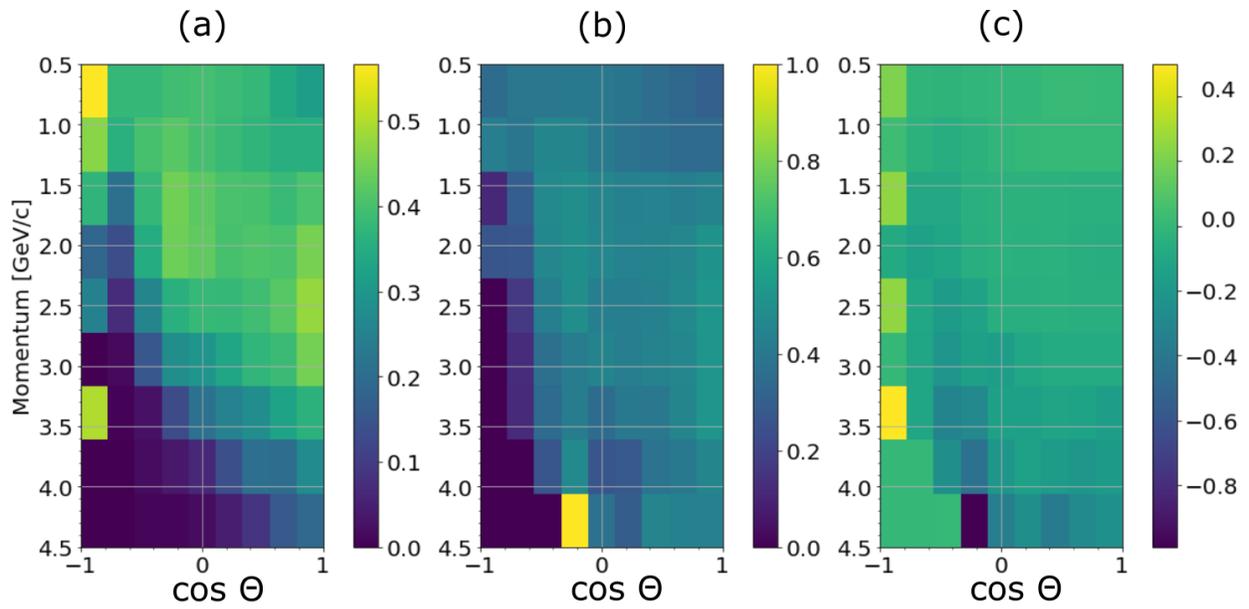


Fig 3-3: A 2D histogram of proton efficiency as a function of  $\cos \Theta$  and particle momentum

