DISSERTATION

Search for muonic dark force and performance studies of lepton identification at Belle II

ausgeführt an der
Fakultät für Physik der Technischen Universität Wien

in Zusammenarbeit mit dem
Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften

zur Erlangung des akademischen Grades
Doktor der Technischen Wissenschaften

eingereicht von

Rajesh Kumar Maiti, Msc
Matrikelnummer: 11938276

Betreuung : Privatdoz. Dipl.-Ing. Dr. techn. Christoph Schwanda
Mitbetreuung : Dr. Gianluca Inguglia

Wien, November 29, 2022

Unterschrift StudentIn

Unterschrift BetreuerIn

Abstract

This thesis presents an advanced search strategy to find a new dark gauge boson Z' and performance studies of particle identification to assist the new physics searches, especially involving flavor-changing charge and neutral currents in semi-leptonic B decays. The data used or planned for these measurements are gathered at the Belle II experiment, located at the SuperKEKB asymmetric e^+e^- collider in Tsukuba, Japan. The existence of Z' has been predicted by extending the standard model lagrangian with a new abelian gauge symmetry, which couples mainly to the second and third generation of leptons. The decay of Z' into muons in the event topology $e^+e^- \to \mu^+\mu^- Z'$, $Z' \to \mu^+\mu^-$ has been investigated model dependently in the prompt decays and model independently in the displaced decays for the first time. The Z' mass phase space 0.212 – 9 GeV has been explored in both cases. A 90% CL on the coupling constant g' has been determined with a targeted luminosity of 178.47 fb^{-1} for the prompt case. These bounds are similar to the BaBar measurement except for the low mass regions, with a $\sim$3 times less dataset. A 90% CL on the production cross-section for the displaced search is also provided at a level of below 0.1 fb for different lifetime scenarios under consideration with a targeted luminosity of 200 fb^{-1}. The expected results are based on simulation only. This thesis also presents an advanced particle identification (PID) method using different machine learning techniques and comparative studies between them. The developed algorithm provided overwhelming improvements in PID, especially for the low momentum particle tracks where backgrounds are more severe. Finally, some trigger studies related to dark sector searches are also provided.

Symbols

e	Electron
μ	Muon
τ	Tau
ν_e	Electron neutrino
ν_μ	Muon neutrino
ν_τ	Tau neutrino
u	Up quark
d	Down quark
c	Charm quark
s	Strange quark
t	Top quark
b	Bottom quark
γ	Phonton
π	Pion
K	Kaon
$\Upsilon(4S)$	$\Upsilon(4S)$ meson
B	B meson
D	D meson
J/ψ	$\psi(1S)$ meson
ϕ	ϕ meson
ϕ_H	Higgs field
g'	Coupling constant
$W^\pm$	W boson
Z^0	Z boson
l	lepton
E_{beam}	Beam Energy in CM frame
Γ	Decay width
$\mathcal{L}$	Lagrangian
A'	Dark photon
h'	Dark Higgs
ϕ_L	Leptophilic dark scalar
Z'	Z' boson

Acyonyms

SM	Standard Model
WIMP	Weakly Interacting Massive Particles
CMB	Cosmic Microwave Background
DM	Dark Matter
MC	Monte Carlo
BASF2	Belle II Analysis Software Framework
CMS	Centre of Mass System
PID	Particle Identification
PDF	Probability Distribution Function
PXD	Pixel Detector
SVD	Silicon Vertex Detector
CDC	Central Drift Chamber
TOP	Time of Propagation Detector
ARICH	Aerogel Ring Imaging Cherenkov Detector
ECL	Electromagnetic Calorimeter
KLM	K_L^0 and μ Detector
ADC	Analog to Digital Converter
MSPS	Million Samples Per Second
MVA	MultiVariate Analysis
NN	Neural Network
MLP	MultiLayer Perceptron
BDT	Boosted Decision Tree
ISR	Initial State Radiation
FSR	Final State Radiation
NLL	Negative Log Likelihood
CL	Confidence Level
UL	Upper Limit
f.o.m/FOM	Figure of merit
TSIM	Trigger Simulation
ROC	Receiver Operating Characteristic
LEE	Look-Elsewhere Effect
InvM	Invariant Mass
1DUML	One dimensional unbinned maximum likelihood
CB	Crystal Ball

Contents

Introduction

Particle physics is the study of elementary particles and forces. The standard model backbone of particle physics is the most precise and accurate mathematical theory describing particles and their interaction at the quantum level. Despite its tremendous success, it's also known to be an incomplete theory. It could not explain the matter-antimatter imbalance; when the universe was formed in the big bang, matter and antimatter should have been created in equal amounts, but today we observe a matter dominated universe. In the standard model, neutrinos are massless. However, they exhibit oscillation between other neutrino states, which suggests they should have mass. Neutrino mass problem is unexplainable by the standard model. The three generations of leptons are identical except for their masses; in particular, electroweak gauge bosons couple similarly to the three leptons generations. This very peculiar symmetry of the standard model is known as Lepton Universality, and the standard model couldn't answer it. In addition, the standard model also doesn't give a unified description of all the known forces. One of the most open issues is dark matter; the standard model doesn't provide a dark matter candidate. Dark matter has been observed through different cosmological measurements but yet to be detected directly. Various experiments in the last few decades put their efforts to detect it experimentally, but no conclusive evidence yet.

The Belle II experiment, located at the SuperKEKB asymmetric e^+e^- accelerator in Tsukuba, Japan, provides the unique opportunity to search dark sector and standard model mediator particles at the intensity frontier. If dark matter particles exist, they could be producible through the decay of standard model particles and the mediators. Different Dark sector and standard model mediator particles have been predicted by extending the standard model lagrangian, and they are light enough to be producible at the e^+e^- collider. The e^+e^- collider's clean environment, well-defined initial states, and high luminosity are best suited for this kind of searches. Dark matter particles don't interact with the detector, and their presence could be verified by the huge missing energy and momentum they carry away. The predicted mediator particles could decay invisibly and visibly, according to the specific models. The invisible decays are very sensitive to direct dark matter production; however, searches involving visible decay are also very important because if invisible decay hints at some new particle, it should also be verified by the visible decay. To perform dark sector searches, one needs

to have excellent particle identification and good triggers system to enhance the search efficiency.

My thesis is structured as follows; I divided the whole thesis into three parts. The first part briefly introduces the standard model and dark matter, followed by the theoretical and experimental motivations behind the proposed dark sector mediator search. It also briefly describes the SuperKEKB and Belle II experiment, which provides an ideal clean environment for this search. The second part summarizes the performance studies of particle identification with different advanced machine learning based methods and comparative studies between them. This part also describes some trigger studies related to the new physics searches. Finally, in the third part, I elaborated on all the techniques and tools developed to perform the proposed dark sector mediator search. For the second and third parts, each chapter provides an individual introduction, conclusion, and future outlook. Finally, in the end, a final conclusion is provided.

Part I.

Preliminaries

1. Theory

1.1. Standard Model

The Standard Model (SM), a rigorous mathematical description of elementary particles and their interaction, is quite successful in describing natural phenomenon at the subatomic level. Different precision measurements performed at the colliders and other experiments established SM as the most precise physics theory at the quantum level known to date. The SM includes several elementary particle classes, categorized into fermions and bosons.

Fermions are spin-1/2 particles and obey Fermi-Dirac statistics. They are further classified into quarks and leptons depending on their interaction. There are six types of quarks (u, d, s, c, t, b) and leptons (e^-, μ^-, τ^-, ν_e, ν_μ, ν_τ) further divided into pairs that exhibit similar physical behavior called generations (see Table 1.1). Quarks carry color charge, hence interacting via the strong interaction. The phenomenon called color confinement helps quarks to bind very strongly with each other and form color-neutral composite particles called hadrons. Hadrons are made of either one quark and one antiquark called mesons or three quarks called baryons. Quarks also carry an electric charge and weak isospin; hence they experience electromagnetic and weak interaction. Leptons don't carry a color charge. Apart from neutrinos, other leptons participate in electromagnetic and weak interaction, while neutrinos exhibit only weak interaction, which makes them very hard to detect experimentally in contrast to the other leptons. Apart from neutrinos, each generation member has a higher mass than the preceding generation. The first-generation charged particles are stable; hence all ordinary matter is made of such particles. Specifically, atoms consist of electrons orbiting around nuclei which consist of up and down quarks. On the other hand, other generation's charge particles decay to the lower generation and are observed only in very high-energy environments like colliders or in cosmic events. Neutrinos of different generations also do not decay. They are considered massless. However, they exhibit oscillation between different flavors which suggests they could have mass, but experimentally not measured yet but will be in the order of a few eV.

Bosons are integer spin particles and obey Bose-Einstein statistics. All the SM gauge bosons are the force carriers and have spin-1 that mediates strong, weak, and electromagnetic interactions between elementary particles. Photons

(γ) mediate electromagnetic force and are massless. They are well described by the relativistic quantum field theory of electrodynamics called quantum electrodynamics (QED). QED is the abelian gauge theory of the $U(1)$ symmetry group. The $U(1)$ gauge field, which mediates the interaction between charged fermionic fields, is called the electromagnetic field, and their force carriers are called photons. The $W^{\pm}$ and Z^0 gauge bosons mediate the weak force, and their theory is well

	Quarks		Leptons	
Generations	Q = $-\frac{1}{3}$	Q = $-\frac{2}{3}$	Q=-1	Q=0
First	down (d) $\sim$ 5 MeV	up (u) $\sim$ 2.5 MeV	electron (e) $\sim$ 0.511 MeV	e neutrino (ν_e) $<$ 1 eV
Second	strange (s) $\sim$ 101 MeV	charm (c) $\sim$ 1270 MeV	muon (μ) $\sim$ 105.7 MeV	μ neutrino (ν_μ) $<$ 1 eV
Third	bottom (b) $\sim$ 4200 MeV	top (t) $\sim$ 172 GeV	tau (τ) $\sim$ 1777 MeV	τ neutrino (ν_τ) $<$ 1 eV

Table 1.1.: The quark and lepton families with their masses and charges Q. The corresponding anti-particles have the same masses as the particles but the opposite charges.

described by the $SU(2)$ gauge field. They are massive, which makes the interaction strength of this force very weak (see Table 1.2). However, mathematically, electromagnetic and weak interactions are unified as a Yang-Mills field with an $SU(2) \times U(1)$ gauge group. The gauge bosons of the $SU(2) \times U(1)$ Yang-Mills field mediate the electroweak interaction, and the symmetry of this field is known as electroweak symmetry. Initially, all the gauges bosons of this field are massless, and one couldn't directly put the mass term in the lagrangian; it would violate the gauge symmetry. In the SM, $W^{\pm}$, Z^0, and the photon are produced through the spontaneous symmetry breaking of electroweak symmetry $SU(2)_L \times U(1)_Y$ (where Y is the weak hypercharge and L indicates coupling to left-handed fermions only) via Higgs mechanism. On the other hand, gluons mediate the strong interactions between the quarks. They are massless, and their theory is described by quantum chromodynamics (QCD), a non-abelian gauge theory of the $SU(3)$ symmetry group. There have been many proposed theoretical models for the unification of

strong and electroweak forces, but none of them experimentally verified yet. Table 1.2 summarizes all the known forces, their relative strengths, and corresponding mediators. The other gauge boson of the SM is the Higgs boson, which provides mass to the particles and is the only fundamental scalar particle discovered to date.

So, in short, including all elementary particles, SM is a quantum gauge theory of symmetry group $SU(3) \times SU(2)_L \times U(1)_Y$ describing the interaction between them.

Fundamental Forces	Strength	Range (m)	Mediators
Strong	1	10^{-15}	gluons
Electromagnetic	$\frac{1}{137}$	Infinite	photon (γ)
Weak	10^{-6}	10^{-18}	$W^{\pm}$, Z^0
Gravity	6×10^{-39}	Infinite	graviton

Table 1.2.: Different fundamental forces and their strengths, ranges and mediators.

1.2. Introduction to dark matter

The SM of particle physics successfully described the interactions of elementary particles in the energy regime accessible by the high energy physics experiments that extend up to a few TeV; despite its success, there are still many unanswered questions. One of the open issues is dark matter. Many astrophysical observations are consistent with the existence of dark matter, which interacts mostly gravitationally with ordinary matter and is completely blind to the strong and electroweak interactions, hence dark.

1.2.1. Technical definition

Einstein's General Theory of Relativity (GTR) provides a mathematical description of gravitation in terms of the interaction between mass, space, and time. Applying GTR to a homogenous, isotropic, and expanding universe leads to the Friedmann equation,

$$H^2 = (\frac{\dot{a}}{a})^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{1.1}$$

Where $H \equiv \frac{\dot{a}}{a}$ is the Hubble parameter (a is the scale factor function of time in the Friedmann–Robertson–Walker metric), ρ is the observed density, k is the curvature of space-time, Λ is the cosmological constant. Parameterizing the Friedman equation in terms of critical density $\rho_c = \frac{3H^2}{8\pi G}$ (the average density of matter required for the universe to halt its expansion, which is ≈ 5 hydrogen-atoms per m^3) one will obtain $\rho = \rho_c - \frac{\Lambda}{8\pi G}$ or $\Omega = 1 - \Omega_\Lambda$, where $\Omega = \frac{\rho}{\rho_c}$ is energy density composed of radiation, matter and curvature while $\Omega_\Lambda = \frac{\Lambda}{8\pi G \rho_c}$ is the energy density coming from cosmological constant. so, in terms of equation,

$$\Omega_r + \Omega_m + \Omega_k + \Omega_\Lambda = 1. \tag{1.2}$$

Now solving the Friedmann equation in an assumption of perfect fluid, one will conclude $\Omega_r \propto a^{-4}$, $\Omega_m \propto a^{-3}$, Ω_Λ is independent of a and $\Omega_k = 0$ due to flat universe.

The total energy budget of the universe comes from the Cosmic Microwave Background (CMB) measurements performed by the PLANCK experiment; they reported $\sim 32\%$ energy density due to matter, which is composed of 4.6% ordinary baryonic matter (or luminous matter) and 27% of unknown "dark matter", while the remaining 68% of the energy density comes from the unknown vacuum energy called "dark energy". So, in general, all components of the universe (apart from missing baryonic matter) which are not visible but still obey the energy density $\propto a^{-3}$ are called "dark matter". Dark matter is necessary to explain the evolution of the universe. Below I briefly describe the different observational evidence and ongoing searches by various experiments to find the dark matter content of the universe.

1.2.2. Observational evidence: why dark matter is needed?

The First Indication of the existence of some missing matter in the galaxy clusters comes from the measurements of Fritz Zwicky in 1930. He studied the Coma Cluster, estimated its mass based on the motions of the galaxies, and compared it to an estimation based on the brightness and number of galaxies. He found that the cluster has about 400 times more mass than actually visible. The gravitational pull of the visible galaxies was far too small for such orbits to hold them, so some missing mass must be there and hidden from view. Based on these, Zwicky concluded some unseen matter provided the extra gravitation pull to hold the cluster together and called them "dark matter."

Further indications of mass-to-light ratio discrepancies come from the measurements of galaxy rotation curves. The arms of a spiral galaxy rotate around their galactic center, and the luminous mass density of the spiral galaxies decreases from the center to the outskirts. If it is made of luminous matter only, one could consider the center as a mass point and test masses orbiting around it. Now applying

Kepler's Law, it is expected that rotation velocities will decrease with distance from the center, similar to our solar system, and it will follow a dependency of $v(r) \propto \sqrt{\frac{1}{r}}$. But this is not observed; instead, it remains flat as distance increases. The velocity distribution cannot be explained with potential from the visible matter only, an additional (non-visible) halo is necessary to explain the total velocity distribution curve. Figure 1.1 shows the galaxy rotation curve for the galaxy NGC 3198 with an extra dark matter halo of mass density $\rho = \frac{\rho_0}{(\frac{r}{R})^\gamma[1+(\frac{r}{R})^\alpha]^{\beta-\gamma/\alpha}}$, where $\alpha = 1.0$, $\beta = 3.0$, $\gamma = 1.0$, R = 20 kpc [1].

