Thesis for the Degree of Master

$\pi^{\pm}$ identification performance study based on $D^{\pm}$ to $K_s^0\pi^{\pm}$ decays in the Belle II detector

$D^{\pm}$가 $K_s^0\pi^{\pm}$ 로 붕괴하여 생성된 데이터를 바탕으로 Belle II 관측기의 $\pi^{\pm}$ 식별 성능을 확증하는 연구

December 2025

Department of Physics

Graduate School of Soongsil University

Gyeongmin Lee

Thesis for the Degree of Master

$\pi^{\pm}$ identification performance study based on $D^{\pm}$ to $K_s^0\pi^{\pm}$ decays in the Belle II detector

$D^{\pm}$가 $K_s^0\pi^{\pm}$ 로 붕괴하여 생성된 데이터를 바탕으로 Belle II 관측기의 $\pi^{\pm}$ 식별 성능을 확증하는 연구

December 2025

Department of Physics

Graduate School of Soongsil University

Gyeongmin Lee

Thesis for the Degree of Master

$\pi^{\pm}$ identification performance study based on $D^{\pm}$ to $K_s^0\pi^{\pm}$ decays in the Belle II detector

A thesis supervisor: Doris Kim

Thesis submitted in partial fulfillment of the requirements for the Degree of Master

December 2025

Department of Physics

Graduate School of Soongsil University

Gyeongmin Lee

To approve the submitted thesis for the

Degree of Master by Gyeongmin Lee

Thesis Committee

Chair	박 세 영	(signature)
Member	소 재 필	(signature)
Member	Doris Kim (김양수)	(signature)

December 2025

Graduate School of Soongsil University

ACKNOWLEDGEMENT

우선 학부 연구생 때부터 석사 과정까지 지도해 주신 김양수 교수님께 감사드립니다. 교수님 밑에서 정말 많은 것들을 배웠고, Belle II실험에서 데이터 분석, 소프트웨어 등을 공부할 수 있었습니다. 앞으로도 이러한 경험을 살려 박사과정에 임하도록 하겠습니다.

다음으로는 논문 심사를 맡아주신 박세영 교수님과 소재필 교수님께 감사드립니다. 박세영 교수님 덕분에 대학원 과정에서 양자역학, 전자기학 등의 4 대 역학에 대해서 깊게 배우게 되었습니다. 소재필 교수님의 양자 광학 수업을 통해 최근 많은 관심을 받는 양자 분야에 대해서 더 자세히 알게 되었습니다.

학부 연구생 과정에서 가장 많은 것을 알려주고 배울 수 있도록 도와준 나이정 선배님께 감사드립니다. 기본적인 리눅스, 파이썬, BASF2 사용법을 비롯한 기초적인 내용들을 배우게 도와 주셔서 감사합니다. 이외에도 같이 학부 연구생, 석사과정을 진행하면서 제가 부족한 부분을 많이 보강해 주신 안지훈 선배님께도 감사합니다.

마지막으로 Belle II charm Group 분들과 KBelle 분들에게도 감사드리고, KEK SSP 과정에서 많은 도움을 준 Miyake Hideki-san, Ueda I-san , Takanori Hara-san, Kalita Dhiraj 등등 Distributed Computing Group 분들에게도 감사합니다.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

$\pi^{\pm}$ identification performance study based on $D^{\pm}$ to $K_s^0\pi^{\pm}$ decays in the Belle II detector

LEE, GYEONGMIN
Department of Physics
Graduate School of Soongsil University

The Belle II experiment focuses on the study of particle flavor through rare decay processes under high-luminosity conditions. In order to observe such rare decay modes, it is crucial to suppress background signals, which require precise particle identification (PID). In this study, the efficiencies and purities of various pion PID packages implemented in the Belle II software framework were evaluated and compared to determine which package provides the most effective identification of charged pion tracks. This work serves as a foundation for future analyses of rare decay modes, as it is based on well-known decay channels to validate the PID performance

국문초록

$D^{\pm}$가 $K_s^0\pi^{\pm}$ 로 붕괴하여 생성된 데이터를 바탕으로 Belle II 관측기의 $\pi^{\pm}$식별 성능을 확증하는 연구

이경민
물리학과
숭실대학교 대학원

Belle Ⅱ 실험은 높은 광도에서 희귀 붕괴를 통한 입자 맛깔 붕괴에 대한 분석을 핵심으로 두고 있는 실험이다. 여기서 희귀 붕괴 모드의 관측을 하기 위해서는 배경신호를 억제하는 게 중요한데 여기서 입자를 정확히 식별하는 것이 중요하다. 입자식별에 대한 여기에 Belle Ⅱ 소프트웨어 프레임워크에 있는 파이온PID 패키지 들의 효율성과 순도에 대한 비교를 측정 및 비교하여 어떤 패키지가 전하를 가지는 피이온 입자식별에 효과적인지 분석하였다. 해당 연구는 다른 희귀 붕괴모드 분석에 대한 기반을 다지게 되는 연구로 잘 알려진 붕괴모드를 통해 연구를 진행하였다.

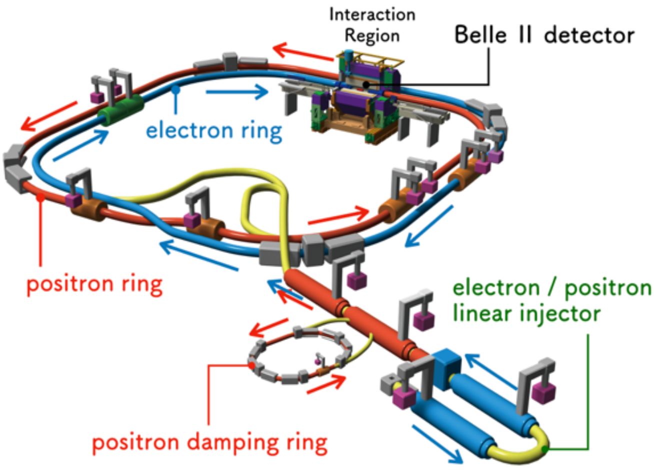
CHAPTER 1 Introduction

Belle II experiment focuses on flavor physics, accurate PID plays on especially important role. Accurate PID is essential for correctly identifying final-state particles and reducing background. Therefore, several particle identification packages in the Belle II software require performance validation. Since particle identification efficiency varies depending on factors such as the particle's momentum, angle, charge and whether it is positively or negatively charged. We specifically compared the probability of a particle being a pion, obtained using the maximum likelihood function of the six candidates with electric charge (electron, muon, pion, kaon, proton, deuterium) with the probability derived through a neural network approach.

1.1 Belle Ⅱ experiment

1.1.1 Super KEKB accelerator

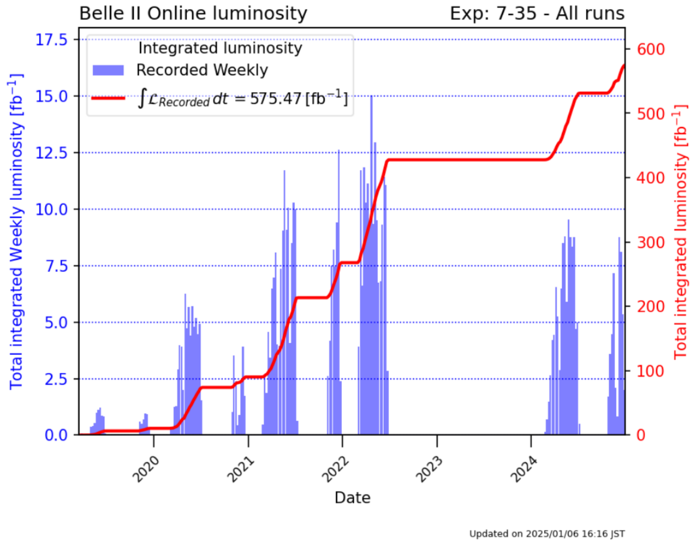
[Figure 1-1] Configuration of SuperKEKB



The Super KEKB accelerator collides electron and positron beams with energies of 7.0 GeV and 4.0 GeV, respectively, asymmetrically. The center-of-mass energy of the collisions is 10.58 GeV, corresponding to the region of the Y(4S) resonance. Belle II experiment is a luminosity-frontier experiment that explores the flavor sector in search of new physics.