Fig 3-3a: Real Data      Fig 3-3b: MC Data      Fig 3-3c: Real - MC difference

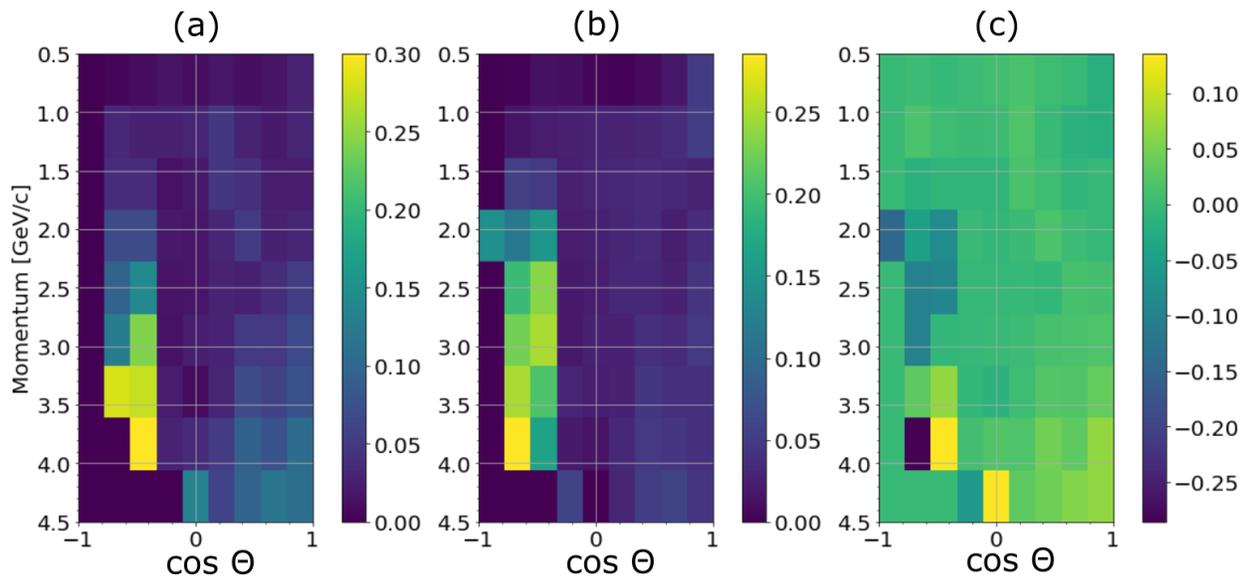


Fig 3-4: A 2-dimensional kaon fake rate as a function of  $\cos \Theta$  and momentum of the particle

Fig 3-4a: Real Data      Fig 3-4b: MC Data      Fig 3-4c: Real - MC difference

These discrepancies were shown to the calibration experts, who made adjustments to simulation and calibration parameters and were able to make the Monte Carlo data more like the real data. Figures 3-5 and 3-6 show only the CDC performance, as opposed to earlier figures which showed all detectors. The global particle identification (globalPID) is the ratio of the likelihood for a track to be a proton, divided by the sum of likelihoods for all particle types. From the legend, proc11 stands for the eleventh official reprocessing of the data, which is reprocessed with new software releases and improved calibrations. MC13b stands for the thirteenth Monte Carlo campaign and includes simulated data of various types that are used to mimic the data. As is evident in Figures 3-5 and 3-6, the data reprocessed with the improvements created after this study (proc12) and a new set of Monte Carlo (MC14a) show an increased similarity between simulations and real data in virtually all areas.

