Search for Heavy QCD Axion-like Particles at BELLE-II

A Dissertation Submitted in Partial Fulfilment of the Requirements
for the Degree of

Bachelor of Science (Research)

by

Aman Sahoo
SR No. 11-01-00-10-91-22-1-22164

भारतीय विज्ञान संस्थान

under the guidance of

DR. MINAKSHI NAYAK

CENTRE FOR HIGH ENERGY PHYSICS
INDIAN INSTITUTE OF SCIENCE
BENGALURU - 560012

April 2026

DECLARATION

I, Aman Sahoo, declare that this project titled, "Search for Heavy QCD Axion-like Particles at BELLE-II" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Name: Aman Sahoo

Signed: Aman

Date:10/04/2026

CERTIFICATE

This is to certify that the Bachelor's project titled "Search for Heavy QCD Axion-like Particles at BELLE-II" submitted by Aman Sahoo (Sr No.11-01-00-10-91-22-1-22164) to Indian Institute of Science, Bangalore towards partial fulfilment of requirements for the award of degree of Bachelor of Science (Research) in Physics is a record of bona fide work carried out by him under my supervision and guidance during Academic Year, 2025–26.

Signature of the faculty

Name of the faculty:

Dr. Minakshi Nayak   Minakshi Nayak

Prof. Sudhir Vempati

Department: Centre for High Energy Physics

IISc Bangalore, India   Date: 10/04/2026

Acknowledgements

I would like to express my sincere gratitude to the Office of Undergraduate Studies and the Centre for High Energy Physics at the Indian Institute of Science (IISc) for providing the academic environment, resources, and institutional support necessary for carrying out this work. Their support has been instrumental in enabling this research.

I am deeply grateful to Dr. Sourav Dey (Kobayashi-Maskawa Institute, Nagoya University) and Prof. Shohei Nishida (KEK) for their valuable insights, guidance, and technical input throughout the course of this thesis. Their expertise and constructive feedback significantly contributed to the development and direction of this work.

I would especially like to thank Dr. Minakshi Nayak (IISc) for her supervision from the very beginning of this project to its completion. Her guidance, encouragement, and ability to provide unique perspectives at critical stages greatly enriched this study. I am particularly appreciative of the opportunities she provided, which were crucial to both the academic and practical aspects of this research.

Finally, I acknowledge all those who contributed directly or indirectly to this work.

Aman Sahoo
Indian Institute of Science
10/04/2026

ABSTRACT

Name of the student: Aman Sahoo
SR No. 22164
Department: **Undergraduate Studies/ Centre for High Energy Physics**
Degree: **BS (Res.)**
Thesis title: **Search for Heavy QCD Axion-like Particles at BELLE-II**
Thesis supervisor: **Dr. Minakshi Nayak**
Date of thesis submission: **12/04/2026**

We present a Monte Carlo feasibility study for searching for heavy QCD axion-like particles (ALPs) at the Belle II experiment via the decay $B^+ \to K^+ a$, $a \to \eta\pi^+\pi^-$, $\eta \to \gamma\gamma$. Heavy QCD axions arise in extensions of the Peccei–Quinn mechanism where additional hidden-sector contributions decouple the axion mass from the canonical QCD relation, permitting masses in the MeV–GeV range while simultaneously resolving the strong CP problem and the axion quality problem. The flavor-changing neutral current transition $b \to sa$ is induced at two loops through the axion-gluon coupling, yielding an estimated branching fraction BR$(B \to Ka) \sim 10^{-6}$–10^{-7} for $f_a \sim$ TeV. Signal Monte Carlo samples were generated using the basf2 framework at the KEK Computing Centre across the kinematically allowed ALP mass range of 0.85–3 GeV, the upper bound being imposed by dense charmonium resonances that preclude reliable signal extraction above this threshold. Standard Belle II selection criteria were applied alongside figure-of-merit-optimized cuts on photon energy ($E_\gamma > 150$ MeV) and the $\eta\gamma\gamma$ opening angle ($\cos\theta > 0$), achieving reconstruction efficiencies of approximately 10–13% across the signal mass points. Standard Model control channels ($B \to KD^0$ and $B \to K\eta'$) sharing the same final-state topology were used to validate the reconstruction strategy. Continuum background from $e^+e^- \to q\bar{q}$ processes was suppressed using Boosted Decision Trees trained on event-shape variables including thrust, Fox–Wolfram moments, and Kakuno–Super-Fox-Wolfram moments, achieving continuum rejection rates exceeding 90% with signal loss below 7%. The energy difference ΔE is established as the primary fitting observable for eventual signal extraction from data.

Contents

List of Figures **ix**

List of Tables **xi**

1 Introduction and Motivation **1**
- 1.1 Overview of Belle II and SuperKEKB . . . 1
- 1.2 Motivation for Searching Heavy QCD Axions at Belle II . . . 3
- 1.3 Groundwork laid by previous experiments and analyses . . . 4

2 Theory of Heavy QCD Axion-like Particles **6**
- 2.1 Strong CP Problem and the Peccei–Quinn (PQ) Mechanism . . . 6
- 2.2 Motivation for Heavy QCD Axions . . . 6
- 2.3 Axion Quality Problem . . . 6
- 2.4 Heavy QCD Axion Effective Theory . . . 7
- 2.5 Flavor-Changing Processes: $B \to Ka$ . . . 8
- 2.6 Target Channel: $B \to K\,[a \to [\eta \to \gamma\gamma]\,\pi^+\pi^-]$ . . . 8
- 2.7 Experimental Signatures . . . 9

3 Monte Carlo Simulations **10**
- 3.1 Introduction . . . 10
- 3.2 Generation-Level and Reconstruction-Level Data . . . 10
  - 3.2.1 Generation-Level Data . . . 10
  - 3.2.2 Reconstruction-Level Data . . . 10
- 3.3 basf2 Framework and Reconstruction Scripts . . . 11
- 3.4 MC Truth Matching . . . 11
- 3.5 Choice of ALP Mass Points . . . 11
- 3.6 Key Kinematic Variables . . . 12
  - 3.6.1 Beam-Constrained Mass . . . 12
  - 3.6.2 Energy Difference . . . 12
  - 3.6.3 ALP Invariant Mass . . . 12
  - 3.6.4 Fitting Variables . . . 12
- 3.7 Initial Reconstruction and Benchmarking . . . 12
- 3.8 Reconstruction Efficiency . . . 13
- 3.9 Figure of Merit Optimization . . . 14
- 3.10 Generic Background . . . 18
  - 3.10.1 Definition . . . 18
  - 3.10.2 Processing . . . 18
- 3.11 Reconstruction Efficiency Across the ALP Mass Range . . . 19
- 3.12 Impact of Standard Model Resonances at High Mass . . . 20
- 3.13 Fitting Strategy . . . 22

4 Continuum Suppression using Multivariate Analysis **23**
4.1 Introduction . . . 23
4.2 Underlying Physical Principle . . . 23
4.3 Continuum Suppression Variables . . . 23
4.3.1 Thrust . . . 23
4.3.2 Fox–Wolfram Moments . . . 24
4.3.3 Modified Super Fox–Wolfram Moments . . . 24
4.4 Boosted Decision Trees (BDT) . . . 25
4.5 Training Strategy . . . 26
4.5.1 Training Dataset . . . 26
4.6 Loose vs Tight Cut Strategies . . . 27
4.7 Variable Selection and Optimization . . . 27
4.7.1 Variable Importance . . . 27
4.7.2 Correlation Analysis . . . 27
4.8 Example of Training Iterations . . . 27
4.9 MVA Output and Weight Files . . . 30
4.10 Optimization using Figure of Merit . . . 31
4.11 Performance Evaluation . . . 31
4.12 Mass-Independent Training . . . 32

5 Outlook and Conclusion **36**
5.1 Outlook: Future Steps of the Analysis . . . 36
5.1.1 Unified Continuum Suppression across the Full Mass Range . . . 36
5.1.2 Signal Shape Estimation and Fitting Strategy . . . 36
5.1.3 Standard Model Particle Vetoes . . . 37
5.1.4 Detector Sensitivity Estimation . . . 37
5.2 Conclusion . . . 38

Appendices **38**

A Derivation of the Cross Section of the decay $B \to Ka$ **39**

B Statistics **40**
B.1 Introduction . . . 40
B.2 Measurement and Uncertainties . . . 40
B.3 Error Propagation . . . 40
B.4 Invariant Mass and Error Propagation . . . 41
B.5 Mass Dependence of ΔE Resolution . . . 41
B.6 Interpretation for ALP Reconstruction . . . 42
B.6.1 Light ALP Regime . . . 42
B.6.2 Heavy ALP Regime . . . 43
B.7 Residual Distributions and Error Cancellation . . . 43
B.8 Energy Difference ΔE Resolution . . . 44
B.9 Detector Resolution in Belle II . . . 45
B.9.1 Vertex Detector (VXD) . . . 45
B.9.2 Central Drift Chamber (CDC) . . . 45
B.9.3 Electromagnetic Calorimeter (ECL) . . . 45
B.10 Correlation Between Observables . . . 45
B.11 Implications for Analysis . . . 45

C Supplementary material for continuum suppression **46**

C.1 ROC curves for characterizing MVA performance 46
C.1.1 Definition 46
C.1.2 Theoretical Significance 46
C.1.3 Interpretation in the Analysis 46
C.2 Plots and data used for iterative training of MVA for ALP mass 1.92 GeV . . 47
C.3 Plots and data used for iterative training of MVA for ALP mass 2.96 GeV . . 50

List of Figures

1.1 Overview of the SuperKEKB accelerator complex [2] 1
1.2 Schematic view of the Belle II detector [4] 2
1.3 Feynman diagrams that arise from Eq. 2.8 5

2.1 Feynman diagrams that arise from Eq. 2.8 7
2.2 Particle Production in target decay channel 8

3.1 Initial attempts at reconstructing the ALP and extracting variable distributions 13
3.2 Applying 3σ cuts to reduce combinatorial background 14
3.3 Applying 3σ cuts to reduce combinatorial background (SM channels) 15
3.4 FoM optimization for E_γ 16
3.6 ΔE distribution for ALP decay channel after std. cuts but before FoM cuts . 16
3.5 FoM optimization for $\cos\theta$ 17
3.7 ΔE distribution for ALP decay channel after std. cuts and FoM cuts 17
3.8 $M_{\eta\pi\pi}$ distribution for ALP decay channel after std. cuts and FoM cuts . . . 18
3.9 ΔE plotted after generic MC events were reconstructed through the same scripts used for reconstructing our signal events. 19
3.10 Reconstruction efficiency vs M_a, which ranges from 0.9GeV to 4.7GeV in the above plot 20
3.11 Reconstructed events that were identified to be particles from Table 3.2. This events are expected to scale vastly and form insuppressible signal peaks as we analyze real data with much greater statistics 21
3.12 Fitting ΔE for various mass points. A mix of Crystal Ball and gaussian functions have been used to fit the signal, while the background is fit by a linear polynomial. 22

4.1 Difference in event shape between continuum and BB events (Credit: Markus Röhrken) 24
4.2 Distribution of variables which are assigned most importance in most continuum suppression MVAs. The above plots are distributions of the input training variables for the mass point $M_a = 0.88\ GeV$ 25
4.4 MVA training iteration 1 on $M_a = 0.88$ GeV: variable importance (left), correlation matrix (top right) and separation power of `KSFWVariable1(hso03)` (centre right), `Thrust0m` (bottom right). 28
4.5 MVA training iteration 2 on $M_a = 0.88$ GeV: variable importance (left), correlation matrix (top right) and separation power of `KSFWVariable3(hso02)` (centre right), `Thrust0m` (bottom right). 29
4.6 MVA classifier Output for ALP mass 0.88 GeV. The large error bars are due to low statistics. Identification error plots reveal a bit of overfitting in the fringes, but within the acceptable 3σ range. Similar results are seen in Appendix C . 30

4.7 Figure of Merit plots for the 3 mass points chosen for continuum suppression, Sub-captions give FoM information in the form (M_a, MVA Output cut) . . . 31
4.8 Applying optimized BDT cuts on ΔE (top) and M_a (bottom) distributions for $M_a = 0.88\ GeV$. Component wise MVA rejections are shown on the right 33
4.9 Applying optimized BDT cuts on ΔE (top) and M_a (bottom) distributions for $M_a = 1.92\ GeV$. Component wise MVA rejections are shown on the right 33
4.10 Applying optimized BDT cuts on ΔE (top) and M_a (bottom) distributions for $M_a = 2.96\ GeV$. Component wise MVA rejections are shown on the right 34
4.11 MVA efficiency vs mass. Efficiency calculated after applying optimal cuts from FoM analyses (Credit: Dr. Sourav Dey, KMI-Nagoya U.) 34
4.12 Number of events supplied to MVA for training as a function of mass (Credit: Dr. Sourav Dey, KMI-Nagoya U.) 35

B.1 Distribution of the reconstructed variables ΔE and M_a for representative masses on the opposite sides of the mass range. Why do the signal peaks become sharper for one variable and vice-versa for the other? 41
B.2 Plots showing the qualitative degree of correlation between momentum and energy of the ALP for the two mass points 43
B.3 Comparison of residuals of the momentum and energy for ALPs of mass 0.9 GeV (left) and 4.6 GeV (right) 44

C.1 ROC curves for the representative mass points in analysis. The high AUC for 0.88 GeV implies it is an almost perfect classifier, as also evidenced by Table 4.2 47
C.2 Importance of continuum suppression varaibles in training MVA for ALP mass 1.92 GeV in iteration 1 (left) and 2 (right) 48
C.3 Correlation between continuum suppression varaibles in training MVA for ALP mass 1.92 GeV in iteration 1 (left) and 2 (right) 49
C.4 Overtraining plot for ALP mass 1.92 GeV 49
C.5 Importance of continuum suppression varaibles in training MVA for ALP mass 2.96 GeV in iteration 1 (left) and 2 (right) 50
C.6 Correlation between continuum suppression varaibles in training MVA for ALP mass 2.96 GeV in iteration 1 (left) and 2 (right) 51
C.7 Overtraining plot for ALP mass 2.96 GeV 51

List of Tables

3.1 Reconstruction efficiencies for control and signal channels after standard selection cuts. 14
3.2 Example resonances and SM particles, along with their PDG IDs and masses, taken from PDG particle list 21

4.1 Relative branching fractions of quark pair production modes in the Standard Model used to construct the continuum Monte Carlo sample for MVA training. 26
4.2 MVA performance for mass points 0.88, 1.92, 2.96 GeV. 32

Chapter 1

Introduction and Motivation

1.1 Overview of Belle II and SuperKEKB

The Belle II experiment is a high-energy physics experiment located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. It is designed to study the properties of B mesons and other particles produced in electron-positron collisions, with the primary goal of searching for new physics Beyond the Standard Model (BSM) of particle physics [1]. Belle II is the successor to the original Belle experiment, which operated from 1999 to 2010 and made significant contributions to our understanding of CP violation and rare B decays.

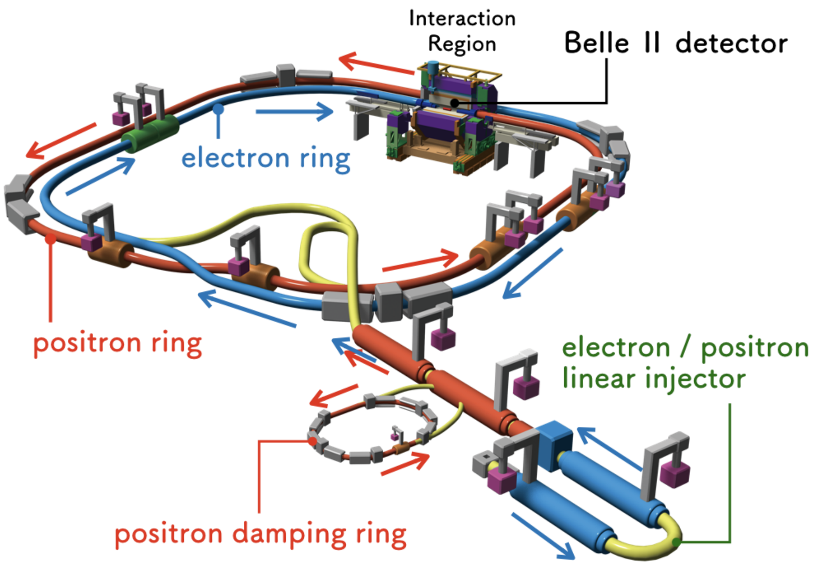


Figure 1.1: Overview of the SuperKEKB accelerator complex [2]

SuperKEKB is the particle accelerator that provides the high-luminosity electron-positron beams for Belle II [3]. It is an upgrade of the KEKB accelerator, with a target luminosity of up to $6.5 \times 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which is about 30 times higher than KEKB. This high luminosity allows Belle II to collect large datasets, enabling precise measurements and searches for rare processes. SuperKEKB operates at the $\Upsilon(4S)$ resonance, where the center-of-mass energy is approximately 10.58 GeV, ideal for producing B meson pairs.

The Belle II detector is a sophisticated apparatus designed to reconstruct and analyze the products of these collisions with high precision. It consists of several subdetectors, each serving specific purposes with detailed working principles and measurement capabilities:

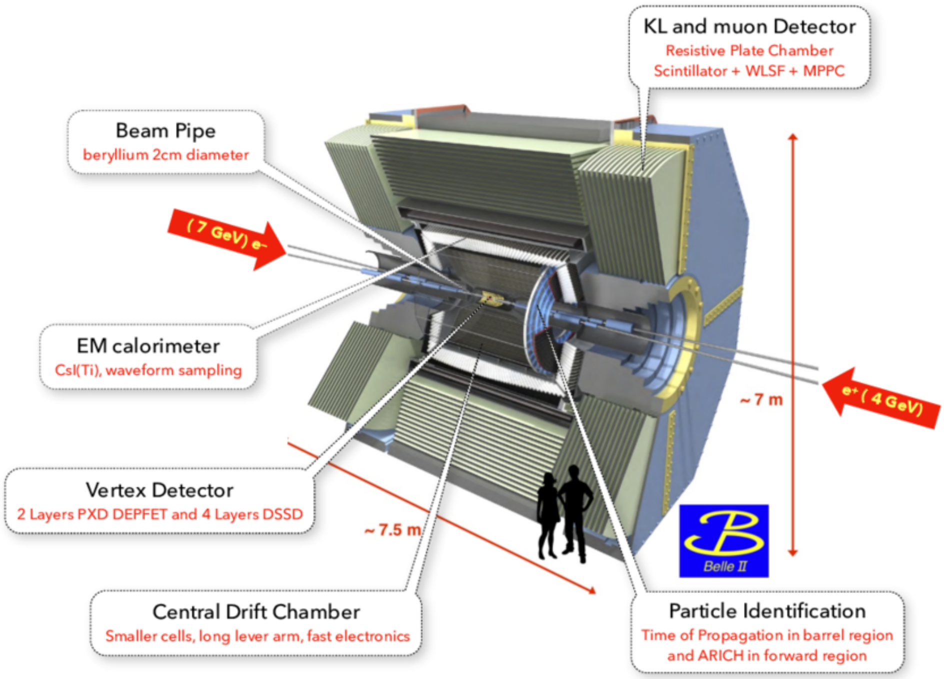


Figure 1.2: Schematic view of the Belle II detector [4]

- **Vertex Detector (VXD)**: This innermost detector comprises pixel and silicon strip sensors that detect charged particles by measuring the ionization they produce as they pass through the silicon. It achieves vertex resolution better than $50\mu m$, enabling precise reconstruction of decay vertices. This is crucial for distinguishing B meson decays from background and measuring lifetimes of short-lived particles.
- **Central Drift Chamber (CDC)**: A helium-ethane gas-filled cylindrical chamber with sense wires, where charged particles ionize the gas, and the resulting electrons drift toward anode wires. The drift time and wire position determine the particle's trajectory with high accuracy (momentum resolution $\delta p/p \approx 0.3\%$ at 1 GeV). Combined with the solenoid's magnetic field, it provides momentum measurements essential for kinematic reconstruction.
- **Time-of-Propagation Counter (TOP)**: Utilizes quartz radiator bars where relativistic particles emit Cherenkov radiation. Photons propagate along the bars and are detected by photomultipliers at the ends. The time-of-flight difference between direct and reflected photons allows particle identification (PID) with π/K separation up to 4 GeV/c, vital for distinguishing kaons from pions in B decays.
- **Aerogel Ring-Imaging Cherenkov Counter (ARICH)**: Employs aerogel as a radiator material to produce Cherenkov light, which is focused by mirrors onto hybrid avalanche photodiodes. It provides PID for particles in the forward region, complementing TOP for comprehensive particle identification across the detector.
- **Electromagnetic Calorimeter (ECL)**: Composed of thallium-doped cesium iodide (CsI(Tl)) crystals that absorb photons and electrons, converting their energy into scintillation light detected by photodiodes. It measures energies with resolution $\delta E/E \approx 1.5\%/\sqrt{E}$ (in GeV), enabling reconstruction of neutral particles like π^0 and η via their photon decays.

- **K_L and Muon Detector (KLM)**: Features resistive plate chambers (RPCs) for muon detection through gas ionization and scintillator strips for K_L identification via hadronic interactions. It provides muon PID and vetoes K_L showers, with penetration depth measurements distinguishing muons from hadrons.
- **Solenoid Magnet**: A 1.5 T superconducting coil that generates a uniform magnetic field, causing charged particles to curve. This curvature, measured by the tracking detectors, allows precise momentum determination via $p = 0.3Br$ (in GeV/c, with B in T and r in m).

These subdetectors work together to achieve excellent particle identification, tracking, and calorimetry, and allows Belle II to perform detailed analyses of complex decay chains with high efficiency and resolution.

1.2 Motivation for Searching Heavy QCD Axions at Belle II

The strong CP problem and the potential existence of axion-like particles (ALPs) motivate searches for new physics [5, 6, 7], but Belle II's unique capabilities make it particularly suited for probing heavy QCD ALPs through rare B decays. Compared to other experiments, Belle II offers distinct advantages:

- **Optimized for B Physics**: Operating at the $\Upsilon(4S)$ resonance, Belle II produces copious B meson pairs in a clean e^+e^- environment, unlike hadron colliders like LHCb, which face higher backgrounds from QCD processes. Belle II's excellent vertex resolution and PID allow precise reconstruction of multi-body decays like $B \to Ka \to K\eta\pi\pi$.
- **High Luminosity and Statistics**: With luminosity 30 times that of its predecessor KEKB, Belle II collects datasets far larger than BABAR (which had similar design but lower integrated luminosity), enabling sensitivity to branching ratios as low as 10^{-6}. This is crucial for rare FCNC processes suppressed in the Standard Model.
- **Clean Experimental Environment**: The asymmetric beam design and hermetic detector provide superior background rejection compared to LHCb, where pile-up and multiple interactions per bunch crossing complicate rare decay analyses. Belle II's calorimetry and tracking excel at reconstructing neutral particles and low-energy photons from ALP decays.
- **Complementary to LHC Experiments**: While LHC probes higher-mass ALPs via different production mechanisms, Belle II accesses the GeV mass range through B decays, filling a gap in parameter space. Its focus on flavor physics complements direct searches at ATLAS/CMS for heavy axions and astrophysical probes for lighter axions.
- **Proven Track Record**: Building on Belle's discoveries in CP violation, Belle II's upgraded detector enhances sensitivity to narrow resonances in invariant mass spectra, making it ideal for ALP signals that appear as peaks in $m_{\eta\pi\pi}$.