During Run1 (2019~2022), a data sample corresponding to an integrated luminosity of 400 fb^{-1}. Following a long shutdown until 2024, Belle II achieved a new world record luminosity of 5.1 $\times 10^{34}\ cm^{-2}s^{-1}$ in December, with a total collected data sample of 575 fb^{-1} Run1 &Run2 (2019 ~ 2022, 2024 January ~) [1]

[Figure 1-2] Belle II online luminosity delivered by SuperKEKB



The luminosity is expressed as,

$$L = \frac{1}{4\pi} \cdot \frac{N_p N_e}{\sigma_x^{effective} \ \sigma_y^*} \cdot \ f_{col} R_L$$

Well-known formula,

$$L = \frac{\gamma_\pm}{2er_e}\left(\frac{I_\pm \zeta_{y\pm}}{\beta_{y\pm}^*}\right)\left(\frac{R_L}{R_{\zeta_y}}\right)$$

The instantaneous luminosity is $6 \times 10^{35}\ cm^{-2}s^{-1}$. An electron bunch completes one revolution in $10\mu s$. Since 2376 bunches are accelerated simultaneously in beam line, collisions can occur at intervals of 250 MHz. The Luminosity per crossing is calculated to be $2.6 \times\ 10^{-6}\ (nb)^{-1}$. The cross section for electron-positron scattering via Bhabha scattering is 125nb, and the probability of scattering occurring in one crossing is approximately $\approx \frac{1}{3100}$. meaning an event happens roughly once every 3100 crossings. [2]

1.1.2 Belle II detector

[Figure 1-3] Belle II detector top view [4]

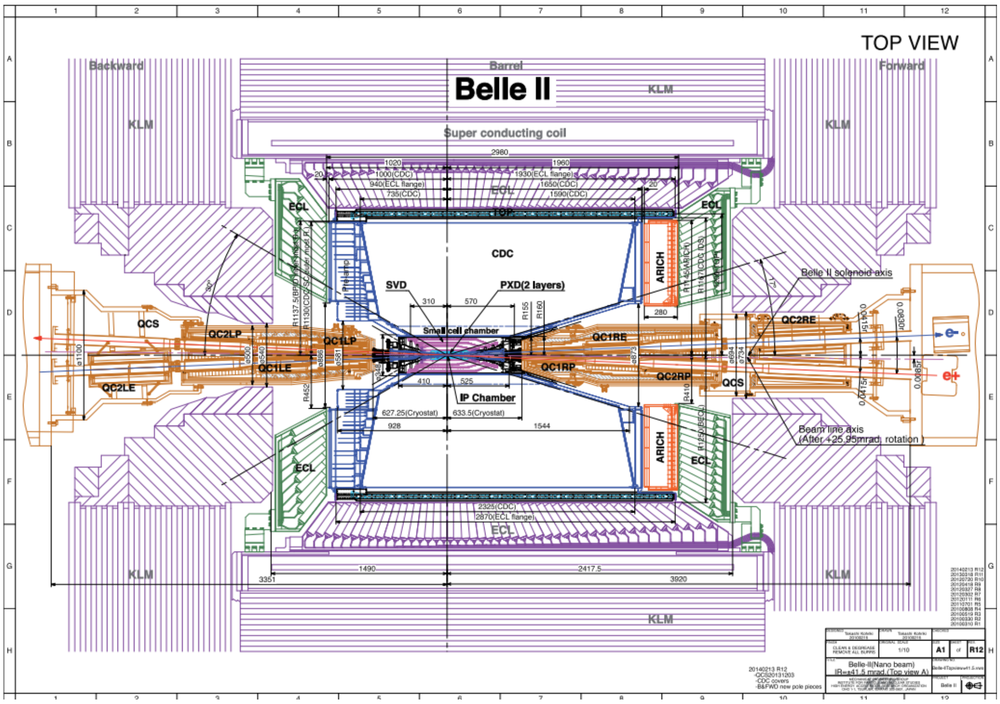
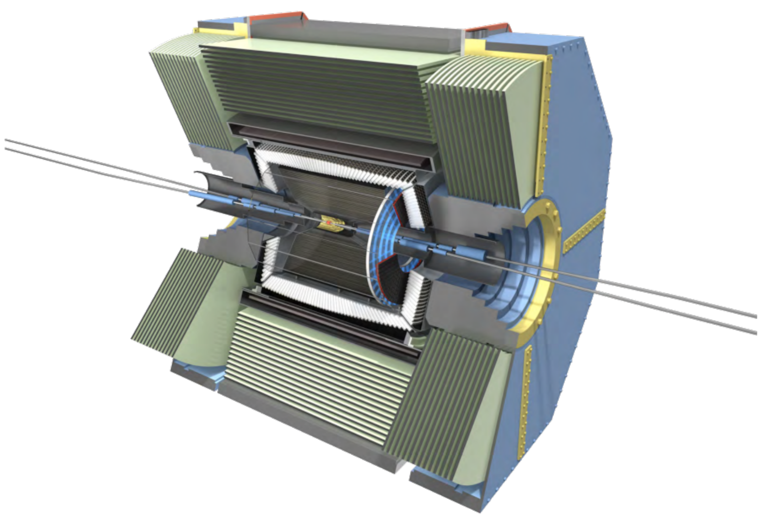


Figure 1-3 show the structure of the Belle II detector. The central region of the. detector consists of silicon detectors known as the PXD and SVD. The PXD (PiXel Dectector) is designed with a focus on precise decay point reconstruction.

The SVD (Silicon Vertex Detector) gathers hit information to reconstruct particle trajectories and obtains information about the energy loss during the process of recording hits for each particle. [3]

The CDC (Central Drift Chamber) consists of 56 layers with a total of 14,336 wires and is filled with a gas mixture of He - C_2H_6 in a 50:50 ratio. Similar to the SVD, the CDC obtains information on energy loss per unit length (dE/dx) and collects hit information to reconstruct particle trajectories. From the trajectory data, it can predict the expected number of hits in other detectors. [4]

[Figure 1-4] Belle II detector image



The ARICH (Aerogel Ring-Imaging Cherenkov) detector is mounted in the forward endcap region. It is designed to identify particles boosted in the forward direction due to the asymmetric beam energies. Cherenkov photons produced in two aerogel layers with different refractive indices ($n_1 = 1.045$ upstream, $n_2 = 1.055$ downstream) from ring patterns on the detector plane, allowing precise particle identification. [5,6]

The TOP (Time of Propagation) detector consists of quartz bars and operates based on the Cherenkov effect. When a particle passes through, Cherenkov photons are emitted at an angle directly related to the particle's velocity. By combining this velocity information with the particle momentum by the tracking detectors, the TOP allows identification of the particle species. [7,8]

The ECL (Electromagnetic Calorimeter) is composed of 8,736 thallium-doped CsI crystal. When a particle enters a crystal, it produces scintillation light proportional to the particle's energy. By measuring this light, the ECL precisely determines photon and electron energies, and – when combined with tracking information – distinguishes electrons from hadrons using the E/p ration and shower shape. [4]

The KLM (K-Long muon) detector is located outside the superconducting solenoid and is composed of alternating layers of 4.7 cm thick iron plates and active detector elements. Muons penetrate deeply through the iron, leaving continuous track-like hit clusters. By distinguishing between track and shower patterns, the KLM identifies muons and K_L^0 mesons. [4]

CHAPTER 2 Motivation

The final states of many of the particles produced at the impact point of the detector are pion (π), photons (γ), and other leptons. To check the event in the high energy physics experiment, we use a method of reconstruction tracks and hits generated by the detector. Particle identification is very important for the reconstruction process to be performed accurately. In the case of rare decay modes, the background should not exceed the signal. Otherwise, leading to distorted results. Pion are the most common final-state particles, but at high momentum, pions can be confused with kions

2.1 PID (Particle identification)

The traces left by a particle passing through the detector are used to infer the identity of the particle. The candidate particles are electron, muon, pion, kaon, proton, and deuterium. The detector is composed of subdetectors such as the SVD, CDC, ARICH, TOP, ECL and KLM. For each detector and particle type, up to 36 log-likelihood functions can be constructed. By summing the log-likelihood functions from each detector for a specific particle, likelihood function for that particle is generated. This process allows the construction of likelihood functions for all six candidate particles

Ex) Pion likelihood

$$\log(\mathcal{L}_\pi) = log(\mathcal{L}_\pi^{SVD} \cdot \mathcal{L}_\pi^{CDC} \cdot \mathcal{L}_\pi^{TOP} \cdot \mathcal{L}_\pi^{ARICH} \cdot \mathcal{L}_\pi^{ECL} \cdot \mathcal{L}_\pi^{KLM})$$

The likelihood function is used to obtain the probability of a particle

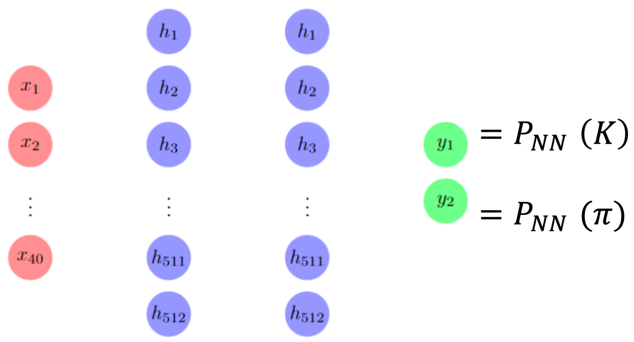
Ex) Probability of a pion

$$P(\pi) \equiv \frac{log(\mathcal{L}_\pi)}{log(\mathcal{L}_\pi \cdot \mathcal{L}_K \cdot \mathcal{L}_e \cdot \mathcal{L}_\mu \cdot \mathcal{L}_p \cdot \mathcal{L}_d)}$$

The probability of a pion, obtained through this method, is called pionID.