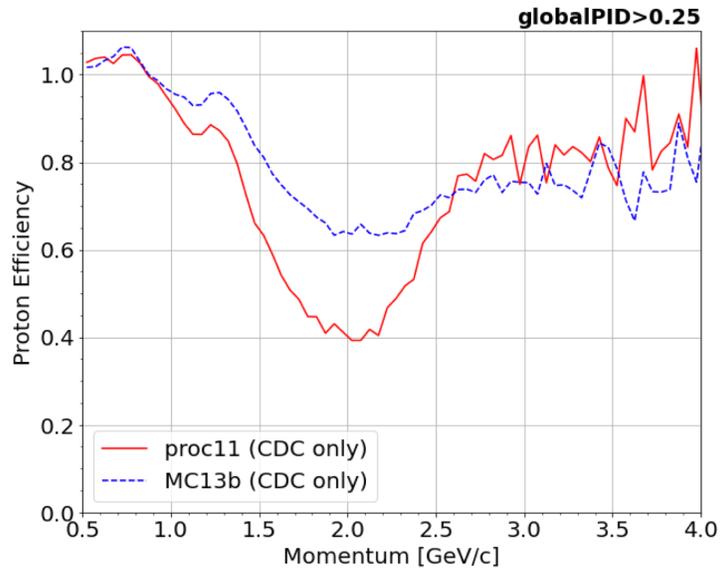


Fig 3-5: CDC Proton ID Efficiency for the CDC before recalibration

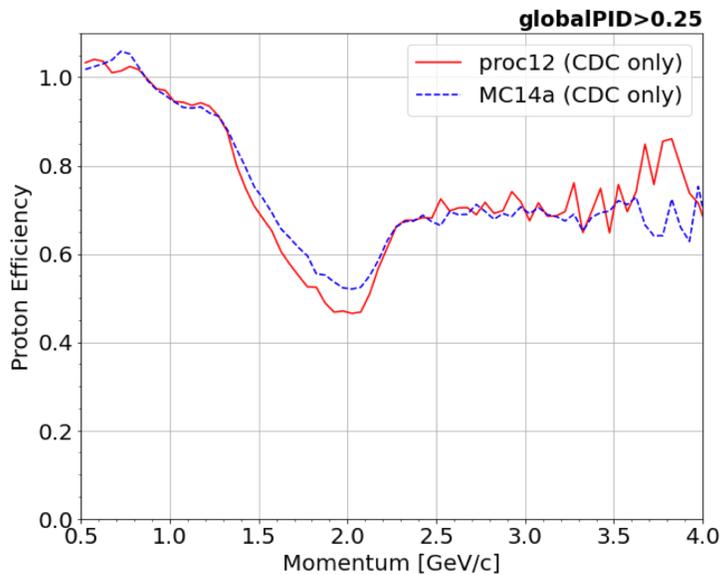


Fig 3-6: Proton ID Efficiency for the CDC after recalibration

## 4: Selection Criteria Optimization

Before getting into later studies, there are a few important concepts to understand. Additionally, the process used in this study when dealing with event-based data must be defined.

### 4.1: Invariant Mass

The invariant mass of a particle does not depend on the reference frame. It can be determined by taking the square-root of the difference of the square of the energy and the square of the momentum of the particle according to the equation  $m^2c^4 = E^2c^2 - p^2c^2$ , where  $E$  is the particle energy,  $p$  is the three-momentum of the particle, and  $c$  is the speed of light. As described below, the invariant mass is used in the *sWeights* process to obtain a clear signal.

### 4.2: Selection Criteria

When looking for a specific process, that process is typically referred to as the *signal*. However, other processes may mimic the signal and are known as the *background*. The ultimate goal for a given analysis is to either eliminate backgrounds or else reduce them to a manageable level. This is accomplished via selection criteria, which applies rules to the sample under study in an attempt to isolate the signal for further study.

Throughout this section, selection criteria may be referred to as *cuts*. These words are synonymous, and can be interchanged freely. When working with a large set of data, most of that data will be of no use for specific tasks. For example, this study is primarily focused on protons,

specifically those resulting from decays of the  $\Lambda^0$  baryon, so events of that type are desired, but not other types of hadrons. Those other hadrons and their resulting particles would present as backgrounds.

The specific method used for background removal in this study is known as sWeights. Figure 4-1 shows the invariant mass of  $\Lambda^0$ , candidate events. This sample is fit with a probability density function (pdf) that includes a signal and background shape. From that pdf, a statistical method called the sPlot technique is used to determine per-event weights called sWeights. The principle is that events in the signal peak have sWeights near one and those in the background region have sWeights that are negative. When the weights are applied to a sample, the negative weights for background events cancel the contribution of the positive weights for background that lands within the signal region, thus removing the effect of the remaining background. Figure 4-2 shows that nearly all of the backgrounds are removed after the sWeights are applied.