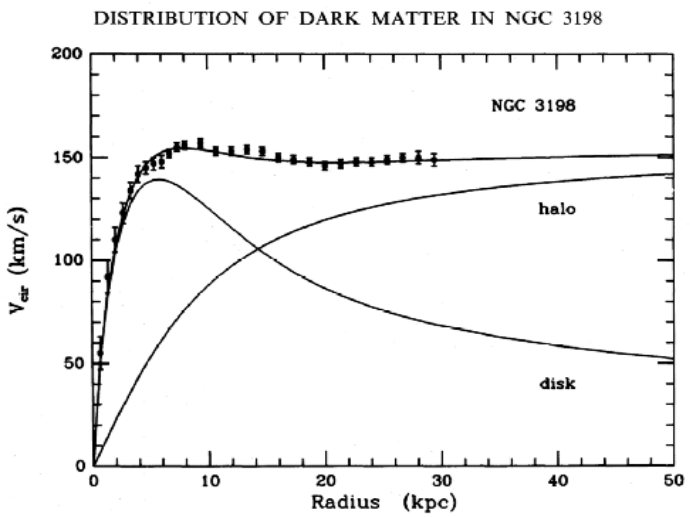


Figure 1.1.: The observed galaxy rotation curves for the stars in the disk of the NGC 3198 galaxy [1]

The other striking evidence of existing dark matter come from gravitational lensing and CMB measurements. A gravitational lens is a massive object or mass distribution between a distant light source and an observer, which causes a bending of light from the distant source. Depending on the light bending, one can measure the mass of the intervening objects. While this has been measured for different galaxy clusters and the mass-to-light ratio is estimated, it is found to be consistent with the dark matter prediction. dark matter doesn't bend the light itself, but it creates additional gravitational pulls and distorts the spacetime, and this information is embedded in the light bending. One can also deduce the map of dark matter distribution by analyzing different lensed images.

CMB is the remnant of electromagnetic radiation from the early stage of the universe embedded all over space. Ordinary matter and dark matter both are matter, although they act very differently. dark matter doesn't interact with radi-

ation directly, but it affects the CMB with its gravitational potential by affecting the density and velocity of ordinary matter. The temperature of the CMB is very close to the perfect black body spectrum, around $\sim$ 2.7 K; however, it contains tiny temperature fluctuations in order of 10^{-5}. While this anisotropy is expressed in terms of the angular power spectrum, a series of acoustic peaks are observed having different heights sensitive to crucial information about the early universe. The Figure 1.2 shows the temperature anisotropy as a function of multipole l, where $\theta = \frac{180^o}{l}$, measured by PLANK collaboration [2]. The first peak is sensitive to the universe's curvature, the second peak corresponds to the universe's baryonic content, and the third peak relates mainly to dark matter density. The observed CMB angular power spectrum provides good evidence of supporting dark matter, as the Λ-cold dark matter model (a model which take in to account normal baryonic matter, dark matter and dark energy to explain the evolution and structure of the universe) fits very well its precise structure.

Some additional non-luminous matter or dark matter is also needed to explain the structure formation of the universe. Dark matter must need to be cold, i.e. its speed significantly slower than the speed of light, to form the structure of the universe.

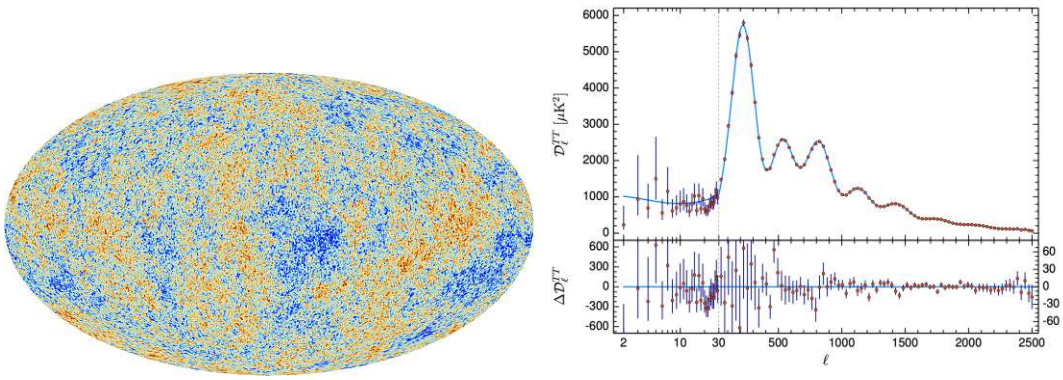


Figure 1.2.: Temperature Anisotropy as a function of multipole l [2].

1.2.3. Particle character of dark matter

Dark matter is observed at different length scales, from galaxies to cosmological scales. All measurements are based on gravitational pull only. The particle character of dark matter could be the best solution. The SM does not offer a dark matter candidate; neutrino is the only non-electromagnetically interacting within SM, but it is "hot", which means it travels with a speed of light inconsistent with

the dark matter requirement on structure formation.

WIMPs

One of the most promising candidates for dark matter is thermally produced Weakly Interacting Massive Particles (WIMPs). The motivation of WIMPs as a dark matter candidate comes from the fact that the obtained relic density from the thermal freeze-out process with electroweak mass and cross-sections are consistent with the observed dark matter relic density. The assumptions are WIMPs interact with the SM particles very weekly, but still sizeable enough to keep them in thermal equilibrium in the early universe, and they are heavier than the SM particles. The thermal equilibrium will be disrupted once the temperature of the thermal bath approaches the WIMP's mass (as the universe is cooling down). In this scenario, the SM particles will not have enough energy to produce the WIMPs, which then annihilate until the universe's expansion decouples them; this is the so-called freeze-out mechanism [3]. This process is depicted in terms of the Boltzmann equation,

$$\frac{dn_\chi}{dt} = -3Hn_\chi - \langle\sigma v\rangle\,(n_\chi^2 - n_{eq}^2) \tag{1.3}$$

where, n_χ is the WIMP density, $\langle\sigma v\rangle$ is the thermally averaged annihilation cross-section, and n_{eq} is density of SM particles in equilibrium. Further rewriting the Boltzmann equation in terms of Y = n/s, where s is entropy and n could be n_{eq} or n_χ, and solving it numerically, one will obtain the cross-section required to match the observed dark matter density leads to a cross-section similar to a cross-section known from the weak interaction, which is in the order of 1 pb ($\sim 10^{-36}$ cm^2) and WIMP mass range is in order of GeV. This coincidence is known as the famous "WIMP miracle" [4].

Figure 1.3 shows the freeze-out mechanism of WIMP mass of 100 GeV and $\langle\sigma v\rangle$ of 3×10^{-26} cm^3 s^{-1} (electroweak scale), the solid black line shows the observed exact relic dark matter density of the universe. The $\langle\sigma v\rangle$ depends on the interaction strength between WIMPs and SM particles. Depending on the different interaction strengths, one can obtain different color bands, as shown in Figure 1.3. Bands below the solid black line imply higher interaction strengths corresponding to low relic dark matter density and vice versa for the bands above the black line.

WIMP miracle predicts the mass of dark matter WIMP particles between a few GeV to TeV range. B. Lee and S. Weinberg calculated the minimum mass of a WIMP to be a 2 GeV, assuming the annihilation cross section of weak interaction [6]. The Lee-Weinberg bound could be avoided if the mediator responsible for WIMP annihilation is something other than the Standard Model gauge bosons

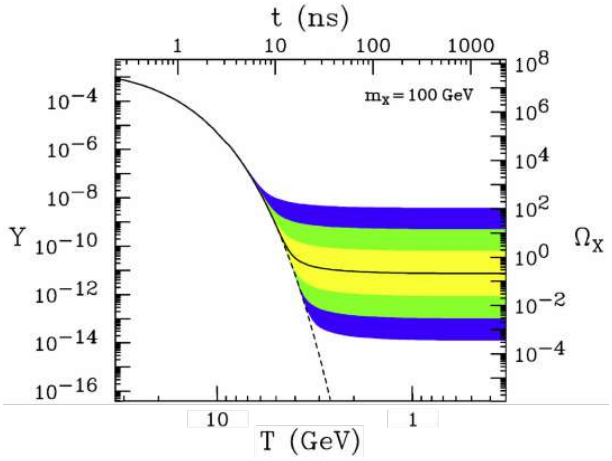


Figure 1.3.: Freeze-out mechanism of WIMPs [5].

(Z and Higgs). In that case, dark matter mass in a range of a few keV is also possible. Dark matter characterized by keV-GeV mass range is called light-dark matter. In the last decades, various intensity and energy frontier experiments put their efforts into finding the WIMPs in GeV-TeV mass range but no interesting signals yet. Strong experimental constraint comes from the measurement of the XENON1T experiment, which almost ruled out Z and Higgs as the mediator between WIMPs and SM particles [7] [8].

Standard Model extensions

As described above, weak interaction gauge bosons as a mediator between WIMPs and SM particles are almost ruled out, and the results from different experiments impose a strong constraint on WIMP mass; mass between 6 GeV to TeV range is almost ruled out [8]. So, there could be some other new force or new kind of interaction mediating between dark matter and SM particles, assuming that dark matter is part of a wider dark sector that includes self-mediators. Various SM extensions with well-motivated theories introduce dark matter weakly coupled with the SM through different possible dark sector mediators without violating the gauge symmetries. Since the thesis focuses on the search for a light-dark sector mediator produced at the Belle II experiment, only SM extensions sensitive to the light-dark sector will be discussed here. Below I discuss different portals through which SM and dark sector could interact with each other:

In **the vector portal**, the SM Lagrangian is extended by introducing a new abelian $U(1)$ gauge symmetry ($U'(1)$). The interaction Lagrangian between the

SM and dark sector is described by, $\mathcal{L} \supset -\frac{\varepsilon}{2} B_{\mu\nu} F'^{\mu\nu}$, where $F'^{\mu\nu}$ is the new tensor field from $U'(1)$ gauge symmetry, $B_{\mu\nu}$ is the tensor field of SM $U(1)$ gauge group, and ε is the kinetic mixing parameter. The symmetry breaking of $U'(1)$ gauge group introduces a new force carrier, having spin-1, similar to the photon of electromagnetism but potentially connected to the dark sector, is called dark photon (A'). Unlike photons, dark photons could have mass due to the spontaneous symmetry breaking of the dark sector Higgs field [9] or due to the Stueckelberg mechanism [10]. The dark photon could mix with the SM photon via a kinetic mixing mechanism. Depending on its mass, it could eventually decay into SM particles or dark matter particles if kinematically accessible.

The scalar portal assumes the existence a new spin 0 boson S, which interacts with SM Higgs field (ϕ_H) as described by the Lagrangian $\mathcal{L} \supset (\mu S + \lambda S^2)\phi_H^\dagger \phi_H$, where μ and λ are appropriate parameters. Through this portal dark sector particles can interact with the SM particles.

The combined charge conjugation and parity (CP) symmetry has been violated in the weak interaction, but strangely strong interaction conserves it. The CP-violating phase appears in the strong interaction through the SM, but the theory does not predict its value. If there has a large CP violation in the strong interaction, it will induce a large electric dipole moment (EDM) to the neutron, but experimental results suggest no evidence of the neutron's EDM that means CP violation at strong interaction is largely suppressed, somehow strong interaction conserve it, this is known as strong CP problem. Peccei–Quinn mechanism provides an elegant solution to the strong CP problem by adding a new global symmetry to the Lagrangian, which compensates the CP violating term in the strong interaction. This new global symmetry is spontaneously broken, and results in a new particle called "Axion" [11]. If "Axion" exits and have low mass in a range μeV- eV could be a potential dark mater candidate. The interaction Lagrangian between SM and Axion field is given by, $\mathcal{L} \supset \frac{a}{f_A} F_{\mu\nu} \tilde{F^{\mu\nu}}$, where $F_{\mu\nu}$ ($\tilde{F^{\mu\nu}}$) is the (dual) field-strength tensor of the SM photon field, and a is the pseudo-scalar "Axion" field couples to the SM model through the dimensional decay constant f_A. This portal known as **the pseudo-scalar portal** through which dark sector particle could interact the SM particles.

The other interesting portal is **the neutrino portal**. The SM neutrinos are left-handed, considered massless (tiny mass in $\sim$ eV), and participate only in the weak interaction. However, they are ruled out as dark matter candidates because they can't form the currently observable universe's structure. The "Type 1 see-saw" mechanism extends the SM and predicts right-handed neutrino fields, which

are inert to the electroweak interaction and have a large mass scale. This model provides a light neutrino and a very heavy neutrino called the sterile neutrino for each of the three known neutrino flavors. The sterile neutrino could act as a potential dark matter candidate, and its mass has to be in the keV region to form the current structure of the universe. The interaction Lagrangian between SM and right-handed sterile neutrinos is given by $\mathcal{L} \supset y_n L \phi_H N$, where N is the fermionic field associated with the right-handed sterile neutrino, L is the $SU(2)_L$ leptonic doublet, ϕ_H is the Higgs field, and y_n is Yukawa coupling.

The $L_\mu - L_\tau$ model

In the framework of the vector portal extension ($U'(1)$), there also exists a new spin-1 massive gauge boson Z'. Search for Z' is particularly interested in the context of a theoretically well motivated model called $L_\mu - L_\tau$ [12] [13]. In this proposed model, Z' mainly couples to second and third generation leptons through the Lagrangian $\mathcal{L} = \sum_l \theta g' \bar{l} \gamma^\mu Z'_\mu l$, where $\theta = +1$ if $l = \mu, \nu_\mu$, $\theta = -1$ if $l = \tau, \nu_\tau$, and $g' \sim 10^{-6} - 10^{-2}$. Such a Z' generally doesn't couple to the first generation of leptons and have mass typically in $\mathcal{O}$(MeV)$-\mathcal{O}$(GeV), which could be easily produced in e^+e^- colliders. The partial decay widths of the decay $Z' \to l^+l^-$ and $Z' \to \nu\bar{\nu}$ is given by [14],

$$\begin{aligned} \Gamma_{Z' \to l^+l^-} &= \frac{(g')^2 M_{Z'}}{12\pi} \left(1 + \frac{2m_l^2}{M_{Z'}^2}\right) \sqrt{1 - \frac{4m_l^2}{M_{Z'}^2}} \\ \Gamma_{Z' \to \nu\bar{\nu}} &= \frac{(g')^2 M_{Z'}}{24\pi} \end{aligned} \tag{1.4}$$

and the corresponding branching fractions are given in Figure 1.4. The branching fraction for one neutrino species is half of the branching ratio to one charged lepton flavor because Z' couples only to left-handed neutrino chiralities, whereas it couples to both left and right handed charged leptons. This postulated Z' boson in the context of anomaly free $L_\mu - L_\tau$ model could be a potential dark matter candidate.