The pionIDNN package, which uses a neural network approach, combines 36 log-likelihood functions along with information on track momentum, magnitude, direction and charge as 40 inputs and processes them through two hidden layers each containing 512 nodes. [9, 10]

[Figure 2-1] PonIDNN for 2 hypotheses [10]



2.2 Deacy mode

This study is a performance study, and therefore, a reference mode has been selected. $D^{\pm}$ to $k_s^0\pi^{\pm}$ decays mode have a relatively large branching fraction and be capable of accumulating numerous events. Additionally, because of the detection performance differences between positively and negatively charged particles, the charge conjugate mode - which is intuitive - was selected. The branching fraction of this decay mode is 1.562% in simulation. The degree of this selection appears reasonable. [11,12,13]

CHAPTER 3 Software and data set

3.1 Software setup

The Belle II software framework, basf2, is developed using both C++ and Python, and it allows users to select various packages and modules. Through this framework. The Belle II workflow performs essential tasks such as particle data generation, detector simulation, event reconstruction, and user analysis. These processes are modularized and can be executed via Python streeting files, providing a flexible and efficient analysis environment. [14]

3.2 MC (Monte-Carlo) data set

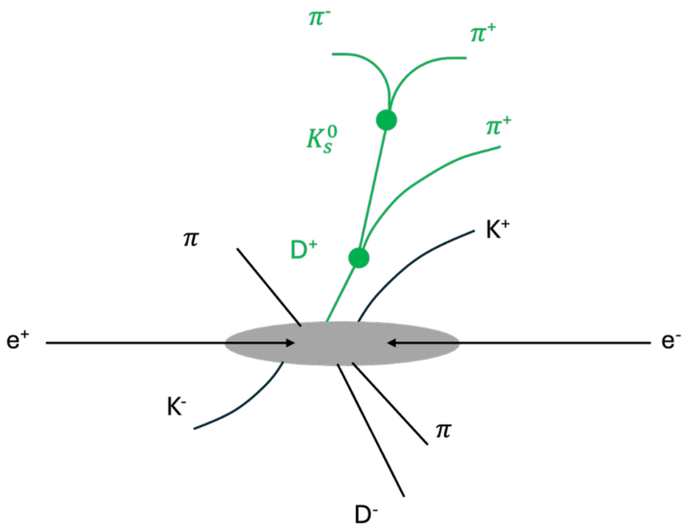
The Monte Carlo includes all known charm decays and accounts for both beam and detector conditions (run - dependent). A total of 554,313 events were generated, and by multiplying by the branching fraction, approximately 6,650 signals are expected.

CHAPTER 4 Event selection

4.1 Reconstruction

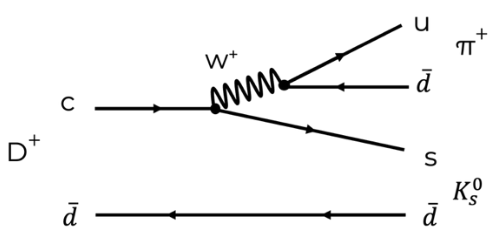
4.1.1 Reconstruction method

[Figure 4-1] Diagram reconstruction structure



The green indicates is signal decay mode. K_S^0 is reconstructed from the tracks of two pions. Approximately 69.2% of K_S^0 decay to $\pi^+\ \pi^-$. The reconstructed K_S^0, combined with an additional charged pion, is used to reconstruct a charged D meson. Ultimately, the detector records three pion tracks. [15]

[Figure 4-2] Feynman diagram of decay mode



4.2 Selection criteria

4.2.1 Pre-selection

[Table 4-1] Pre-selection table

particle	Value name and cut
$\pi^{\pm}$	abs(dr) < 3.0cm, abs(dz) < 5.0, thetaCDCAceeptance (17~150 degrees), χ^2 < 0.001 (pvalue)
k_s^0	Standard Kshorts package (0.480 < M <0.520 GeV/C^2)
$D^{\pm}$	1.75 < M < 2 GeV/C^2 , 2 < P* < 5 [GeV/C]

For pion candidates, the track selection requires abs(dr) < 3.0cm, abs(dz) < 5.0, corresponding to the transverse and longitudinal impact parameters, respectively. These cuts ensure that the pion track are consistent with originating from the interaction point, rather than from secondary vertices or detector background. The thetaCDCAceeptance (17~150 degrees) cut is applied because the CDC is designed to detect particles within the polar angle range of approximately 17~ 150 degrees. Particles outside this range travel into regions such as the beam pipe, where no tracking detectors are present, and therefore cannot be reconstructed. A requirement of χ^2 probability (p-value) > 0.001 is applied to ensure the quality of the track fit, rejecting poorly reconstructed tracks that are inconsistent with the fitted trajectory

“Standard K shorts package “is standard particle package in Belle II software framework. In this package, a vertex fit is performed, and only candidates with an invariant mass in the range 0.450 GeV < M < 0.550 GeV, and for which the vertex fit did not fail, are kept. [15]

For the D meson, selection cuts were applied based on its basic mass and the momentum in the center-of-mass frame (P*)

As a result of these cuts, a total of 47,499 events were reconstructed, with 8,141 identified as signal events. The reconstruction efficiency, calculated as the ratio of the number of signal events (8,141) to the expected number of signal events (17,316), is 0.47

4.3 Fitting result

For the fitting, the signal was modeled using a double-sided-Crystal ball function, while the background was described by a second-order polynomial.

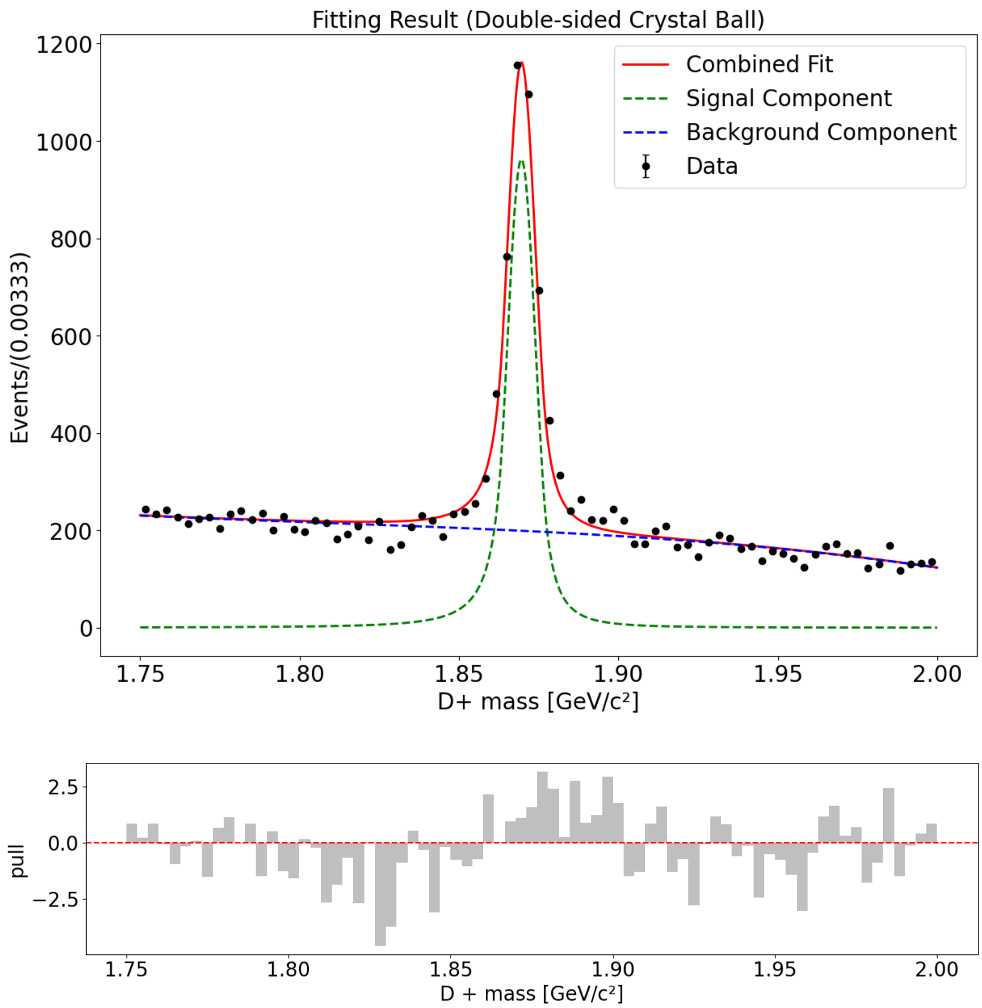
Double-sided-Crystal ball function. [16]