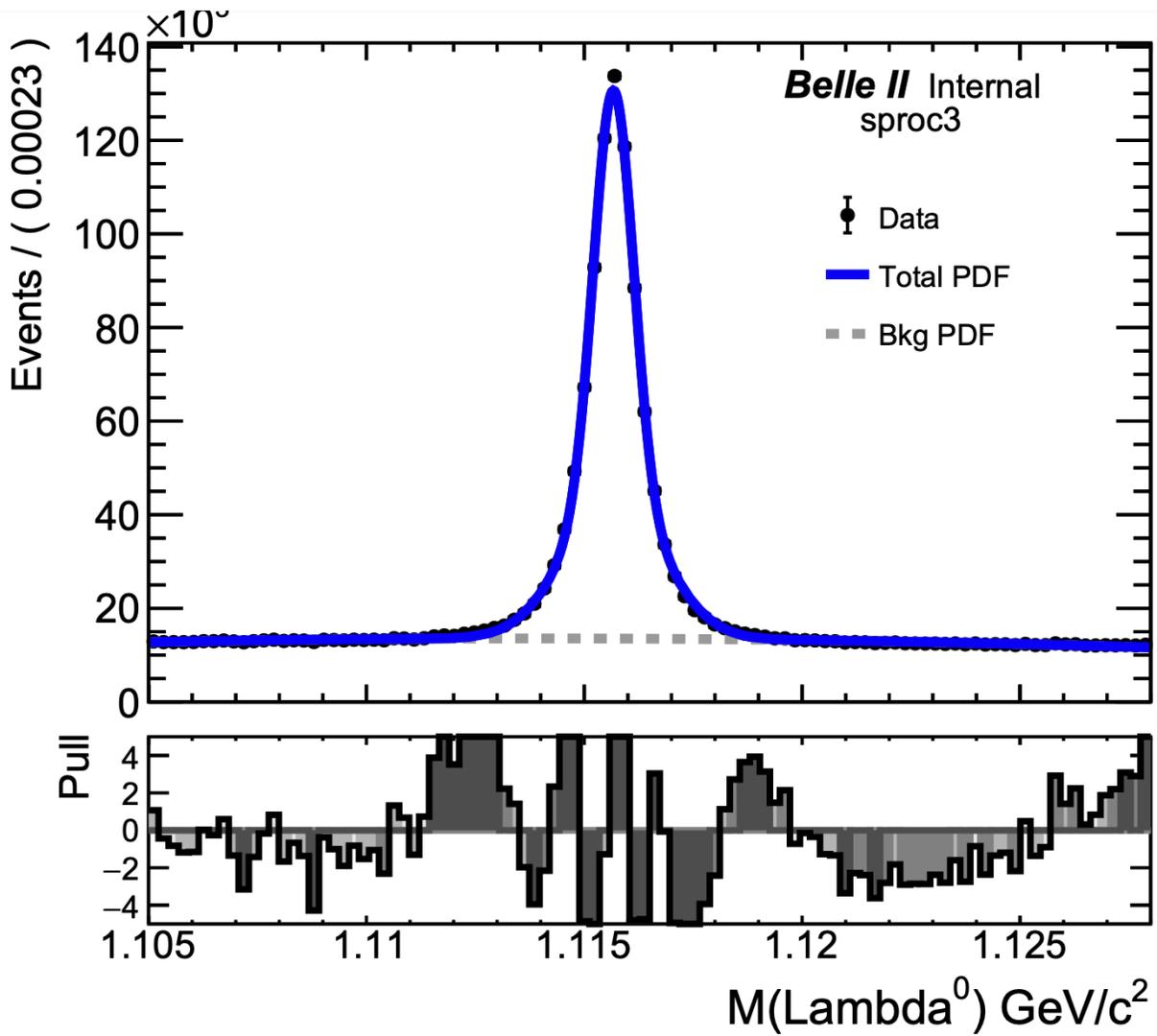
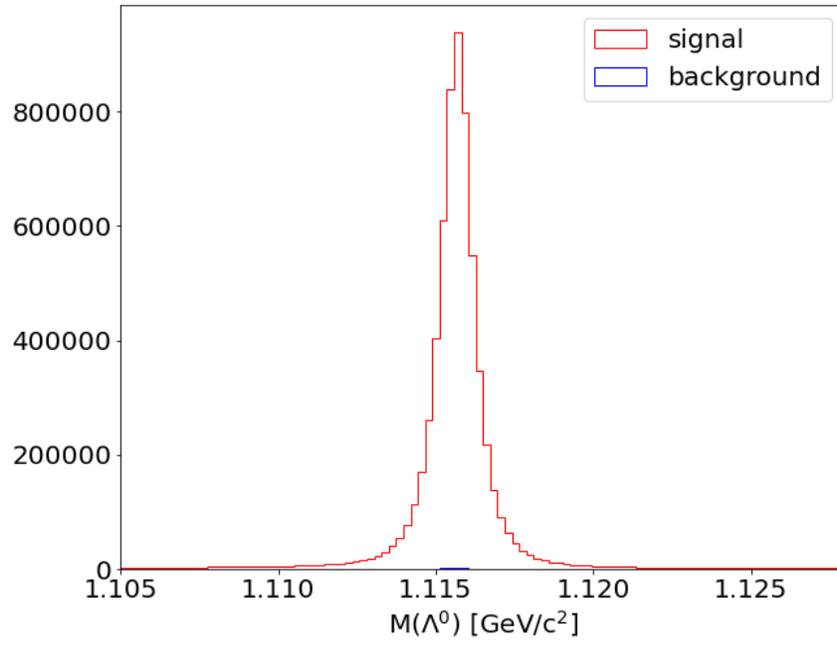


Fig 4-1: Events vs Invariant Mass plot of the  $\Lambda^0$  baryon for the sample of interest. The fit to the sample is shown in blue and is used to determine the  $sWeights$ .