Thus, Belle II's combination of high statistics, precise reconstruction, and B-focused design positions it as the premier experiment for discovering or constraining heavy QCD ALPs via flavor-changing B decays.

Heavy QCD axions arise when additional contributions generate the axion mass, decoupling it from the canonical QCD relation and allowing masses in the GeV range. In this scenario, axion-like particles can be produced in rare B decays such as $B \to Ka$, with the subsequent decay $a \to \eta\pi\pi$ giving a narrow resonance in the $m_{\eta\pi\pi}$ invariant mass. Previous Belle and BABAR studies have already demonstrated the power of this approach in constraining heavy ALP production [8, 9]. This channel offers several advantages for Belle II:

- **Rare FCNC Process**: The $b \to sa$ transition is not allowed in the SM, making it sensitive to Beyond SM physics.
- **Clean Experimental Signature**: The narrow resonance in $m_{\eta\pi\pi}$ distinguishes the signal from background, and the final state particles ($\eta\pi\pi$) are well-reconstructed by Belle II's subdetectors.
- **Complementary to Other Experiments**: While light axions are probed by astrophysical observations, heavy ALPs require collider experiments like Belle II, which can access the GeV mass range through B decays.
- **Existing Sensitivity**: Analyses on analogous decay channels from previous experiments like Belle and BABAR have set limits, and Belle II's improved detector and luminosity position it to explore new regions, potentially discovering or constraining heavy QCD ALPs.

1.3 Groundwork laid by previous experiments and analyses

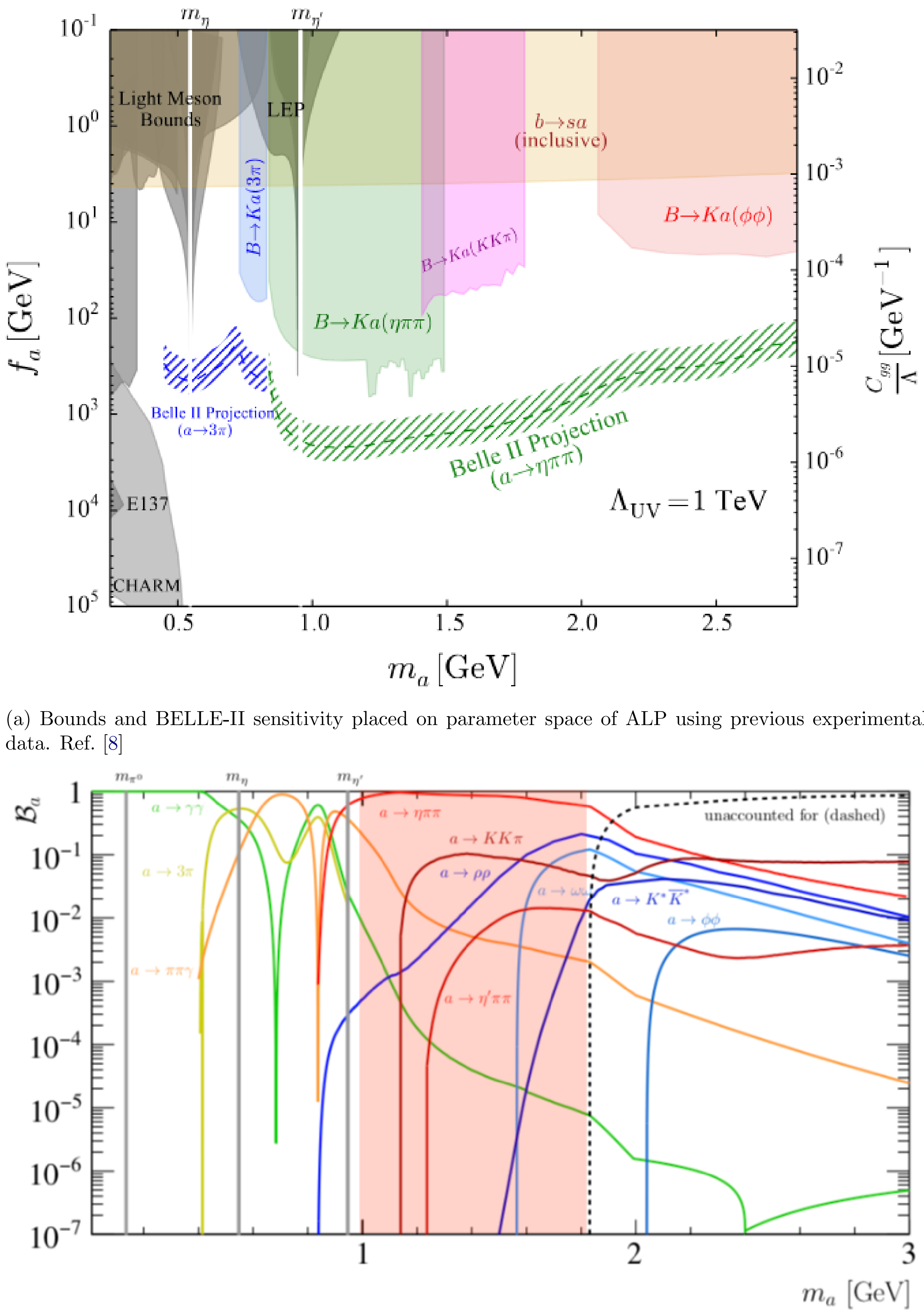
Analyses using data from BELLE and BABAR have looked into decay channels analogous to our target channel, which have been used to predict the sensitivity of the BELLE-II experiment to our decay channel $B^+ \to K^+a$, $a \to \eta\pi^+\pi^-$.

Ref. [8] estimates several $B \to Ka$ decays using some experimental data. Figure 1.3a shows the bounds for $B \to Ka$ decays while considering the branching fraction of the heavy QCD axion (Fig. 1.3b). The horizontal axis shows the axion mass, and the left vertical axis represents the limits of the axion decay constants. To compare with other decay modes, the value of the decay constant is converted to the gluon coupling c_{gg}/Λ (the right vertical axis in Fig. 1.3a). The green and blue shaded regions show the limit on $a \to \eta\pi^+\pi^-$ and $a \to \pi^0\pi^+\pi^-$ decays, respectively.

Ref. [10] shows the results of the search for the excited η in $B^+ \to K^+\eta_X$ by BaBar with $\sim 400\ \mathrm{fb}^{-1}$ data. This result is reinterpreted as a limit for the heavy QCD axion study in the $a \to \eta\pi^+\pi^-$ decay channel from $0.9\ \mathrm{GeV}/c^2$ to $1.5\ \mathrm{GeV}/c^2$ (and up to $3\ \mathrm{GeV}/c^2$ for the expected limit in Belle II).

Ref. [11] shows the reconstructed $M_{3\pi}$ data in Belle ($\sim 700\ \mathrm{fb}^{-1}$) used to search for $B \to K\omega$ decay. This was used to evaluate $a \to \pi^0\pi^+\pi^-$ decay below $800\ \mathrm{MeV}/c^2$.

Thus, the limits in Fig. 1.3a are estimated by recalculating from other measurements, and an experimental analysis to directly search for heavy QCD axions has not been performed so far. In this work, the analysis of $B^+ \to K^+a$, $a \to \eta\pi^+\pi^-$ decay using Monte Carlo simulated data is presented.



(a) Bounds and BELLE-II sensitivity placed on parameter space of ALP using previous experimental data. Ref. [8]

(b) Summary of the branching fractions for the heavy QCD axions. [8]

Figure 1.3: Feynman diagrams that arise from Eq. 2.8

Chapter 2

Theory of Heavy QCD Axion-like Particles

2.1 Strong CP Problem and the Peccei–Quinn (PQ) Mechanism

Quantum Chromodynamics (QCD) admits a CP-violating term in the Lagrangian,

$$\mathcal{L}_\theta = \theta \frac{g_s^2}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{a\mu\nu}, \tag{2.1}$$

which induces observable CP violation such as a neutron electric dipole moment (EDM). Experimental limits constrain $|\theta| \lesssim 10^{-10}$, posing the strong CP problem [12, 13].

A dynamical solution was proposed by Peccei and Quinn [5, 14], introducing a global $U(1)_{\mathrm{PQ}}$ symmetry that is spontaneously broken. The associated pseudo-Nambu–Goldstone boson is the axion [6, 7], which dynamically relaxes the effective θ parameter to zero.

The axion couples to gluons via

$$\mathcal{L} \supset \frac{\alpha_s}{8\pi f_a} a\, G^a_{\mu\nu} \tilde{G}^{a\mu\nu}, \tag{2.2}$$

leading to the relation

$$m_a f_a \simeq m_\pi f_\pi, \tag{2.3}$$

in the minimal QCD axion scenario [12].

2.2 Motivation for Heavy QCD Axions

If the Peccei–Quinn scale is near the electroweak scale ($f_a \sim \mathrm{TeV}$), the above relation predicts $m_a \sim \mathrm{keV}$. Such parameter space is excluded by astrophysical constraints and beam dump experiments [15, 16, 17].

Conversely, very large $f_a \sim 10^9$–$10^{13}\,\mathrm{GeV}$ yields ultra-light axions, motivated as dark matter candidates but extensively constrained experimentally [18].

This motivates scenarios where the relation

$$m_a f_a \simeq m_\pi f_\pi \tag{2.4}$$

is violated by additional contributions to the axion mass [8].

2.3 Axion Quality Problem

A central theoretical issue is the **axion quality problem**. Global symmetries such as $U(1)_{\mathrm{PQ}}$ are expected to be violated by quantum gravity, leading to Planck-suppressed operators of the form

$$\Delta\mathcal{L} \sim \frac{\phi^n}{M_{\mathrm{Pl}}^{n-4}} + \text{h.c.}, \tag{2.5}$$

where ϕ carries PQ charge.

These operators generically induce shifts in the axion potential, reintroducing CP violation unless highly suppressed [19, 20, 21, 22].

Quantitatively, unsuppressed Planck-scale corrections spoil the axion solution unless

$$f_a \lesssim 10^7\text{–}10^{10}\,\mathrm{GeV}, \tag{2.6}$$

depending on model assumptions [8].

This creates tension with traditional invisible axion models requiring large f_a. Heavy QCD axions provide a natural resolution: if the axion mass receives dominant contributions from a hidden sector, then the CP-violating shift induced by Planck-suppressed operators becomes negligible relative to the total potential [23, 8].

Thus, heavy axions with

$$f_a \sim \mathrm{TeV}, \qquad m_a \sim \mathrm{MeV\text{–}GeV} \tag{2.7}$$

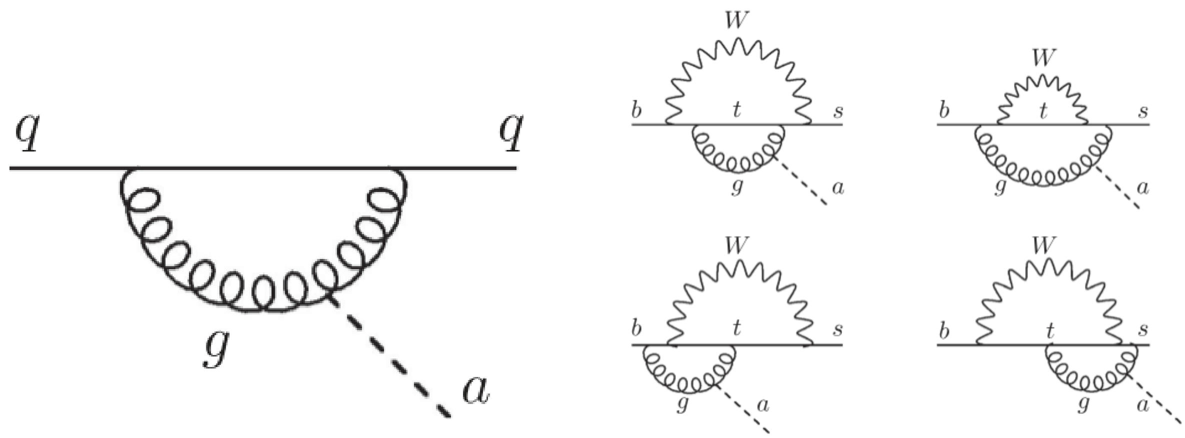
can simultaneously solve the strong CP problem and evade the quality problem.

2.4 Heavy QCD Axion Effective Theory

In this framework, m_a and f_a are independent parameters. The dominant interaction remains the axion-gluonic coupling inside the effective Lagrangian,

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \frac{\alpha_s}{8\pi}\frac{a}{f_a}G_{a\mu\nu}\tilde{G}^{a\mu\nu} + \frac{1}{2}(\partial_\mu a)^2 - \frac{m_a^2}{2}a^2 \tag{2.8}$$

giving rise to Feynman diagrams of the form Fig. 2.1. Couplings to photons and fermions arise via mixing and loops, but hadronic decays dominate once $m_a \gtrsim \mathcal{O}(100\,\mathrm{MeV})$.



(a) Axion-gluon-gluon coupling

(b) Flavor changing neutral current processes as a result of 2.1a

Figure 2.1: Feynman diagrams that arise from Eq. 2.8

A data-driven approach to hadronic decays was developed in [24], allowing reliable predictions in the GeV regime.

The dominant decay modes are:

- $a \to \pi\pi\pi$ for $m_a \gtrsim 3m_\pi$,

- $a \to \eta\pi\pi$ for $m_a \gtrsim m_\eta + 2m_\pi$,

with the latter becoming dominant above ~ 0.85–1 GeV [23].

2.5 Flavor-Changing Processes: $B \to Ka$

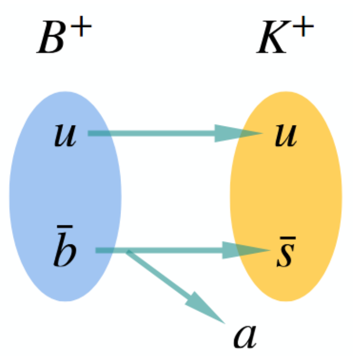
The gluonic coupling induces flavor-changing neutral currents at two loops, leading to the transition $b \to sa$ as shown in Fig. 2.1b (taken from [8]).

The effective Lagrangian (Appendix A) can be written as

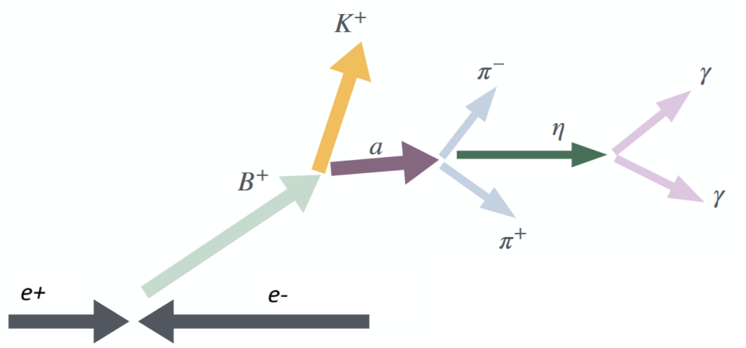
$$\mathcal{L}_{\text{eff}} \sim \frac{\partial^\mu a}{f_a}\,\bar{s}\gamma_\mu\gamma_5 b. \tag{2.9}$$

This generates the decay

$$B \to Ka. \tag{2.10}$$



(a) Decay B→ Ka acc. to Eq. 2.10



(b) Full Decay Reaction

Figure 2.2: Particle Production in target decay channel

The branching ratio is estimated as (Appendix A)

$$\mathrm{BR}(B \to Ka) \sim 10^{-5}\left(\frac{100\,\mathrm{GeV}}{f_a}\right)^2, \tag{2.11}$$

up to model-dependent factors [23].

For $f_a \sim \mathrm{TeV}$, this yields

$$\mathrm{BR}(B \to Ka) \sim 10^{-6}\text{–}10^{-7}. \tag{2.12}$$

2.6 Target Channel: $B \to K\,[a \to [\eta \to \gamma\gamma]\,\pi^+\pi^-]$

For axion masses above the $\eta\pi\pi$ threshold, the dominant decay is

$$a \to \eta\pi\pi, \tag{2.13}$$

leading to our target decay chain (Fig. 2.2b)

$$B \to K\,[a \to [\eta \to \gamma\gamma]\,\pi^+\pi^-]. \tag{2.14}$$

The total branching fraction is therefore

$$\mathrm{BR}(B \to K\,[a \to [\eta \to \gamma\gamma]\,\pi^+\pi^-]) \approx \mathrm{BR}(B \to Ka) \times \mathrm{BR}(a \to \eta\pi\pi) \sim 10^{-7}, \tag{2.15}$$

since $\mathrm{BR}(a \to \eta\pi\pi) \sim \mathcal{O}(1)$ in this mass range [23].

2.7 Experimental Signatures

The signal is characterized by:

- A narrow resonance in $m_{\eta\pi\pi}$,
- A rare FCNC transition $b \to sa$,
- Prompt or displaced decays depending on f_a.

The clean topology and low Standard Model background make this channel particularly suitable for Belle II searches [23].

Chapter 3

Monte Carlo Simulations

3.1 Introduction

Monte Carlo (MC) simulations form the backbone of modern high energy physics analyses. In the context of the Belle II experiment, MC simulations are used to model e^+e^- collisions, the subsequent particle decays, and the detector response in a fully controlled environment. These simulations allow one to study signal characteristics, optimize selection criteria, estimate efficiencies, and understand backgrounds before analyzing real data.

In this work, Monte Carlo samples were generated using the KEK Computing Centre (KEKCC) to simulate e^+e^- collisions producing B mesons that undergo the target decay channel:

$$B \to Ka, \quad a \to \eta\pi^+\pi^-. \tag{3.1}$$

To maximize statistics and isolate the signal topology, the branching fraction for this decay was set to 100% at generation level. For each channel analyzed, 2×10^5 events were generated.

3.2 Generation-Level and Reconstruction-Level Data

Monte Carlo simulation in Belle II proceeds in two conceptually distinct stages:

3.2.1 Generation-Level Data

At the generation level, events are produced using physics generators that simulate the underlying interaction and decay processes. The complete information of all particles in the event, including their four-momenta, decay chains, and particle identities, is stored. This information is often referred to as **MC truth**.

In this analysis, the generated dataset contains the full decay chain with perfect knowledge of all intermediate and final-state particles.

3.2.2 Reconstruction-Level Data

The generated particles are then passed through a detailed detector simulation, followed by event reconstruction. This stage mimics the actual Belle II detector response, including inefficiencies, finite resolution, and detector effects.

The output of this stage is stored in **mDST (mini Data Summary Table)** files, which contain the reconstructed quantities such as:

- Particle momenta and energies
- Vertex information

- Particle identification variables

This reconstructed information represents the **machine readout** and is what is used in analysis, analogous to real experimental data.

NOTE: Reconstructed information is different from reconstructed particles, which are intermediate particles on the decay chain, created by 4-momentum addition of the final state particles. It is these reconstructed particles that is used for the majority of this analysis.

3.3 basf2 Framework and Reconstruction Scripts

The Belle II analysis software framework, **basf2**, is used for both simulation and reconstruction. It provides modular tools for:

- Event generation
- Detector simulation
- Particle reconstruction
- Analysis variable calculation

A **reconstruction script** in basf2 is a Python-based workflow that defines:

- Which particles to reconstruct
- The decay chains to be built
- Selection criteria (cuts)
- Variables to be stored for analysis

In this work, reconstruction scripts were developed to identify the decay chain:

$$B \to K[a \to \eta(\to \gamma\gamma)\, \pi^+\pi^-]. \tag{3.2}$$

3.4 MC Truth Matching

A crucial concept in MC-based analyses is **MC truth matching**. This procedure associates reconstructed particles with their true generated counterparts.

A reconstructed candidate is labeled as signal if:

$$\texttt{isSignal} = 1, \tag{3.3}$$

indicating that all its daughters match the correct generated decay chain.
If a B-candidate is reconstructed incorrectly, it is given $\texttt{isSignal} = 0$
MC truth matching allows:

- Validation of reconstruction performance
- Calculation of reconstruction efficiencies
- Separation of correctly reconstructed signal from combinatorial background

3.5 Choice of ALP Mass Points

The allowed ALP mass range spans approximately:

$$0.8 \text{ GeV} \lesssim m_a \lesssim 4.76 \text{ GeV}. \tag{3.4}$$

Three representative mass points were chosen:

- $m_a = 0.957 \text{ GeV} = m_\eta$

- $m_a = 2.3$ GeV
- $m_a = 4.6$ GeV

These were used to study the dependence of reconstruction performance and kinematic distributions across the accessible phase space.

3.6 Key Kinematic Variables

The analysis relies on several kinematic variables:

3.6.1 Beam-Constrained Mass

$$M_{\text{bc}} = \sqrt{E_{\text{beam}}^2 - |\vec{p}_B|^2} \tag{3.5}$$

This variable exploits the precisely known beam energy and is highly effective in identifying correctly reconstructed B mesons.

3.6.2 Energy Difference

$$\Delta E = E_B - E_{\text{beam}} \tag{3.6}$$

This measures the consistency of the reconstructed B candidate energy with the beam energy.

3.6.3 ALP Invariant Mass

$$M_{\eta\pi\pi} \tag{3.7}$$

This is used to reconstruct the ALP candidate and verify signal formation.

3.6.4 Fitting Variables

The primary fitting variables are:

- ΔE
- M_{bc}

These are chosen because:

- They have well-defined signal peaks
- Background distributions are smooth
- They are largely independent of ALP mass assumptions

The ALP invariant mass is used as a validation variable rather than a primary fit observable.

3.7 Initial Reconstruction and Benchmarking

Initial attempts at ALP reconstruction revealed 3.1:

- Poor reconstruction of ALP invariant mass - The distribution was extremely asymmetrical, biased
- Significant background contamination

To refine and validate the reconstruction strategy, Standard Model control channels were used:

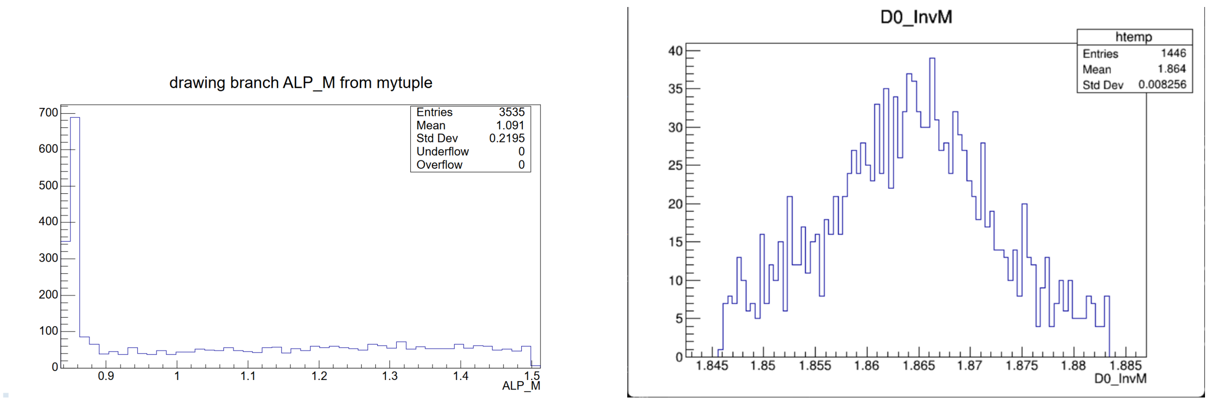


Figure 3.1: Initial attempts at reconstructing the ALP and extracting variable distributions

- $B \to K\,[D \to [\eta \to \gamma\gamma]\,\pi^+\pi^-]$
- $B \to K\,[\eta' \to [\eta \to \gamma\gamma]\,\pi^+\pi^-]$

These channels share identical final states and topology, allowing direct comparison and benchmarking through the same reconstruction strategy.