$Z' \to \mu\mu$ decay (prompt)

The main topic of this thesis is to search for a Z' boson in its decay to the muonic final state final (muonic dark force), i.e., $Z' \to \mu^+\mu^-$, in the Belle II experiment using $L_\mu - L_\tau$ model (see chapter 6). In the Belle II experiment, Z' could be easily producible from the following process $e^+e^- \to \mu^+\mu^- Z', Z' \to \mu\mu$, where Z' is radiatively emitted by one of the two muons, at the center-of-mass energy of $\Upsilon(4S)$ (10.58 GeV) and decays to $\mu^+\mu^-$ pair. In this model Z' has no finite lifetime and

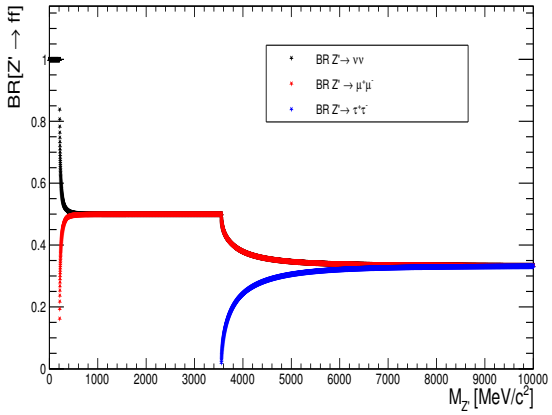


Figure 1.4.: $BR(Z' \rightarrow ff)$ as a function of the Z′ mass. The $\nu\bar{\nu}$ BR includes both ν_μ and ν_τ

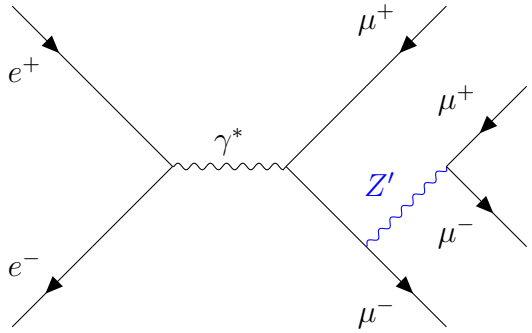


Figure 1.5.: Feynman diagram of the process $e^+e^- \rightarrow \mu^+\mu^- Z', Z' \rightarrow \mu^+\mu^-$.

decays promptly to second or third generation of leptons. The feynman diagram of the process is given in figure 1.5. The $L_\mu - L_\tau$ model is theoretically very well motivated, In this section I will discuss about some theoretical and experimental motivations behind the proposed search,

- *solution to the muon's anomalous magnetic moment*: the anomalous magnetic moment "a" is defined as $a = \frac{g-2}{2}$, where g is the g-factor that characterizes an atom's magnetic moment and angular momentum. The SM prediction for muon's anomalous magnetic moment consists of 3 parts, $a_\mu = a_\mu^{QED} + a_\mu^{EW} + a_\mu^{hadron}$, where a_μ^{QED} represents photon and lepton loops, and a_μ^{EW} comes from the W, Z and Higgs boson loops. Both of these predictions can be calculated precisely, but the hadronic contribution can't be calculated accurately from theory and is very sensitive to new physics [15]. The current experimental measurement from the Fermilab exceeds the SM prediction of

around 4.2σ [16]. The observed deviation could be explained by the additional exchange term due to the dark photon or a dark Z' [17]. However, recent theoretical calculation performed by the BMW lattice QCD collaboration by considering the leading hadronic vacuum polarization and reduces the discrepancy to 1.56σ [18]

- *solution to the discrepancy observed in the angular observable measurement in the rare decay* of $\boldsymbol{B \to K^* \mu^+ \mu^-}$: The LHCb collaboration observed a deviation of around 3.7σ from theoretical predictions in angular distributions of the final state particles in the decay mentioned above [19], which can lead to a viable way to search for new physics. Therefore, the new physics models that generate the vector coupling to muons are particularly interesting. The anomaly-free $L_\mu - L_\tau$ model, which introduces the Z′ boson, is one of the most promising candidates for explaining the discrepancy observed by LHCb collaboration [20].

- In addition, in the framework of $L_\mu - L_\tau$ model, the existing Z' could decay to sterile neutrino in the early universe in a sufficient amount to explain the observed dark matter abundance [21].

The experimental search for the $Z' \to \mu^+\mu^-$ decay in the process $e^+e^- \to \mu^+\mu^-\mu^+\mu^-$ final state has already been performed by the BaBar and Belle experiment using a data sample of 514 fb^{-1} and 643 fb^{-1} respectively [22] [23]. Both of the performed searches didn't observe any signal in the mass range of 0.212 GeV (dimuon threshold) to 10 GeV and set an upper limit on the coupling constant g' as shown in Figure 1.6. Our aim is to improve this limit through an aggressive background reduction and obtain the same (or better) performance despite the lower luminosity (see chapter 6).

$Z' \to \mu\mu$ decay (displaced)

We don't have any theoretical model available for the finite lifetime Z'. The motivation behind performing this search originates from the BaBar Dark Leptophilic Scalar search [24].

In the framework of scalar portal extension of the SM, there exist a new scalar that predominantly mixes with the SM Higgs and have mass in the MeV-GeV range and could mediate interactions between SM and dark matter [25] [26]. This new scalar generally couples to the heavy-flavor quarks but strongly constrained by the searches for rare flavor-changing neutral current decays, such as $B \to K\phi$ or $K \to \pi\phi$ [27]. However, this restriction could be avoided if the coupling to the quarks is suppressed and it interacts preferentially with heavy-flavor leptons. This kind of scalar is called Dark Leptophilic Scalar (ϕ_L). It's interaction Lagrangian

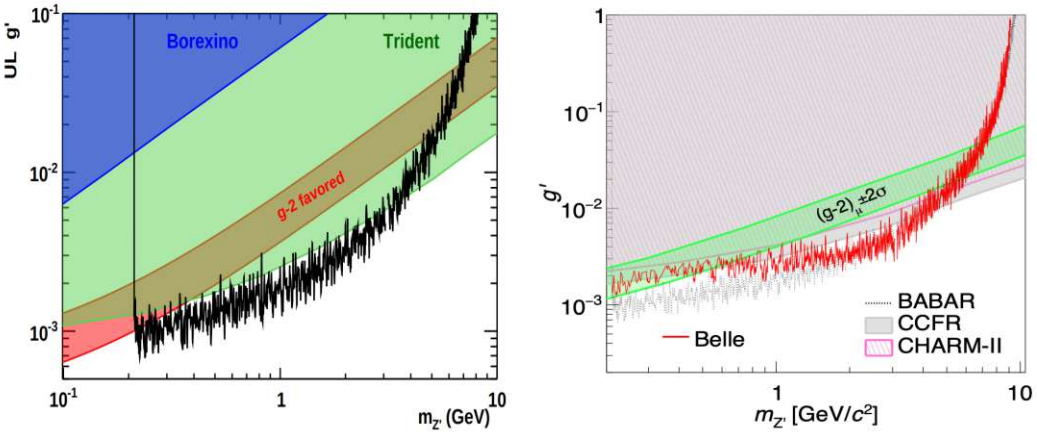


Figure 1.6.: 90% CL upper limits on the gauge coupling g′ as a function of the Z′ mass set by the BaBar (left) and Belle (right) experiment.

with leptons is given by $\mathcal{L} = -\xi \sum_{l=e,\mu,\tau} \frac{m_l}{v} \bar{l}\phi_L l$, where ξ denotes the coupling strength to the leptons.

BaBar performed the Dark Leptophilic Scalar search model independently in the following process $e^+e^- \to \tau^+\tau^-\phi_L,\ \phi_L \to l^+l^-,\ l = e, \mu$, where one of the τ is radiating ϕ_L, which decays into the pair of electrons or muons depending on the mass. If m_{ϕ_L} is between the di-electron and di-muon threshold, it predominantly decays to pair of electrons and could lead to displaced vertices due to the small values of coupling constant. Three different lifetimes of ϕ_L, 0.1, 1, and 10 cm, as well as prompt decay, have been tested for the $2m_e < m_{\phi_L} < 2m_\mu$ and the rest of the ϕ_L mass spectrum the prompt decay of ϕ_L into muons is searched. No significant signal has observed and set an upper limit to the cross-section, see Figure 1.7 [24].

Our idea is to follow the same strategy for the Z' case, give it some finite lifetime, and perform the $e^+e^- \to \mu^+\mu^- Z', Z' \to\ \mu\mu$ search model independently at the Belle II experiment for the first time (see chapter 7).

1.2.4. Different approaches for dark matter detection

The last decade has been very exciting for dark sector searches; many new methods and tools have been developed to detect the dark matter particle, but no evidence of dark matter is still up-to-date. A lot of theoretically well-motivated models got strong constraints and eventually ruled out. There are three possible scenarios to detect the dark matter experimentally (non-gravitational measurement): direct detection, indirect detection, and collider searches. All of these approaches have

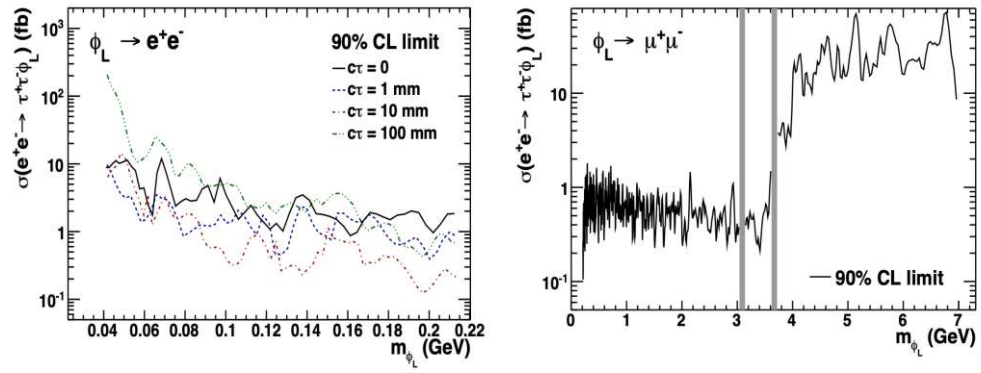


Figure 1.7.: 90% CL upper limit on the cross-section for $e^+e^- \to \tau^+\tau^-\phi_L$ process, left is for electronic final state and right is for muonic. The gray bands indicate the regions excluded from the search.

their own strengths and weaknesses.

Direct detection method based on elastic dark matter-nucleus scattering, in this approach, it is assumed that our galaxy is filled with dark matter particle WIMPs, which undergo elastic scattering with the nucleus of detector's target material and transfer energy. One can measure the dark matter mass and cross-section by measuring the nuclear recoil energy and event rate. Direct detection experiments have excellent sensitivity for $\mathcal{O}$(10 - 100) GeV dark matter particles. Depending on the dark matter and nucleus interaction, sensitivity increases for the spin-independent interaction than spin-dependent interaction (requires a spin to couple unpaired nucleon pair). The sensitivity also reduces for the low nuclear recoil energy. The most intriguing result based on this approach comes from the DAMA/LIBRA experiment, which uses NaI(Tl) as a target crystal and detects the scintillation light from the interaction. They measure the annual modulation of dark matter flux and claimed the observation of WIMP with a mass of 60 GeV and nuclei scattering cross section of $\sim 10^{-41}$ cm^{-2} [28], but their results have not been confirmed by other experiments yet. The other next-generation experiments, such as CRESST-II, COSINUS, and SuperCDMS, depending on the different procedures of measuring recoil energy, could confirm their result and eventually give a hint of dark matter detection in the near future.

Indirect detection experiments are based on the annihilation of dark matter particles in a dense environment such as the galactic center or stars to the standard model particles. All the SM particles could act as a decay product, but

the photon and neutrinos are more interesting because they don't feel any deflection in an interstellar magnetic field and could travel long. The decay of dark matter particles to the charged particles is also interesting because it would increase the anti-particle fraction; by measuring the excess of anti-particles, one could confirm dark matter detection. All the detection methods strongly depend on astrophysical models describing the dark matter density at the annihilation location. The indirect detection method has the best sensitivity towards high dark matter masses ($>\sim$ 100 GeV). The Fermi LAT, PAMELA, AMS-02, IceCube, and Super-Kamiokande are the experiments based on this approach. The most interesting result comes from the Fermi LAT experiment measurement of the photon energy from cosmic rays. They observed a 133 GeV γ peak but not significant enough to claim as a discovery [29]. All the experiments provided crucial inputs regarding the dark matter but no conclusive evidence yet.

Collider search is the pathway to experimentally detect the dark sector and SM mediator particles. If there exists any interaction between SM and dark matter particles (expected to be weakly interacting), dark matter particles could be producible from the decay of SM particles and hence the mediators. The search for hidden particles mediating the interaction between dark matter and SM has been actively performed by different energy and intensity frontier experiments. Since the thesis focuses on the search for a light dark sector mediator produced at the Belle II experiment, I briefly describe some of the crucial searches performed at the e^+e^- colliders. The advantages of e^+e^- colliders are the high luminosity, controlled laboratory environment, well defined initial state with low multiplicity final states. As dark matter particles don't interact electromagnetically and therefore leave no signature in the detector, but they carry energy and momentum, so large missing energy could be an indication of a dark matter signal. e^+e^- collider experiment has excellent sensitivity for low mass dark matter particles in $\mathcal{O}$(few MeV - GeV).

The **pseudo scalar portal** mediator Axion Like Particles (ALPs), a generalized version of axion [30], has been searched by the Belle II experiment using 0.445 fb^{-1} of data collected during 2018. The analysis explored the transition $e^+e^- \to \gamma a,\ a \to \gamma\gamma$ in the mass range $0.2 < m_a < 9.7$ GeV/c^2, and didn't observe any evidence of ALPs and set a 95% upper limit to the coupling between ALPs to photons at the level of 10^{-3} GeV^{-1} [31]. This is the best realistic result available to date and will improve with more data available shortly.

The **vector portal** mediator A' has already been searched by BaBar in its visible and invisible decays via initial state radiation. They explored the transitions $e^+e^- \to \gamma A',\ A' \to invisible,\ e^+e^-,\ \mu^+\mu^-$, and didn't observe any hint of signal and set an upper limit to the production cross-section and mixing strength between SM photon and dark photon [32] [33]. Belle II didn't search for this pro-

cess yet but expects to obtain better sensitivity due to the usage of the advanced trigger to trigger the photon. But Belle II searched for the dark photon with simultaneous production of dark Higgs (h') and its decay to a muonic final state with an integrated luminosity of 8.34 fb^{-1}. However, the analysis didn't observe any signal and set an upper limit to the effective coupling, the combined effect of mixing between SM photon and dark photon, and coupling between dark photon and dark Higgs. This analysis first explored the mass region between 1.65 - 10.51 GeV/c^2 and excluded a much larger region than previously covered by other experiments [34]. The analysis expects to get much better sensitivity with a higher dataset.

The other **vector portal** mediator Z', in the context of $L_\mu - L_\tau$ model, decaying to an invisible final state has already been explored by Belle II using 0.276 fb^{-1} of data. It was Belle II's first physics paper, and this decay was explored for the first time. The invisible decay of Z' is very sensitive to dark matter production. However, this search didn't observe any signal, set an upper limit to the coupling constant (g'), and excluded the region above $g' \sim 5 \times 10^{-2}$ [35]. The other decays of Z' in the visible final states, such as muons, as already mentioned, were explored by BaBar and Belle. They also didn't observe any signal and set an upper limit to the coupling constant. My thesis also concentrates on the same final state but aims to improve the sensitivity compared to the existing results (see Chapter 6).

The **scalar portal** mediator leptophilic dark scalar, in the event topology $e^+e^- \to \tau^+\tau^-\phi_L,\ \phi_L \to \mu^+\mu^-,\ e^+e^-$ was searched by BaBar. They investigated the displaced signature of ϕ_L but could not observe the signal. In my thesis, I also followed the same strategy as BaBar but for the Z' with muonic final state for the first time in the Belle II experiment (see Chapter 7). This study just not only allows us to do this measurement for this specific final state but also helps us to understand Belle II's capability for doing displaced vertex searches in the near future.

The next chapter briefly describes the Belle II experiment, which provides a unique environment to perform these kinds of searches at the intensity frontier.

2. Belle II and SuperKEKB

The Belle II experiment is a general-purpose spectrometer installed at the interaction point of the SuperKEKB accelerator, a high-intensity frontier asymmetric e^+e^- collider hosted by the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. Belle II is the successor of the Belle experiment with advanced technologies, and slightly different beam energies (7 GeV of e^- and 4 GeV of e^+) [36]. The data sample used or planned for the thesis is gathered at the Belle II experiment. Belle II aims to collect a dataset of 50 ab^{-1} in the near future, which is 50 times more than its predecessor Belle. Belle II's high statistics sample will provide an important and unique source of information on the details of new physics processes that are expected to be uncovered at high-energy frontier hadron colliders in the coming years. Belle II's rich physics program includes B and charm physics, quarkonium, and dark sector physics [37]. The following chapter will describe the main features of SuperKEKB and Belle II.

2.1. SuperKEKB

SuperKEKB [38], an upgrade of the KEKB accelerator [39], consists of a 7 GeV electron ring called the high energy ring (HER) and a 4 GeV positron ring called the low energy ring (LER), connected to an injector linear accelerator (linac) with a 1.1 GeV positron damping ring, as shown in Figure 2.1. The electron beam is generated in the linac through a short-pulse photon laser irradiating a cold cathode. Positrons are produced by irradiating electrons to a fixed tungsten target. Then the produced electron and positron are accelerated to the desired energies and collide at Interaction Point (IP) where the Belle II detector is situated. The beam energies are chosen so that the resulting center-of-mass energy is 10.58 GeV, equal to the $\Upsilon(4S)$ mass. The vast majority of data are collected at this resonance; that's why it's also called a B-factory. But the flexibility of the beam energies will allow covering the full range from just below the $\Upsilon(1S)$ (9.46GeV) to just above the $\Upsilon(6S)$ (11.24GeV) for different physics operations. The asymmetric energy of the electron and positron beams produces a Lorentz boost $\beta\gamma$ =0.28 in the Center of Mass (CM) system. It allows separating the decay vertices of B mesons with an aim to perfrom time-dependent CP measurements.