$$f(x;\mu,\sigma,\alpha_L,n_L\alpha_R,n_R) = \begin{cases} A_L\cdot\left(B_L-\frac{x-\mu}{\sigma}\right)^{-n}, for\,\frac{x-\mu}{\sigma} < -\alpha_L \\ exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ A_R\cdot\left(B_R-\frac{x-\mu}{\sigma}\right)^{-n}, for\,\frac{x-\mu}{\sigma} > -\alpha_R \end{cases}$$

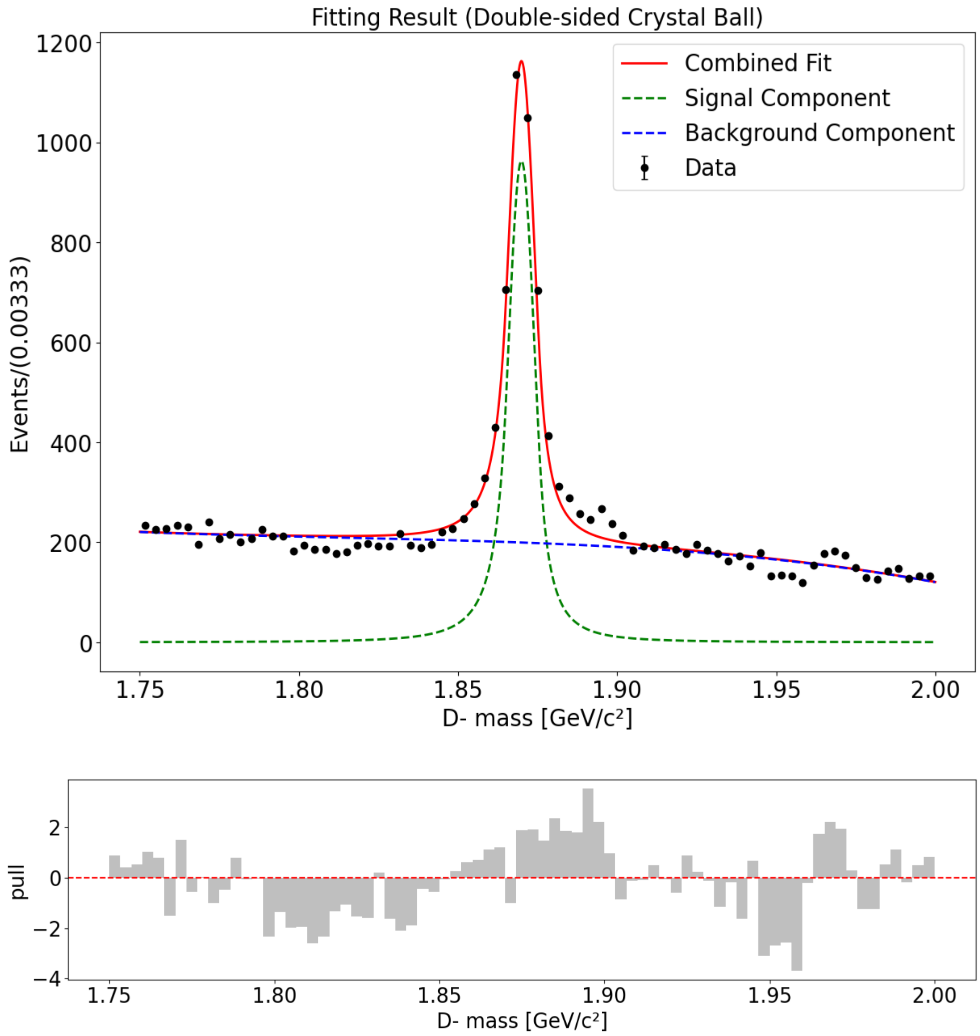
$$A_{L/R} = \left(\frac{n_{L/R}}{\alpha_{L/R}}\right)^{n}_{L/R} \cdot exp\left(-\frac{|\alpha_{L/R}|}{2}\right)$$

$$B_{L/R} = \frac{n_{L/R}}{|\alpha_{L/R}|} - |\alpha_{L/R}|$$

[Figure 4-3] Mass fitting Result of $D^{+} to\ k_s^0\ \pi^{+}$



[Figure 4-4] Mass fitting Result of $D^{-} to\ k_s^0\ \pi^{-}$



[Table 4-2] Fitting parameters table

Parameter	$D^{+} to\ k_s^0\ \pi^{+}$	$D^{-} to\ k_s^0\ \pi^{-}$
mean	1.869	1.869
sigma	4.881×10^{-3}	4.815×10^{-3}
α_L	5.267	9.908
n_L	0.1003	0.4611
α_R	1.364	1.274
n_R	2.756	2.321
quadratic	-0.2545	-0.2259
linear	-0.7201×10^{-1}	-0.9842×10^{-1}
constant	-0.2611×10^{-1}	-0.3913×10^{-1}

The number of degrees freedom (ndf) for this fit is 66, obtained by subtracting the number of fit parameter (9) from the number of bin (75). The resulting χ^2 values are 173.02 for the positive mode and 156.09 for the negative mode, yielding χ^2/ndf ratios of 2.622 and 2.365, respectively

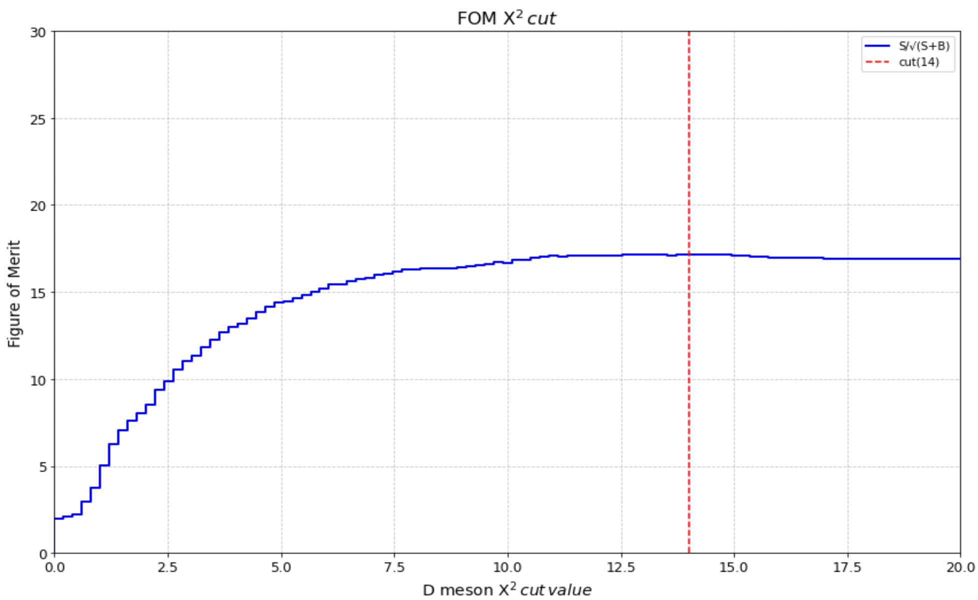
The pull is a graph used to check the deviation at each bin, and the formula for the pull is given as follows.