3.8 Reconstruction Efficiency

Reconstruction for a given set of cuts is defined as:

$$\epsilon = \frac{\text{Reconstructed signal events (isSignal==1)}}{\text{Generated events}} \tag{3.8}$$

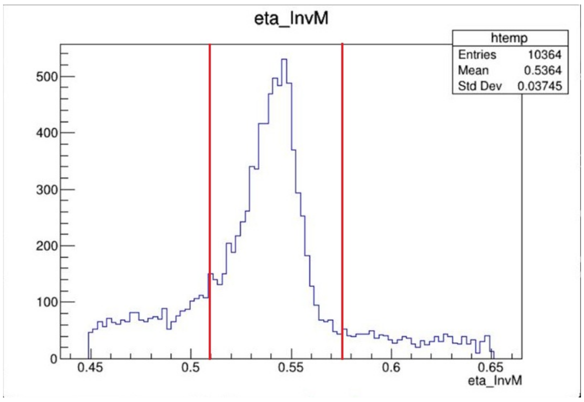
Efficiencies were evaluated for the SM channels as well as the 3 ALP mass points using the below cuts

The event selection cuts for $B^+ \to K^+ a,\ a \to \eta\pi^+\pi^-,\ \eta \to \gamma\gamma$ decay are summarized as follows:

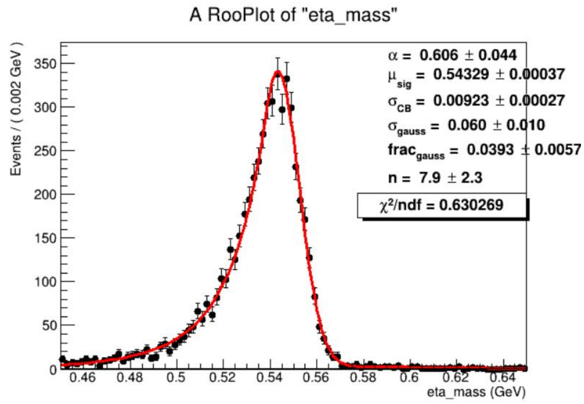
- B selection: $M_{bc} > 5.27\ \mathrm{GeV}/c^2$ and $|\Delta E| < 0.15\ \mathrm{GeV}$ (standard cuts recommended by BELLE-II experts)
- Good pion and kaon tracks from the interaction point: $|dz| < 2\ \mathrm{cm}$ and $|dr| < 0.5\ \mathrm{cm}$ (standard cuts recommended by BELLE-II experts)
- Selection of η mass: $509\ \mathrm{MeV} < M_{\gamma\gamma} < 577\ \mathrm{MeV}$ (most of the signal is concentrated within the 3σ region. Cut made accordingly after fitting the η mass distribution Figure 3.2)
- Global pion ID: $\mathrm{pionID} > 0.4$.
- Global kaon ID: $\mathrm{kaonID} > 0.6$.

The above cuts are so called standard cuts that are generic to every BELLE-II analysis. Figure 3.3 shows the reduction in combinatorial background as opposed to loss in signal for the SM channels. We also introduced a few cuts specific to our analysis by performing a Figure of Merit and MVA analysis.

- Gamma energy cut to reject low energy gamma background: $E_\gamma > 150\ \mathrm{MeV}$ (Based on Figure of Merit analysis, see section 3.9)
- Cut on opening angle between final two γ daughters: $\cos\theta > 0$ (Encodes information of ALP mass and provides good separation from background. FoM analysis, see section 3.9)
- MVA cut for continuum suppression (chapter 5).



(a) η mass distribuition, and the 3σ cut in red



(b) Fitting η mass distribuition using a mix of Crystall Ball and gaussian functions, and using fit parameters σ_{CB} and σ_{gauss} to extract 3σ cut

Figure 3.2: Applying 3σ cuts to reduce combinatorial background

Table 3.1: Reconstruction efficiencies for control and signal channels after standard selection cuts.

Channel	**Reconstruction Efficiency (%)**
$B \to KD^0$ (control)	4.8
$B \to K\eta'$ (control)	12.85
$B \to Ka$ ($m_a = 0.95$ GeV)	12.84
$B \to Ka$ ($m_a = 2.3$ GeV)	10.82
$B \to Ka$ ($m_a = 4.6$ GeV)	12.9

Table 3.1 shows the reconstruction efficiency after applying the standard cuts. Since η' is a promptly decaying particle, with almost no discernible qualities from an ALP at the same mass, the reconstruction efficiency remains same. An analogous effect is not observed for D^0 since it has a finite decay lifetime, unlike our target ALPs.

3.9 Figure of Merit Optimization

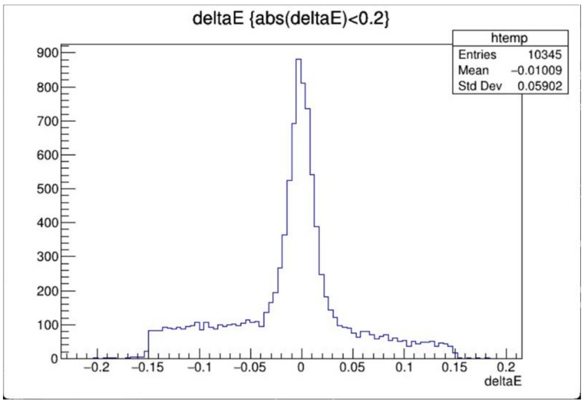
As seen in previous section, the standard cuts were effective to reduce the combinatorial background (that is, wrongly reconstructed events strictly from the signal Monte Carlo). However, the real data will contain generic Standard Model events with much higher cross sections than our signal channel. This is called the **generic background**. To optimize selection cuts that reduce generic background, a **figure of merit (FoM)** approach was used to discriminate between signal and background.
The Figure of merit for a given cut is defined as

$$\text{FoM} = \frac{S}{\sqrt{S+B}} \tag{3.9}$$

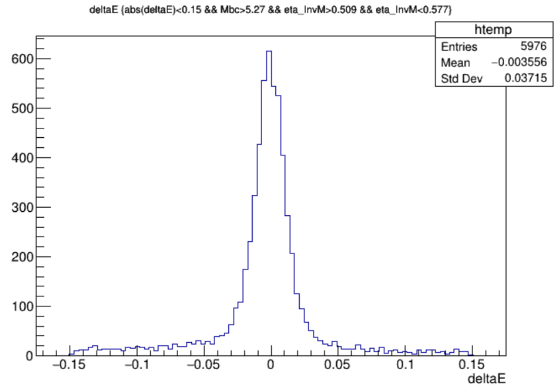
where:

- S = number of signal events that pass the cut
- B = number of background events that pass the cut

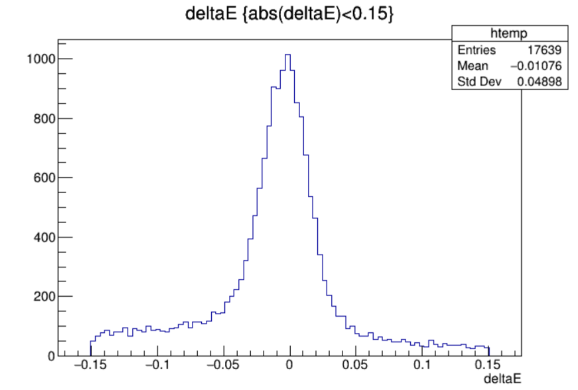
This metric balances signal efficiency against background rejection, and the cut where FoM maximizes is selected as the most optimal cut for a given variable For uniformity, FoM was



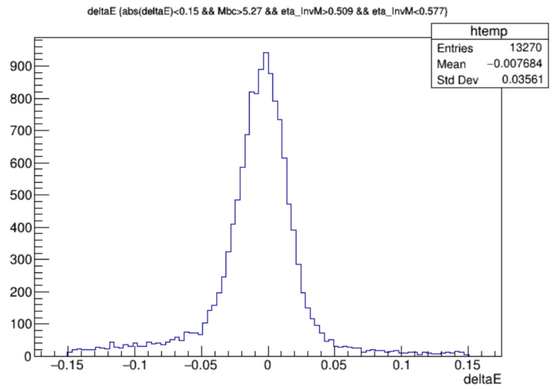
(a) D^0 no cuts



(b) D^0 std. cuts



(c) η' no cuts



(d) η' std. cuts

Figure 3.3: Applying 3σ cuts to reduce combinatorial background (SM channels)

maximized for a single mass point (2.3 GeV) and the extracted optimal cut was applied to all mass points (this is motivated by the observation that the optimal cut had very low mass dependence, despite the generic background having different distributions Figure 3.4,Figure 3.5).

Photon-related variables were optimized:

- Photon energy E_γ (Figure 3.4)
- Opening angle $\cos\theta$ (Figure 3.5)

Optimal cuts were found to be:

- $E_\gamma > 0.15$ GeV
- $\cos\theta > 0$

Other than significantly reducing the generic background, these cuts also had the added effect of further reduced combinatorial background while retaining signals Figs. 3.6, 3.7, 3.8.

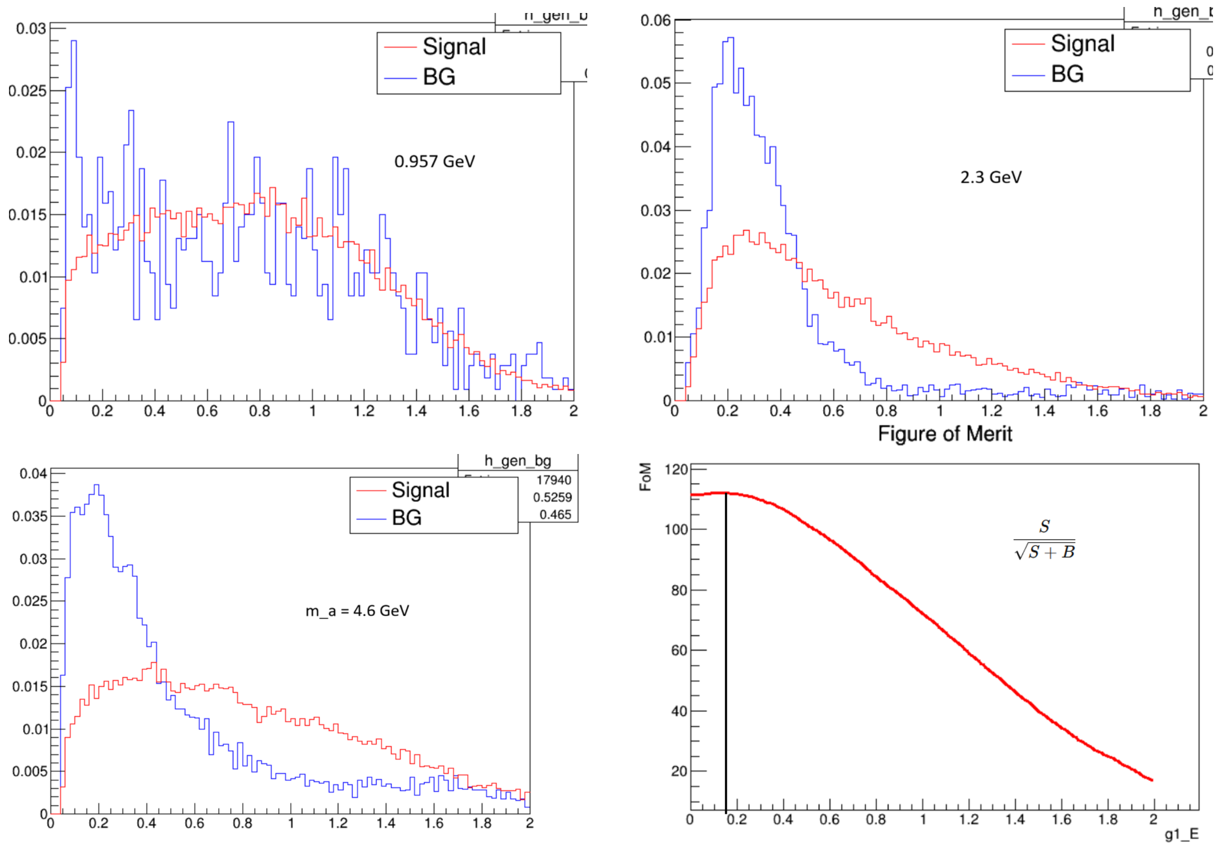
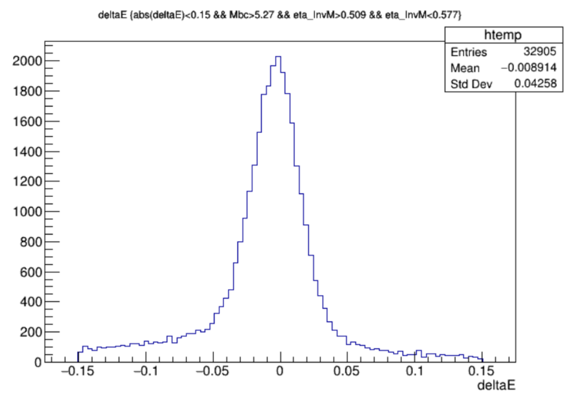
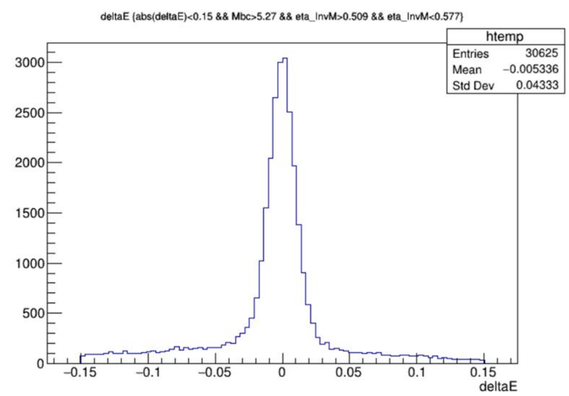


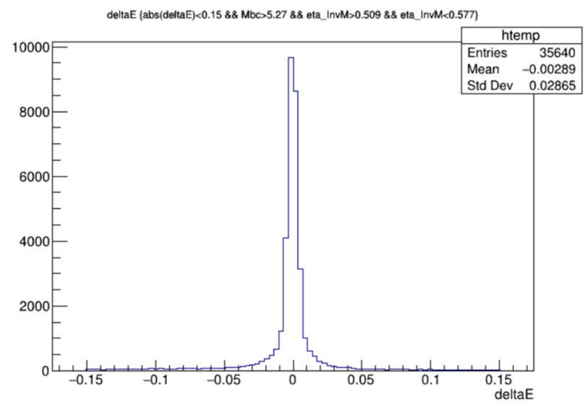
Figure 3.4: FoM optimization for E_γ



(a) 0.957 GeV



(b) 2.3 GeV



(c) 4.6 GeV

Figure 3.6: ΔE distribution for ALP decay channel after std. cuts but before FoM cuts

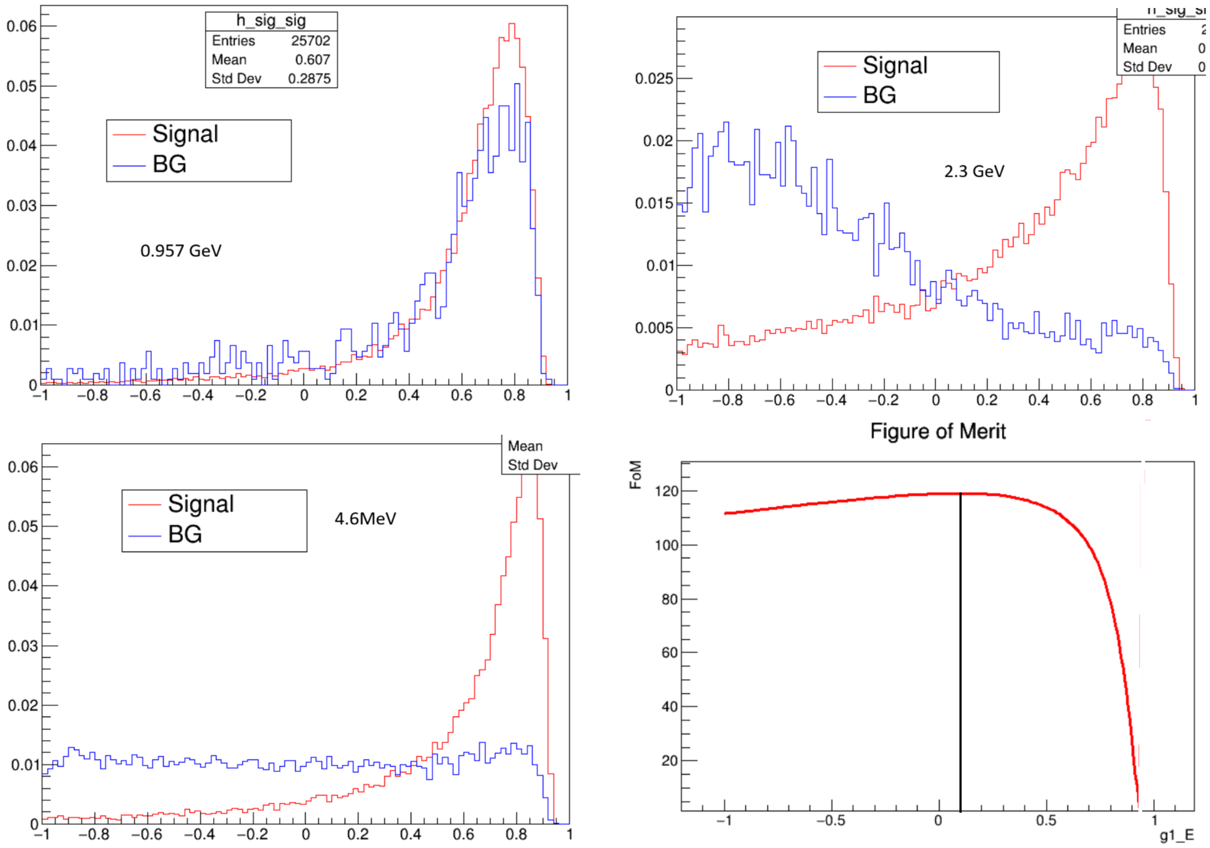
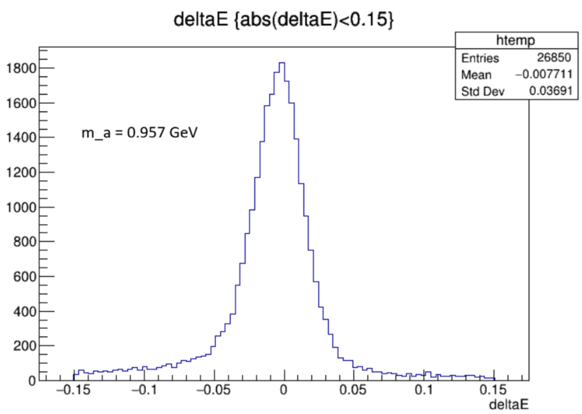
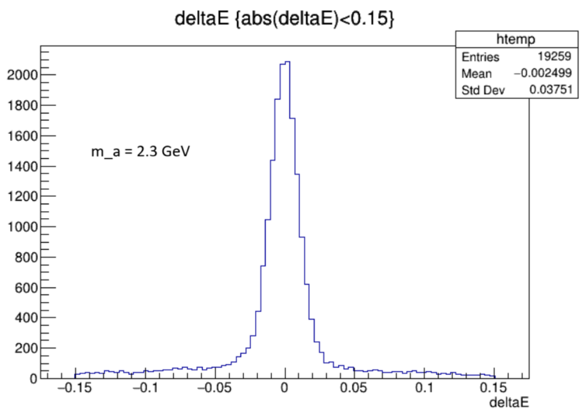


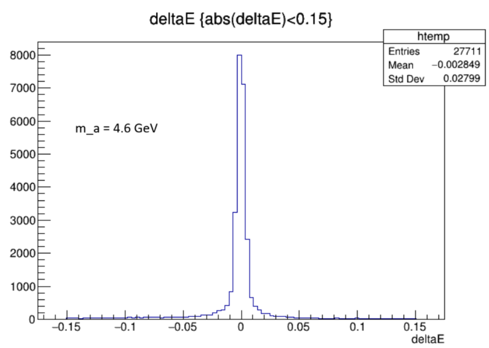
Figure 3.5: FoM optimization for $\cos\theta$



(a) 0.957 GeV

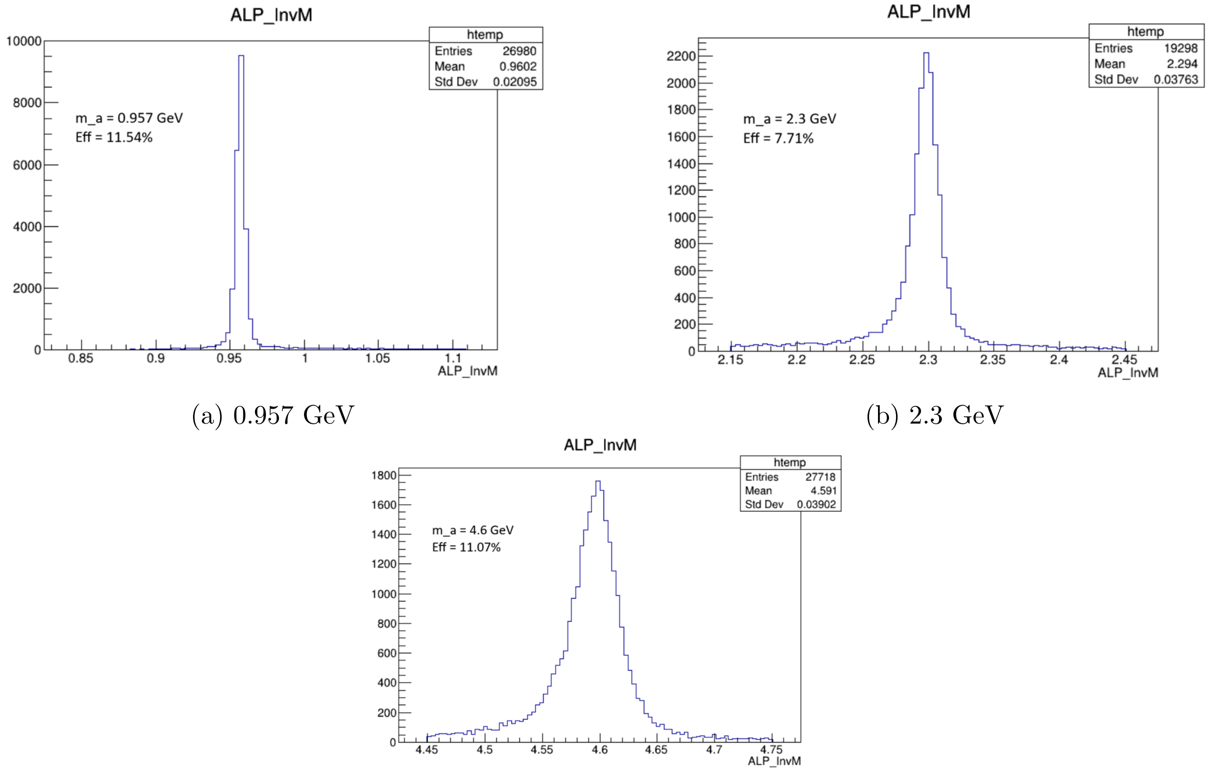


(b) 2.3 GeV



(c) 4.6 GeV

Figure 3.7: ΔE distribution for ALP decay channel after std. cuts and FoM cuts



(a) 0.957 GeV

(b) 2.3 GeV

(c) 4.6 GeV

Figure 3.8: $M_{\eta\pi\pi}$ distribution for ALP decay channel after std. cuts and FoM cuts

3.10 Generic Background

3.10.1 Definition

Generic background refers to events from all Standard Model processes that can mimic the signal topology, including:

- $B\bar{B}$ decays (mixed, charged)
- Continuum $q\bar{q}$ production $(u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c})$

3.10.2 Processing

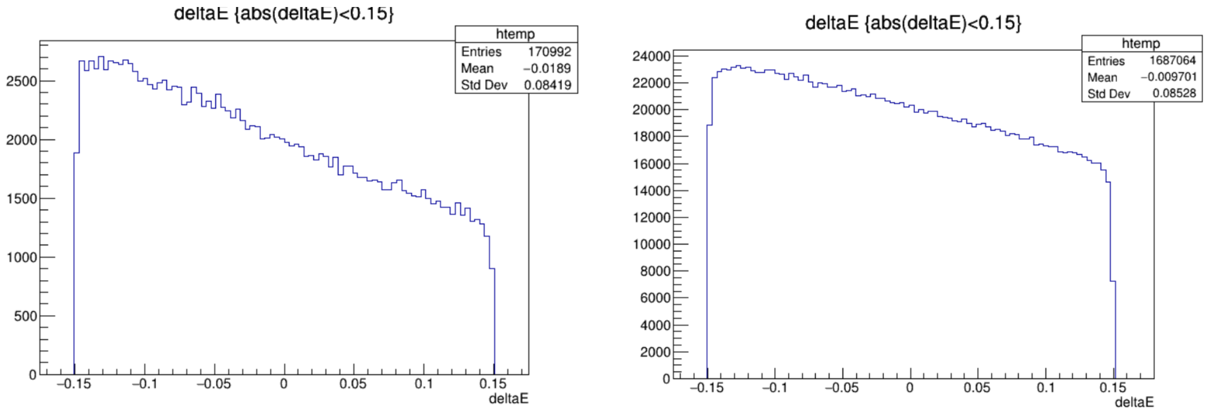
For each hadronic channel above, generic MC samples are provided by Belle II. For our analysis, we used:

- $\sim 100\ \mathrm{fb}^{-1}$ at KEKCC
- $\sim 500\ \mathrm{fb}^{-1}$ from the Belle II grid collections

These samples were passed through the same reconstruction pipeline, but with a broader invariant mass window:

$$0.9\ \mathrm{GeV} < M_{\eta\pi\pi} < 4.8\ \mathrm{GeV} \tag{3.10}$$

Unlike signal MC, which uses a narrow window (± 200 MeV), this ensures full background coverage. The full range is also used to train MVA models to perform continuum suppression. Figure 3.9 shows the ΔE distribution for generic SM events. Since the ALP is a BSM particle, there are no peaks around 0, as expected. There may be other particles in the allowed mass range (eg.D^0, η') but the cross sections of those channels are so low compared to SM hadronic events that their peaks do not show up in the distribution.