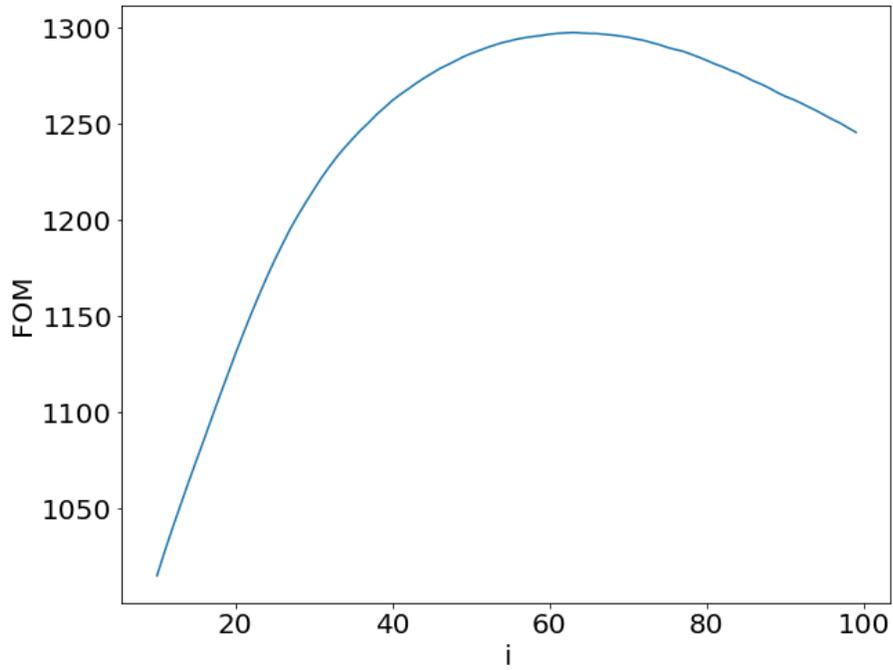
*Fig 4-2: Events vs Invariant Mass plot of the  $\Lambda^0$  baryon after  $sWeights$  are applied.*

### 4.3: Fits

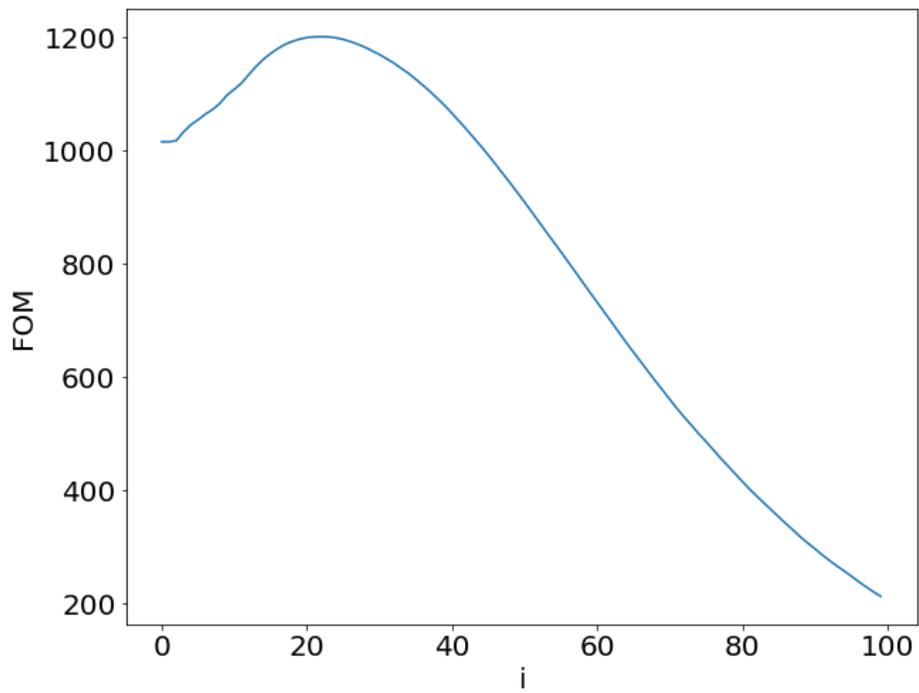
As previously mentioned, the desired shape for the background is a flat line. This is because the *fit* we used relies on a method which requires a flat shape for the background in the signal region that can be quantified with a *polynomial* fit. The signal uses a different function known as a double Gaussian.