$$Pull = \frac{N_{data} - N_{fit}}{\sqrt{N_{fit}}}$$

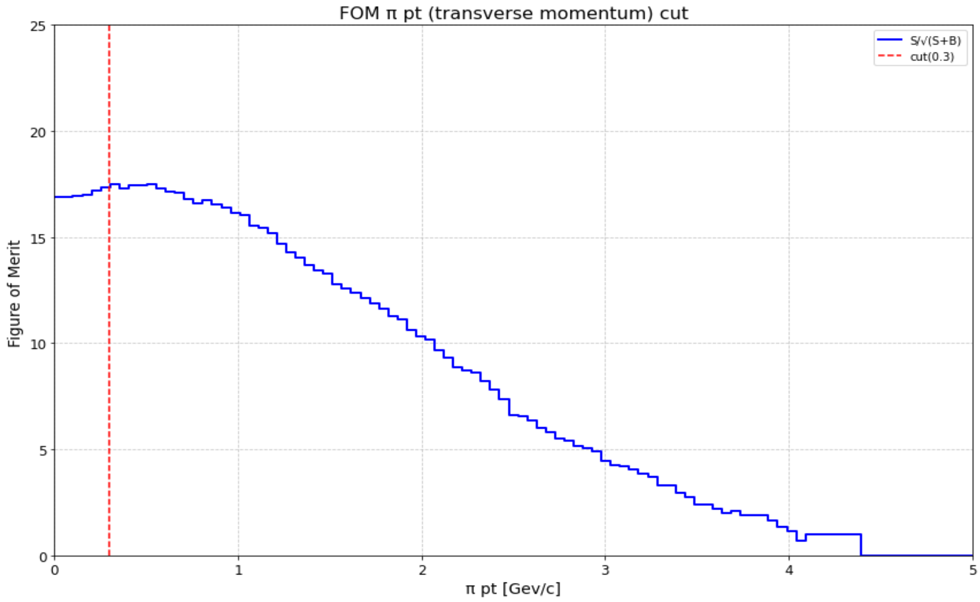
4.4 FOM (Figure of Merit) for after selection

To improve the purity of the tuples, additional cuts were applied to remove background events. The figure of merit, defined as $\frac{S}{\sqrt{S+B}}$ (S: signal events, B: background events), was used to identify cuts that maximize signal events while minimizing background. Using this FOM, cuts on the pion's transverse momentum and χ^2 value were introduced. [11]

[Figure 4-5] FOM of D meson χ^2 value



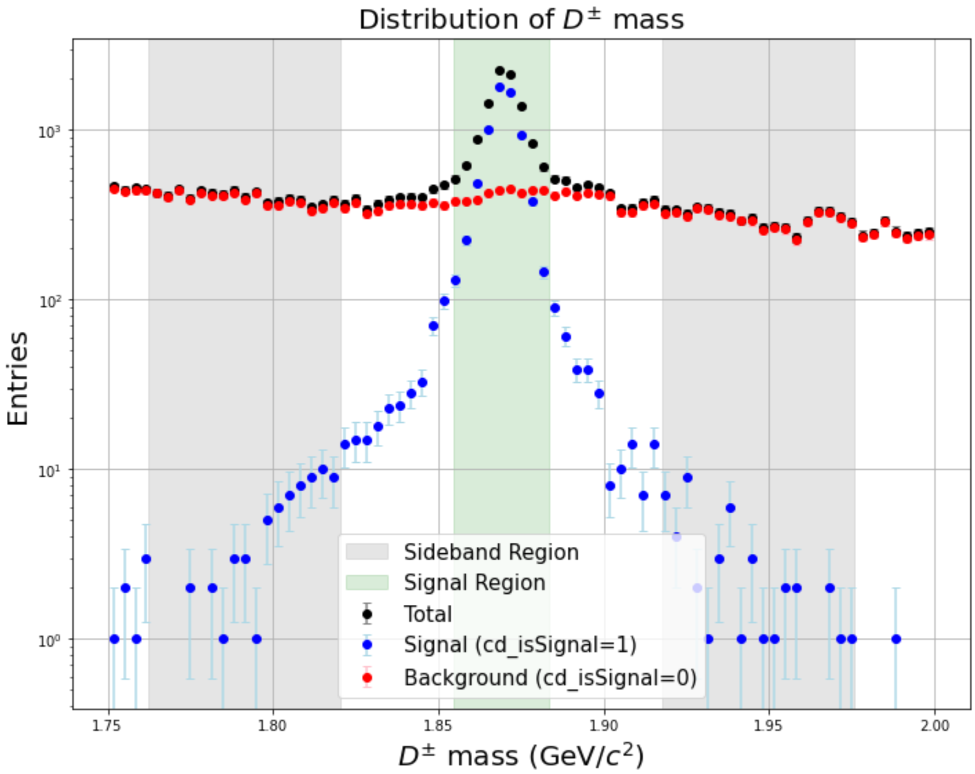
[Figure 4-6] FOM of pion's transverse momentum value



4.5 Sideband subtraction

The sideband subtraction method is a. technique used to remove background events within the signal region. First, the signal region is defined, and region outside the signal area are designated as sideband. These sidebands are regions where signal events are nearly absent, and only background events are present. The observed events in the sideband regions are used to estimate the expected background within the signal region. By subtracting the estimated background from the total events in the signal region is chosen as the 3-sigma interval around the mean, obtained from the average and sigma values determined during the fitting process, while regions outside this are designated as the sidebands.

[Figure 4-7] Distribution of D meson mass, signal and sideband region



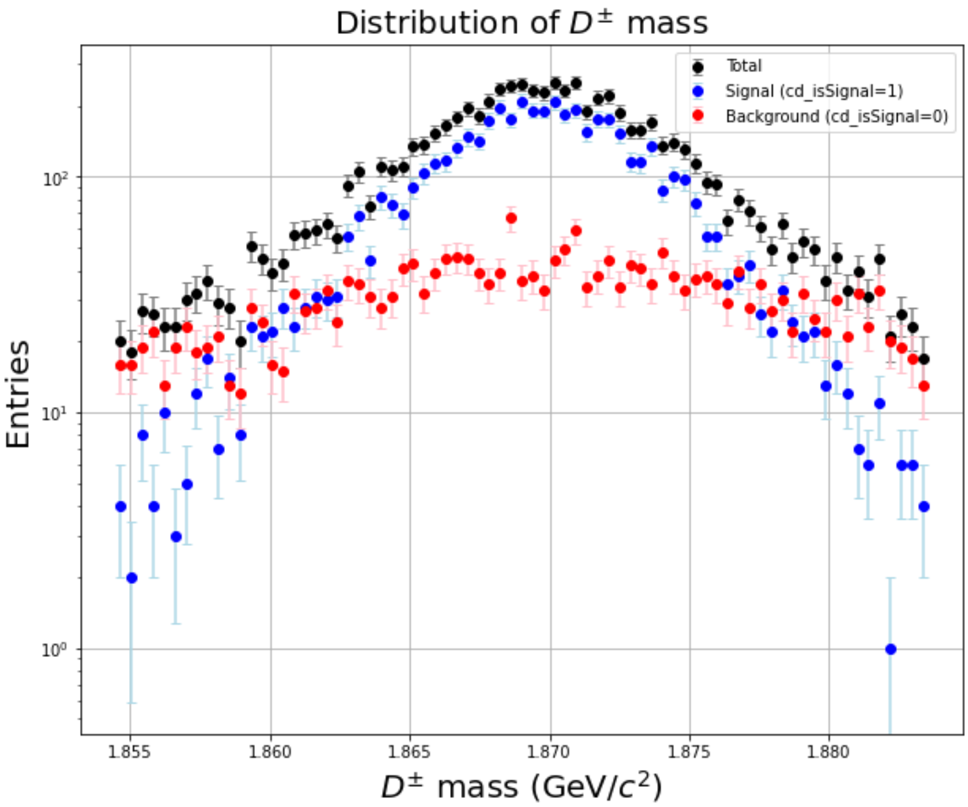
Signal region: mean $\pm 3\sigma$ [1.844 ~ 1.893] GeV/c^2

Background region:

Left: mean -22σ ~ mean -10σ [1.762 ~ 1.821] GeV/c^2

Right: mean $+22\sigma$ ~ mean $+10\sigma$ [1.918 ~ 1.976] GeV/c^2

[Figure 4-8] After sideband subtraction



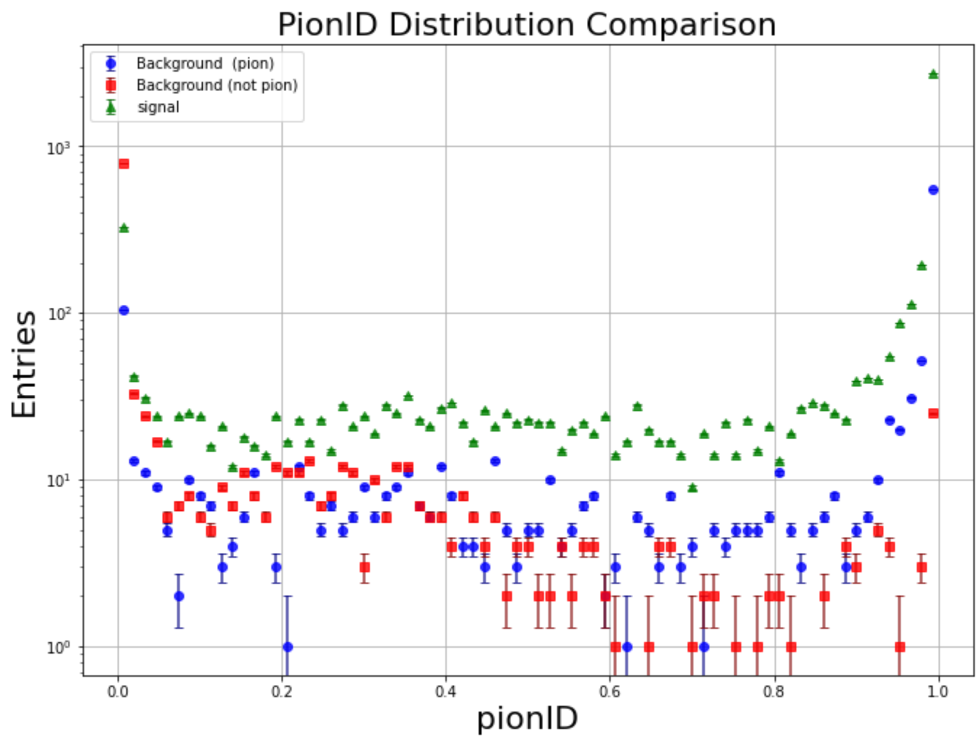
After sideband subtraction, the total number of remaining events is 4,433 with 3,379 identified as signal events. The purity is therefore 0.76

CHAPTER 5 PID performance

5.1 PionPID

This is the distribution of the pionID from the final data. Values close to 0 indicate particles classified as not pions, while values near 1 indicate particles identified as pions. The color of the points corresponds to their origin: green points represent pions from $D^{\pm}$ to $k_s^0\pi^{\pm}$ decays mode, blue points are background pions misreconstructed from other decay modes, and red points are particles that are not pions but were reconstructed as such.

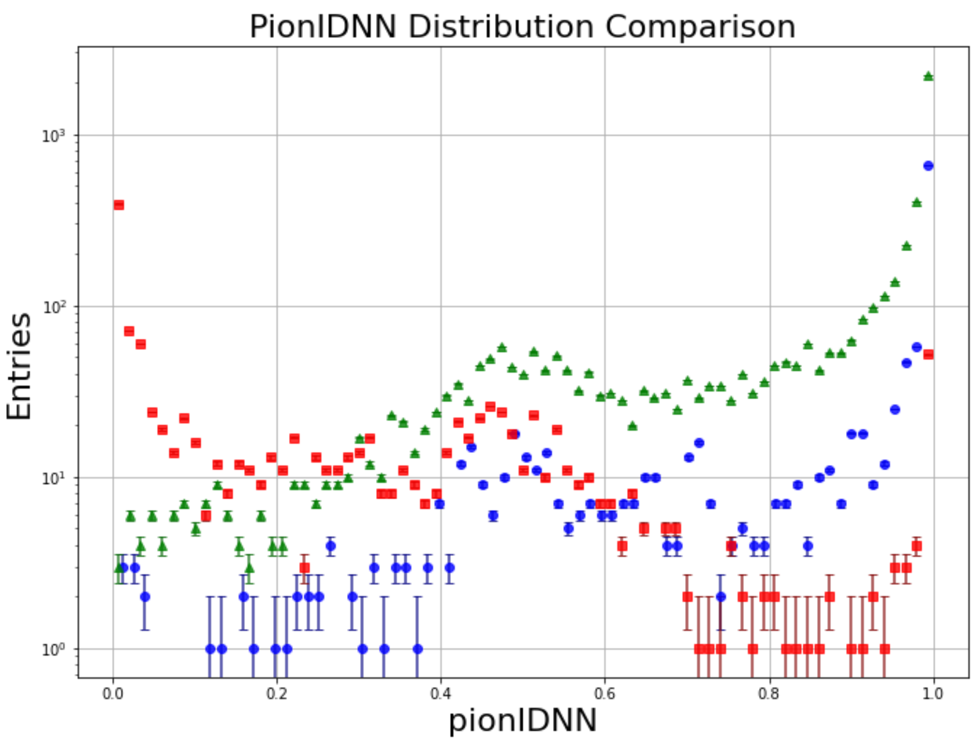
[Figure 5-1] PionID Distribution



5.2 PionPIDNN

The pion identification package obtained using a neural network exhibits the following distribution. Unlike the pionID, which overall mixes signal and background ratios and shows distinct differences near 0 and 1, the neural network show a clustering of events around 0.5. Additionally, in the region below 0.2, the neural network distinguishes pions defectively based on their presence or absence.

[Figure 5-2] PionIDNN Distribution



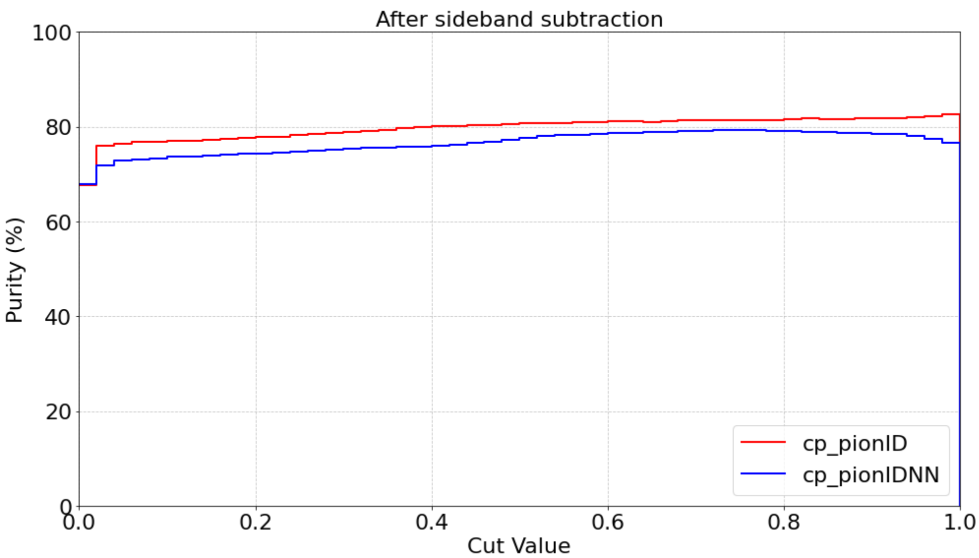
5.3 PID comparison

When using the PID value as a cut, the number of remaining events changes, and the graphs below show the purity and efficiency of the remaining events. The formulas for calculating purity and efficiency are given as follows.