(a) ΔE for generic $B\bar{B}$ events

(b) ΔE for generic $q\bar{q}$ events

Figure 3.9: ΔE plotted after generic MC events were reconstructed through the same scripts used for reconstructing our signal events.

3.11 Reconstruction Efficiency Across the ALP Mass Range

Now that we have decided on the signal cuts for our reconstruction, a study on the dependence of the reconstruction efficiency on the ALP mass can be done. The reconstruction efficiency is defined as:

$$\epsilon(m_a) = \frac{N_{\text{reco}}(\texttt{isSignal==1})}{N_{\text{gen}}}, \tag{3.11}$$

where N_{reco} is the number of correctly reconstructed signal candidates and N_{gen} is the number of generated events.

To study this dependence systematically, Monte Carlo simulations were performed across the full kinematically allowed mass range:

$$0.85 \text{ GeV} \lesssim m_a \lesssim 4.8 \text{ GeV}. \tag{3.12}$$

The analysis strategy consists of:

1. Generating signal Monte Carlo (2×10^5 events each) at discrete mass points from 0.9 GeV to 4.8 GeV in steps of 0.1 GeV.
2. Processing these samples through the reconstruction pipeline.
3. Evaluating the reconstruction efficiency at each mass point.
4. Constructing an efficiency curve $\epsilon(m_a)$ as a function of the ALP mass.

The variation of efficiency with m_a is driven primarily by kinematic effects:

1. At low masses, the ALP is highly boosted, leading to collimated decay products, particularly in the $\eta \to \gamma\gamma$ channel. This can reduce reconstruction efficiency due to overlapping clusters in the electromagnetic calorimeter and increased sensitivity to detector resolution.
2. It also easier for the $b\bar{b}$ quarks to hadronize into lighter particles and hence can lead to loss of detector sensitivity to the ALP decay channel.
3. At higher masses, the decay products are less boosted and more isotropically distributed, improving reconstruction stability. However, phase space suppression and detector acceptance effects may counteract this improvement.

The above effects interact and evolve w.r.t. ALP mass in a complex manner and results in the efficiency curve Figure 3.10.

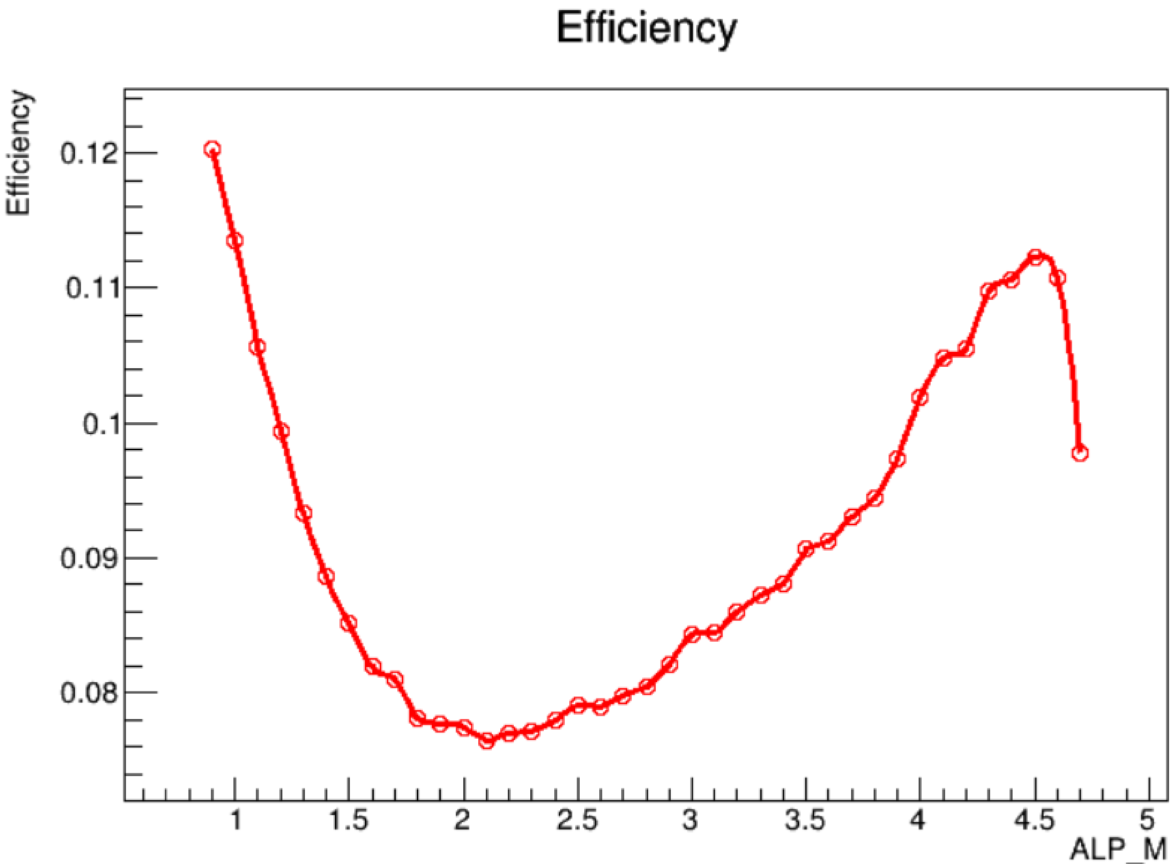


Figure 3.10: Reconstruction efficiency vs M_a, which ranges from 0.9GeV to 4.7GeV in the above plot

3.12 Impact of Standard Model Resonances at High Mass

For ALP masses above approximately 2.9 GeV, the invariant mass region becomes densely populated with Standard Model resonances, particularly charmonium ($c\bar{c}$) states. The number of accessible resonant modes per unit mass increases significantly in this region, leading to a loss of sensitivity for BELLE-II to the ALP decay channel, assuming the ALP has $M_a > 3GeV$.

Prominent resonances in this mass region are shown in Table 3.2

These resonances can decay into final states similar to the signal topology, including multi-hadron and $\eta\pi\pi$ final states (Figure 3.11). As a result:

- The background becomes highly structured rather than smooth.
- Narrow resonant peaks can mimic or obscure a potential ALP signal.
- Interference effects and overlapping resonances complicate the extraction of a signal.

Furthermore, theoretical predictions for the ALP production cross section in the literature are typically limited to masses below 3 GeV (Figure 1.3b,[24]).

Resonance	PDG ID	Mass (GeV)
$\eta_c(1S)$	441	2.9839
$\chi_{c0}(1P)$	10441	3.4147
$\eta_c(2S)$	100441	3.6375
$J/\psi(1S)$	443	3.0969
$h_c(1P)$	10443	3.5254
$\chi_{c1}(1P)$	20443	3.5107
$\psi(2S)$	100443	3.6861
$\psi(3770)$	30443	3.7737
$\psi(4040)$	9000443	4.039
$\psi(4160)$	9010443	4.191
$\psi(4415)$	9020443	4.421
$\chi_{c2}(1P)$	445	3.5562
$\chi_{c2}(3930)$	100445	3.927

Table 3.2: Example resonances and SM particles, along with their PDG IDs and masses, taken from PDG particle list

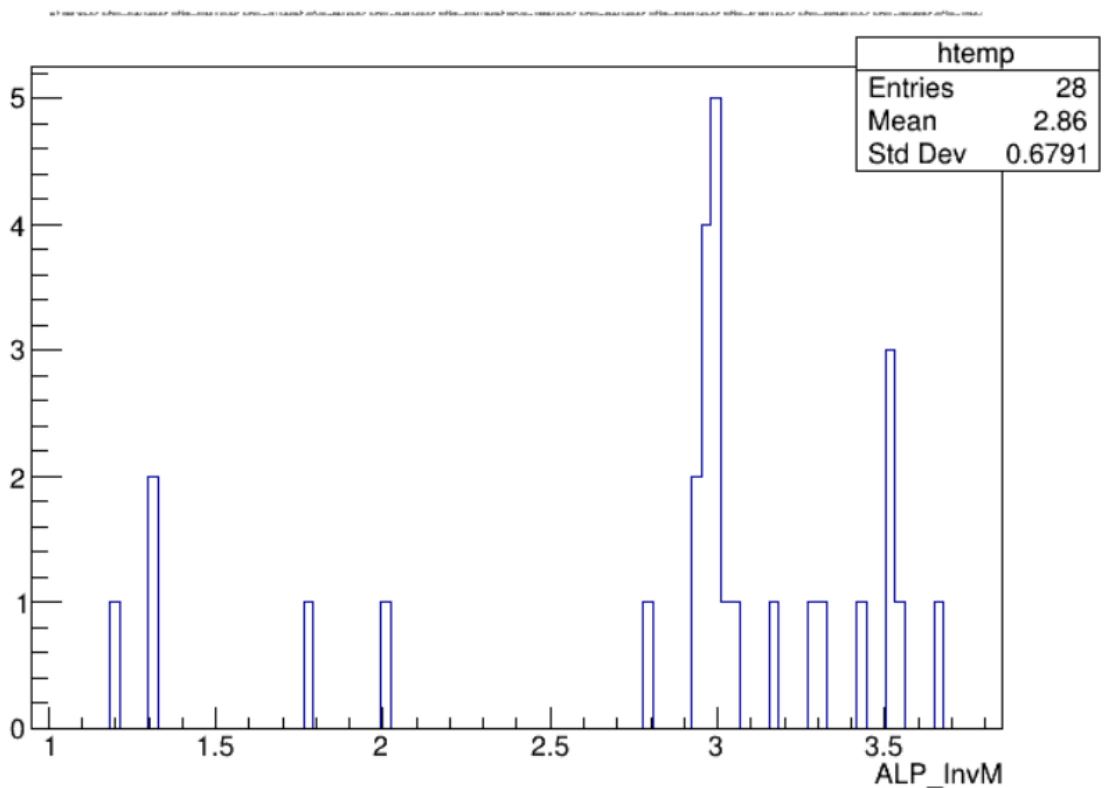


Figure 3.11: Reconstructed events that were identified to be particles from Table 3.2. This events are expected to scale vastly and form insuppressible signal peaks as we analyze real data with much greater statistics

For these reasons, the upper limit of the search window is restricted to:

$$m_a \lesssim 3 \text{ GeV}, \tag{3.13}$$

even though the kinematic limit extends to approximately 4.8 GeV.

This choice ensures that:

- The analysis is performed in a region with manageable background.
- Theoretical predictions can be validated.
- The sensitivity to potential ALP signals is maximized.

3.13 Fitting Strategy

The ΔE distribution is chosen as the primary observable for signal extraction due to:

- Strong peak at zero for signal
- Smooth background distribution
- Stability across ALP mass points

Fits are performed independently for each mass hypothesis to extract signal yields and shape parameters. Although fits are generally performed after full continuum suppression and background rejection, a few example plots for fitting have been shown in Figure 3.12. The parameters of the distribution extracted from such a fit distribution are used to fit the true measured data to look for signals and confirm the presence (or absence) of our hypothetical ALP.

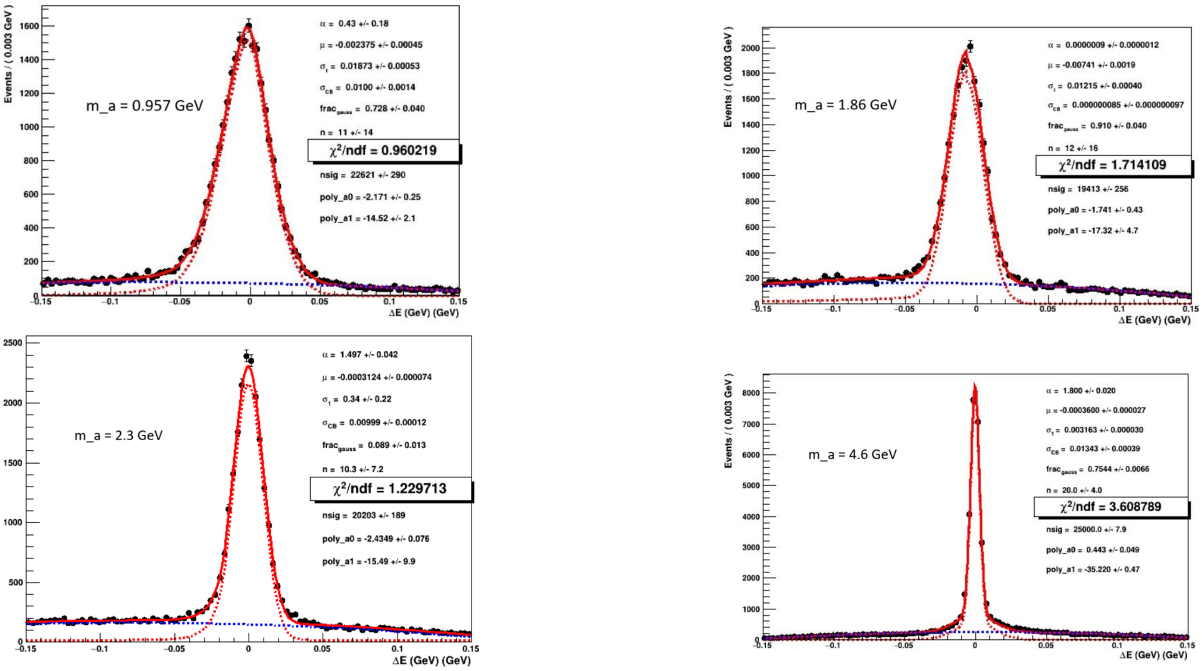


Figure 3.12: Fitting ΔE for various mass points. A mix of Crystal Ball and gaussian functions have been used to fit the signal, while the background is fit by a linear polynomial.

Chapter 4

Continuum Suppression using Multivariate Analysis

4.1 Introduction

One of the dominant sources of background in Belle II analyses arises from continuum processes:

$$e^+e^- \to q\bar{q}, \quad (q = u, d, s, c), \tag{4.1}$$

which produce hadronic final states that can mimic signal decays. These events differ fundamentally from signal $B\bar{B}$ events produced via the $\Upsilon(4S)$ resonance. The process of discriminating between these two classes of events is known as **continuum suppression**.

4.2 Underlying Physical Principle

The key physical distinction between signal and continuum events lies in their topology:

- **Signal ($B\bar{B}$ events):** Produced near threshold, the B mesons are almost at rest in the center-of-mass frame, leading to **spherical, isotropic** decay topologies.
- **Continuum ($q\bar{q}$ events):** Produced via hard scattering, resulting in **jet-like, back-to-back** structures.

This difference in event shape forms the basis for continuum suppression [25] Fig. 4.1.

4.3 Continuum Suppression Variables

Several event-shape observables are used to quantify this difference:

4.3.1 Thrust

The thrust axis is defined as the direction $\hat{n}$ that maximizes:

$$T = \frac{\sum_i |\vec{p}_i \cdot \hat{n}|}{\sum_i |\vec{p}_i|}. \tag{4.2}$$

- $T \to 1$: jet-like (continuum)
- $T \to 0.5$: spherical (signal)

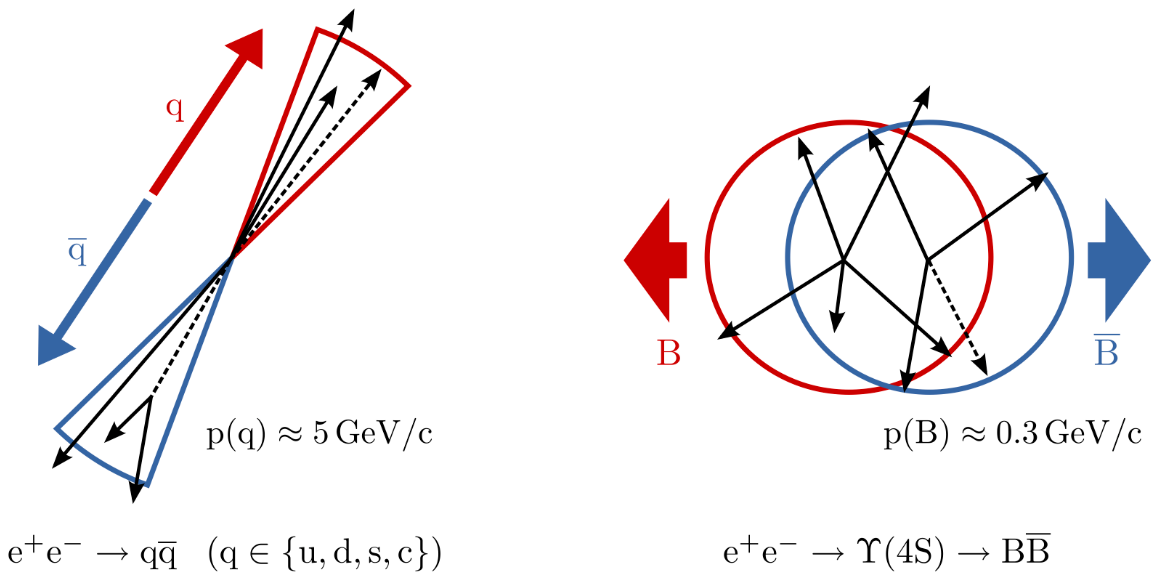


Figure 4.1: Difference in event shape between continuum and BB events (Credit: Markus Röhrken)

4.3.2 Fox–Wolfram Moments

Defined as:

$$H_l = \sum_{i,j} |\vec{p}_i||\vec{p}_j| P_l(\cos\theta_{ij}), \tag{4.3}$$

where P_l are Legendre polynomials. Ratios such as:

$$R_2 = \frac{H_2}{H_0} \tag{4.4}$$

are commonly used.

4.3.3 Modified Super Fox–Wolfram Moments

Belle II employs modified versions optimized for separating signal and continuum [25].The the Kakuno-Super-Fox-Wolfram (KSFW) moments are more refined version of the FW moments.

These variables collectively encode the global event topology.

Some of the variables assigned highest importance (Equation 4.8) after training the MVA are plotted in Figure 4.2 to show how the various differences between continuum and $B\bar{B}$ events are encoded.

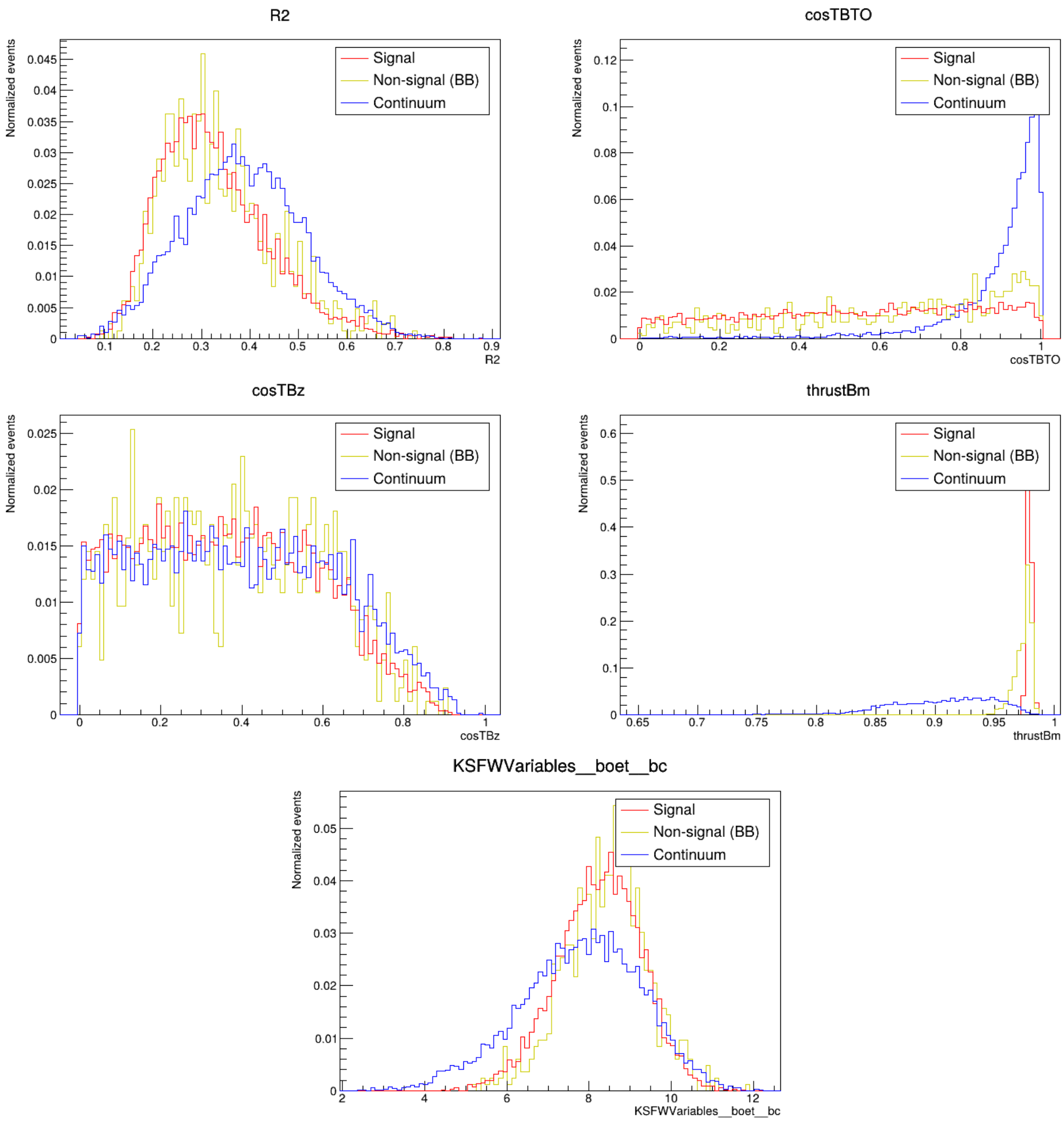


Figure 4.2: Distribution of variables which are assigned most importance in most continuum suppression MVAs. The above plots are distributions of the input training variables for the mass point $M_a = 0.88\ GeV$

4.4 Boosted Decision Trees (BDT)

A Boosted Decision Tree (BDT) is a supervised machine learning algorithm that combines multiple decision trees to improve classification performance [26].