### 4.4: Figure of Merit (FOM)

The figure of merit is a useful tool for determining where a cut is likely to be most effective. In this study, the figure of merit is the number of signal events divided by the square root of the total number of events. Applying this formula over very fine intervals, a plot was created for available selection criteria, resulting in plots such as the ones found in Figures 4-3 and 4-4, below. For Figure 4-3, a good cut would be somewhere around the sixtieth bin, while for Figure 4-4, a good cut might be around 25-30. Using these figures of merit provides a consistent means to motivate a particular selection.



*Fig 4-3: Figure of Merit graph for flight significance*



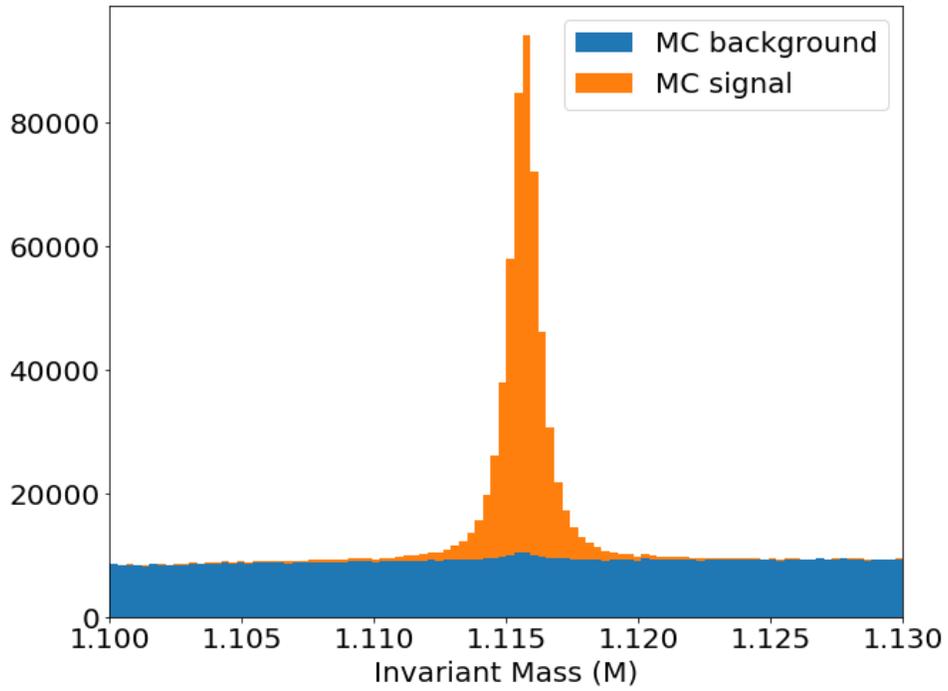
*Fig 4-4: Figure of Merit graph for proton momentum*

#### 4.5: Presentation of a Problem

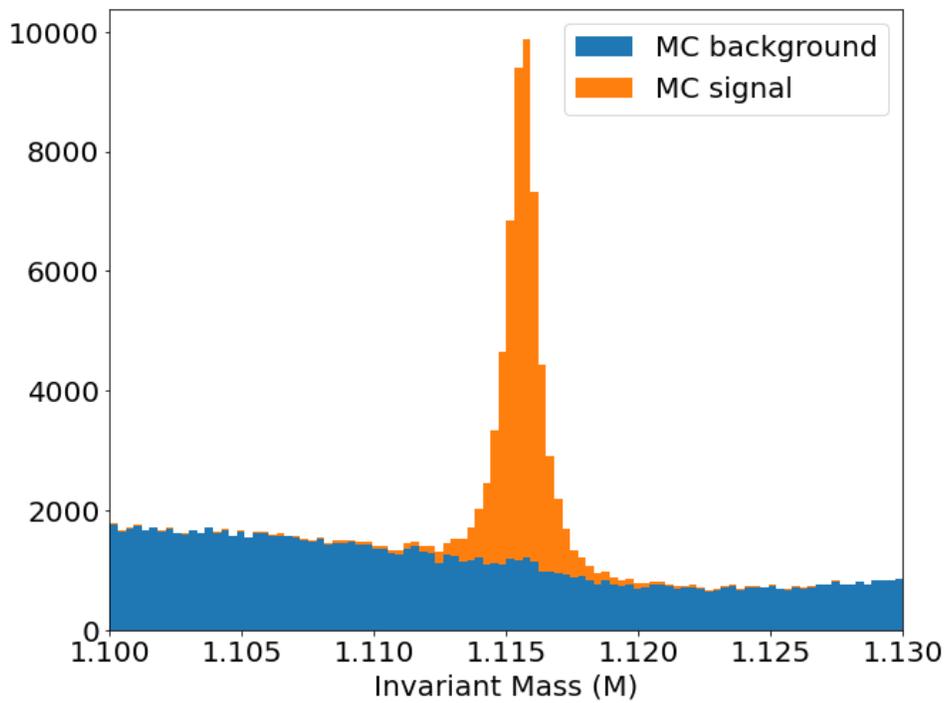
In later studies associated with this work, the main focus was to experiment with different selection criteria in an effort to create a background that could be easily removed. The background looked flat when all possible momenta were considered, as shown in Figure 4-5a, but when we look closer at specific regions of momentum, specifically in lower momentum regions, we find that the backgrounds vary significantly from a flat line, as shown in Figure 4-5b. This is a problem since the sWeights that we apply to the dataset as a whole function poorly when trying to eliminate non-flat backgrounds. Therefore, the background contamination in the signal region is not fully canceled and the efficiency can exceed 100%.