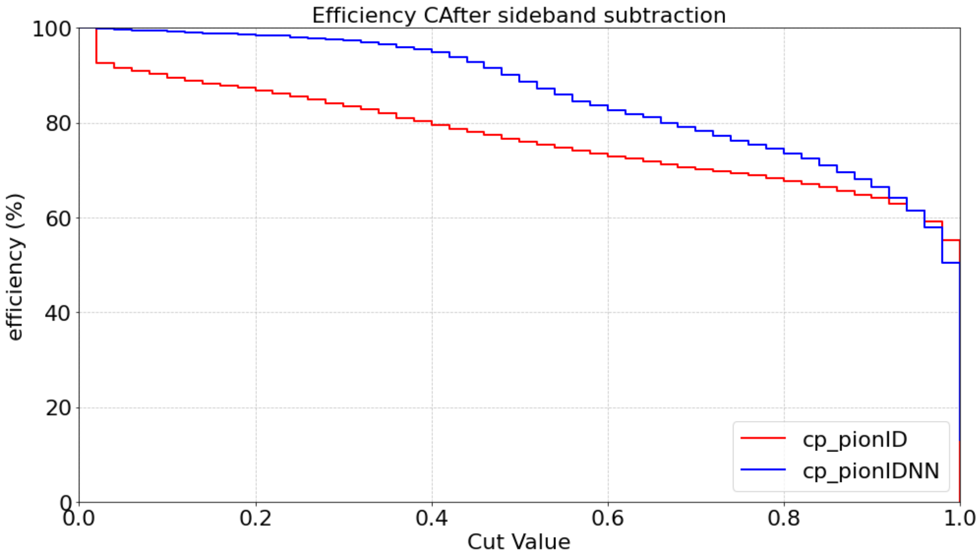
$$Purity = \frac{N(signal\ after\ selections)}{N(candidates\ after\ selections)}\ ,$$

$$Efficiency = \frac{N(signal\ after\ PIDcut)}{N(signal\ before\ PIDcut)}$$

[Figure 5-3] Purity

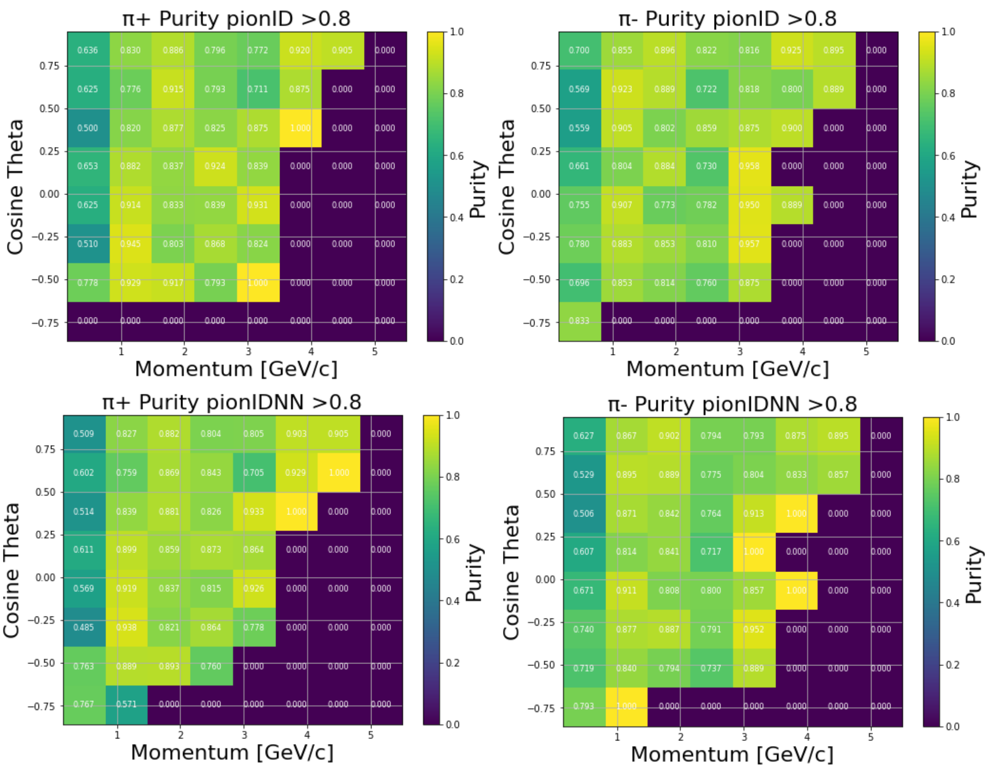


[Figure 5-4] Efficiency

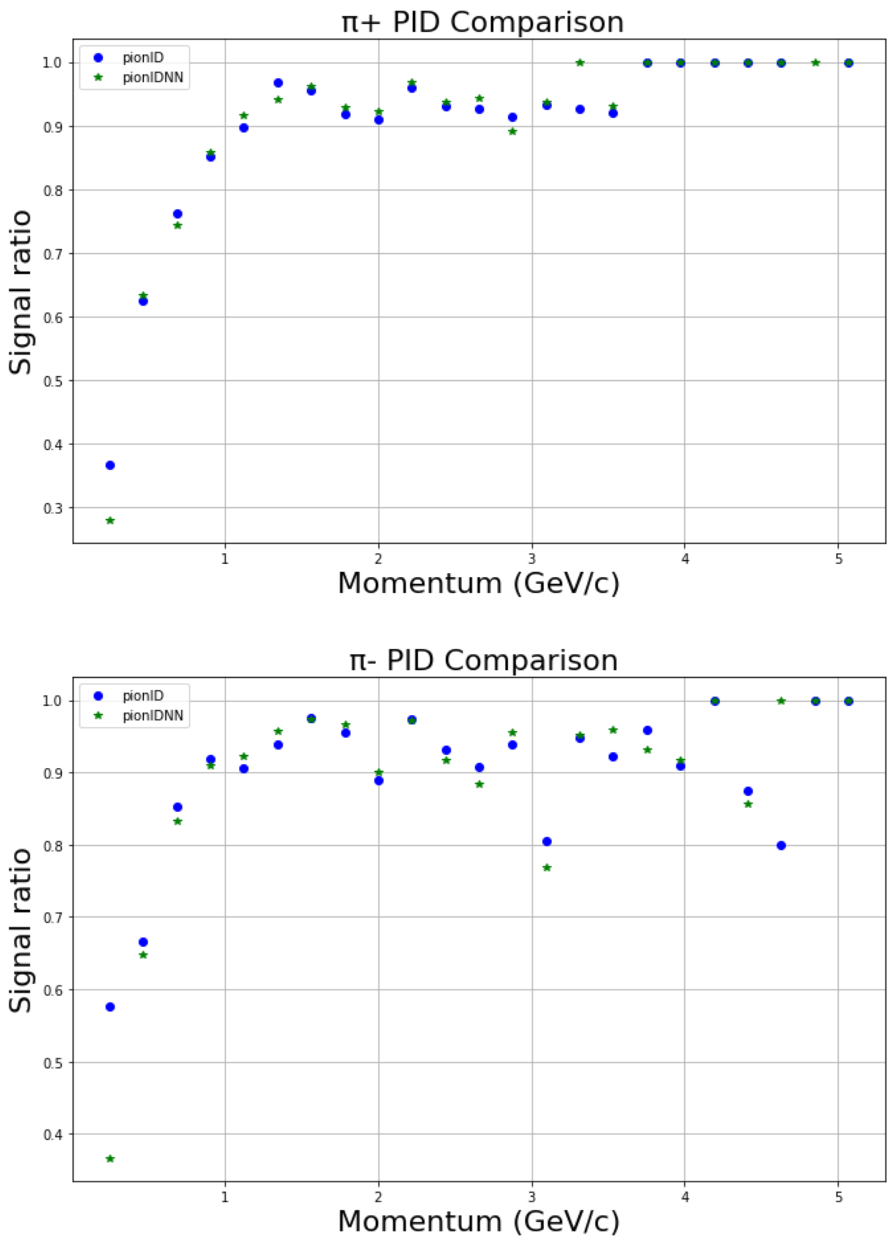


Since the particle identification efficiency varies depending on the particle's trajectory, heatmap graphs of the particle's track path and momentum were created for each charge state. These graphs show the ratio of signal to background events for each bin. Additionally, to verify identification by momentum, graphs of the signal – to - background ratio as a function of momentum were also plotted. Only events with a PID value greater than 0.8 were retained for these graphs.