A **decision tree** recursively partitions the feature space using binary splits:

$$x_i < \text{threshold} \tag{4.5}$$

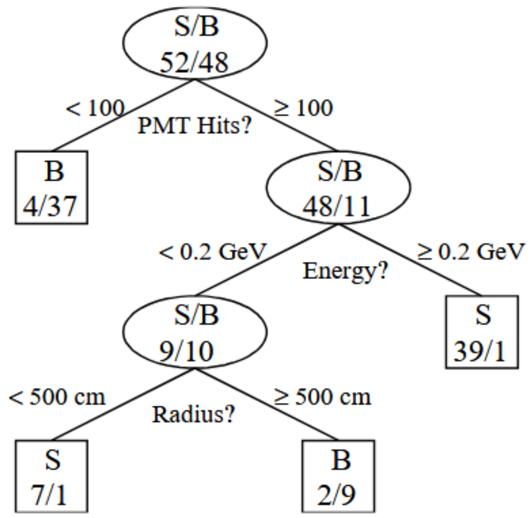
Each leaf node assigns a classification score.

Boosting combines multiple weak learners into a strong classifier by iteratively reweighting misclassified events.

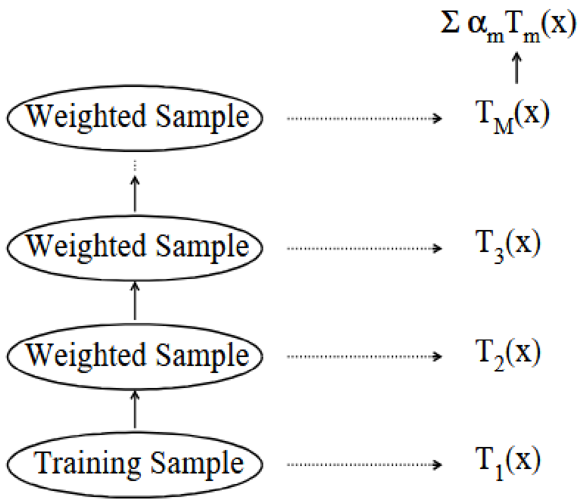
The final output is a classifier score:

$$y_{\text{BDT}} \in [0, 1], \tag{4.6}$$

representing the probability that an event is signal-like.



(a) Schematic of a decision tree. S for signal, B for background. Terminal nodes(called leaves) are shown in boxes. If signal events are dominant in one leave, then this leave is signal leave; otherwise, background leave. [26]



(b) Schematic of a boosting procedure. [26]

4.5 Training Strategy

Separate MVAs were trained for three representative ALP mass points:

$$m_a = 0.88,\ 1.92,\ 2.96\ \text{GeV}. \tag{4.7}$$

4.5.1 Training Dataset

For each mass point:

- For signal events, 2×10^5 events were generated and underwent reconstruction for each mass point.
- For continuum events, 500 fb^{-1} MC data was used for each quark mode $(u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c})$ and reconstructed on the BELLE-II grid, for the entire mass range.
- The reconstructed continuum events were mixed in a ratio corresponding to their relative branching ratios, such that the total number of continuum and signal events for each mass point is equal (Table 4.1)
- Dataset split: 50% training, 50% testing

Table 4.1: Relative branching fractions of quark pair production modes in the Standard Model used to construct the continuum Monte Carlo sample for MVA training.

Quark Mode	Fraction
$u\bar{u}$	0.350836
$d\bar{d}$	0.0725795
$s\bar{s}$	0.135152
$c\bar{c}$	0.441433

The target variable was `isNotContinuumEvent`, where signal events are labeled as 1 and continuum as 0.

4.6 Loose vs Tight Cut Strategies

Two approaches were considered:

- **Tight cuts:** (section 3.9)
  - Background already reduced
  - No bias between training and analysis
  - Reduced statistics
- **Loose cuts:**
  - Final state $E_\gamma > 0.075$ GeV, $M_{bc} > 5.2$GeV
  - Larger statistics for (possibly) better training results
  - Possible bias introduced when applied to tightly selected data

In this work, the tight-cut strategy was adopted for consistency with the signal analysis.

4.7 Variable Selection and Optimization

After initial training, the following steps were performed:

4.7.1 Variable Importance

Each variable is assigned an importance score based on its contribution to classification:

$$I_i = \sum_{\text{splits}} \Delta G, \tag{4.8}$$

where ΔG is the reduction in impurity. Variables with zero importance were removed after verifying their separation power.

4.7.2 Correlation Analysis

Correlation matrices were computed for all variables.

- If $|\rho_{ij}| > 0.8$, variables are considered highly (positively or negatively) correlated

Highly correlated variables are undesirable because:

- They provide redundant information
- They increase model variance
- They lead to overfitting

The less important variable in such pairs was removed. Two iterations of this optimization were performed.

4.8 Example of Training Iterations

In Figure 4.4, we see the results of the first iteration of MVA training on data supplied for ALP mass 0.88 GeV. The importance plots assigns importance 0 to the variables `KSFWVariab1-13`, `ThrustOm`. As can be seen, the distribution of the variable KSFW variable `hso03` is almost identical, and hence can be safely removed. But the variable `ThrustOm`, despite having importance 0 has quite distinct separation, and can still be used for training. Hence, it shall

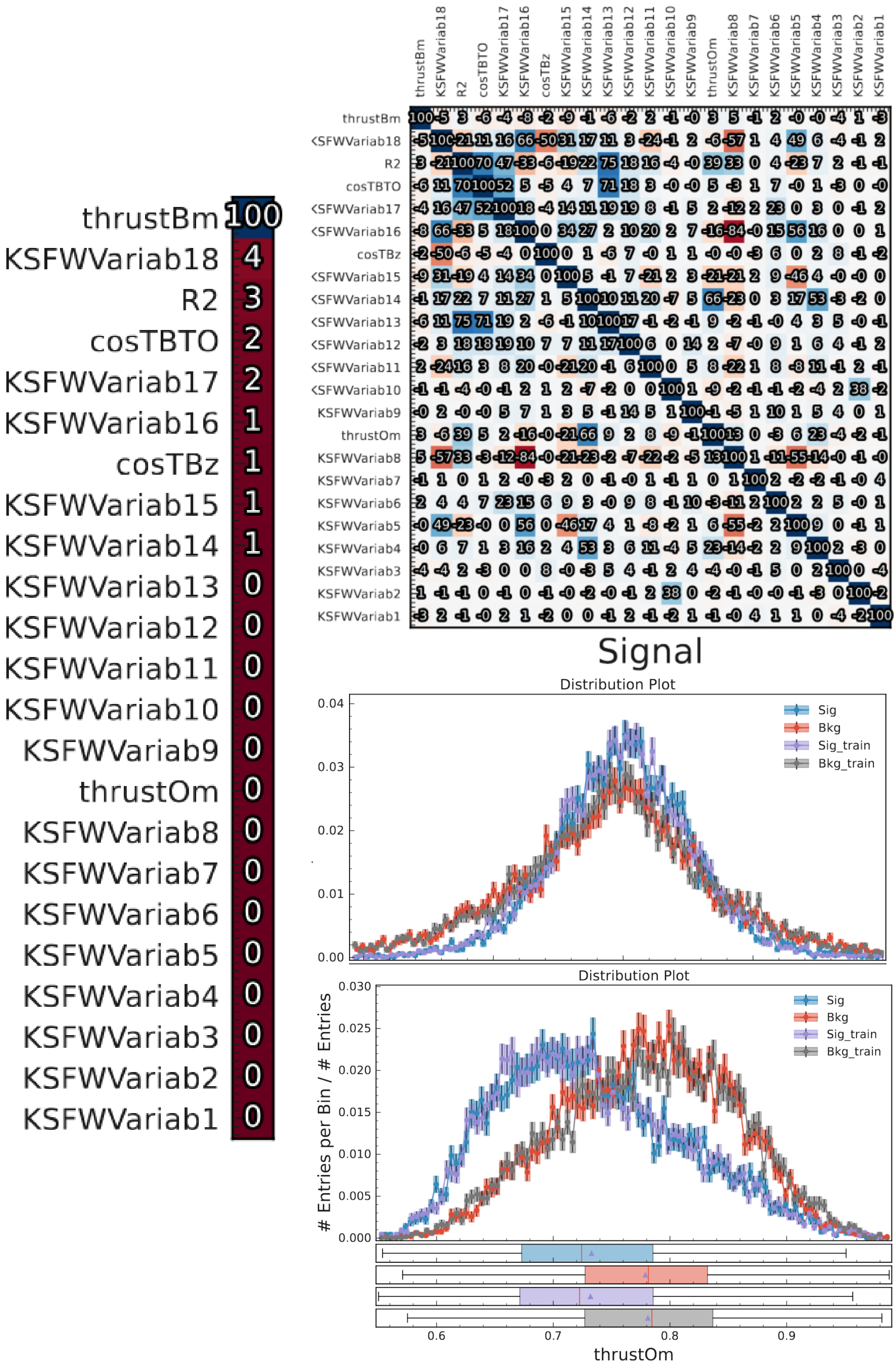


Figure 4.4: MVA training iteration 1 on $M_a = 0.88$ GeV: variable importance (left), correlation matrix (top right) and separation power of `KSFWVariable1(hso03)` (centre right), `ThrustOm` (bottom right).

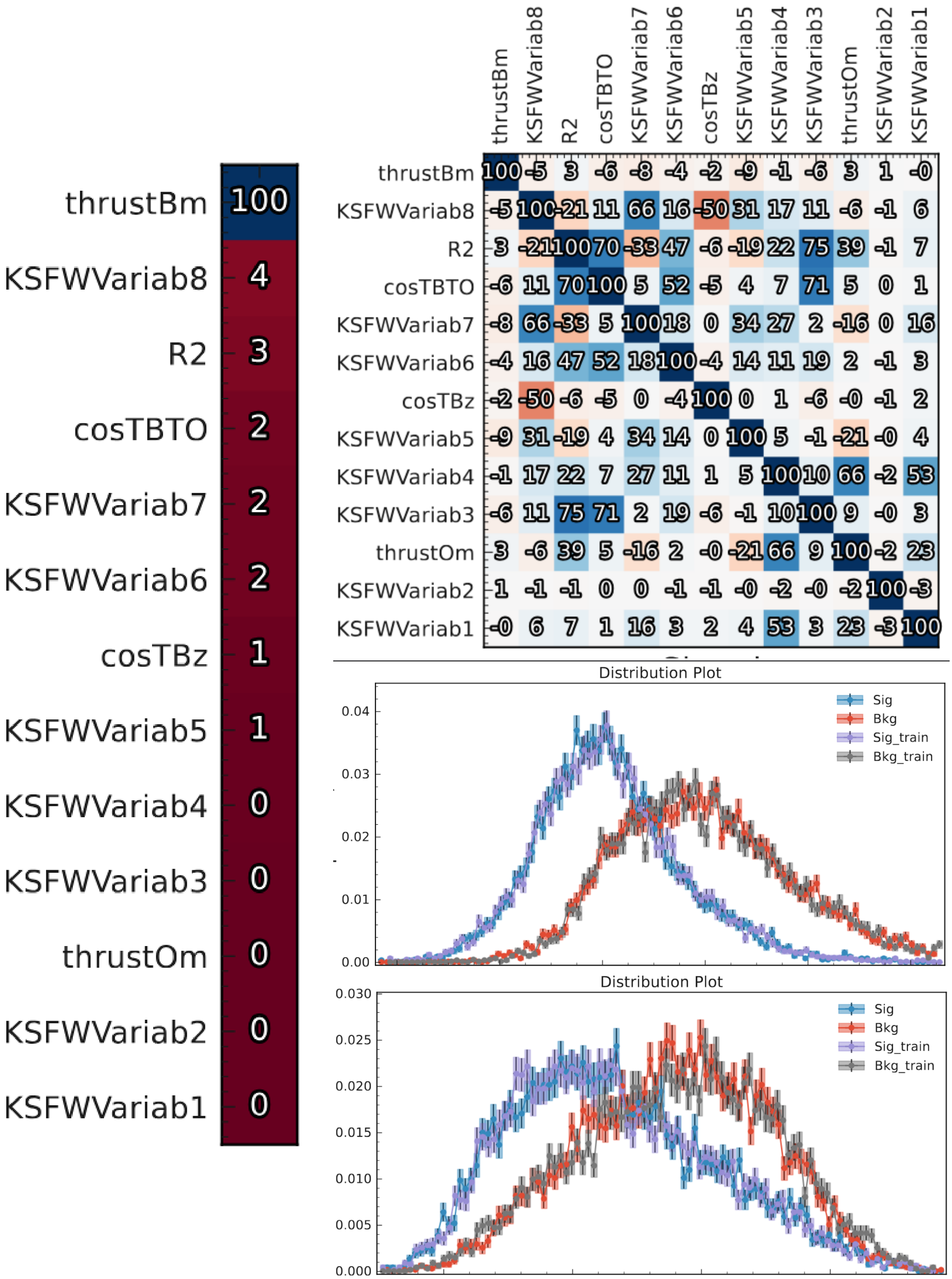


Figure 4.5: MVA training iteration 2 on $M_a = 0.88$ GeV: variable importance (left), correlation matrix (top right) and separation power of `KSFWVariable3(hso02)` (centre right), `ThrustOm` (bottom right).

be preserved for the next training iteration. The correlation plot tells us that `KSFWVariab8` and `KSFWVariab16` have a very high negative correlation. Since, `KSFWVariab16` is assigned a higher importance, it shall be preserved for the next iteration.

Figure 4.5 shows the same plots after the second training iteration. Many variables now have importance 0 but show good separation power. Based on similar observations for the other 2 mass points, two iterations was set as the limit for the sake of avoiding a huge number of recursive iterations.

Finally, the MVA classifer output is plotted separately for signal and background to check for overtraining. Overtraining is checked by studying any significant deviation between the plotted train and test data. If not, then the MVA has not overfit it's parameters and works as intended.

Appendix C shows all relevant plots for ALP masses $M_a = 1.92,\ 2.96\ GeV$

NOTE: The KSFW variables are re-labeled during each training iteration. For eg. $KSFW3_{iter1} \neq KSFW3_{iter2}$

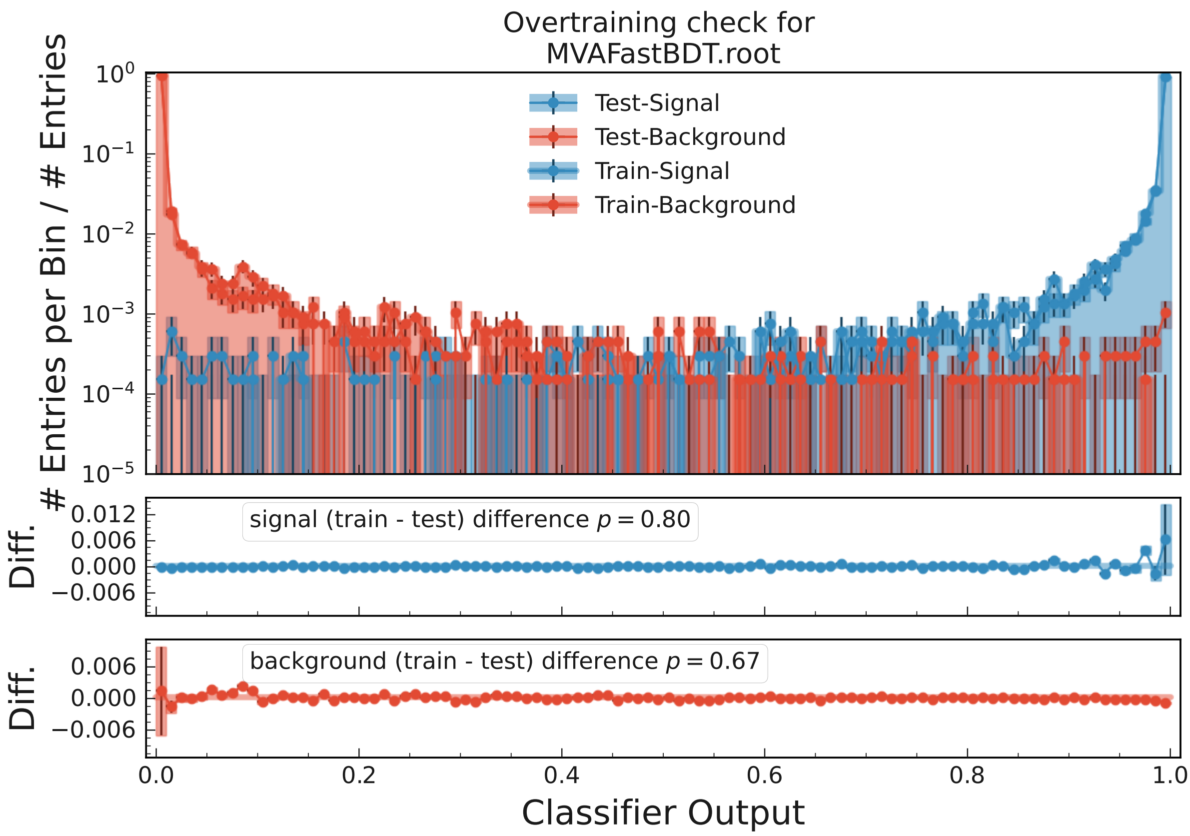


Figure 4.6: MVA classifier Output for ALP mass 0.88 GeV. The large error bars are due to low statistics. Identification error plots reveal a bit of overfitting in the fringes, but within the acceptable 3σ range. Similar results are seen in Appendix C

4.9 MVA Output and Weight Files

After training, the MVA produces a **weight file**, which contains:

- Decision tree structure
- Node thresholds
- Boosting weights

This file is applied to new testing data to compute the classifier output for each event:

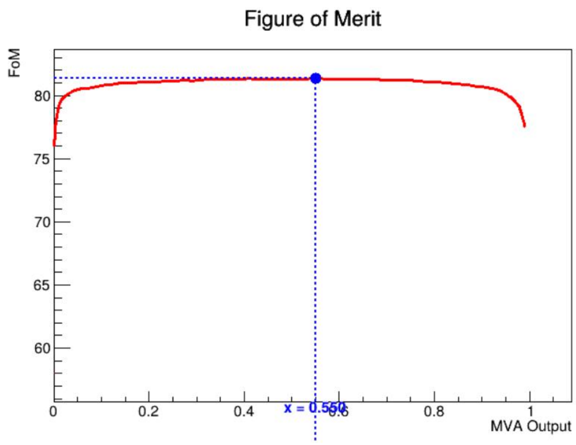
$$y_{\mathrm{BDT}} = \text{MVA classifier output.} \tag{4.9}$$

4.10 Optimization using Figure of Merit

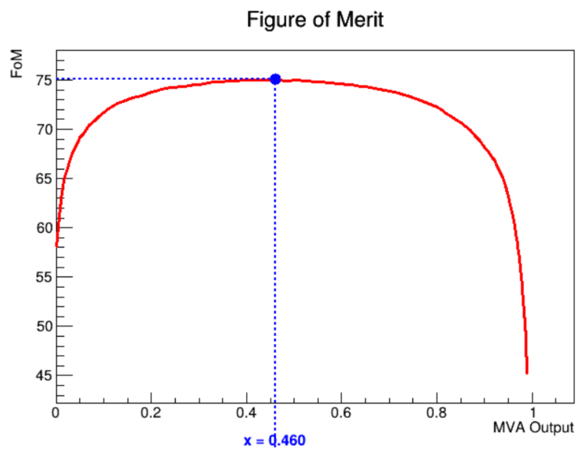
Given the classifier output for a given set of real data, where would be the most optimal cut for this classifier output be in order to maximize signal retention and continuum rejection? This is decided by performing a Figure of Merit analysis on y_{BDT} that is produced by applying the MVA to a set of Monte Carlo data. The same cut is then applied when the MVA is used on real data.

$$\mathrm{FoM} = \frac{S}{\sqrt{S+B}}. \tag{4.10}$$

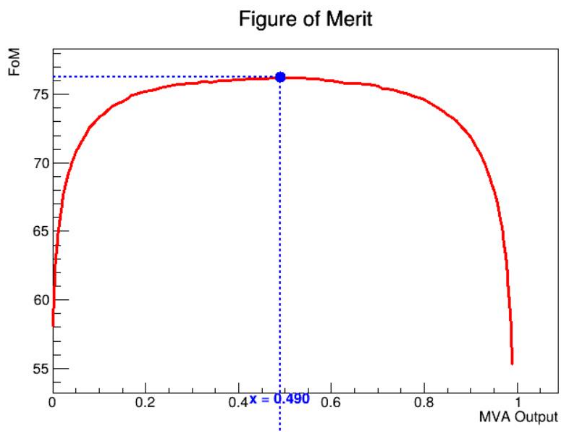
S = # events that have truth-matched information `isNotContinuumEvent` = 1
B = # events that have truth-matched information `isNotContinuumEvent` = 0



(a) 0.88GeV, 0.55



(b) 1.92GeV, 0.46



(c) 2.96GeV, 0.49

Figure 4.7: Figure of Merit plots for the 3 mass points chosen for continuum suppression, Subcaptions give FoM information in the form (M_a, MVA Output cut)

4.11 Performance Evaluation

Performance is evaluated using signal efficiency, background rejection.

Table 4.2 shows the MVA performance at the different mass points chosen for continuum suppression analysis. As can be seen, the background rejection rate remains above 90% and the signal rejection is minimal at $\sim 7\%$ or lower. Since the 3 mass points were chosen to represent the low, middle and high sections of the ALP mass range $0.85 < M_a < 3\ GeV$, similar performances can be expected throughout the mass range, though more mass points must be studied for finer understanding.

Table 4.2: MVA performance for mass points 0.88, 1.92, 2.96 GeV.