#### 4.6: Attempts at Resolution

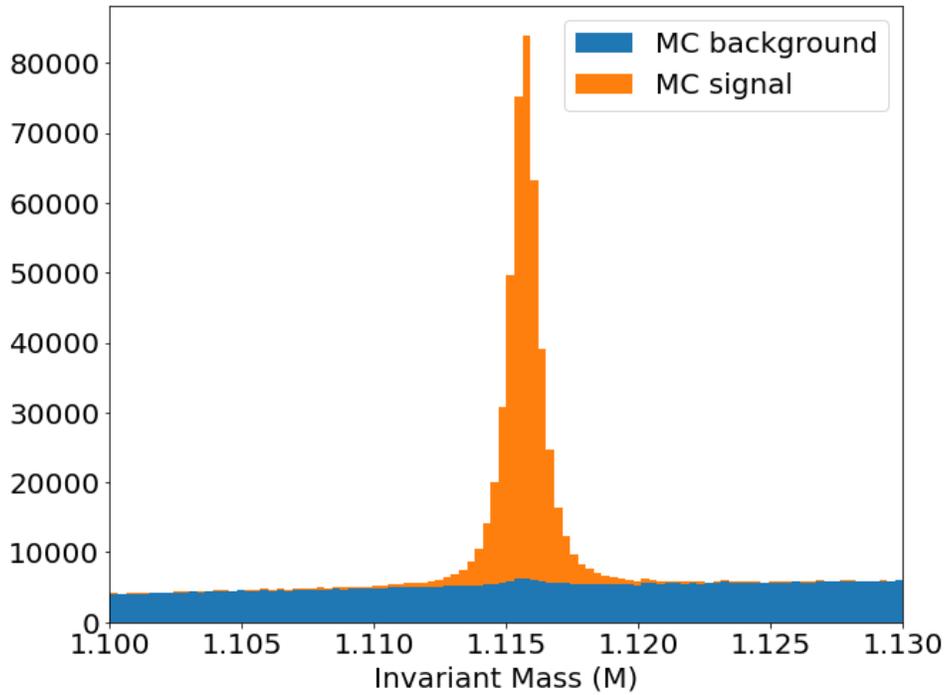
After obtaining Monte Carlo data, cuts were made using different values for a range of variables which were then plotted to see if the background had been flattened. As shown in Figures 4-6a and 4-6b, some signal events were lost along with the cuts. The plot in 4-6b shows a less severe background shape, but it still was not enough to flatten it for the fit.