The designed targeted instantaneous luminosity of SuperKEKB is 6×10^{35} $cm^{-2}s^{-1}$,

a factor 30 times higher than what was achieved by its predecessor KEKB. To accomplish this, the KEKB accelerator underwent many upgrades in its components, and more interestingly, a nano-beam scheme strategy was adopted [40]. The basic idea of the nano-beam scheme is to reduce the vertical betatron function β_y^* at the IP to improve the instantaneous luminosity $\mathcal{L}$ of the accelerator with only a moderate increase of beam currents. The $\mathcal{L}$ depends on β_y^* as $\mathcal{L} \sim (\beta_y^*)^{-1}$ through equation,

$$\mathcal{L} = \frac{\gamma}{2er_e}(1 + \frac{\sigma_y^*}{\sigma_x^*})\frac{I_{\pm}\xi_{y\pm}}{\beta_{y\pm}^*}\frac{R_L}{R_{\xi_y}} \tag{2.1}$$

where + and - subscripts are, respectively, for the LER and the HER, γ is the Lorentz factor, e is the electron charge, r_e is the classical electron radius, I is the total beam current, $\xi_{y\pm}$ is the vertical-beam parameter and $\beta_{y\pm}^*$ is the vertical betatron function. The R_L and R_{ξ_y} parameters are reduction factors for the luminosity and the vertical beam-beam parameter. The reduction of $\beta_{y\pm}^*$ is possible by minimizing the size of the overlapped region d of the beams. The overlap region d depends on the angle ϕ and the horizontal size of the beam σ_x^* as shown in Figure 2.2 and given by equation,

$$d \cdot sin(2\phi) = 2\sigma_x^* \rightarrow d \sim \frac{\sigma_x^*}{\phi} \tag{2.2}$$

To achieve this, SuperKEKB is equipped with a final focus superconducting magnet, called QCS, composed of four quadrupole magnets very close to the IP to squeeze the $\beta_{y\pm}^*$ up to 0.3 mm.

In SuperKEKB, beam energies have been changed from the values used in KEKB, from 3.5 and 8.0 GeV to 4.0 and 7.0 GeV. In the Nano-Beam scheme, emittance growth due to intra-beam scattering and the short beam lifetime due to the Touschek effect [41] are severe problems, particularly in the LER. The increase in the beam energy of the LER helps to solve these problems. In addition, the decrease in the beam energy of the HER is beneficial in obtaining a lower emittance.

SuperKEKB was commissioned from February to June 2016 without the Belle II detector and QCS. This is the so-called "Phase 1", which is succeeded by different phases of data taking, called Phase-2 and Phase-3. Phase 2 started in February 2018 and ended in July 2018. During this period, 0.5 fb^{-1} of data was collected, and SuperKEKB delivered a peak luminosity of $0.5\times10^{34}cm^{-2}s^{-1}$. This phase was mainly dedicated to tuning the machine parameters, studying the detector response, and measuring the beam background levels. During this period, no vertex detector was installed to avoid possible damages. Instead, the BEAST II detector [42] was installed. Moreover, data collected during this period is used for dark sector analyses, which do not require the vertexing system and can also

be done with low statistics, the search for an invisible Z' [35] and the search for axion-like particles [31] are the first physics results of the Belle II experiment. With the full detector installed, Phase 3 started in March 2019. Up to now, ∼400 fb^{-1} of data has been collected, while SuperKEKB was able to set a new record in the instantaneous luminosity of $4.7\times10^{34}cm^{-2}s^{-1}$.

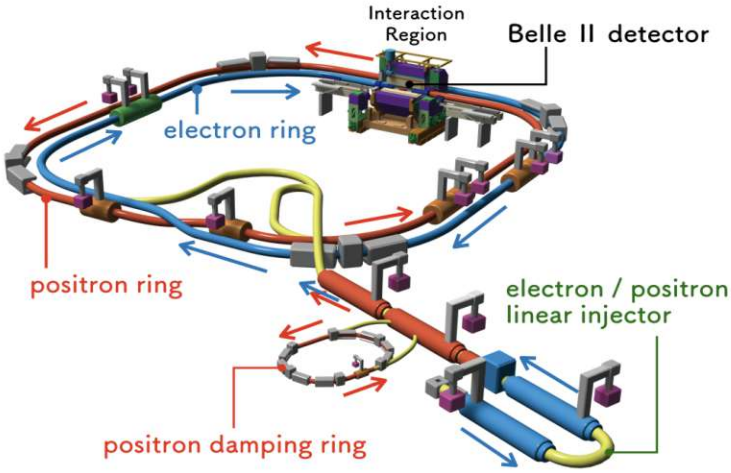


Figure 2.1.: Schematic view of the SuperKEKB collider.

2.2. Belle II detector

The Belle II detector surrounding the HER and LER interaction region (IR) is designed to maintain high performance and precision in an environment characterized by high background levels with respect to the Belle detector. Because of higher currents, smaller beam size, and modified IR, the background hit (∼ 20 times more) and event rate (∼ 50 times more) are expected to be higher than Belle's. Belle II detector consists of different sub-detector components dedicated to detecting different elementary particles traversing the detector material. Figure 2.3 shows a schematic of the Belle II detector. The exact arrangement of the different sub-detectors can be seen in Figure 2.6. Below I describe the different sub-detector components in more detail and their role in detecting different elementary particles.

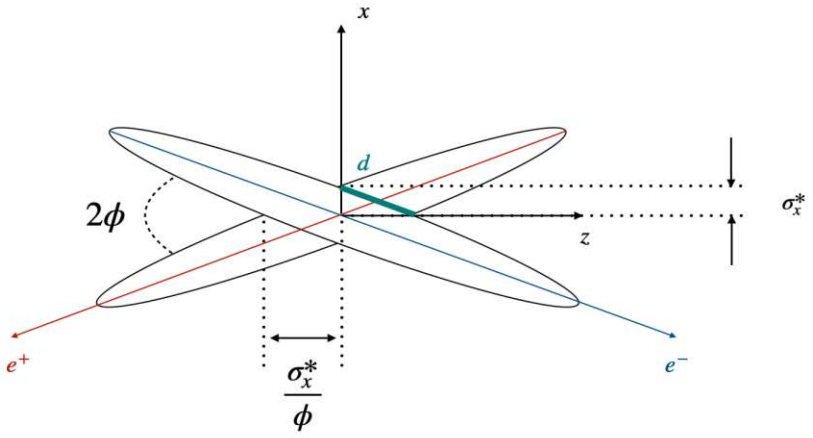


Figure 2.2.: Representation of the nano-beam scheme: σ_x^* is the horizontal beam size, d is the size of the overlap region and ϕ is half of the horizontal crossing angle.

2.2.1. Vertex detector (VXD)

The sub-detector closest to the interaction point is the vertex detector (VXD). VXD consists of two layers of pixel detector (PXD) and four layers of double-sided silicon vertex detector (SVD) [43]. The fundamental role of vertex detectors is to reconstruct primary and secondary decay vertices od B, D meson and τ leptons and measure the impact parameters of the tracks.

PXD: It consists of two layers, coaxial with the beam pipe and located at 14 mm and 22 mm from the IP, respectively. The innermost layer comprises eight planar modules, called ladders, and the outermost layer contains 12 ladders. The ladders are oriented in the ϕ plane in such a way that the one layer covers the insensitive area of the other layer. The geometric region covered by the sensitive sensors is in the range $17^o \lesssim \theta \lesssim 155^o$, where θ is the polar angle. The PXD consists of around 8 million pixels in total, organized into arrays. PXD sensors are based on the DEPleted Field Effect Transistor (DEPFET) technology [44]. A DEPFET is a semiconductor-based device that detects and amplifies signals by itself; hence excellent for minimizing the material budget. Due to a smaller Lorentz boost factor than the KEKB, tracking detectors are kept closer to the IR to compensate the boost factor and maintain a good vertex resolution. The background levels and event hit rates significantly increase at the small radius, and silicon strip-based vertex detectors are not usable due to large occupancy. The PXD can survive with a higher background rate keeping a lower occupancy due to high granularity. The

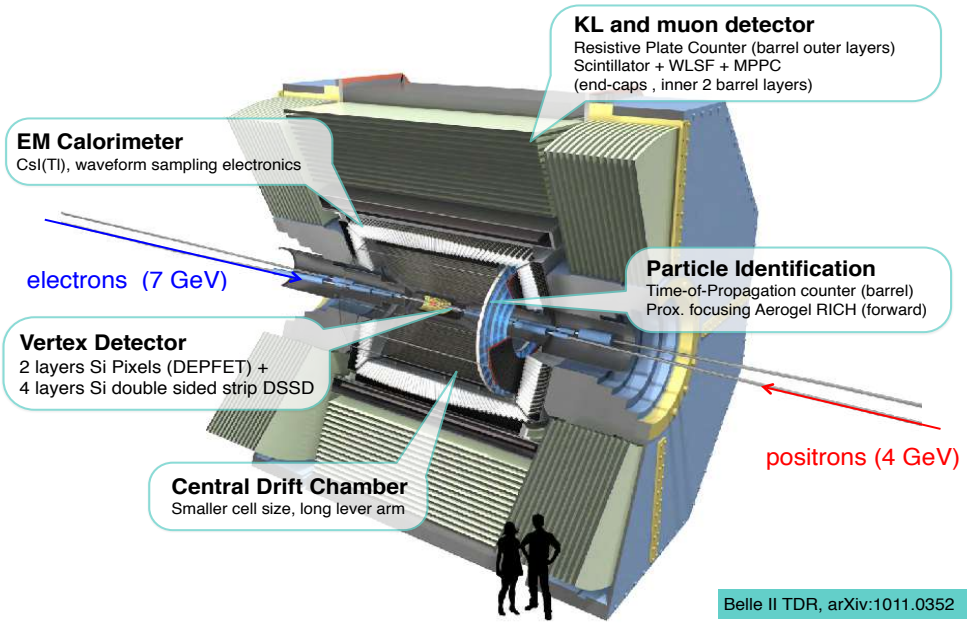


Figure 2.3.: A schematic of the Belle II detector and its different sub-detector components.

observed hit efficiency is above 98% in all modules, and preliminary measurements on data show an impact parameter resolution of about 14 μm [45]. Currently, only two of the twelve ladders of the second PXD layer are installed. Full PXD installation is expected to be completed by 2023 during its first prolonged shutdown.

SVD: It comprises four layers, at a distance of 3.9 cm, 8.0 cm, 10.4 cm, and 13.5 cm from the IP. Each layer consists of different modules, called ladders, arranged around the interaction point to form a nearly cylindrical geometry. Layers 1-4 are composed of 7, 10, 12, and 16 ladders supported by carbon fiber ribs. Each ladder is equipped with Double-Sided Silicon Strip Detectors (DSSD). The geometrical acceptance covered by SVD goes from 17 degrees, in the forward region, to 150 degrees, in the backward area. Ladders are built with different geometric shapes of sensors: ladders of 1st layer consist of rectangular sensors of size 123 mm $\times$ 38 mm. In contrast, ladders of other layers have rectangular sensors of size 123 mm $\times$ 58 mm and one trapezoidal sensor in the forward region to improve the angular acceptance and optimize the incident angle on the sensor of particles

coming from the interaction point. SVD has total 172 DSSD sensors, covering a sensible area of 1.2 m^2. SVD also provides particle identification information using the energy loss information of the particle tracks by measuring the quantity $\frac{dE}{dx}$ and could perform standalone reconstruction of low momentum particles that do not reach the CDC. The SVD detector was installed in 2018 and had been operating since 2019. Results of the first physics run showed a good performance of the SVD sensor efficiencies at a level of 99% and stable with time. An excellent signal-to-noise ratio is also observed for all the sensors [46] [47]. Figure 2.4 shows the 3D representation of VXD.

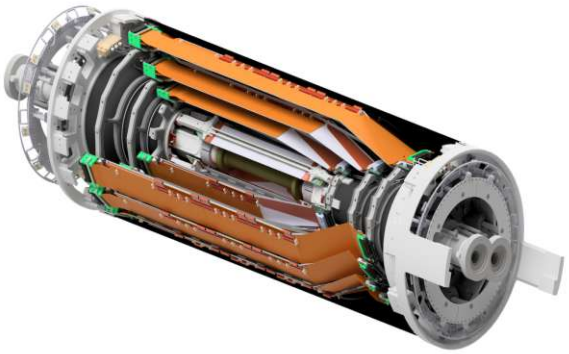


Figure 2.4.: The Belle II vertex detector, composed of the PXD and silicon strip (SVD) detectors. Pictures are taken from [47].

2.2.2. Central Drift Chamber (CDC)

The Central Drift Chamber (CDC) is the outermost tracking device of the Belle II detector. CDC plays three crucial roles,

1. reconstructing charged particle tracks and measuring their momentum precisely.

2. provides particle identification (PID) information using energy loss measurements within its gas volume. The Low-momentum tracks, which do not reach the PID device (ARICH, TOP, ECL, KLM), can be identified using the CDC alone.

3. provides efficient and reliable 2D and 3D trigger signals for charged particles.

Belle II CDC follows the global structure of its predecessor Belle for the material of the major parts, the superlayer wire configuration, the cell structure, the

wire material, and the gas mixture, as Belle CDC works pretty well without any serious issues [48]. The Belle II CDC comprises 9 superlayers; apart from the first superlayer, other superlayers are composed of 6 layers, and the former has 8 layers, so in total, 56 layers. The innermost and outermost superlayers contain axial ("A") layers to match the shape of the inner and outer cylinders. The intermediate superlayers alternate between stereo ("U" or "V") and axial layers. The radius of the inner cylinder is 160 mm, and the radius of the outer cylinder is 1130 mm w.r.to IP, and they are increased compared to Belle due to the high background level and to provide more space to SVD for the inner cylinder, while Belle II has a very compact Barrel PID device that offers more space to the outer layer. The layers are immersed in a gas composed of 50% helium and 50% ethane, providing a high drift speed. The geometrical acceptance of the CDC goes from $\theta = 17^o$ to $\theta = 150^o$. The measured spatial resolution on the individual hit is around 100 μm and the measured CDC tracking efficiency found to be >99% for events with $p_t > 1$ GeV/c. The different wires configuration is shown in Figure 2.5, where the axial and the stereo wires are represented respectively in blue and red colors and compared with Belle.

In the chapter 4 of this thesis I briefly described the procedure of accessing the likelihood information from the CDC and its role on charge particle identification. In that chapter, I also tried to develop a machine learning-based PID algorithm concentrating only on CDC with an aim of obtaining better PID performance than likelihood. The developed algorithm is also helpful in separating low-momentum tracks that cannot reach other PID devices.

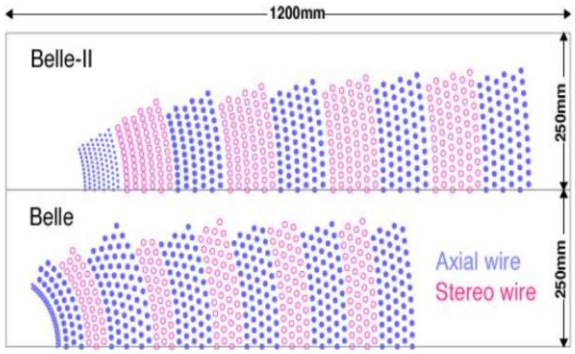


Figure 2.5.: Wire configuration of Belle-II CDC and comparison with Belle CDC.