Mass (GeV)	**Category**	**Total**	$B\bar{B}$	**Signal**	**Continuum**
	No BDT	13254	6714	5902	6540
0.88	BDT	6706	6649	5892	57
	Rejection Rate	–	0.00968	0.001694	0.99128
	No BDT	13258	6718	5360	6540
1.92	BDT	6837	6210	5003	627
	Rejection Rate	–	0.07562	0.066604	0.90413
	No BDT	13272	6732	5533	6540
2.96	BDT	6711	6251	5176	460
	Rejection Rate	–	0.07145	0.064522	0.92966

The effect of the BDT cuts on the signal and continuum background distribution can be observed clearly in Figs. 4.8, 4.9 and 4.10. The non-flat areas in the ALP mass distributions may indicate Standard Model particles that have the same final state products (like D^0 , η') and have passed through the selection cuts. Nevertheless, they are also suppressed by the MVA's optimized selection cut.

4.12 Mass-Independent Training

If a thorough search is to be performed in the entire allowed mass range of the ALP, a large number of mass points will have to be simulated and analyzed and hence, a vast number of MVAs will have to be trained, each with their own iterations of variable optimization. In order to find a more convenient and general approach, an attempt was made to train a single MVA across the full signal mass range. The objective was to have stable MVA identification efficiency acroos the mass range so the MVA can reliably identify and reject the continuum at any given mass point.

However, the performance was found to be extremely mass-dependent (Figure 4.11). We hypothsize that this is due to the extremely irregular number of events (w.r.t. ALP mass) supplied for the MVA training. This is both an artefact of

1. The mass dependent reconstruction efficiency (Figure 3.10)
2. The sharpness of the ALP mass peak at each point, which leads to drop-offs when the mass points are merged to form a range (Figure 4.12).

As a result, independent MVAs were trained for each mass point for the current analysis. Efforts persist to understand and apply single MVA approaches for the future of this analysis.

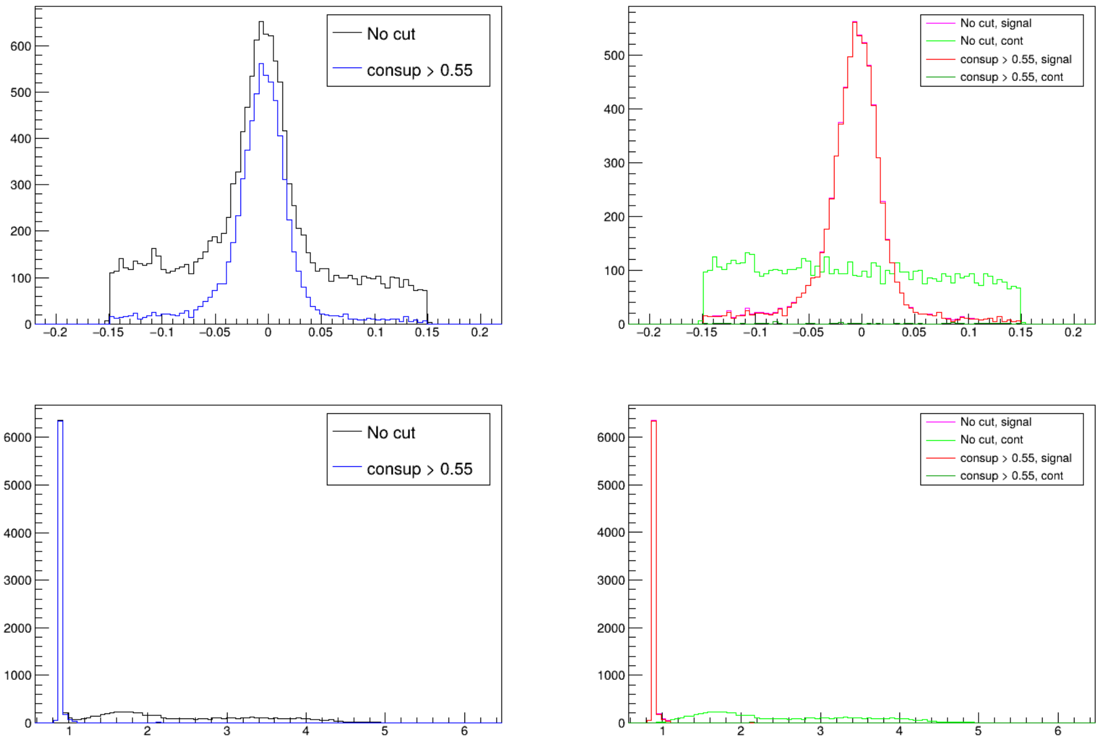


Figure 4.8: Applying optimized BDT cuts on ΔE (top) and M_a (bottom) distributions for $M_a = 0.88\ GeV$. Component wise MVA rejections are shown on the right

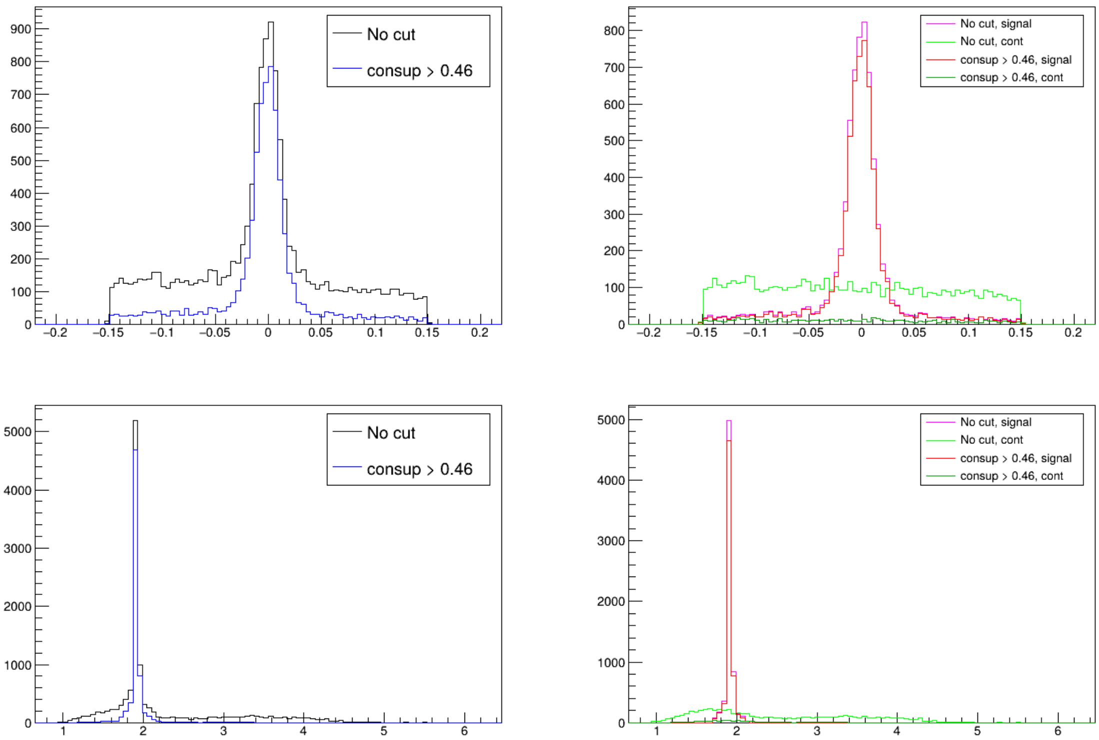


Figure 4.9: Applying optimized BDT cuts on ΔE (top) and M_a (bottom) distributions for $M_a = 1.92\ GeV$. Component wise MVA rejections are shown on the right

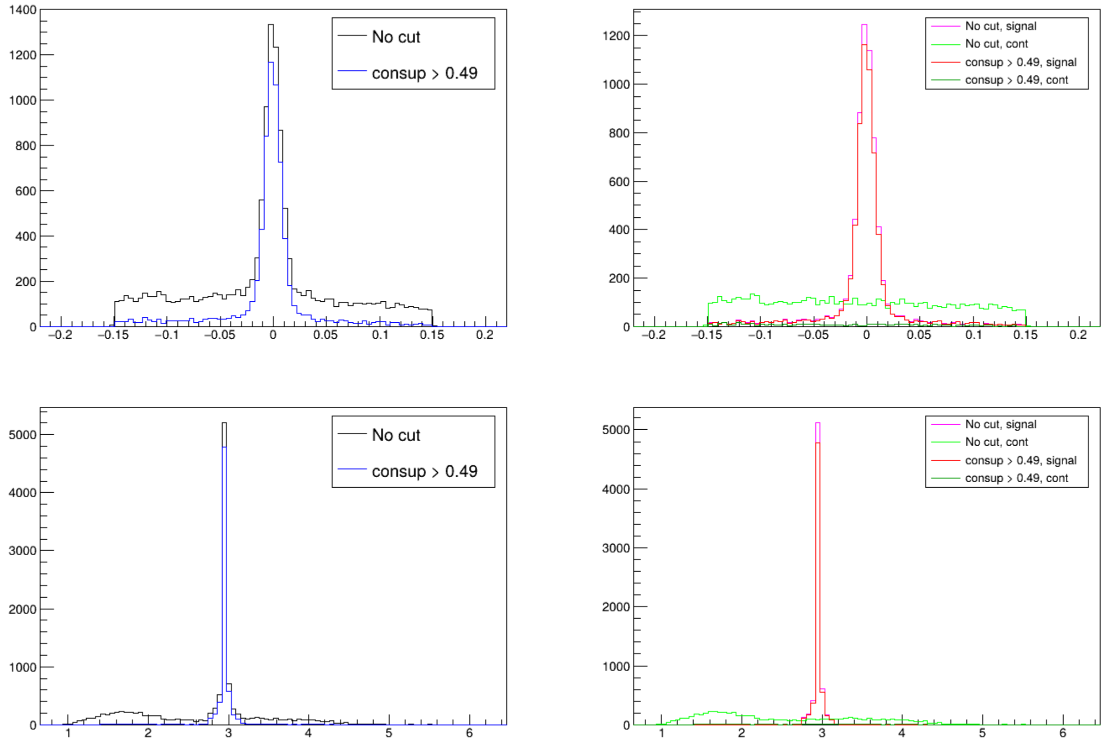


Figure 4.10: Applying optimized BDT cuts on ΔE (top) and M_a (bottom) distributions for $M_a = 2.96\ GeV$. Component wise MVA rejections are shown on the right

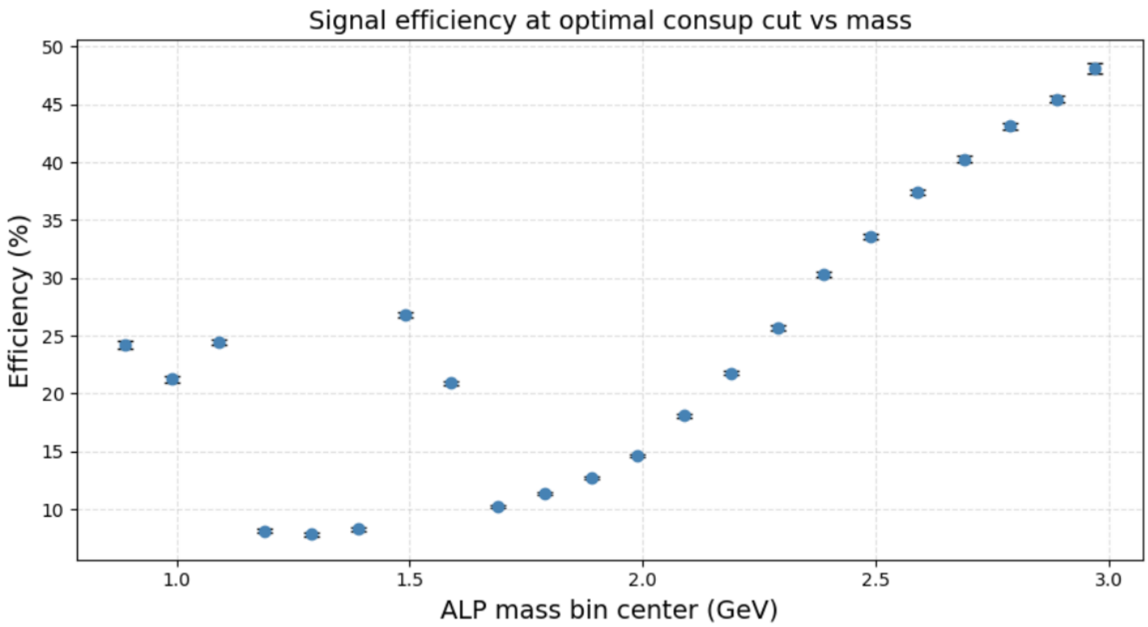


Figure 4.11: MVA efficiency vs mass. Efficiency calculated after applying optimal cuts from FoM analyses (Credit: Dr. Sourav Dey, KMI-Nagoya U.)

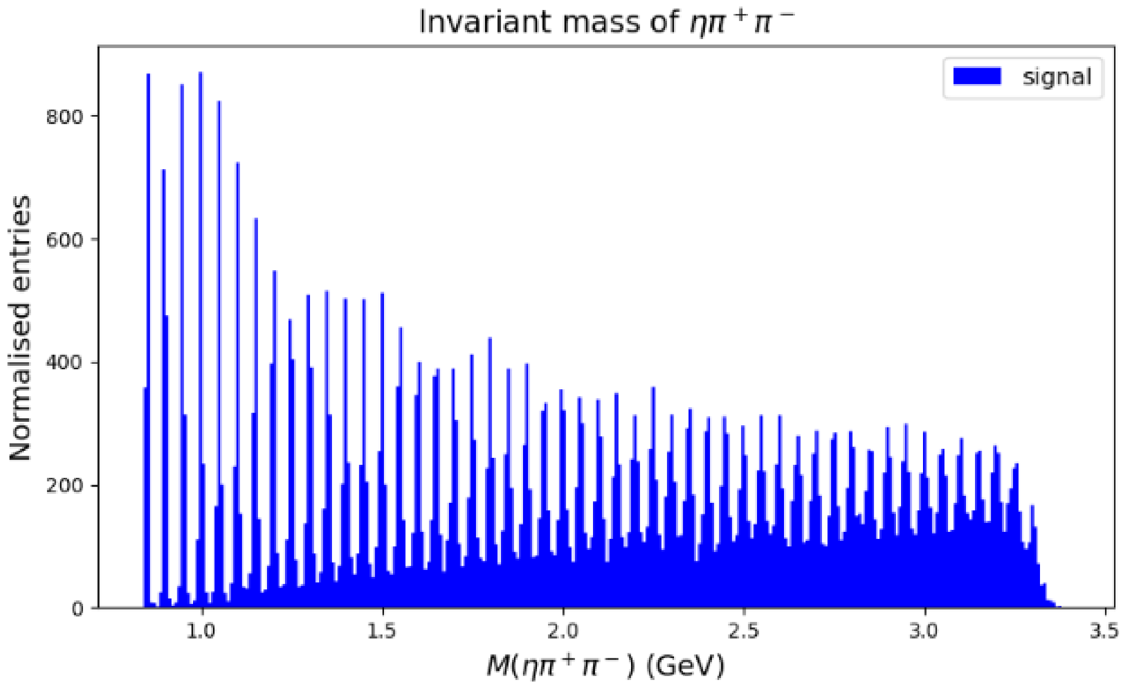


Figure 4.12: Number of events supplied to MVA for training as a function of mass (Credit: Dr. Sourav Dey, KMI-Nagoya U.)

Chapter 5

Outlook and Conclusion

5.1 Outlook: Future Steps of the Analysis

The work presented in this thesis constitutes the foundational stage of a full search for heavy QCD axion-like particles at Belle II via the decay $B^+ \to K^+ a$, $a \to \eta\pi^+\pi^-$. Several important steps remain before a statistically meaningful result can be extracted from real data. This section outlines the roadmap for advancing the analysis.

5.1.1 Unified Continuum Suppression across the Full Mass Range

As discussed in Chapter 4, the current continuum suppression strategy trains an independent BDT for each of the three representative ALP mass points ($M_a = 0.88,\ 1.92,\ 2.96$ GeV). While this achieves excellent background rejection rates exceeding 90% at each individual point, it does not scale efficiently to a comprehensive scan across the full mass range $0.85 \lesssim M_a \lesssim 3$ GeV.

A primary goal of future work is to develop a single, unified MVA that maintains stable identification efficiency across the entire signal mass range. The primary challenge, as identified in Chapter 4, is that the highly irregular distribution of training events as a function of mass — arising from mass-dependent reconstruction efficiency and the sharpness of the ALP invariant mass peaks — leads to highly mass-dependent MVA performance when a combined training is naively attempted.

Once a satisfactory unified continuum suppression method is established, it will be validated on all mass points in the full range.

5.1.2 Signal Shape Estimation and Fitting Strategy

Following the establishment of a reliable background suppression method, the next step is to develop a robust fitting strategy for signal extraction. This involves:

1. **Signal shape parametrization:** The ΔE distribution for the signal is currently described empirically using a combination of Crystal Ball and Gaussian functions (Chapter 3). Fits will be performed across all simulated mass points to extract the shape parameters — means, widths, tail parameters — and study their evolution as functions of M_a. Smooth interpolation functions will be constructed so that a signal shape can be generated for any arbitrary mass hypothesis without requiring a dedicated simulation at every point.
2. **Background shape modelling:** The background ΔE distribution from generic $B\bar{B}$ and $q\bar{q}$ events is currently modelled by a linear polynomial. After full continuum suppression, the residual background shape will be re-evaluated, and higher-order polynomial

or non-parametric models may be considered if the background structure is found to be non-trivial.

3. **Fit validation:** The full fit model will be validated using toy Monte Carlo studies. Ensembles of pseudo-experiments will be generated from the background-only model, and signal injection tests will verify that the fit correctly recovers known signal yields with appropriate uncertainties.
4. **Fitting strategy for real data:** A blind analysis strategy will be adopted when moving to real Belle II data. The signal region in ΔE will remain masked until all selection criteria, background models, and systematic uncertainties are finalized. Fits will be performed simultaneously over the ALP mass range in a scan, with each mass hypothesis tested independently.

5.1.3 Standard Model Particle Vetoes

As noted in Chapter 3, the reconstructed $M_{\eta\pi\pi}$ spectrum contains contributions from Standard Model particles with the same final state topology, like D^0 and η'. While these peaks are suppressed by the selection cuts and MVA, they can still mimic or overlap with an ALP signal at specific mass hypotheses.

Before applying fits to data, dedicated vetoes will be introduced for the dominant SM resonances in the accessible mass range. This will involve:

- Identifying all SM particles and charmonium resonances that can produce $\eta\pi^+\pi^-$ final states (e.g. D^0, η', η_c, χ_{c0}, J/ψ decay products) and characterizing their contributions in the reconstructed $M_{\eta\pi\pi}$ distribution.
- Defining veto windows around the nominal masses of these resonances, with window widths determined from the reconstructed mass resolution at each point.
- Evaluating the signal efficiency loss due to each veto and incorporating the resulting systematic uncertainties into the final sensitivity estimate.

The dense population of charmonium states above $M_a \gtrsim 2.9$ GeV, as catalogued in Chapter 3, motivates careful treatment in this region, and may lead to the exclusion of sub-ranges of the search window where vetoes are insufficient to control the structured background.

5.1.4 Detector Sensitivity Estimation

Once the fitting framework and background model are in place, the detector sensitivity to the target channel will be quantified. This will be done by:

1. Computing the expected upper limit on the branching fraction BR$(B^+ \to K^+a) \times$ BR$(a \to \eta\pi^+\pi^-)$ as a function of M_a, using the CL$_s$ method or a profile likelihood approach.
2. Incorporating the mass-dependent reconstruction efficiency curve (Chapter 3) and MVA efficiency to translate raw event yields into branching fraction limits.
3. Propagating systematic uncertainties arising from the signal efficiency, background modelling, particle identification, tracking, and photon reconstruction.
4. Translating the branching fraction limits into constraints on the axion decay constant f_a and the gluon coupling c_{gg}/Λ, and comparing the resulting exclusion regions in ALP parameter space with existing bounds from BaBar and Belle (Chapter 1).

This will yield the first direct experimental limits on the heavy QCD axion in the $\eta\pi^+\pi^-$ decay channel from Belle II data, extending and improving upon the indirect reinterpretations of earlier measurements.

5.2 Conclusion

This thesis has presented a comprehensive Monte Carlo study in preparation for the first direct search for heavy QCD axion-like particles at the Belle II experiment through the decay chain $B^+ \to K^+ a$, $a \to \eta(\to\gamma\gamma)\pi^+\pi^-$.

The theoretical motivation for heavy QCD axions was reviewed in Chapter 2. The Peccei–Quinn mechanism provides an elegant dynamical solution to the strong CP problem, but traditional invisible axion models face tension with the axion quality problem and astrophysical constraints. Heavy QCD axions, in which additional hidden-sector contributions decouple m_a from f_a, resolve both problems simultaneously for $f_a \sim$ TeV and $m_a \sim$ MeV–GeV. The gluonic coupling of the axion induces the flavor-changing neutral current transition $b \to sa$ at two loops, making rare B decays a particularly sensitive probe.

Chapter 3 described the Monte Carlo simulation campaign carried out using the basf2 framework at KEKCC. Signal samples were generated at representative mass points spanning 0.85–4.8 GeV and processed through the full detector simulation and reconstruction pipeline. Standard Model control channels sharing the same final-state topology ($B \to KD^0$ and $B \to K\eta'$) were used to benchmark and validate the reconstruction strategy. A suite of selection criteria was established, combining standard Belle II cuts on vertex quality, particle identification, and B kinematics with analysis-specific optimizations derived from figure-of-merit studies on the photon energy and $\eta\gamma\gamma$ opening angle. Reconstruction efficiencies of 10–13% were achieved across the signal mass range. The search window was restricted to $M_a < 3$ GeV, motivated by the proliferation of charmonium resonances at higher masses that would render signal extraction unreliable. The ΔE distribution was identified as the primary observable for signal extraction, with signal shapes well described by a combination of Crystal Ball and Gaussian functions.

Chapter 4 addressed the dominant background from continuum $e^+e^- \to q\bar{q}$ processes. The physical distinction between the spherical topology of $B\bar{B}$ signal events and the jet-like topology of continuum events was exploited through event-shape variables including thrust, Fox–Wolfram moments, and Kakuno–Super-Fox-Wolfram moments. Boosted Decision Trees were trained and optimized at three representative ALP mass points, with careful variable importance and correlation analyses performed iteratively to avoid redundancy and overfitting. The resulting classifiers achieve continuum rejection rates exceeding 90% with signal loss below 7%. An attempt to train a single mass-independent MVA revealed significant mass-dependent performance, motivating continued development of unified approaches.

Together, these results establish a complete analysis framework for the signal channel and lay the quantitative groundwork for a full sensitivity estimate and eventual application to real Belle II data. With the fitting strategy, SM vetoes, and unified continuum suppression steps remaining, the analysis is well-positioned to deliver the first direct experimental constraints on heavy QCD axions from Belle II, contributing meaningfully to the broader landscape of axion and new physics searches at B factories.