[Figure 5-5] Purity by momentum and cosθ (PID > 0.8 and charge)

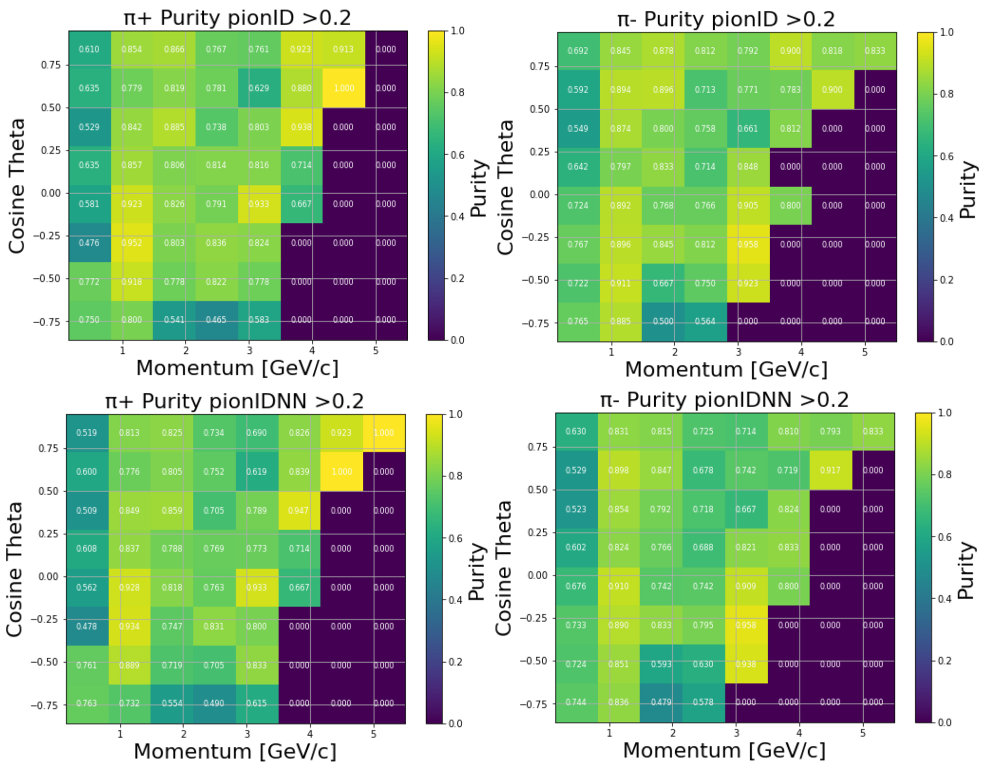


[Figure 5-6] Purity by momentum (PID > 0.8 and charge)

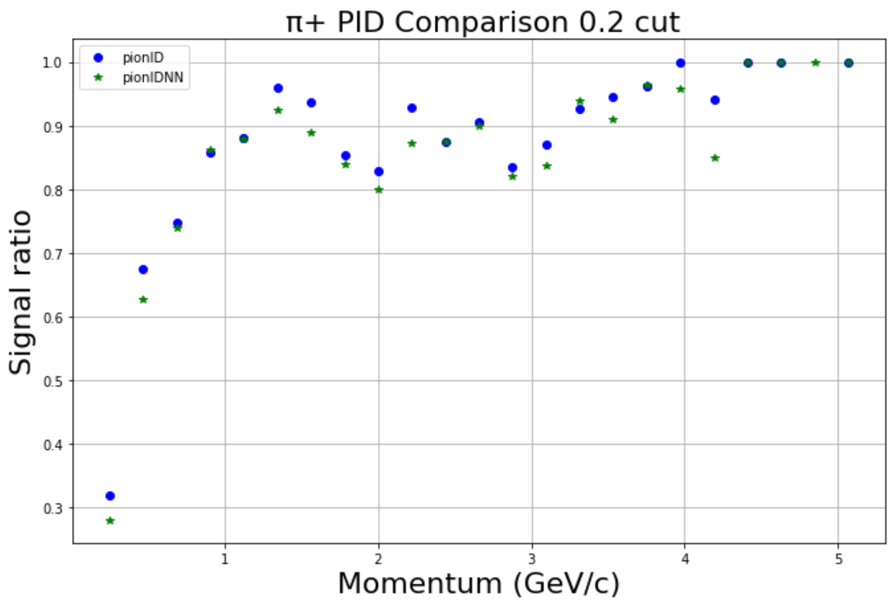


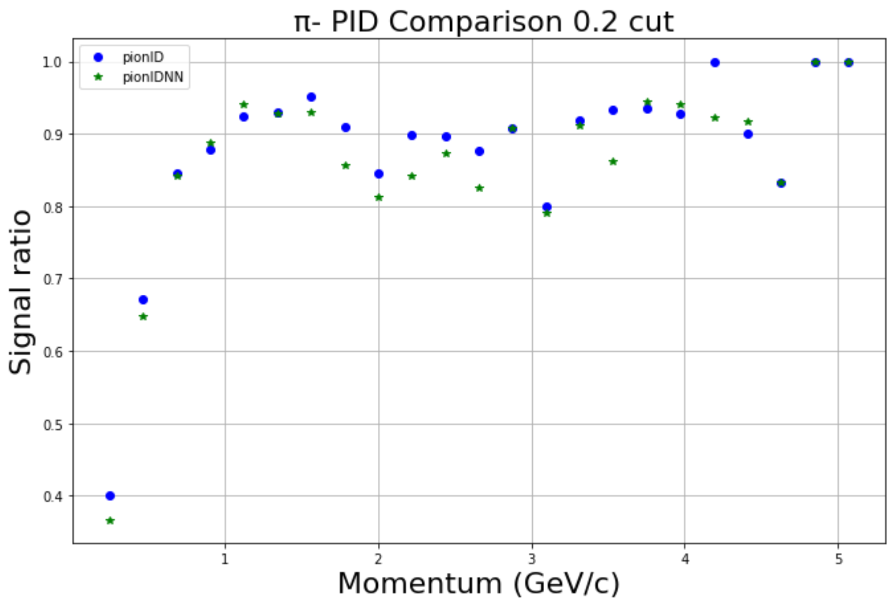
For the pion identification using the neural network method, pionIDNN, it was expected that applying a cut at values below 0.2 would significantly reduce the background while retaining signal events. Therefore, a cut of 0.2 was applied.

[Figure 5-7] Purity by momentum and cosθ (PID > 0.2 and charge)



[Figure 5-8] Purity by momentum (PID > 0.2 and charge)





CHAPTER 6 Conclusion

In the Belle II experiment, because many final states include pions, it is crucial to accurately identify pions. The amount of signal and background events varies depending on the particle identification package used and the cut values set for the PID, which is especially important for rare decay modes. After examining the types of information available from the detector, we obtained the signal – to – background event ratios for each parameter used in the PID. Using these results, we compared the differences between the PID based on neural networks and the traditional PID based on likelihood functions.

REFERENCES

[1] Belle II Collaboration, "Belle II Online luminosity"

https://www.belle2.org/research/luminosity/

[2]$e^{+}e^{-}$ $collision$, Belle II Collaboration.

"Belle II Online Documentation 3.2.2. data taking"

https://training.belle2.org/online_book/fundamentals/02-datataking.html

[3] T. Abe, et al., arXiv:1011.0352 [physics.ins-det] [Search inSPIRE].

[4] Progress of Theoretical and Experimental Physics – The Belle II Physics Book

[5] T. Iijima et al., Nucl. Instrum. Meth. A **548**, 383 (2005) [arXiv:physics/0504220]

[6] P. Krizan et al., Nucl. Instrum. Meth. A **565**, 457 (2006) [arXiv:physics/0603022]

[7] M. Akatsu et al., Nucl. Instrum. Meth. A **440**, 124 (2000) [arXiv:physics/9904009]

[8] M. Staric et al., Nucl. Instrum. Meth. A **595**, 252 (2008).

[9] PID variables, Belle II Collaboration.

"Belle II Online Documentation 7.3. Variables"

https://training.belle2.org/analysis/doc/Variables.html

[10] pionIDNN, Belle II Collaboration.

BELLE2-TALK-CONF-2023-023

[11] Search for CP violation in the decay $D^{\pm}$ to $k_{s}^{0}\pi^{\pm}$ (BaBar)

PHYSICAL REVIEW D 83, 071103(R) (2011)

[12] Evidence for CP Violation in the decay $D^{\pm}$ to $k_{s}^{0}\pi^{\pm}$ (Belle)

Phys. Rev. Lett. 109, 021601

[13] Particle Data Group. “2025 Review of Particle Physics-$D^{\pm}$ Listings”, page 3

[14] basf2, Belle II Collaboration.

“Belle II Online Documentation 3.4.1 The basics”.

https://software.belle2.org/development/sphinx/online_book/basf2/introduction.html

[15] Standard Particles, Belle II Collaboration.

“Belle II Online Documentation 7.1.1 Standard Particles”.

https://software.belle2.org/development/sphinx/analysis/doc/StandardParticles.html

[16] Double-sided-Crystal ball function.

“DoubleCB Zfit 0.6.4 Documentation”

https://zfit.readthedocs.io/en/0.6.4/user_api/_generated/pdf/zfit.pdf.DoubleCB.html