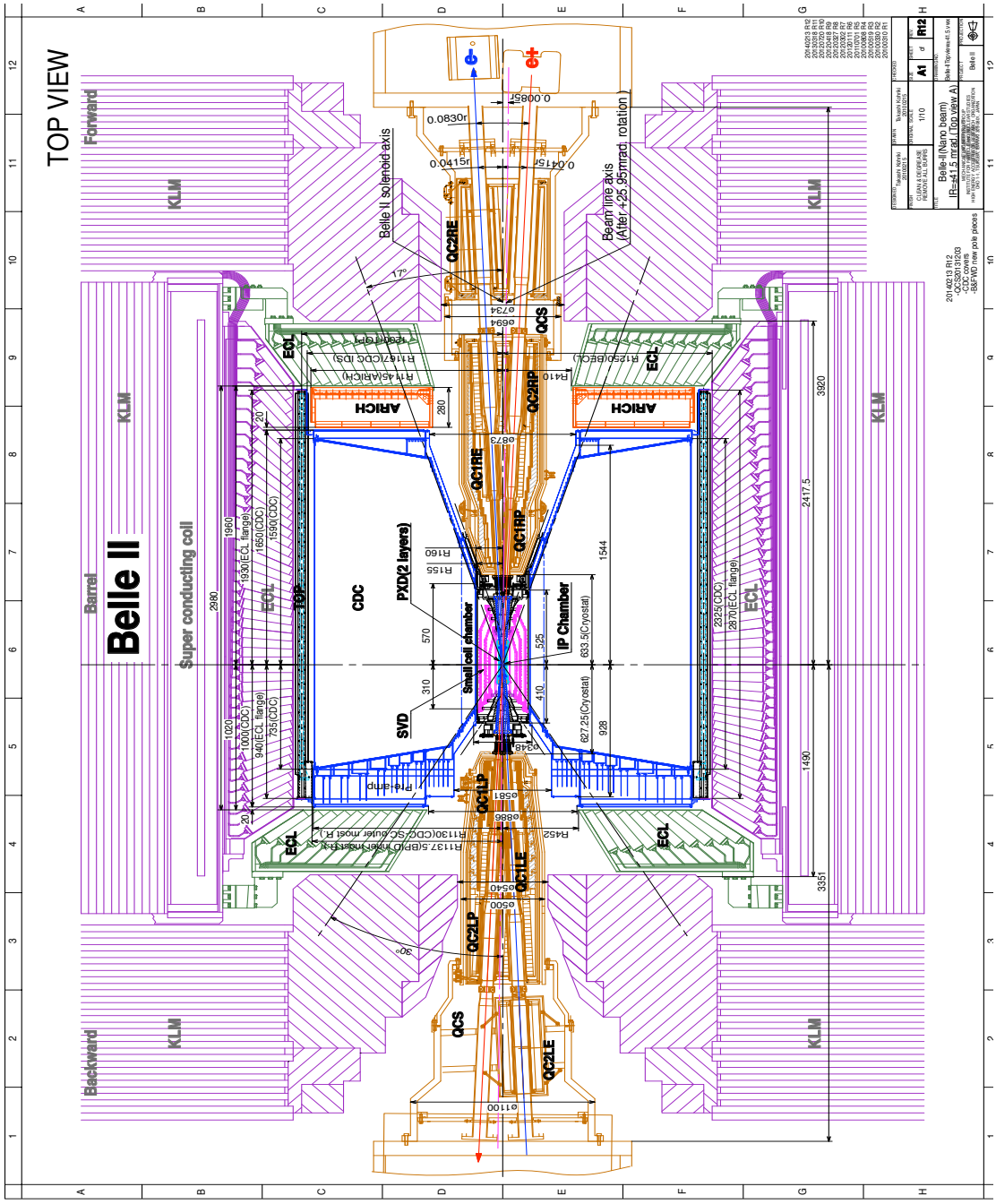


Figure 2.6.: Cross section view of the Belle II detector.

2.2.3. Particle identification system

The particle identification (PID) system consists of two Cherenkov radiation detectors, the Time-of-propagation counter (TOP) [49] and the Aerogel Ring Imaging Cherenkov counter (ARICH) [50]. The basic principle of Cherenkov detectors is to measure the Cherenkov angle θ_C between the direction of the photon produced by relativistic charged particles and its momentum traversing the radiator material. Depending on this angular information, TOP provides crucial information about the particle position and time of propagation in the detector material, and ARICH provides different ring images for different particles; hence help to distinguish different particle hypothesis. Below I briefly describe each of these sub-detector components and their working principle.

Time-Of-Propagation (TOP) counter: The TOP counter is installed in the barrel region of the spectrometer between the ECL inner support and the CDC outer cover, whose conceptual overview is shown in Figure 2.7. It is composed of 16 modules surrounding the CDC, and the radius of the TOP is around 1.24 m. Each module is made of a long quartz bar, readout with micro-channel plate photo-multipliers (MCP-PMTs) [51] installed at the one end, and a spherical focusing mirror is installed on the other end of the bar. When a particle crosses the quartz bar, they produce Cherenkov photons internally reflected inside the quartz radiator's walls. Cherenkov photons are focused and directed toward the MCP-PMTs by the focusing mirror. The focusing mirror minimizes the chromatic effect and increases the flight time resolution. An expansion prism is used to expand the Cherenkov ring image before the MCP-PMTs. Finally, the MCP-PMTs measure the time of propagation, t_{TOP}, of the Cherenkov photons and provide information on the position coordinates (x, y) of the photons. Then Cherenkov ring image is reconstructed from the 3-dimensional information (t_{TOP}, x, y) provided by the MCP-PMTs.

The particle identification information is extracted by comparing the distribution of the time of arrival of the photons (t_{TOP}) in each of the MCP-PMT channels with the expected PDFs for the six standards charged particle (e, μ, π, K, p, d) hypotheses, where the expected PDFs are calculated analytically given the mass hypothesis and the particle's track parameters [52]. The ratios of the six corresponding likelihood values are then used to assign identification probabilities. The TOP has been designed with the primary goal of providing identification for hadrons (mainly π's and K's) with momentun > 1.5 GeV/c, where energy loss information is not discriminating enough. However, it also plays a crucial role in lepton and pion identifcation for momentum 0.3 GeV/c to 1 GeV/c [53].

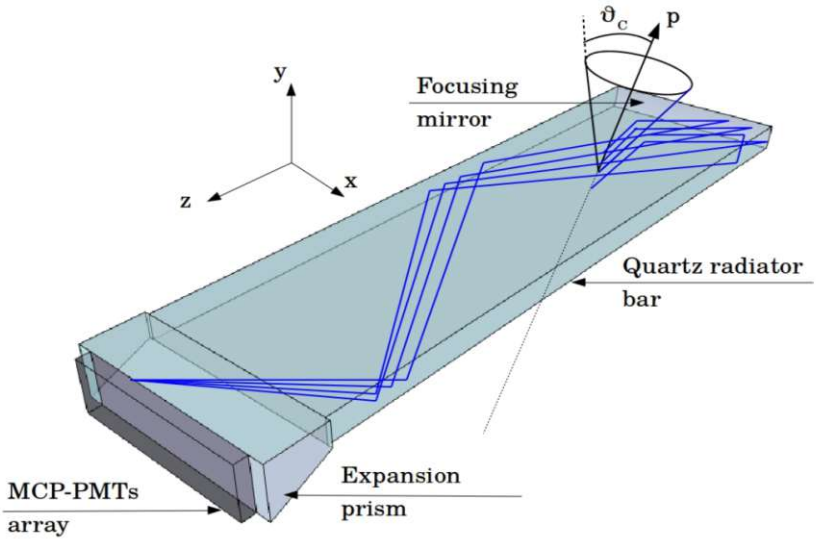


Figure 2.7.: A schematic view of one of the module of the TOP detector, taken from [52]

Aerogel Ring-Imaging Cherenkov detector (ARICH): ARICH, the other PID detector located at the forward region of the Belle II detector, helps to distinguish different particle hypotheses based on the ring imaging technique by detecting the Cherenkov photons. It has been designed mainly to separate kaons from pions over most of their momentum spectrum, but it also provides good discrimination between lepton and pions below 1 GeV/c. We don't have any PID devices at the backward region of the spectrometer; due to the boost, most events go in the forward direction. The performance of a RICH counter depends on the quantity $\sigma_{track} = \frac{\sigma_\theta}{\sqrt{N}}$ for each charged track, where N is no of detected photons and σ_θ is the Cherenkov angle resolution. The number of detected photons increases with a longer radiator, but the photon resolution degrades due to the emission point uncertainty. The broadness of the emission point is improved by implementing a non-homogeneous aerogel radiator [50]. By appropriately choosing the refractive indices of consecutive aerogel radiator layers, one could perfectly overlap the corresponding Cherenkov rings on the photon detector [54], as shown in Figure 2.8. Such a variable of refractive indices for individual layers is only possible with aerogel, which could be produced with any desired refractive index in the range between 1.01-1.2 [55]. In Belle II, the optimal thickness of the aerogel radiator is found to be 2 cm [56]; two 2cm thick layers of silica aerogel with refractive indices of 1.055 and 1.065 are implemented as Cherenkov radiators.

Particle identification in ARICH is based on comparing the observed spatial distribution of photon hits on the photo-detector plane and the probability density function describing the expected distribution. The likelihood function for each particle hypothesis is constructed as a product of probabilities of individual pixels (in the photo detector plane) being in the observed state, i.e., the probability of a pixel being hit by photons and the PDF is constructed from the follwoing components: Cherenkov photons emitted in the aerogel, random hits describing contributions from the electronics noise, and beam backgrounds. The PDF is parametrised as a function of the reconstructed Cherenkov angle and projected onto the photo-detector plane. Detailed detector geometry and its properties are also taken into account in the PDF.

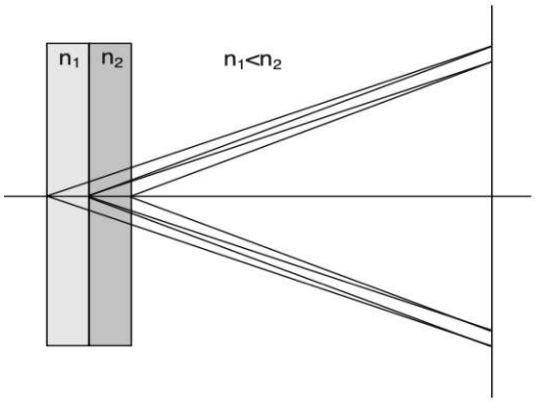


Figure 2.8.: The focusing configuration of ARICH with an inhomogeneous areogel radiator. Image is taken from [57]

2.2.4. Electromagnetic Calorimeter (ECL)

The main tasks of ECL are:

1. Identify photons with high efficiency and precisely determine their energy and angular coordinates,
2. electron identification,
3. On-line and off-line luminosity measurement,
4. Generate a high-efficiency trigger signal.

In Belle II, it reuses the Belle's CsI(Tl) scintillation crystal material due to its high performance, but some changes have been implemented in the readout electronics. ECL splits into three different regions barrel, forward and backward endcaps. It covers an angular acceptance from $\theta = 17^o$ to $\theta = 150^o$ except for two gaps of about 1^o wide between the barrel and (BECL) the endcaps. The total number of ECL crystals is 8736, divided into 6624 in the BECL and 2212 in the endcaps. The energy resolution of crystals are around 20 MeV. The scintillation light detection is done using two silicon photodiodes glued in the back of crystals. A preamplifier is connected to the photodiode, and two independent outputs for each crystal are obtained. During data-taking, the two signals emitted by both photodiodes are first summed and then digitized. The resulting waveform is processed online by field-programmable-gate-array (FPGA) to measure the deposited energy magnitude and time. According to simulation, this new electronic reduces in the fake rate by a factor of 7, maintaining efficiency on photon detection of 97%.

The baseline method for charged particle identification at ECL relies on the E/p ratio, where E is the particle's energy deposited in the calorimeter crystal ($E_{cluster}$) and p is the reconstructed momentum of the topologically matched charged track to the ECL, where the momentum measurement comes from the tracking systems. Templates of E/p are generated from simulated samples for each particle hypothesis, and PDFs are extracted via adaptive Gaussian Kernel Density Estimation (KDE) fits [58]. Different PDFs are used as a function of bins of polar angle, momentum, and charge(q). The charge dependency is also introduced to account for different ECL energy deposition patterns in hadronic interactions. Depending on the E/p ratio, particle identification works very well in Belle and Belle II. But in Belle II, as we expect the most robust background conditions, one would need a more efficient PID algorithm that could vastly reduce the fake rate and improve the identification efficiency. One way to achieve this is to use Martivarite Machine learning techniques. In Belle II, Boosted Decision Trees (BDT) based particle identification algorithm is already developed and observed overwhelming improvements compared to traditional E/p based particle identification. This algorithm mainly exploits observables sensitive to the different physics governing interactions of hadrons, electrons, and muons with the calorimeter crystals. In chapter 3, I tried to develop a Neural network-based PID algorithm concentrating on ECL and relying on Likelihoods from other detectors. I observed potential improvements over the BDT-based PID at the low momentum region.

2.2.5. K_L^0 and μ detector (KLM)

KLM is the outermost detector of the Belle II. Its main task is to identify muons and K_L's. Apart from the inner two layers, the KLM's barrel region (BKLM) is composed of Resistive Plate Chambers (RPCs) [59]. The endcap KLM (forward

and backward) and the inner two layers of the barrel are made of scintillator strips coupled with a silicon photomultiplier (SiPM). This kind of configuration has been decided based on the previous experience of Belle because RPC efficiency degraded in the endcaps due to the high level of backgrounds. In SuperKEKB, the background rate in the endcaps is expected to be 20 to 40 times higher, and the endcap RPCs are replaced with scintillators. The barrel region covers an angular acceptance from $\theta = 45^o$ to $\theta = 125^o$ and that is extended by endcaps from $\theta = 20^o$ to $\theta = 125^o$.

The RPCs are composed of two electrodes (2 parallel planes 3 mm thick) made by high resistivity glass spacing of 3 mm. The space between electrodes is filled with a gas mixture of HFC, freon, argon, and butane. The outer surface of each electrode is coated with a carbon-doped paint that allows the distribution of high voltages to electrodes so that a uniform electric field of 4.3 kV/mm is generated in the gas-filled region. When a charged track passes, it induces a signal on 5-cm-wide metal strips on each side of the RPCs, used for the readout. A dielectric foam layer separates the metal strips from an external ground plane. Two RPCs are coupled to form a single super layer to prove particles' detection efficiency. Figure 2.9 shows a section of an RPC super layer.

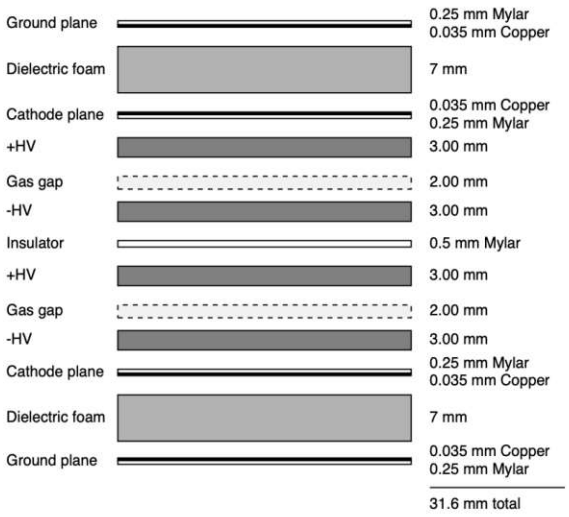


Figure 2.9.: Cross-section of an RPC superlayer. Image is taken from [57]

Since muon identification [60] is very important for our analysis (see chapter 6, 7), I briefly describe the process of muon identification in KLM. Muon identification happens in two steps, 1. track extrapolation and 2. likelihood estimation. Track extrapolation begins in the CDC, assuming it is a pion track,

and this assumption benefits from the PID detectors and ECL (helps to identify pions). The extrapolation only considers the mean energy loss of particles in determining the range of the track. The extrapolated track is accepted within the KLM acceptance if it crosses at least one RPC layer; this requires a minimum momentum of 0.6 GeV/c. Suppose a KLM hit is found within 5σ from crossing an extrapolated track. In that case, it is declared a matching hit, and further extrapolation (re-extrapolation) begins assuming the muon hypothesis– using the Kalman filter and fitting technique. Hits in the KLM layer are again associated with the track accounting 5σ deviation of the re-extrapolated crossing point. In this re-extrapolation, the associated hits in preceding KLM layers provide feedback to the track extrapolation to the next layer. The track extrapolation ends if kinetic energy falls below a defined threshold or escapes the KLM geometry. Now two quantities are used to test the hypothesis that a track is a muon rather than a hadron (π or K); the difference between the measured and expected range of the extrapolated track (ΔR), and the goodness of fit of the transverse deviations of all hits associated with the re-extrapolated track (χ^2). Probability density functions of ΔR and χ^2 are constructed before using simulated single-track events containing a muon, pion, or kaon. The joint probability density is defined as a product of these separate probability densities: $p(\Delta R, \chi^2) = p(\Delta R)\ p(\chi^2)$. For a given track characterized by ΔR, χ^2 and the track whether predicted to stop inside or escape from either the barrel or endcap KLM, different probability densities are assigned to the particles (p_μ, p_π, p_K) and the muon likelihood is defined by, $L_\mu = \frac{p_\mu}{p_\mu + p_\pi + p_K}$.