Appendix A

Derivation of the Cross Section of the decay $B \to Ka$

Ref. [8] provides detailed calculation, and it is briefly reproduced in this section. Starting from the Lagrangian in Eq. 2.8, the leading contribution to $B \to Ka$ ($b \to sa$) arises at 2-loop in Fig. 2.1b. Cancelling UV divergences in these diagrams requires the following additional interactions to be further included in the Lagrangian:

$$\mathcal{L} = \cdots + C_{qq} \sum_q \frac{\partial a}{f_a} \bar{q}\gamma^\mu \gamma_5 q + C_{bs} \frac{\partial a}{f_a} \bar{s}_L \gamma^\mu \gamma_5 b_L + \text{h.c.} \tag{A.1}$$

where $\cdots$ represents the Lagrangian in Eq. 2.8. The parameters C_{qq} and C_{bs} at the cutoff Λ_{UV} are

$$C_{qq}(\Lambda_{\text{UV}}) \equiv A\, C_F \left(\frac{\alpha_s}{4\pi}\right)^2, \tag{A.2}$$

$$C_{bs}(\Lambda_{\text{UV}}) \equiv B\, C_F \left(\frac{\alpha_s}{4\pi}\right)^2 \frac{\alpha_W}{4\pi} \sum_k V_{ik} V^*_{kj} \xi_k, \tag{A.3}$$

where A and B are $\mathcal{O}(0\text{–}1)$ parameters that depend on the unknown UV completion of the Lagrangian. The decay rate of $B \to Ka$ is

$$\Gamma_{B\to Ka} = |C_W|^2 \frac{m_B}{64\pi f_a^2} \left(1 - \frac{m_K^2}{m_B^2}\right)^2 \lambda_{Ka} \, [f_0(m_a^2)]^2. \tag{A.4}$$

$$\lambda_{Ka} = \left[\left(1 - \frac{(m_K + m_a)^2}{m_B^2}\right)\left(1 - \frac{(m_K - m_a)^2}{m_B^2}\right)\right]^{1/2}, \tag{A.5}$$

and $f_0(m_a^2)$ is the form factor as

$$f_0(m_a^2) = \frac{0.330}{1 - m_a^2/(37.5\ \text{GeV})}. \tag{A.6}$$

The factor C_W is determined by C_{qq} and C_{bs} in Eqs. A.2 and A.3, with $\mu_W \sim M_W$ and the contributions from integrating out t and M_W as

$$C_W = C_{bs}(\mu_W) + \frac{\alpha_W}{4\pi} C_{qq}(\mu_W) g(\mu_W) + \frac{1}{2}\frac{\alpha_W}{4\pi}\left(\frac{\alpha_s}{4\pi}\right)^2 f(\mu_W), \tag{A.7}$$

where f and g are 1- and 2-loop matching functions defined in the appendix section in Ref. [8]. The approximate branching fraction is

$$\text{BF}(B \to Ka) \approx 1.1\ (7.6) \times 10^{-5} \left[\frac{100\ \text{GeV}}{f_a}\right]^2, \tag{A.8}$$

for $\Lambda_{\text{UV}} = 1\ (10)\ \text{TeV}$ assuming $A = B = 0$.

Appendix B

Statistics

B.1 Introduction

A precise understanding of statistical uncertainties and detector resolution is essential for interpreting reconstructed observables in high energy physics. In this analysis, quantities such as invariant mass, energy difference ΔE, and beam-constrained mass M_{bc} are derived from measured particle momenta and energies, each subject to finite detector resolution.

This chapter develops the statistical framework necessary to quantify measurement uncertainties, propagate errors, and understand correlations between observables. It is essential to understand why our distributions look the way they do.

B.2 Measurement and Uncertainties

A physical observable x measured by a detector can be expressed as:

$$x_{\mathrm{meas}} = x_{\mathrm{true}} + \delta x, \tag{B.1}$$

where δx represents the measurement error, typically modeled as a random variable with zero mean:

$$\langle \delta x \rangle = 0, \qquad \sigma_x^2 = \langle (\delta x)^2 \rangle. \tag{B.2}$$

In collider experiments, these uncertainties arise from:

- Finite detector resolution
- Calibration errors
- Reconstruction ambiguities

B.3 Error Propagation

For a function $f(x_1, x_2, \dots)$ of measured variables, the variance is given by:

$$\sigma_f^2 = \sum_i \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_i^2 + 2\sum_{i<j} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\mathrm{Cov}(x_i, x_j). \tag{B.3}$$

The covariance is defined as:

$$\mathrm{Cov}(x, y) = \langle (x - \bar{x})(y - \bar{y}) \rangle. \tag{B.4}$$

This becomes crucial when observables depend on correlated quantities such as energy and momentum.

B.4 Invariant Mass and Error Propagation

The invariant mass of a particle is given by:

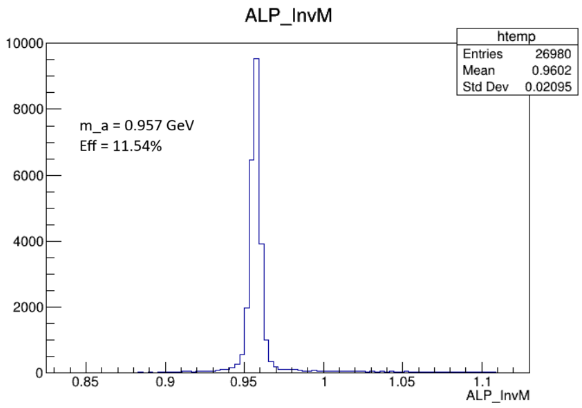
$$m^2 = E^2 - |\vec{p}|^2. \tag{B.5}$$

Applying the error propagation formula (delta method), we obtain:

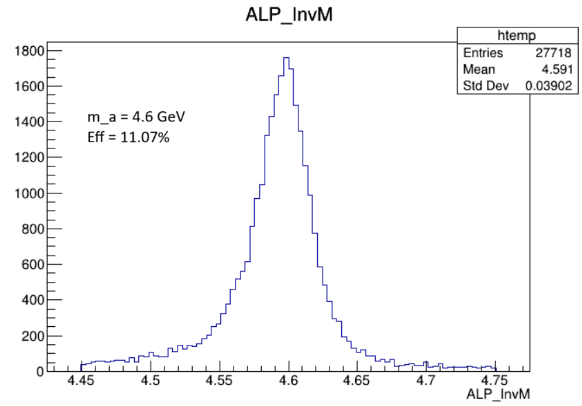
$$\sigma_{m^2}^2 = \left(\frac{\partial m^2}{\partial E}\right)^2 \sigma_E^2 + \left(\frac{\partial m^2}{\partial p}\right)^2 \sigma_p^2 + 2\frac{\partial m^2}{\partial E}\frac{\partial m^2}{\partial p}\mathrm{Cov}(E,p) \tag{B.6}$$

$$= 4E^2\sigma_E^2 + 4p^2\sigma_p^2 - 8Ep\,\mathrm{Cov}(E,p). \tag{B.7}$$

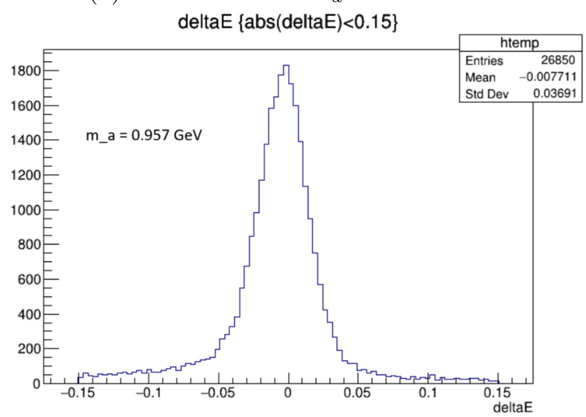
This expression shows that the invariant mass resolution depends critically on both individual uncertainties and their correlation.



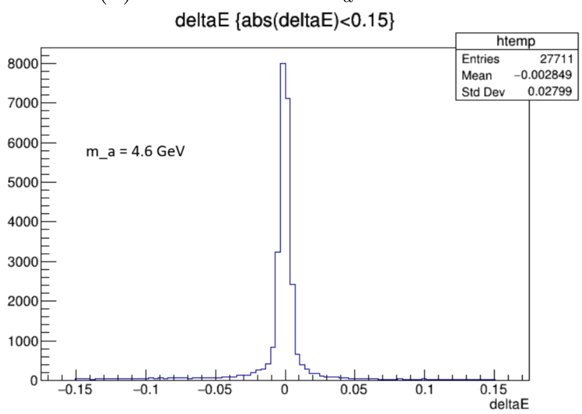
(a) ALP InvM for $M_a = 0.957\ GeV$



(b) ALP InvM for $M_a = 4.6\ GeV$



(c) ΔE for $M_a = 0.957\ GeV$



(d) ΔE for $M_a = 4.6\ GeV$

Figure B.1: Distribution of the reconstructed variables ΔE and M_a for representative masses on the opposite sides of the mass range. Why do the signal peaks become sharper for one variable and vice-versa for the other?

B.5 Mass Dependence of ΔE Resolution

An important feature observed in the reconstructed distributions is that the ΔE distribution is significantly broader for low-mass ALP hypotheses compared to higher masses. This behavior can be understood from kinematics and error propagation.

The observable ΔE is defined as:

$$\Delta E = E_B - E_{\mathrm{beam}}, \qquad E_B = E_K + E_a, \tag{B.8}$$

and its variance is given by:

$$\sigma^2_{\Delta E} = \sigma^2_{E_K} + \sigma^2_{E_a} + 2\,\mathrm{Cov}(E_K, E_a). \tag{B.9}$$

Unlike invariant mass observables, ΔE is a linear combination of energies and does not benefit from cancellation between correlated quantities. As a result, measurement uncertainties add constructively.

The origin of the mass dependence lies in the two-body kinematics of the decay:

$$B \to Ka, \tag{B.10}$$

where the momentum of the decay products in the B rest frame is:

$$p = \frac{1}{2m_B}\sqrt{\lambda(m_B^2, m_K^2, m_a^2)}, \tag{B.11}$$

with λ being the Källén function.

For smaller ALP masses, the available phase space increases, resulting in larger momenta for both the kaon and the ALP. In this regime, the ALP is highly boosted, and its energy satisfies:

$$E_a \approx p_a. \tag{B.12}$$

The propagation of uncertainties from momentum to energy is given by:

$$\sigma_E = \frac{p}{E}\,\sigma_p. \tag{B.13}$$

Thus, for a light ALP where $p/E \approx 1$, the energy uncertainty is maximal:

$$\sigma_E \approx \sigma_p. \tag{B.14}$$

In contrast, for heavier ALPs, the boost is reduced and $p/E < 1$, leading to suppressed energy uncertainties:

$$\sigma_E < \sigma_p. \tag{B.15}$$

Since ΔE depends linearly on the reconstructed energies, these uncertainties accumulate without cancellation. Consequently, the ΔE distribution is broader for low-mass ALPs and narrower for higher masses.

This behavior is consistent with the observed distributions (Figure B.1)

B.6 Interpretation for ALP Reconstruction

As observed in the reconstructed distributions (Figure B.1), the behavior differs significantly from the trends in M_a. An explanation follows:

B.6.1 Light ALP Regime

For a highly boosted light ALP:

$$E \approx p, \tag{B.16}$$

and typically:

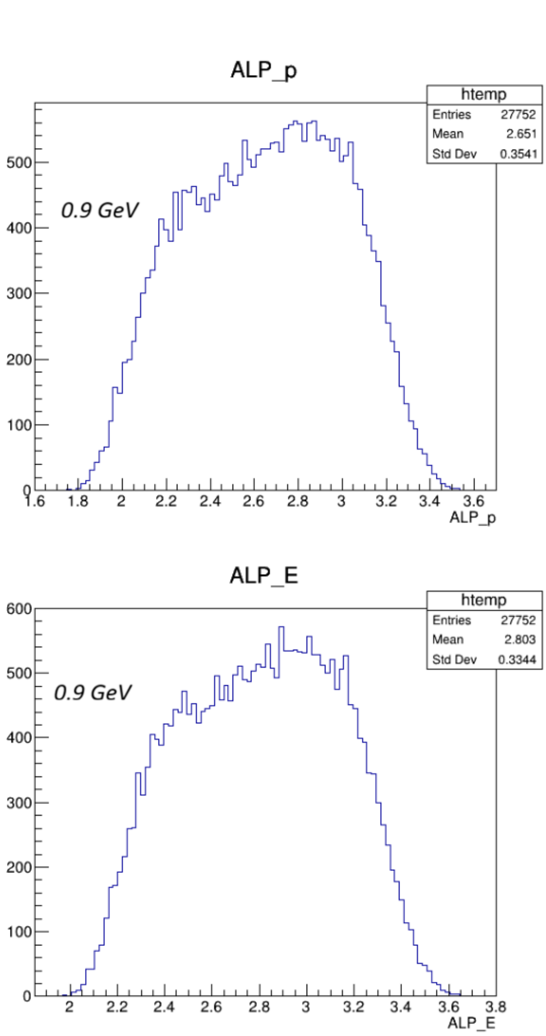
$$\mathrm{Cov}(E, p) \approx \sigma_E \sigma_p \approx \sigma_E^2 \tag{B.17}$$

since the two measurables are highly correlated (Figure B.2a).

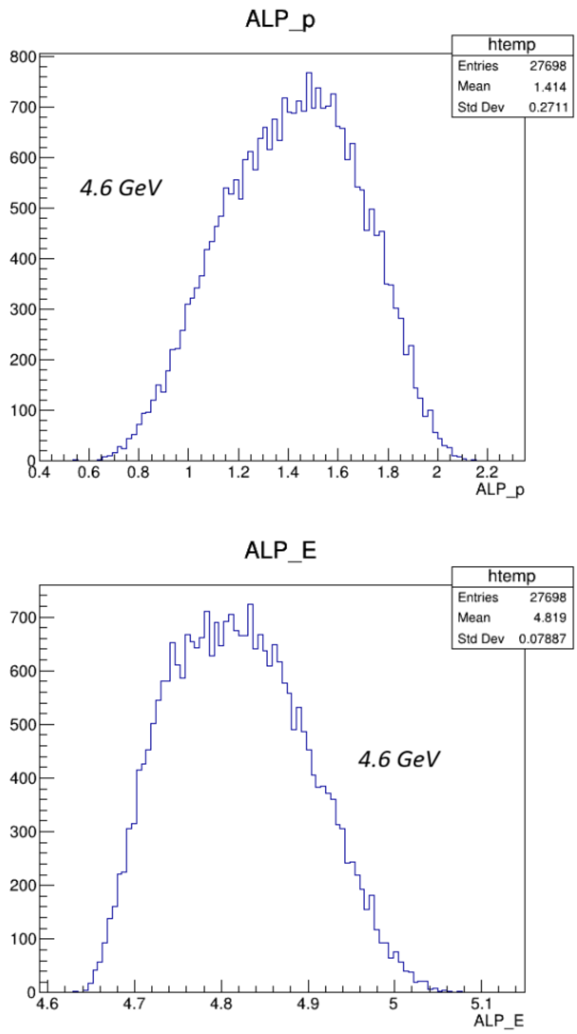
Substituting into the variance expression begin(Equation B.7) leads to strong cancellation:

$$\sigma^2_{m^2} = 4E^2\sigma_E^2 + 4p^2\sigma_p^2 - 8E^2\sigma_E\sigma_p \approx 4E^2\sigma_E^2 + 4E^2\sigma_E^2 - 8E^2\sigma_E^2 = 0 \tag{B.18}$$

This results in a significantly reduced mass variance, explaining the relatively sharp invariant mass distributions observed for low-mass ALPs.



(a) 0.9 GeV(light)



(b) 4.6 GeV(heavy)

Figure B.2: Plots showing the qualitative degree of correlation between momentum and energy of the ALP for the two mass points

B.6.2 Heavy ALP Regime

For heavier ALPs:

$$E \gg p, \tag{B.19}$$

and the correlation term is less effective (in Figure B.2b it is apparent that the distributions are quite different) at canceling the variance contributions. Consequently, the invariant mass resolution worsens, leading to broader distributions.

B.7 Residual Distributions and Error Cancellation

The residuals, defined as:

$$\delta x = x_{\text{reco}} - x_{\text{MC}}, \tag{B.20}$$

show that for low-mass ALPs, energy and momentum errors tend to cancel more effectively (Figure B.3). This implies the measurement errors that propagate in case of light ALPs are much lower than those of massive ALPs.

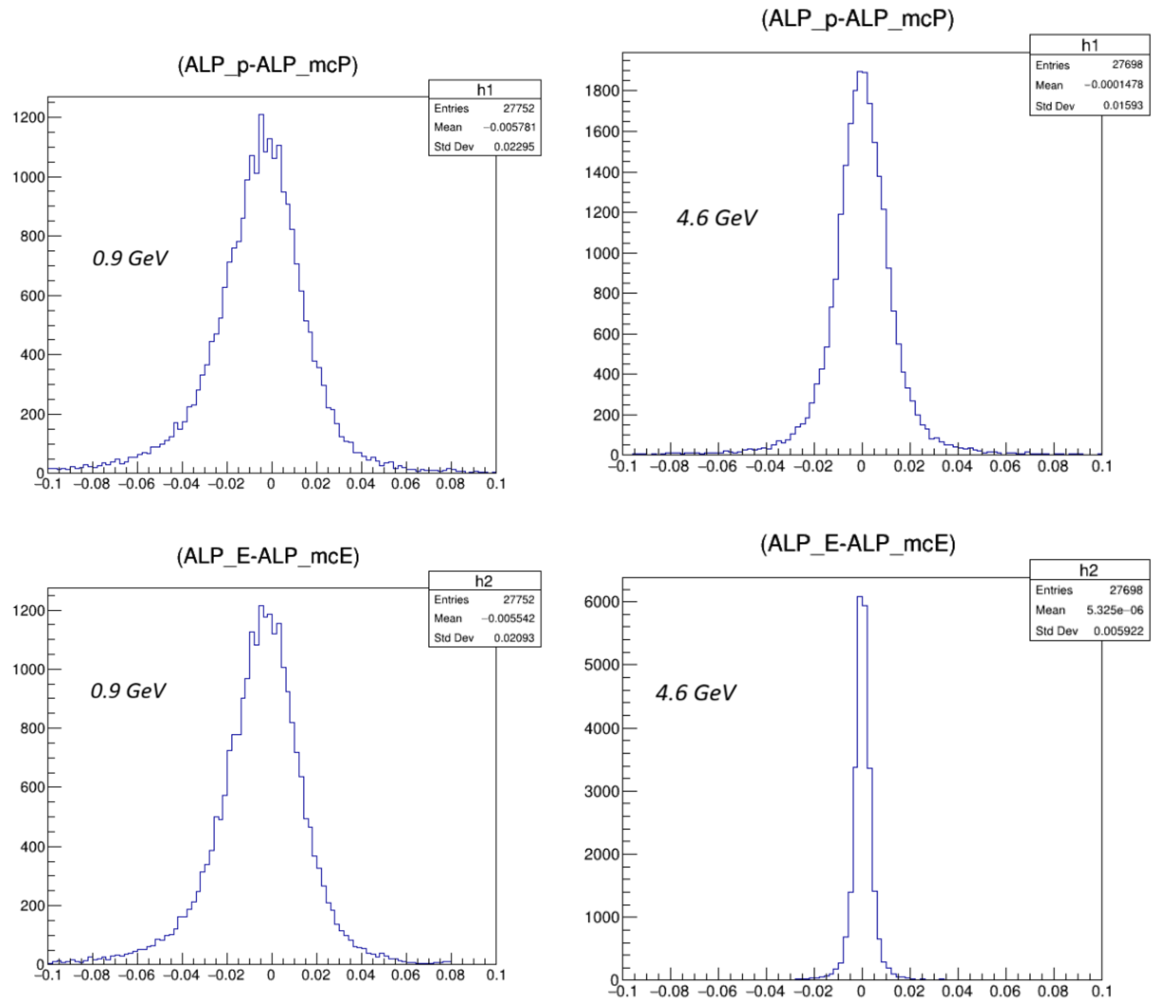


Figure B.3: Comparison of residuals of the momentum and energy for ALPs of mass 0.9 GeV (left) and 4.6 GeV (right)

B.8 Energy Difference ΔE Resolution

The observable ΔE is defined as:

$$\Delta E = E_B - E_{\text{beam}}. \tag{B.21}$$

Since:

$$E_B = E_a + E_K, \tag{B.22}$$

the uncertainty propagates as:

$$\sigma^2_{\Delta E} = \sigma^2_{E_a} + \sigma^2_{E_K} + 2\,\mathrm{Cov}(E_a, E_K). \tag{B.23}$$

Unlike invariant mass, this is a linear combination, and therefore:

- No cancellation mechanism analogous to the $E^2 - p^2$ structure
- Errors add constructively

This explains why ΔE distributions are broader for low-mass ALPs (Figure B.1c,d).

B.9 Detector Resolution in Belle II

The Belle II detector consists of several subdetectors, each contributing differently to measurement precision:

B.9.1 Vertex Detector (VXD)

- Excellent spatial resolution ($\sim 10\ \mu$m)
- Crucial for vertex reconstruction

B.9.2 Central Drift Chamber (CDC)

- Momentum resolution:

$$\frac{\sigma_p}{p} \sim 0.3\% \oplus \frac{0.5\%}{p} \tag{B.24}$$

- Dominant for charged particle momentum measurement

B.9.3 Electromagnetic Calorimeter (ECL)

- Energy resolution:

$$\frac{\sigma_E}{E} \sim \frac{1.6\%}{\sqrt{E(\mathrm{GeV})}} \oplus 0.9\% \tag{B.25}$$

- Critical for photon measurements

B.10 Correlation Between Observables

Correlations arise naturally due to:

- Shared detector components
- Kinematic constraints
- Reconstruction algorithms

In particular, strong correlations between E and p for photons and charged particles significantly impact invariant mass resolution.

B.11 Implications for Analysis

The statistical behavior of reconstructed observables leads to several key insights:

- Invariant mass resolution improves for boosted systems due to covariance cancellation
- ΔE resolution is dominated by additive uncertainties
- Detector resolution directly impacts sensitivity across ALP mass range
- Correlations must be properly accounted for in fitting procedures

Appendix C

Supplementary material for continuum suppression

C.1 ROC curves for characterizing MVA performance

The performance of a multivariate classifier, such as a Boosted Decision Tree (BDT), is commonly evaluated using the Receiver Operating Characteristic (ROC) curve. The ROC curve provides a quantitative measure of the classifier's ability to distinguish between signal and background events as a function of the applied decision threshold.

C.1.1 Definition

The ROC curve is defined as a plot of the True Positive Rate (TPR) versus the False Positive Rate (FPR), where

$$\text{TPR} = \frac{N_{\text{signal, selected}}}{N_{\text{signal, total}}} \equiv \epsilon_S, \tag{C.1}$$

$$\text{FPR} = \frac{N_{\text{background, selected}}}{N_{\text{background, total}}} \equiv \epsilon_B. \tag{C.2}$$

Here, ϵ_S and ϵ_B denote the signal and background efficiencies, respectively. Each point on the ROC curve corresponds to a specific threshold applied to the BDT output.

C.1.2 Theoretical Significance

The ROC curve encapsulates the trade-off between signal efficiency and background rejection. An ideal classifier would achieve $\epsilon_S = 1$ and $\epsilon_B = 0$, corresponding to perfect separation between signal and background. In contrast, a classifier with no discriminating power produces a diagonal ROC curve, equivalent to random classification.

A commonly used scalar metric derived from the ROC curve is the Area Under the Curve (AUC). The AUC provides a global measure of classifier performance:

- AUC $= 1$ corresponds to perfect discrimination,
- AUC $= 0.5$ corresponds to random guessing.