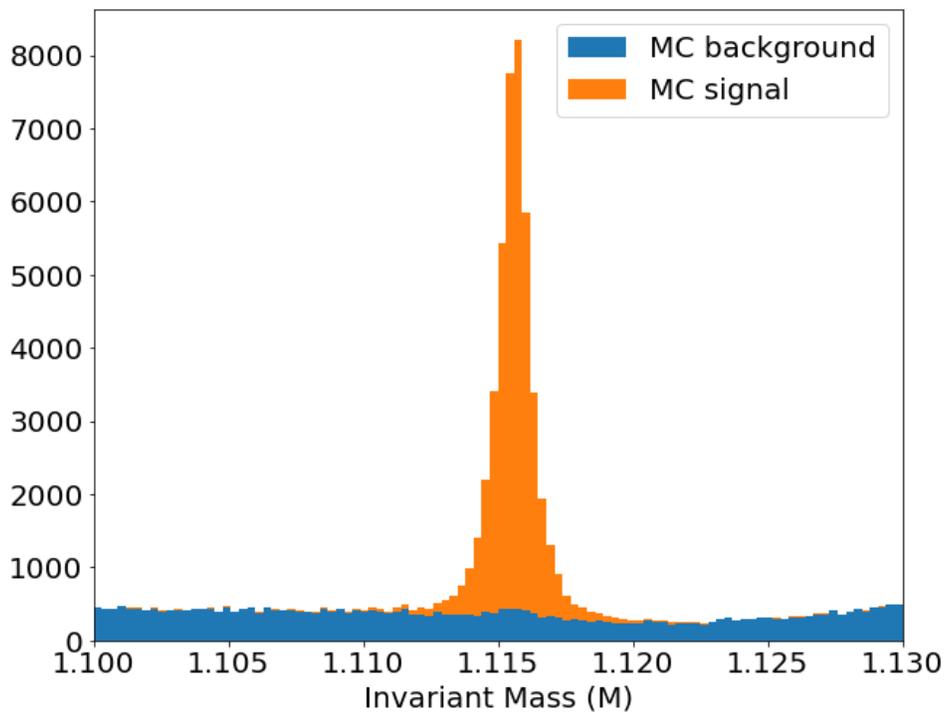
*Fig 4-5a: MC events across all momenta*



*Fig 4-5b: MC events across low momentum region only*



*Fig 4-6a: MC events (with new selection criteria) across all momenta*



*Fig 4-6b: MC events (with new selection criteria) across low momentum region only*

Selection criteria were not arbitrarily chosen. Using the figure of merit (section 1.6), we were able to find where our cuts would be most effective. Looking back at Figure 4-6b, the inability to flatten it may have to be resolved with other methods. One such method is a specific fit for momentum regions that refuse the general fit. Instead of the flat polynomial fit for the background, one with more curve may help to counter the curve of the troublesome background.

## 5: Discussion

This analysis of the proton detection efficiency of the Belle II detector is important because many decays result in protons. If the detector is performing poorly with regard to proton detection, this can negatively affect all studies of decays that include protons. Conversely, any improvements in the proton detection efficiency will positively affect any proton-focused research.

Monte Carlo simulations are helpful because they allow us to “create” the physics that we want to observe, and this ideal environment is used as a reference for what the data should look like. When there are discrepancies between Monte Carlo and real data, however, this will skew the value of our Monte Carlo reconstruction efficiency. In cases such as these, corrections are in order; this is the job of analysts. Section 3 of this study is one of many necessary to determine the correction factors for those analysts. These correction factors were found using the Systematic Corrections Framework, a toolkit that is available to Belle II analysts and to which we contributed during this study. By comparing the MC and real data through this framework, the two datasets were able to be corrected to agree better. Later studies used the corrections from this study to collect better data

## **Conclusion**

This study provides valuable information on the proton identification performance of Belle II data. In particular, results from early studies were provided to calibration experts, who were able to identify problems in their calibration code and improve the agreement between data and simulations in subsequent samples. This agreement is extremely important to reduce the need for complicated corrections in analyses that include protons as corrections introduce additional systematic uncertainties. Future studies may be made using the results of this study to further improve the calibration and simulation at Belle II. While there are still some strange features that need to be better understood, this analysis provided important insight to the underlying issues. With this insight, new approaches may be taken, such as a different fit for backgrounds that cannot be flattened.

## References

- [1] T. Abe, Belle II collaboration, Belle II Technical Design Report, arXiv:1011.0352 [physics.ins-det].
- [2] Corning® HPFS® 7979, 7980, 8655 Fused Silica (2015).
- [3] C. Booth, Coherent Effects for Charged Particles (2006).