2.2.6. Trigger System

The Belle II trigger system consists of hardware-based Level 1 trigger (L1) [61] and a software-based High-Level Trigger (HLT) [62]. Belle II trigger system designed mainly to achieve high efficiency for hadronic events from $\Upsilon(4S) \to B\bar{B}$ and from the continuum. As in Belle II we expect more backgrounds than Belle due to high instantaneous luminosity it is required to have an effective and sophisticated trigger system to reduce the high L1 trigger rate. The Belle II trigger system adopts the Belle triggering scheme [59] with new technologies. Mainly old components are replaced by new ones, some new trigger lines are added and each component has Field Programmable Gate Array (FPGA) so that the trigger logic is configurable rather than hard-wired. The Belle II trigger system consists of sub-triggers from sub-detectors and one final-decision logic. A sub-trigger system gathers trigger information from its sub-system and sends it to the global-decision logic to make a final decision. HLT trigger mainly rejects the physically uninteresting decays online to sort mainly the storage issues. The schematic overview of the Belle II trigger system is shown in Figure 2.10. The CDC sub-trigger provides the 2D and 3D charged tracks information using the Track Segment Finding method. The

ECL-based triggers are for neutral and charged tracks oriented physics events. ECL-based triggers select events using the energy deposited information on the total or isolated ECL cluster. The Barrel PID (BPID) and the Endcap PID (EPID) sub-trigger give precise timing and hit topology information. The KLM sub-trigger gives information about muon tracks. The Global Decision Logic (GDL) receives all the sub-trigger information and makes the final decision. The right decision send as a trigger signal.

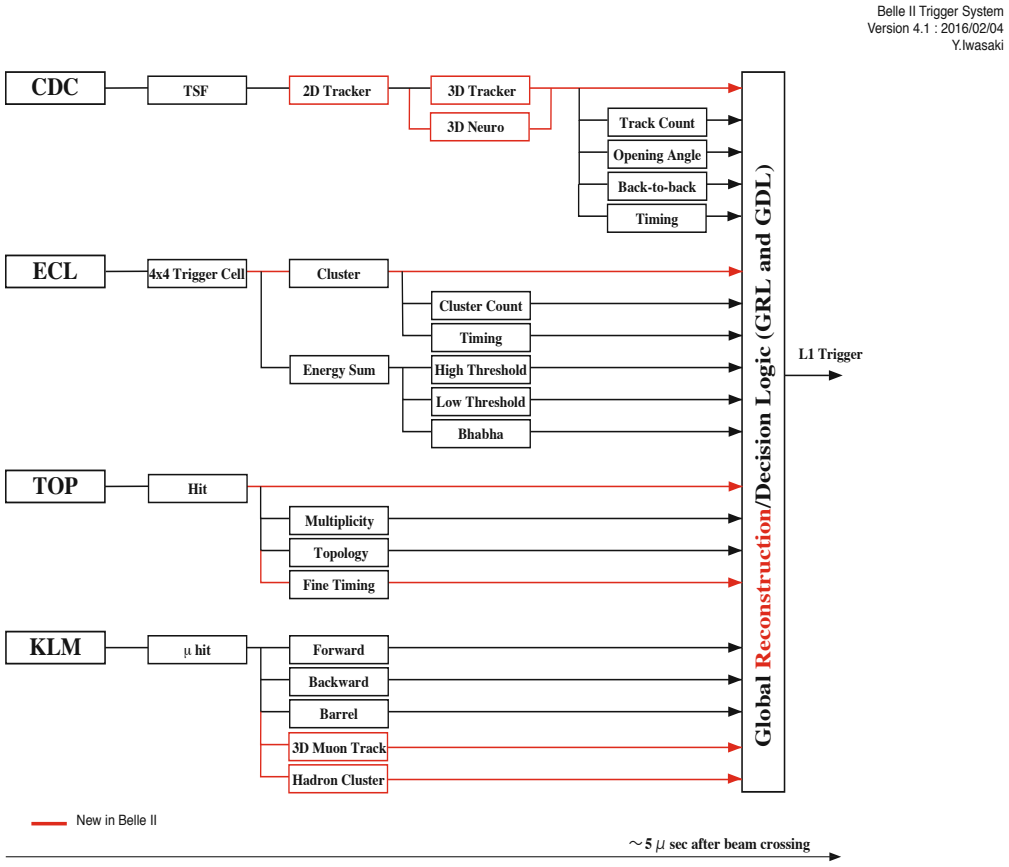


Figure 2.10.: Schematic overview of the L1 trigger system. The output from the sub-trigger systems goes to the Global Decision Logic (GDL). The final trigger decision is made in the GDL. The red lines are newly added information paths relative to Belle. Image is taken from [57]

2.3. Summary

Belle II is a next-generation B-factory experiment located in Tsukuba, Japan. It is the successor of Belle with advanced technologies and slightly different beam energies and aims to collect 50 ab^{-1} of data in the near future. I briefly described above all its sub-detector components, its crucial role in the particle identification, and some significant changes compared to the Belle. Below Table 2.1 summarizes different sub-detector components that help to separate different particle hypotheses interested in this thesis. Likelihood based particle identification worked very well in Belle and Belle II. Still, some physics processes sensitive to new physics effects require better identification efficiency and lower fake rate than the existing likelihood based method. In the next chapter, I briefly describe a newly developed tool for particle identification at Belle II, and it was found that machine learning based algorithms provide overwhelmingly better performance than the likelihood.

Particle	Energy	Momentum	Position	Particle Identification
$e^+(e^-)$	ECL	CDC	PXD, SVD, CDC	ECL, ARICH, TOP, CDC
$\mu^+(\mu^-)$		CDC	PXD, SVD, CDC	KLM, ARICH, TOP, CDC
$\pi^+(\pi^-)$		CDC	PXD, SVD, CDC	ARICH, TOP, CDC
γ	ECL		ECL	ECL, CDC

Table 2.1.: Different sub-detectors help to detect different particles concerned for the thesis.

Part II.

Physics Performance Studies

3. Lepton-pion identification using Multivariate techniques at Belle II detector

The following work was done solely by the author, while additional guidance was provided by Gianluca Inguglia.

3.1. Introduction

A key topic of the Belle II physics program is to study the semileptonic B decays, specially decays involving τ leptons for the test of Lepton Universality (a universal gauge coupling to different generations of leptons) both in the inclusive ($B \to X\tau\nu$) and exclusive ($B \to D^*\tau\nu$) measurements. The sensitivity of this analysis mainly depends on the capability of separating low momentum lepton candidates ($l = e, \mu$) in the $\tau \to l\nu_l\nu_\tau$ decay from the hadronic backgrounds. An excellent low momentum lepton-hadron separation is also desirable for precisely measuring the CKM matrix elements $\mid V_{ub} \mid$ and $\mid V_{cb} \mid$ [63].

For μ's momentum $\geq$ 600 MeV/c, there is a dedicated KLM detector to identify them. However, low momentum μ's ($\leq$ 600 MeV/c) is easily mimicked by hadrons. Furthermore, at low momenta, electrons also suffer energy losses due to bremsstrahlung, making it hard to separate from hadrons. Therefore, an identification method must be developed depending on the ECL. A Multivariate analysis (MVA) based algorithm could be exploited to combine measurements from the ECL governing different physics processes related to hadrons and leptons. These inputs could also be combined with other sub-detectors lower level information to give the global description of particle identification. Boosted Decision Tree (BDT) based particle identification algorithm has already been developed by the Belle II collaboration [64] and observed a factor of 10 and 2 improvements in the misidentification probability for electron-pion and muon-pion compared to the likelihood-based separation in the low momentum region.

In this chapter, for the first time, I tried to develop a Neural Network (NN) based binary lepton-hadron particle identification in an aim to obtain better performance than the existing BDT-based algorithm, at least for the low momentum regions.

Before going to the actual physics part of this topic, I would like to describe the used MVA methods considered for this study briefly.

3.2. Boosted Decision Tree (BDT)

A decision tree is a binary tree-structured classifier widely used in experimental particle physics [65]. Repeated yes or no decision is taken on a single variable at a time until a stop criterion is fulfilled. The entire phase space is divided this way into many regions, eventually classified as signal or background, depending on most training events in the final leaf node. The boosting of a decision tree extends this concept from one tree to several trees which form a forest. In boosting, a single classifier is formed by combining the weighted average of the individual decision trees. Boosting stabilizes the response of the decision trees and can considerably enhance the performance w.r.t. a single tree.

3.2.1. Gradient boosting

In this study, I use the gradient boosting technique [66], available in the TMVA package [67].
The idea of predictor function ($F(x)$) estimation through boosting could be understood by considering an additive expansion approach. The function $F(x)$ is assumed to be a weighted sum of "weak learners" (parametrized base functions $f(x; a_m)$). Any TMVA classifier could act as a weak learner, but decision trees benefit most from gradient boosting. Thus each base function in the expansion below corresponds to a decision tree,

$$F(x; P) = \sum_{m=1}^{M} \beta_m f(x; a_m); P \in \beta_m, a_m. \tag{3.1}$$

Where m corresponds to the number of decision trees, β_m is the weight, and a_m is the parameter from the weak learner.
Now the boosting procedure adjusts the parameters P such that the deviation between the model response $F(x)$ and the true value y obtained from the training sample is minimized. The deviation is measured by the so-called loss-function $L(F, y)$. The current TMVA implementation of GradientBoost uses the binomial log-likelihood loss

$$L(F, y) = ln(1 + e^{-2F(x)y}), \tag{3.2}$$

for classification. There has no straightforward way to obtain the boosting algorithm for the loss function, and one has to apply the steepest-gradient decent approach to do the minimization. This is done by calculating the current gradient

of the loss function and then growing a regression tree whose leaf values are adjusted to match the mean value of the gradient in each region defined by the tree structure. Iterating this procedure yields the desired set of decision trees, which minimizes the loss function. The whole optimization procedure is summarized below,

If we have a training data set $(x_i, y_i)_{i=1}^N$, a differentiable loss function $L(F(x), y)$, and number of iteration M,

1. The optimization starts with initializing a constant value to the predictor function $F(x)$,

$$F_0(x) = \arg\min_{\eta} \sum_{i=1}^{N} L(y_i, \eta) \tag{3.3}$$

2. Now for each iterations $m = 1$ to M

   a) compute the pseudo-residuals (defined by $y_i - F_m(x_i)$) for each training data set; residual is proportional to the gradient of the loss function.

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right] \tag{3.4}$$

   b) Now fit the residuals with a regression tree $h_m(x)$, i.e fit to the training data set $(x_i, r_{im})_{i=1}^N$.

   c) compute the multiplier η_m by optimizing the loss:

$$\eta_m = \arg\min_{\eta} \sum_{i=1}^{N} L\left(y_i, F_{m-1}(x_i) + \eta h_m(x_i)\right) \tag{3.5}$$

   d) Then update the predictor function by,

$$F_m(x) = F_{m-1}(x) + \nu \cdot \eta_m h_m(x), \ 0 < \nu \leq 1 \tag{3.6}$$

   where ν is the learning rate or shrinkage.

3. Finally one will get the desire predictor function $F_M(x)$ at the Mth iteration.

In this study, I split according to the variable that optimizes the gain in cross-entropy between a given node and the two daughter nodes. Cross entropy is defined as $p \cdot \log p - (1-p)\log(1-p)$, where $p = \frac{n_s}{n_s+n_b}$ and n_s, n_b is no of signal and background events at that node.

3.2.2. Bagging

In some instances, GradientBoost may also benefit from introducing a bagging-like resampling procedure. The bagging procedure uses random sub-samples of the training events for growing the trees. This is called stochastic gradient boosting. Resampling is implemented by applying random Poisson weights to each occurrence of the parent sample.

3.2.3. BDT hyper-parameters

The optimal choice of the algorithm hyper-parameters is summarised in Table 3.1. I am using 80% events for training and 20% events for testing purposes.

Parameters	Values	Description
NTrees	500	Number of trees in the forest.
Shrinkage	0.2	Learning rate for the gradient boosting algorithm.
MaxDepth	2	Maximum depth of each decision tree
MinNodeSize	2.5%	Minimum fraction of training events to define a leaf node.
SeparationType	CrossEntropy	The F.O.M. for the node splitting
nCuts	20	optimal cut in node splitting.
BaggedSampleFraction	0.5	For bragging like resampling procedure.

Table 3.1.: List of BDT hyper-parameters.

3.3. Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) is a collection of interconnected neurons, with each neuron producing a definite response to a given set of input signals. By applying an external signal to the input neurons, the network goes to a defined state that could be measured from the response of one or several output neurons. One can therefore view a neural network as a mapping from a set of input variables (x_1,..., $x_{n_{var}}$) onto a one or multi-dimensional space of output variables (y_1,..., $y_{m_{var}}$) depending on desire classification.

There are four types of ANN available in the TMVA [67]. This study uses the recommended MLP implemented in the TMVA package.

3.3.1. MultiLayer Perceptron (MLP)

MLP is a feedforward ANN that consists of several layers of neurons. The first layer is the input layer, the last one is the output layer, and all others are hidden

layers. This kind of arrangement is known as Multilayer Perceptron (Figure 3.1). For a binary classification problem with n input variables, the input layer consists of n neurons that hold the input values $x_1,...,x_n$, and one neuron in the output layer that contains the output variable, the neural net estimator y_{ANN}.

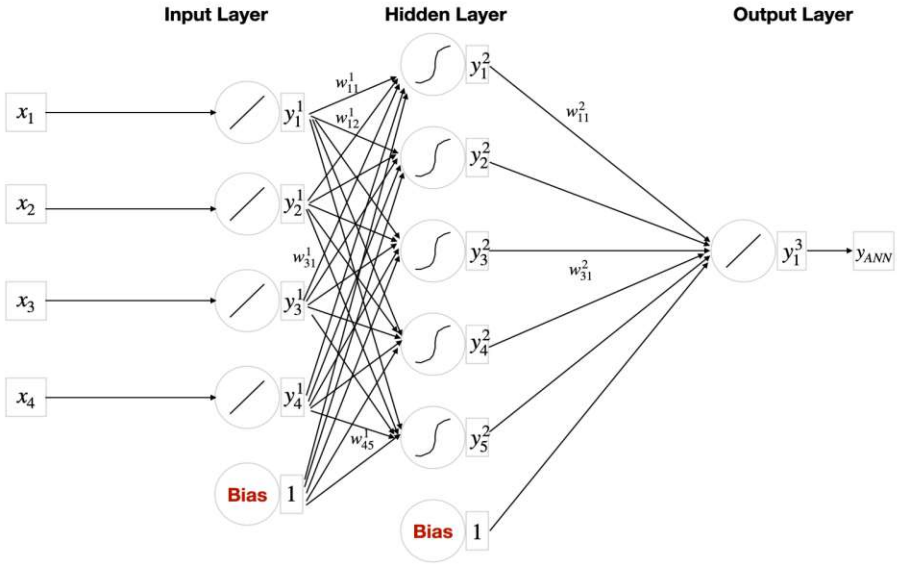


Figure 3.1.: Multilayer perceptron with one hidden layer.

3.3.2. Neuron response function

The neuron response function ρ maps the neuron input $x_1,...,x_n$, onto the neuron output y_{ANN}. Often it can be separated into two functions, one synapse function (κ) which maps $R^n \to R$ and a $R \to R$ neuron activation function (α), so that $\rho = \alpha$ o κ. The α and κ are available in the following forms.

$$\kappa : (y_1^{(l)}, ..., y_n^{(l)} | w_{0j}^{(l)}, ..., w_{0n}^{(l)}) \to \begin{cases} w_{0j}^{(l)} + \sum_{i=1}^{n} y_i^{(l)} \cdot w_{ij}^{(l)} & \text{Sum} \\ w_{0j}^{(l)} + \sum_{i=1}^{n} (y_i^{(l)} \cdot w_{ij}^{(l)})^2 & \text{Sum of squares} \\ w_{0j}^{(l)} + \sum_{i=1}^{n} \mid y_i^{(l)} \cdot w_{ij}^{(l)} \mid & \text{Sum of absolutes} \end{cases} \tag{3.7}$$

$$\alpha : x \rightarrow \begin{cases} x & Linear \\ \frac{1}{1+e^{-kx}} & Sigmoid \\ \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} & Tanh \\ e^{-x^2/2} & Radial \end{cases} \tag{3.8}$$

The synapse function (κ) and activation function (α) chosen for this study are Sum and Tanh, respectively. I also investigated the other functions, but they didn't seem to improve the MLP performance.