C.1.3 Interpretation in the Analysis

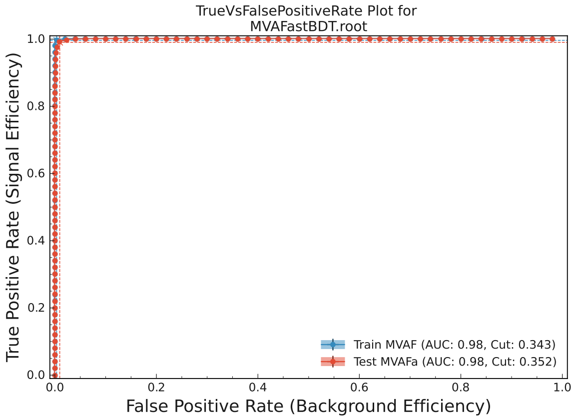
In the context of this analysis, the ROC curve is used to evaluate the performance of the BDT in separating signal events from both continuum and combinatorial background. Each point on the curve corresponds to a particular choice of BDT output cut, which determines the fraction of retained signal and background events.

The selection of an optimal operating point is guided by maximizing a figure of merit (FoM), typically defined as

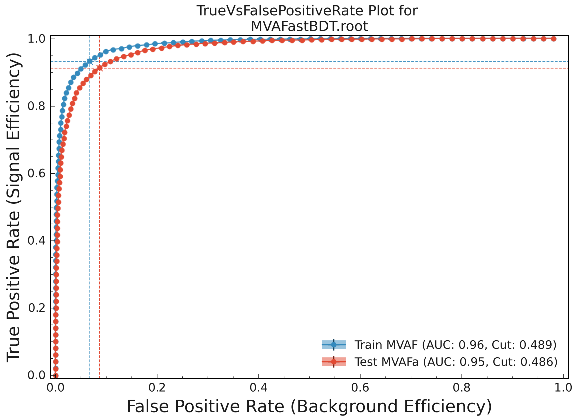
$$\mathrm{FoM} = \frac{S}{\sqrt{S+B}}, \tag{C.3}$$

where S and B represent the expected signal and background yields after the application of the BDT cut.

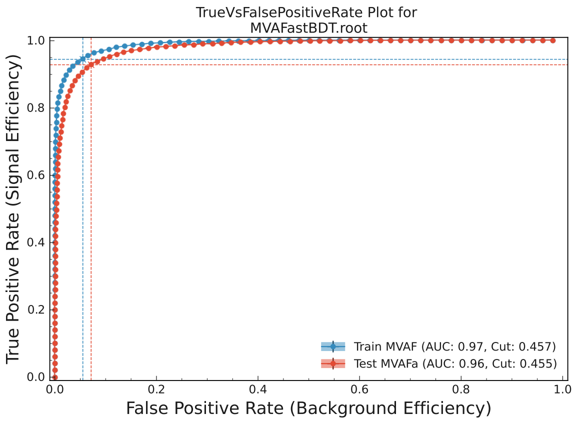
The ROC curve also provides insight into the robustness of the classifier. A stable ROC shape across different mass hypotheses indicates that the BDT maintains consistent discriminating power throughout the phase space of interest. Deviations from this behavior may signal overtraining or dependence on specific kinematic regions.



(a) 0.88GeV, AUC = 0.98



(b) 1.92GeV, AUC = 0.95



(c) 2.96GeV, AUC = 0.96

Figure C.1: ROC curves for the representative mass points in analysis. The high AUC for 0.88 GeV implies it is an almost perfect classifier, as also evidenced by Table 4.2

C.2 Plots and data used for iterative training of MVA for ALP mass 1.92 GeV

Below are the importance plots, correlation matrices and final overtraining plot used to optimize and validate the MVA trained for 2 iterations.

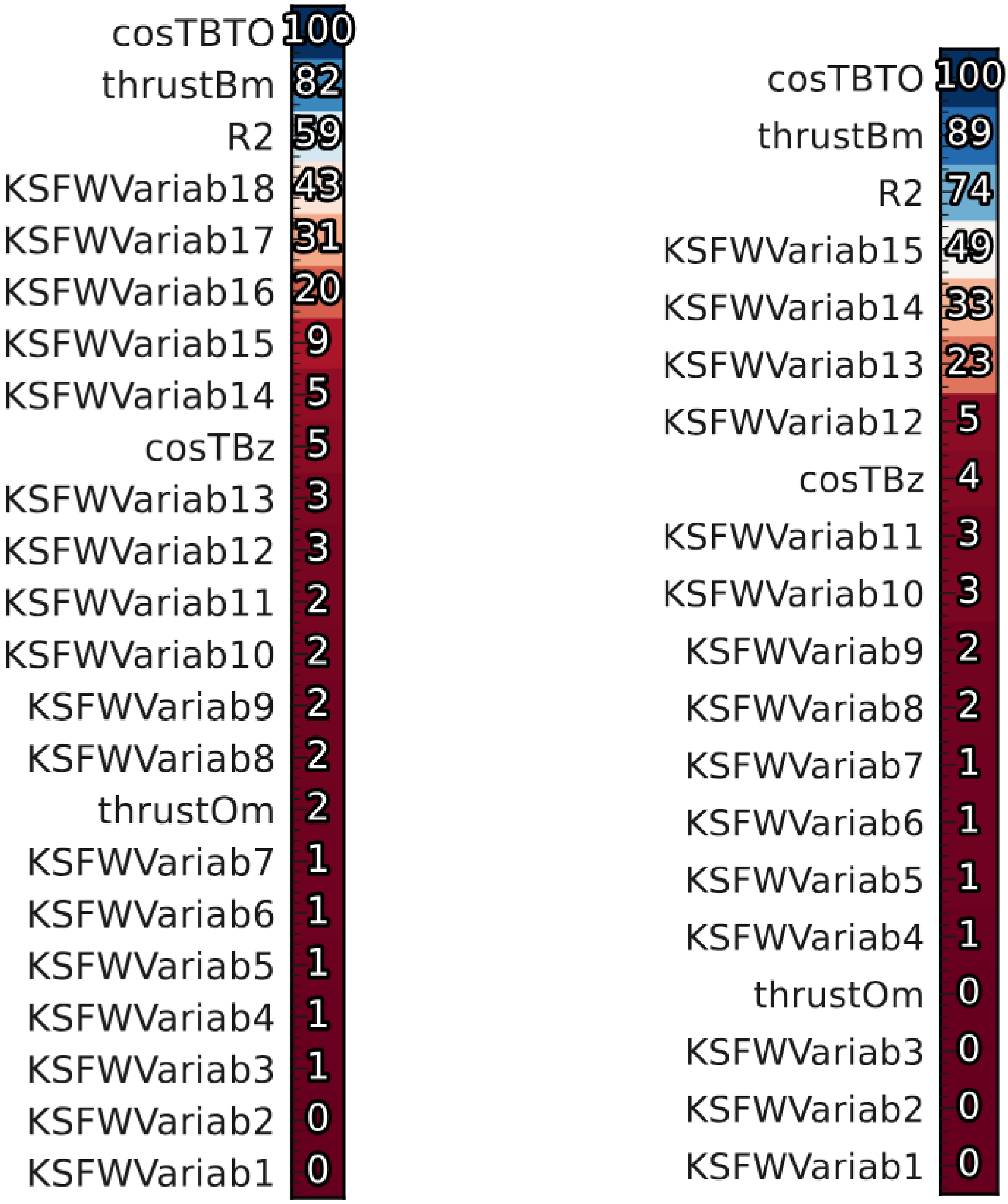


Figure C.2: Importance of continuum suppression varaibles in training MVA for ALP mass 1.92 GeV in iteration 1 (left) and 2 (right)

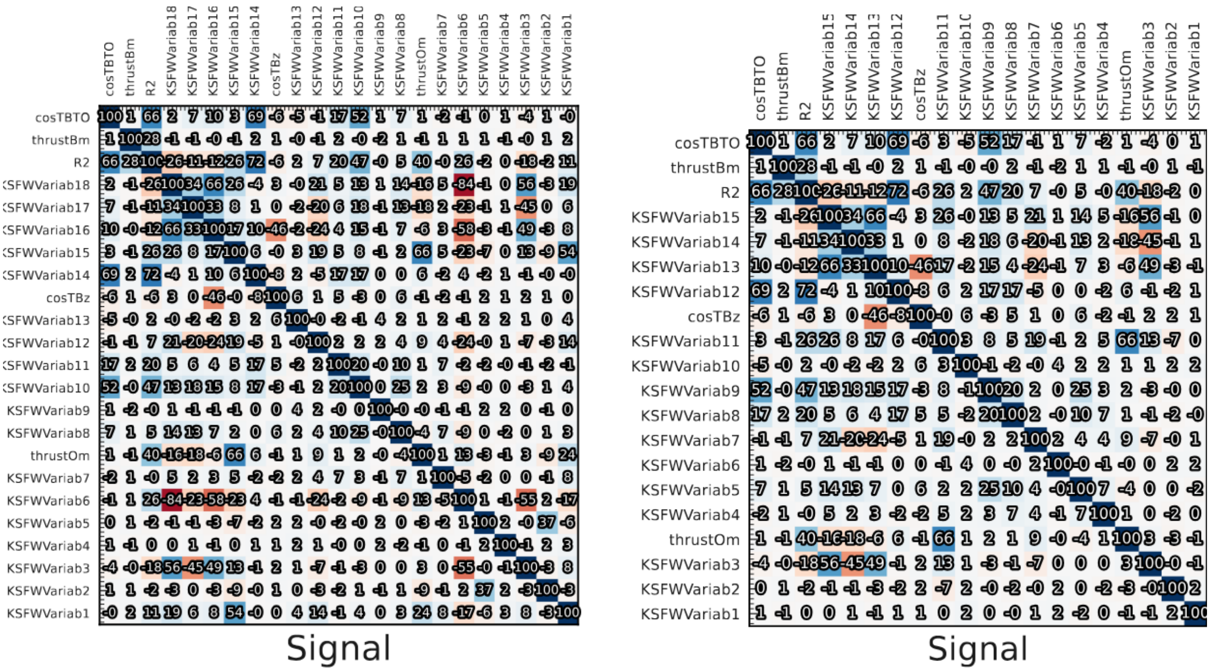


Figure C.3: Correlation between continuum suppression varaibles in training MVA for ALP mass 1.92 GeV in iteration 1 (left) and 2 (right)

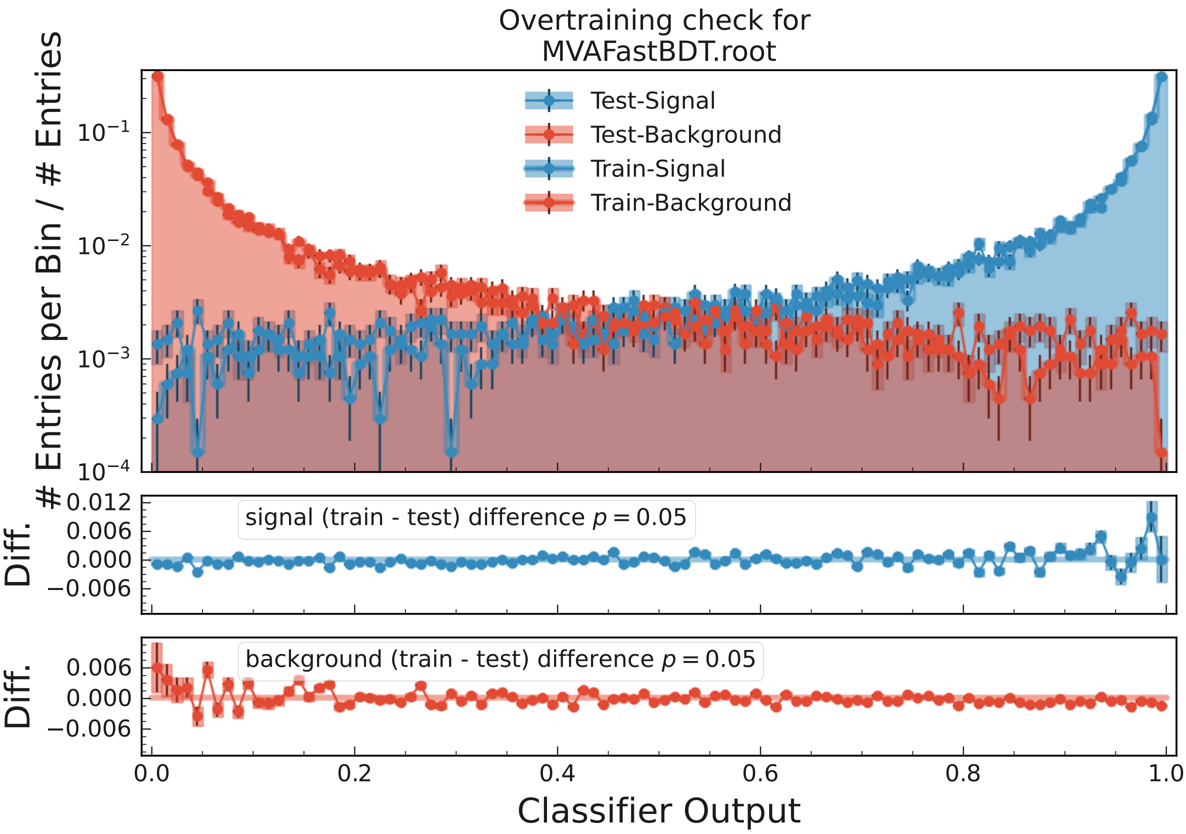


Figure C.4: Overtraining plot for ALP mass 1.92 GeV

C.3 Plots and data used for iterative training of MVA for ALP mass 2.96 GeV

Below are the importance plots, correlation matrices and final overtraining plot used to optimize and validate the MVA trained for 2 iterations.

Figure C.5: Importance of continuum suppression varaibles in training MVA for ALP mass 2.96 GeV in iteration 1 (left) and 2 (right)

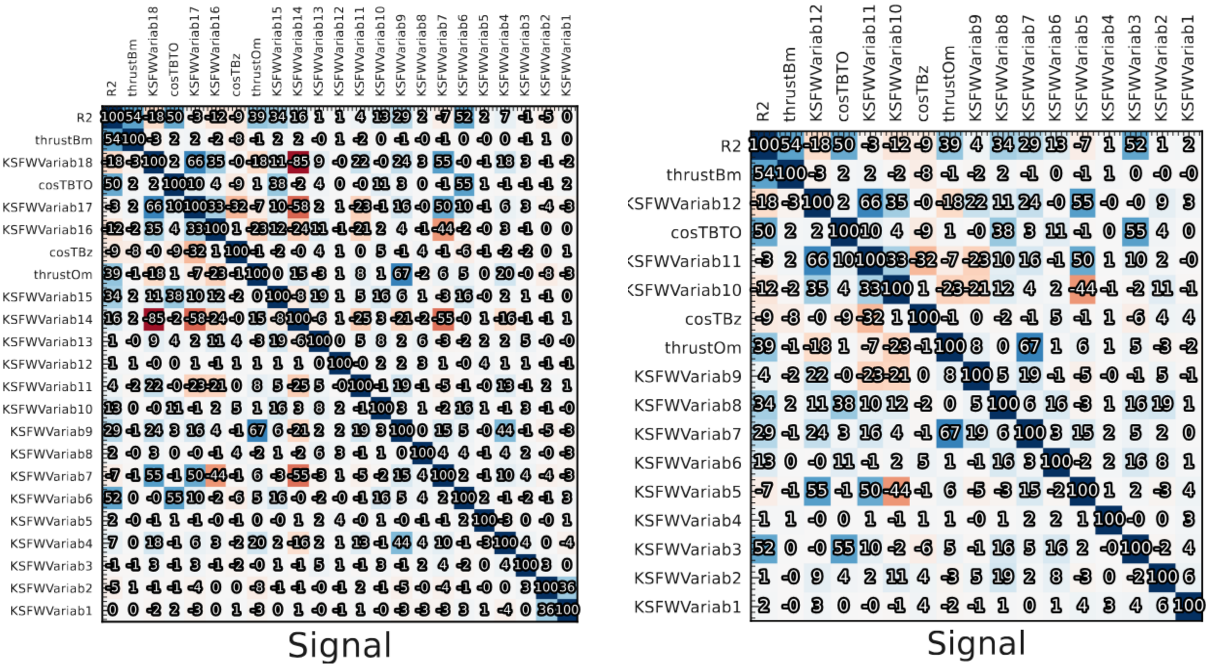


Figure C.6: Correlation between continuum suppression varaibles in training MVA for ALP mass 2.96 GeV in iteration 1 (left) and 2 (right)

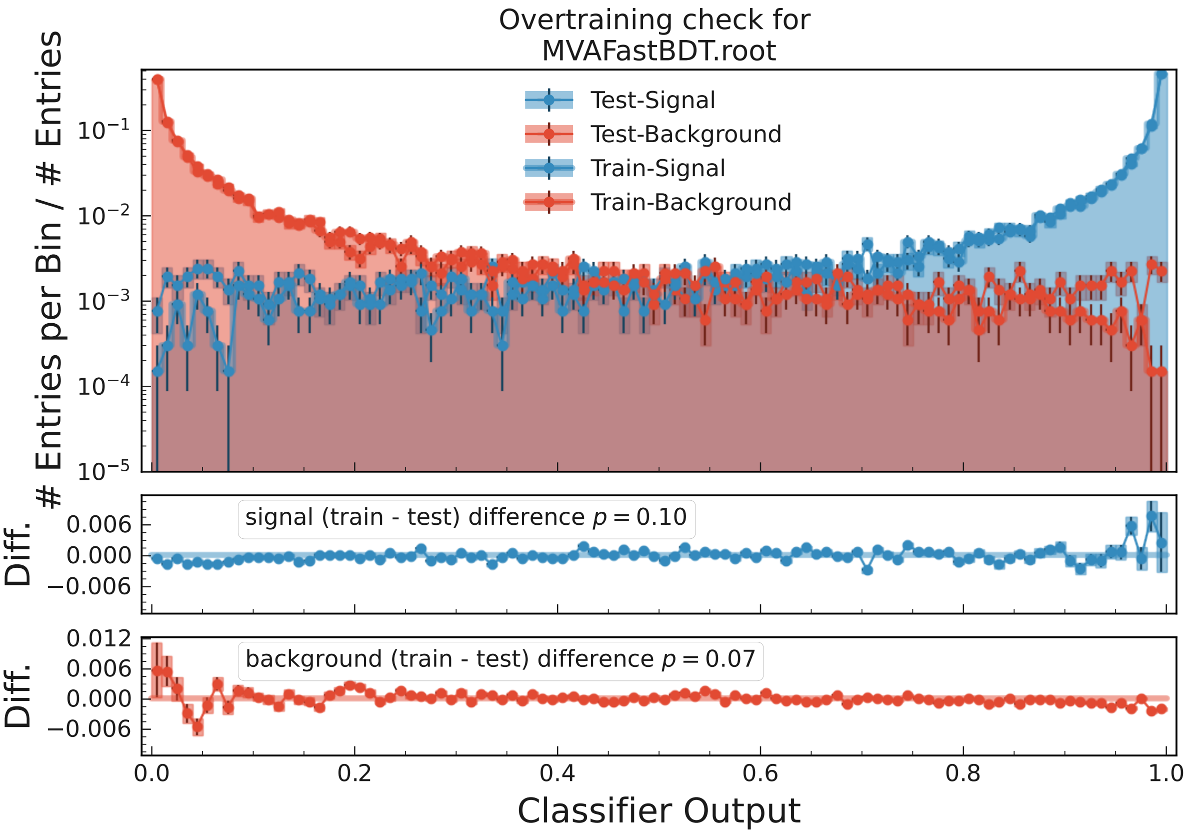


Figure C.7: Overtraining plot for ALP mass 2.96 GeV

Bibliography

[1] T. Abe et al. "The Belle II experiment". In: *Nucl. Instrum. Meth. A* 764 (2014), p. 1.

[2] July 2024. URL: https://www.kek.jp/en/topics/202402291200.

[3] S. Kurokawa et al. "SuperKEKB accelerator". In: *Nucl. Instrum. Meth. A* 764 (2014), p. 1.

[4] Jitendra Kumar. "Charm status and prospects at Belle II". In: Sept. 2021, p. 050. DOI: 10.22323/1.385.0050.

[5] Roberto D. Peccei and Helen R. Quinn. "CP Conservation in the Presence of Instantons". In: *Phys. Rev. Lett.* 38 (1977), p. 1440.

[6] S. Weinberg. "A New Light Boson?" In: *Phys. Rev. Lett.* 40 (1978), p. 223.

[7] F. Wilczek. "Problem of Strong P and T Invariance". In: *Phys. Rev. Lett.* 40 (1978), p. 279.

[8] Sabyasachi Chakraborty et al. "Heavy QCD axion in $b \to s$ transition: Enhanced limits and projections". In: *Phys. Rev. D* 104 (2021), p. 055036. DOI: 10.1103/PhysRevD.104.055036.

[9] Emilie Bertholet et al. "Heavy QCD axion at Belle II: Displaced and prompt signals". In: *Phys. Rev. D* 105 (2022), p. L071701. DOI: 10.1103/PhysRevD.105.L071701.

[10] B Aubert et al. "Study of B meson decays with excited eta and eta' mesons". en. In: *Phys. Rev. Lett.* 101.9 (Aug. 2008), p. 091801.

[11] V. Chobanova et al. "Measurement of branching fractions and CP violation parameters in $B \to \omega K$ decays with first evidence of CP violation in $B^0 \to \omega K_s^0$". In: *Phys. Rev. D* 90 (1 July 2014), p. 012002. DOI: 10.1103/PhysRevD.90.012002. URL: https://link.aps.org/doi/10.1103/PhysRevD.90.012002.

[12] J. E. Kim and G. Carosi. "Axions and the Strong CP Problem". In: *Rev. Mod. Phys.* 82 (2010), p. 557.

[13] Particle Data Group. In: *Prog. Theor. Exp. Phys.* 2020 (2020), p. 083C01.

[14] R. D. Peccei and H. R. Quinn. "Constraints Imposed by CP Conservation". In: *Phys. Rev. D* 16 (1977), p. 1791.

[15] G. Raffelt. *Stars as Laboratories for Fundamental Physics.* 1996.

[16] J. Bjorken. In: *Phys. Rev. D* 38 (1988), p. 3375.

[17] J. Blumlein. In: *Z. Phys. C* 51 (1991), p. 341.

[18] CAST Collaboration. In: *Nature Phys.* 13 (2017), p. 584.

[19] M. Kamionkowski and J. March-Russell. In: *Phys. Lett. B* 282 (1992), p. 137.

[20] R. Holman et al. In: *Phys. Lett. B* 282 (1992), p. 132.

[21] S. M. Barr and D. Seckel. In: *Phys. Rev. D* 46 (1992), p. 539.

[22] S. Ghigna et al. In: *Phys. Lett. B* 283 (1992), p. 278.

[23] E. Bertholet et al. “Heavy QCD Axion at Belle II: Displaced and Prompt Signals”. In: *Phys. Rev. D* 105 (2022), p. L071701. DOI: 10.1103/PhysRevD.105.L071701.

[24] Daniel Aloni, Yotam Soreq, and Mike Williams. “Coupling QCD-Scale Axionlike Particles to Gluons”. In: *Phys. Rev. Lett.* 123 (2019), p. 031803. DOI: 10.1103/PhysRevLett.123.031803.

[25] T. Keck et al. “The Full Event Interpretation: An exclusive tagging algorithm for the Belle II experiment”. In: *Comput. Softw. Big Sci.* 3 (2019), p. 6.

[26] B. P. Roe et al. “Boosted decision trees as an alternative to artificial neural networks for particle identification”. In: *Nucl. Instrum. Meth. A* 543 (2005), p. 577.