3.3.3. Network architecture

The performance of a network mainly depends on the configuration of the hidden layers and the available neurons on them. Weierstrass theorem [68] states that an MLP having a single hidden layer is sufficient to provide the optimal performance given that a sufficiently large number of neurons is used in that hidden layer. If the available computing resources and the size of the training data sample are sufficient, then one can increase the number of neurons in the hidden layer until the optimal performance is achieved. The same performance could be possible to achieve with a network of more than one hidden layer and a much smaller total number of neurons on them. This would lead to a shorter training time and a more robust network. For this study, I use one hidden layer with N+5 neurons on them, where N is no of the input variables. This chosen hidden layer configuration provided optimal performance.

3.3.4. Training of the neural network (Back-propagation (BP))

The so-called back-propagation is the most common algorithm for adjusting the weights that optimize a neural network's classification performance. It belongs to supervised learning, where the desired output of every input is known. For simplicity, let's assume we have a Neural Network (NN) with a single hidden layer (see Figure 3.1) with a Tanh activation function and a linear activation function in the output layer; the output of the network is given by,

$$y_{ANN} = \sum_{j=1}^{n_h} y_j^{(2)} w_{j1}^{(2)} = \sum_{j=1}^{n_h} tanh(\sum_{i=1}^{n} x_i w_{ij}^{(1)}) \cdot w_{j1}^{(2)} \tag{3.9}$$

,

where n, n_h are the number of neurons in the input and hidden layers, $w_{ij}^{(1)}$ are the weights between input layer neuron i and hidden layers neuron j, and $w_{j1}^{(2)}$ is

the weight between the hidden layer neuron j and the output neuron. A simple sum is used for the synapse function.

During the training, the network is given N training events with a feature vector $\mathbf{x_a} = (x_1, ..., x_n)_a$, $a = 1, ..., N$. for each training event a, the output of the network $y_{ANN,a}$ is computed and compared with the desired outcome $\hat{y}_a \in 0, 1$ (1 is for signal events, and 0 is for background events). An Error function E measuring the network response is defined by,

$$E(\mathbf{x_1}, ..., \mathbf{x_n}|\mathbf{w}) = \sum_{a=1}^{N} E_a(\mathbf{x_a}|\mathbf{w}) = \sum_{a=1}^{N} \frac{1}{2}(y_{ANN,a} - \hat{y}_a)^2 \tag{3.10}$$

where $\mathbf{w}$ denotes the ensemble of adjustable weights in the network. The set of weights that optimize the error function can be found using the steepest or gradient descent method, provided that the neuron response function is differentiable concerning the input weights. Starting from a random set of weights $\mathbf{w}^\rho$, weights are updated by,

$$\mathbf{w}^{(\rho+\mathbf{1})} = w^{(\rho)} - \eta \nabla_w E \tag{3.11}$$

where η is the learning rate. The weights related to the output layer are updated by,

$$\Delta w_{j1}^{(2)} = -\eta \sum_{a=1}^{N} (y_{ANN,a} - \hat{y}_a) y_{j,a}^{(2)} \tag{3.12}$$

and weights concerning the hidden layers are updated by,

$$\Delta w_{ij}^{(1)} = -\eta \sum_{a=1}^{N} (y_{ANN,a} - \hat{y}_a) y_{j,a}^{(2)} (1 - y_{j,a}^{(2)}) w_{j1}^{(2)} x_{i,a} \tag{3.13}$$

This back-and-forth up-gradation procedure continues until optimal performance is achieved.

3.3.5. Variable ranking

Variable ranking in the MLP Neural is given by the sum of the weights-squared of the connections between the variable's neuron in the input layer and the first hidden layer. The importance of the input variable I_i of the input variable i is given by,

$$I_i = \bar{x_i}^2 \sum_{j=1}^{n} (w_{ij}^{(1)})^2, \;\; i = 1, ..., n \tag{3.14}$$

where $\bar{x_i}$ is the sample mean of input variable i.

3.3.6. MLP hyper-parameters

Here I summarised the list of hyperparameters used for the MLP. Like BDT, I also used 80% of the total events for training and 20% for testing.

Parameters	Values	Description
NCycles	600	Number of training cycles.
HiddenLayers	N+5	Specification of hidden layer architecture.
NeuronType	tanh	neuron activation function type.
EstimatorType	CE (Cross- Entropy)	Error estimator.
NeuronInputType	sum	neuron input function type.
TrainingMethod	BP	MLP training algorithm.
LearningRate	0.02	learning rate parameter.
UseRegulator	True	This feature is used to avoid over-training.

Table 3.2.: List of MLP hyper-parameters.

3.4. Inputs from ECL and other sub-detectors

As mentioned, separating the low momentum lepton-pion tracks with improved precision is one of our primary goals. Our MVA algorithm mainly exploits the variables from the ECL governing different physics processes related to tracks and the likelihoods from other sub-detectors. When particles traverse through ECL clusters, they deposit their energy and form showers. Different particles produce different showers—for example, electromagnetically interacting particles like γ and e form a radially symmetric shower. Electron-induced showers are very similar to photons but lead to more fuzzy shower shapes due to Bremsstrahlung loss. Neutral hadrons, mostly K_L^0's and neutrons, don't interact with ECL (punch through). However, a fraction of these particles interact with ECL and form irregular and non-radially symmetric shower shapes. If a charged hadron ($\pi^\pm$ or $K^\pm$) interact with ECL clusters, they form irregularly shaped showers in addition to a tilted entry angle due to the magnetic field. If they don't interact (punch through), they leave the same signal as muons. Minimal ionizing particles (muons) and some hadrons don't interact inelastically and usually leave energy in just one or two crystals. Therefore, shower shape variables from ECL are very important for charge particle identification [69]. Below I have briefly described the variables considered for this study.

$E_{cluster}$:

Energy deposited by the particles in the ECL cluster. Different energy distributions are expected for electrons, muons, and pions depending on the particle's interaction with ECL cluster material (CsI).

E/p:

Ratio between the cluster energy and the track momentum. It is also expected to differ between electromagnetically interacting and minimally ionizing particles.

E_1/E_9:

Ratio of energies of the central crystal, E_1, and 3×3 crystals, E_9, around the central crystal,

$$E_1/E_9 = \frac{E_1}{\sum_{3\times3} E_i} \tag{3.15}$$

ratio is ≤ 1 as $E_1 \leq E_9$, and the ratio tends towards larger values for photons and smaller values for hadrons.

E_9/E_{21}:

Ratio of energies of the inner 3×3 crystals, E_9, and 5×5 crystals around the central crystal without corners,

$$E_9/E_{21} = \frac{\sum_{3\times3} E_i}{\sum_{5\times5} E_i - \sum_{4}^{corner} E_i} \tag{3.16}$$

ratio is ≤ 1 as $E_9 \leq E_{21}$, and the ratio tends towards larger values for photons and smaller values for hadrons.

Zernike moments $\mid Z_{40} \mid$, $\mid Z_{51} \mid$, and $\mid Z_{MVA} \mid$:

Zernike first introduced Zernike polynomials [70]. They are useful in optics and play a vital role in diffraction theory [71]. They are also helpful for image recognition [72]. BaBar and ZEUS experiments use Zernike moments to distinguish between electromagnetically and hadronically interacting particles [69]. Belle didn't use Zernike moments for PID, but in Belle II, they are available in the Belle II ECL software for analysis use [73].

Zernike moments for the ECL cluster is given by,

$$\mid Z_{mn} \mid = \frac{n+1}{\pi} \frac{1}{\sum_i w_i E_i} \mid \sum_i R_{nm}(\rho_i) e^{-im\alpha_i} w_i E_i \mid, \tag{3.17}$$

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s!(\frac{(n+|m|)}{2}-s)!(\frac{(n-|m|)}{2}-s)!} \rho^{n-2s} \tag{3.18}$$

Where n, m determine the moment rank, E_i is the energy of i-th crystal in the shower, ρ_i is the radial distance of the i-th crystal in the perpendicular plane. As one crystal could be related to multiple showers, w_i is the fraction of the energy of the i-th crystal associated with the shower.

One could have different Zernike moments depending on the values of n and m. Saving all the moments for each ECL cluster at the detector simulation level is not feasible, as it would take up too much disk space. An MVA study was performed for K_L^0 and γ separation using different sets of Zernike moments. It was concluded that $\mid Z_{40} \mid$ and $\mid Z_{51} \mid$ are the most useful and saved; the output score $\mid Z_{MVA} \mid$ is also kept at the detector level and available for analysis [74]. Although this performance is for $K_L^0 - \gamma$ separation, it could provide some lower-level information and be useful for charged particle identification.

Longitudinal Shower Depth (Δ L):

The longitudinal shower depth is also a crucial variable for charged particle identification. BaBar experiment first introduced this variable for particle identification, and they observed higher identification efficiency and lower fake rate for e, μ, and π separation [75]. It is defined as the distance between the trajectory into the ECL from the track entry point in the calorimeter to the point of closest approach to the cluster centroid. For a more accessible illustration, a schematic view of how ΔL is defined is given in figure 3.2

Lateral moment (LAT):

Lateral shower moment is defined as,

$$LAT = \frac{\sum_{i=3}^{N} w_i E_i r_i^2}{\sum_{i=3}^{N} w_i E_i r_i^2 + (w_1 E_1 + w_2 E_2) r_0^2} \tag{3.19}$$

where N is the total number of crystals associated with a shower, E_i is the energy deposited in the i-th crystal such that $E_1 > E_2 > ... > E_N$, w_i is the crystal weight. r_i is the distance of the i-th crystal to the shower center projected to a plane perpendicular to the shower axis, and $r_0 = 5$ cm, which is approximately the average distance between two crystals.

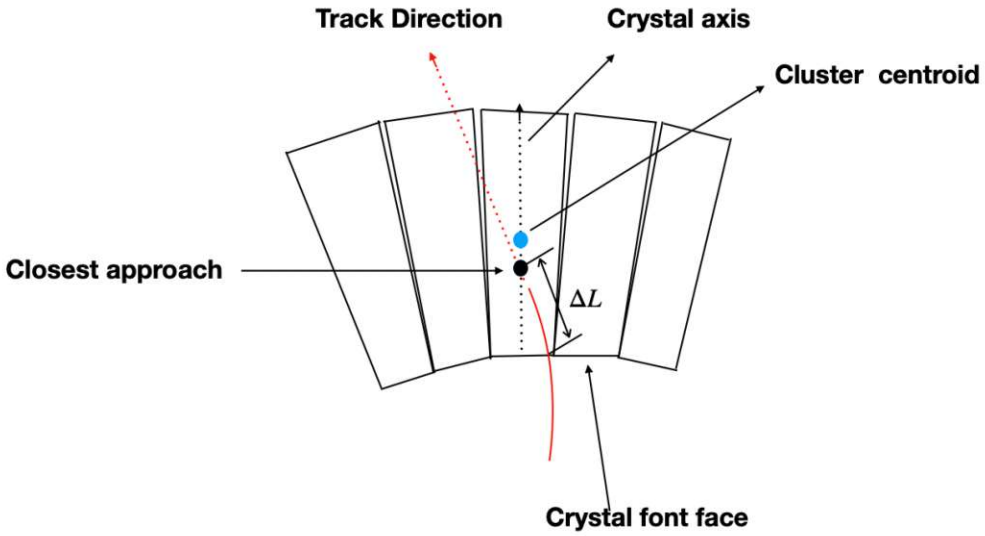


Figure 3.2.: Schematic view of how ΔL is defined.

Likelihoods from other sub-detectors ($\Delta log\mathcal{L}_i^{detector}$, $i \in (e, \mu, \pi)$ and $detector \in (CDC, ARICH, TOP, KLM)$):

As described previously, our study concentrates mainly on ECL; however, we have other important sub-detector components dedicated to the tracking (CDC, SVD) and PID (TOP, ARICH, KLM) for covering all the momentum and angular phase space of the detector. We took likelihoods from other sub-detector components. Actually, we gave a $\Delta log\mathcal{L}_i^{detector}$ discriminator as an input to the MLP, which is defined as follows,

$$\Delta log\mathcal{L}_i^{detector} = log\mathcal{L}_i^{detector}(x|e,\mu) - log\mathcal{L}_i^{detector}(x|\pi) \tag{3.20}$$

for the signal hypothesis (e, μ) and the background hypothesis (π).

Which eventually turns our MVA into a global MVA (ECL + all other subdetectors likelihood). In chapter 4, I performed a dedicated MVA study considering lower-level information from CDC only in an attempt to replace the likelihood information used in the global MVA.

3.5. Variable pretreatment

MLP is very sensitive to the input variables; it can't handle the arbitrarily large number and missing values, so a variable pre-treatment is desirable. I used MinMaxScaler algorithm [76] to map the input variable in the range (0, 1), and then the available missing values are assigned to -1. In the section appendices A, normalized plots of all the variables are shown for the Barrel region only.

3.6. Algorithm

3.6.1. Event generation and Simulation

- Signal particle samples $e^{\pm}$, $\mu^{\pm}$, and $\pi^{\pm}$ are generated with ParticleGun Module available in the Belle II Analysis Simulation Framework (BASF2) [77].
- 2 Million events per particle generated with uniform momentum (p) $\in$ (0.05 - 5.5) Gev/c^2, $\theta \in (0^o, 180^o)$, and $\phi \in (0^o, 360^o)$ distribution.
- GEANT4 [78] is used to reproduce interactions of particles traversing the Belle II detector, considering the varying detector conditions and beam backgrounds.
- Then particle to particle reconstruction has performed. All generation, simulation and reconstruction has performed using BASF2 release-05-02-00.

3.6.2. Reconstruction selection

- Impact parameter cut : | dr |$<$ 2.0 cm and | dz |$<$ 5.0 cm.
- In order to exploit calorimetric info, only tracks that have a matching ECLCluster are retained, which corresponds to minimum energy of 0.2 GeV.
- For electrons, Bremsstrahlung correction has been done using the correctBrems Module in basf2. For that we are using γ energy(E)$<$ 0.1 GeV.

3.6.3. Momentum and Angular binning

- The ECL shower shapes generally depend on a particle's momentum and geometrical effects related to the calorimeter structure. Furthermore, the other sub-detectors are often defined only in specific subsets of the full detector acceptance. Therefore, a categorization is performed in reconstructed track momentum (p_{lab}) (three regions of low, medium, and high momentum) and

ECL cluster polar angle ($\theta_{cluster}$) (three regions of ECL forward, backward endcaps, and barrel region), as outlined in Table 3.3. A pictorial view is given in Figure 3.3

Polar angle $\theta_{cluster}$ [rad]	Track momentum p_{lab} [GeV/c^2]
0.21 - 0.56 (ECL FWD endcap)	0.2 - 0.6 (low momentum)
0.56 - 2.24 (ECL Barrel)	0.6 - 1.0 (medium momentum)
2.24 - 2.70 (ECL BWD endcap)	$>$ 1.0 ((high momentum))

Table 3.3.: The ECL cluster polar angle and track momentum ranges that define the 9 independent categories where BDTs and MLPs are trained.

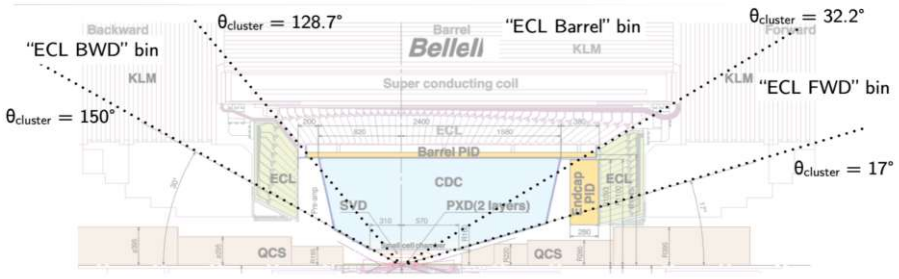


Figure 3.3.: Graphical view of different angular regions.

- As our primary focus is on the ECL, we need a minimum of 0.2 GeV momentum to reach the ECL, which explains the reason of starting low momentum bin from 0.2 GeV. For momentum below 0.2 GeV, we have to rely on the tracking detectors (CDC, SVD), by exploiting the information about the energy loss one could extend the momentum region below 0.2 GeV, studies from chapter 4 are very helpful on that purpose.

  There is no apparent reason why I chose the other bins like that, and it depends on the analysis; maybe some other angular and momentum could be useful. We did some study by looking at the 2D distribution of features and particle momentum for different angular regions, but we didn't observe any problems with these chosen regions.

3.6.4. Set of variables for different training regions

We are considering the tracks must reach the ECL. The variables related to the ECL are included in all the momentum and angular training region; for the likelihoods depending on the momentum and angular bins, available sub-detector likelihoods are included. Table 3.4 summarises the included sub-detector components for training depending on momentum and angular bins.

separation	$\theta_{cluster}$	$0.2 < p_{lab} < 0.6$	$0.6 < p_{lab} < 1.0$	$p_{lab} > 1.0$
e-π, μ-π and e - μ	ECL Barrel	CDC, TOP, ECL	CDC, TOP, ECL, KLM	CDC, TOP, ECL, KLM
	ECL FWD	CDC, ARICH, ECL	CDC, ARICH, ECL, KLM	CDC, ARICH, ECL, KLM
	ECL BWD	CDC, ECL	CDC, ECL, KLM	CDC, ECL, KLM

Table 3.4.: The included sub-detector components for training depend on momentum and angular bins.

3.7. Performance

Here I present a comparative performance between BDT and MLP-based particle identification. Figure 3.4 and Figure 3.5 show the signal efficiency vs. background rejection (ROC) curve depending on different identification hypotheses and training regions. The BDT and MLP-based algorithms provide similar performance in most training regions apart from the low momentum ECL Barrel region (which covers most of the detector phase-space), where we observed 1-2% better performance for $e-\pi$ and 2-3% improvement for $\mu-\pi$ than BDT (Figure 3.4). Figure 3.5 shows the ROC curve for some other regions where the performance of BDT and MLP are similar. The observed improved performance could be due to the Neural network's (a lon-linear discriminator) better handling of non-linear features (curved low momentum tracks in CDC) than BDT.

Figure 3.6, 3.7 and 3.8 show the linear co-relation matrix among $e-\pi$, $\mu-\pi$, and $e-\mu$ separation variables for low momentum ECL barrel region (where we observed improved performance). As one can see, most of the variables are uncorrelated (they describe entirely different properties of particle track) apart from cluster energy and E/p (expected); however, they help to get some lower-level non-linear information.

The section A.2 describes the variable ranking (in decreasing order of importance) for all the training regions. From the tables, one can see that for $e-\pi$, and $e-\mu$ separation, E/p, and $E_{cluster}$ is the most crucial variable (as electrons are identifiable by ECL). Different detector likelihoods also provide strong separation power depending on the angular regions. For low momentum, $\mu-\pi$ separation, $E_{cluster}$ and the likelihood from CDC delivers essential information. In contrast, for high momentum (able to reach KLM), most of the separation power comes from KLM likelihood. Tables also compare MLP and BDT features ranking, although a direct comparison is not possible because ranking is very method specific.

The improved performance may be the effect of overtraining, but as shown in Figure 3.9, no overtraining is observed (training and testing events are perfectly superimposed).

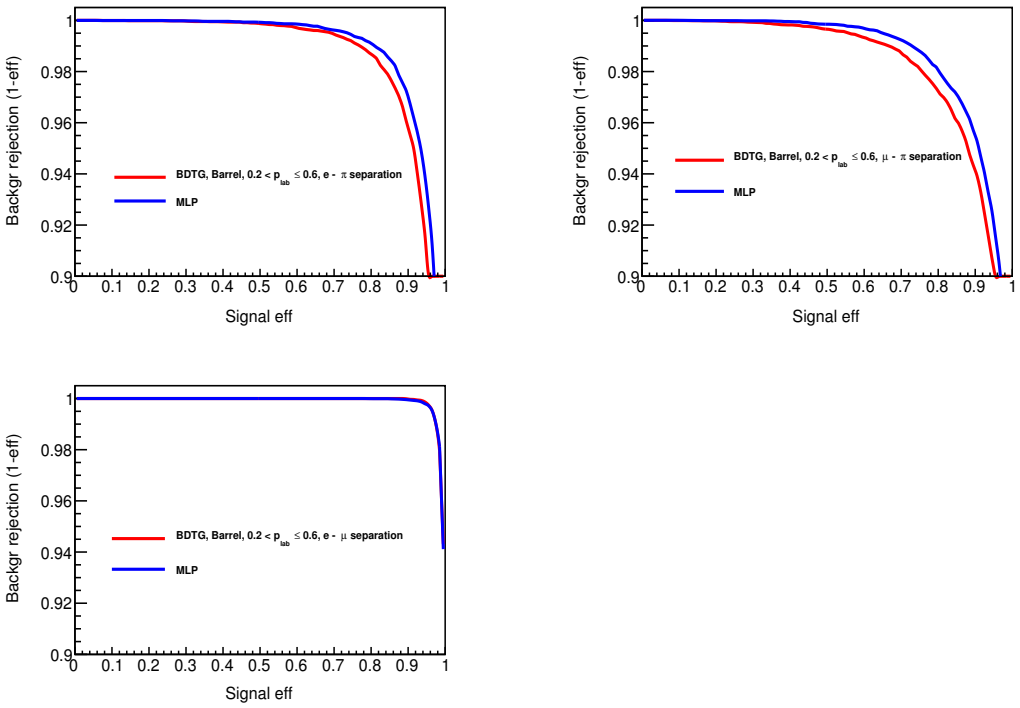


Figure 3.4.: From top left → top right → bottom left, $e-\pi$, $\mu-\pi$, $e-\mu$ separation performance for low momentum ECL barrel region.

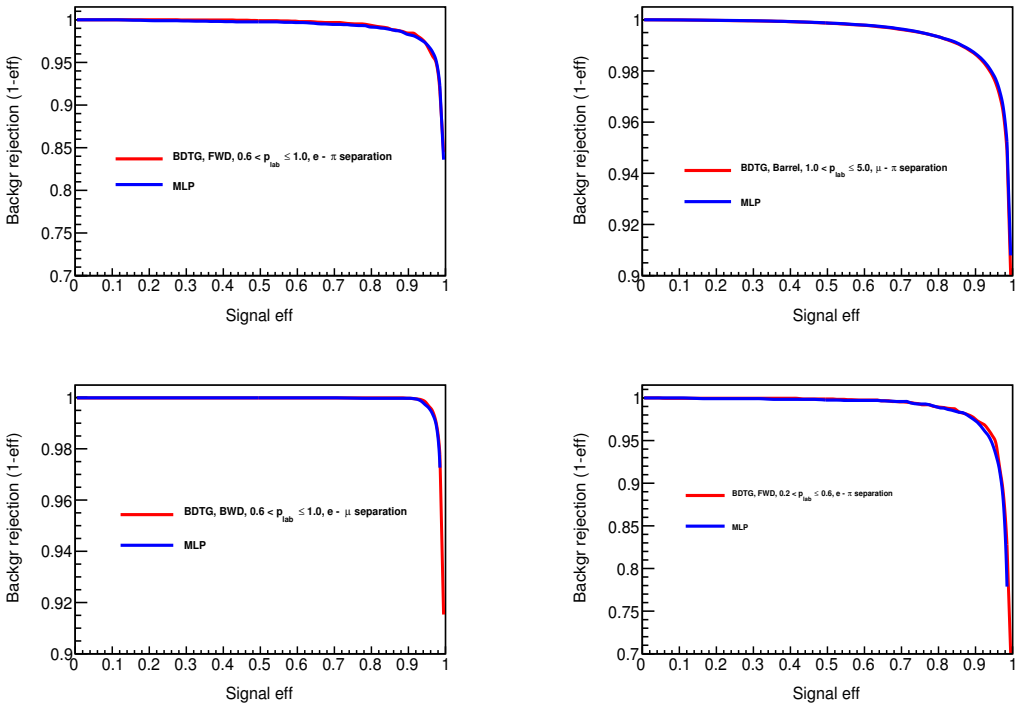


Figure 3.5.: From top left → top right → bottom left → bottom right, $e - \pi$, $\mu - \pi$, $e - \mu$ and $e - \pi$ separation performance for medium momentum FWD region, high momentum Barrel region, medium momentum BWD region and low momentum FWD region respectively.

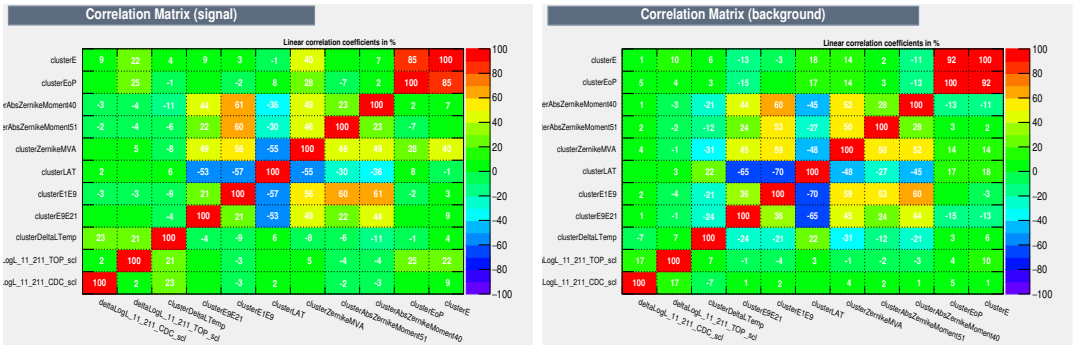


Figure 3.6.: Co-relation matrix of $e - \pi$ separation for low momentum ECL Barrel region. Here signal $\implies$ e, background $\implies$ π.

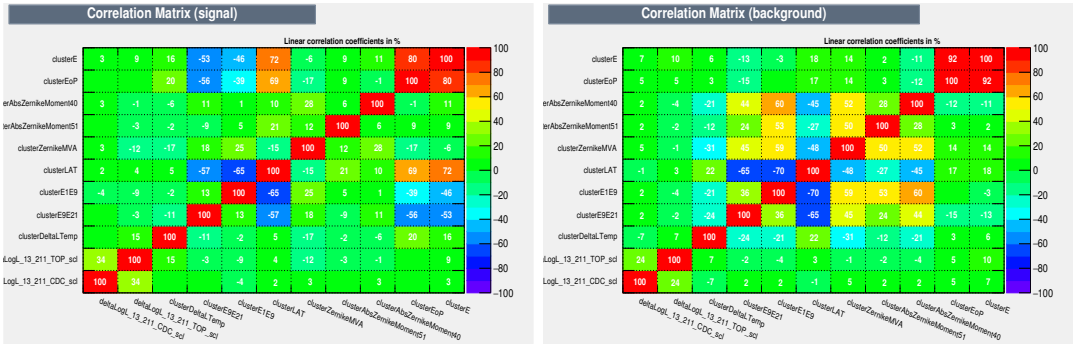


Figure 3.7.: Co-relation matrix of $\mu - \pi$ separation for low momentum ECL Barrel region. Here signal $\implies \mu$, background $\implies \pi$.

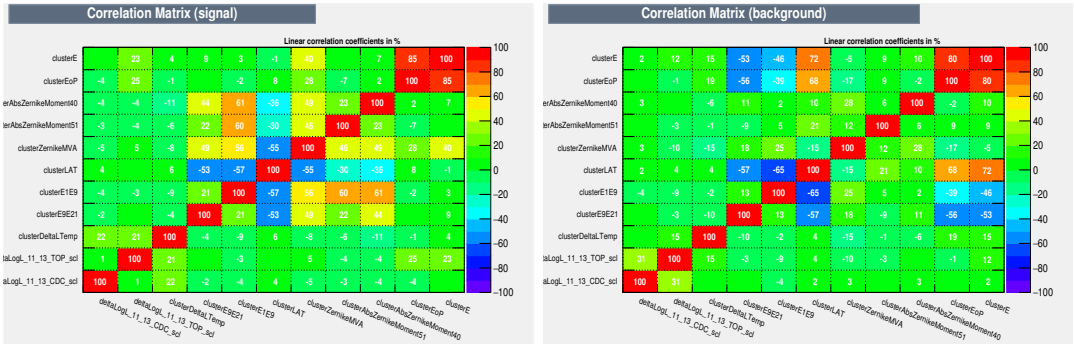


Figure 3.8.: Co-relation matrix of $e - \mu$ separation for low momentum ECL Barrel region. Here signal $\implies e$, background $\implies \mu$.

3.8. Conclusions

BDT-based particle identification algorithm is already implemented in BASF2 and is available for analysis use. I showed here that combining several calorimetric measurements and particle likelihoods from other sub-detectors in a neural network (MLP) gives promising improvements in the Belle II lepton identification performance, especially in the critical low momentum region, where we observed 1-2% better performance for $e - \pi$ and 2-3% improvement for $\mu - \pi$ than BDT. This clearly indicates that a deep neural network could eventually be helpful in gaining higher performance (by adding more hidden layers in the network architecture). Belle II aims to gather 50 ab^{-1} of data in the near future, which comes along with higher backgrounds. In those conditions, higher identification performance would be achievable by a Deep Neural Network.

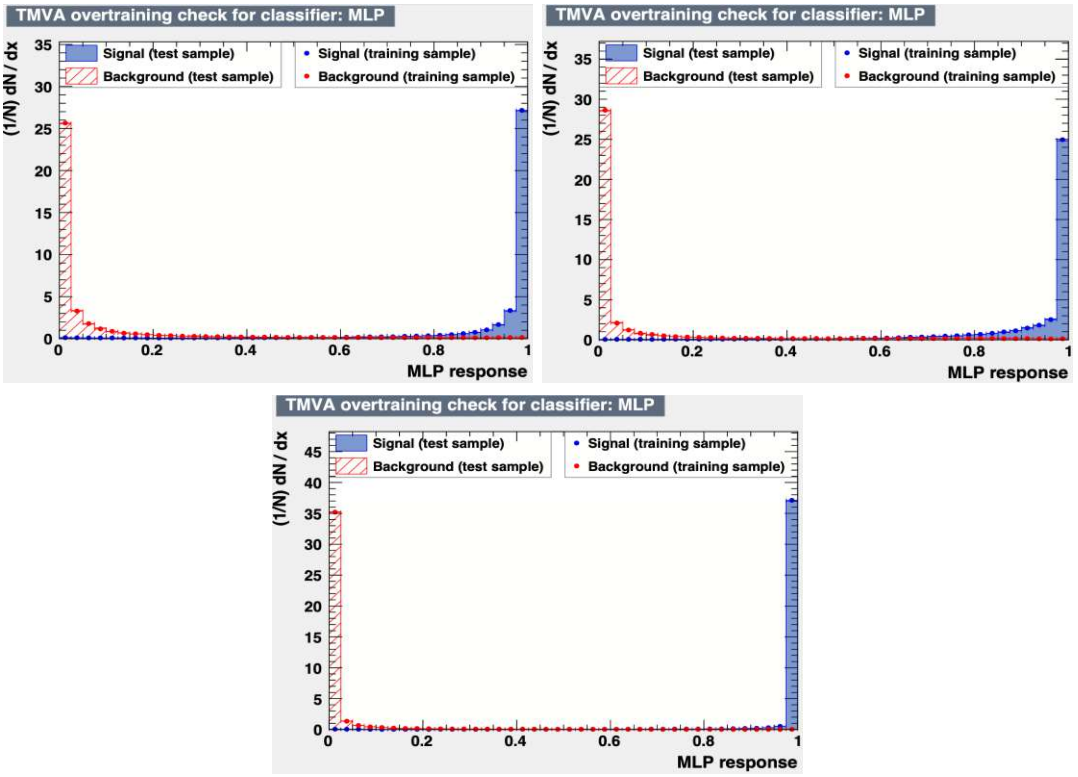


Figure 3.9.: Over train check for low momentum $e - \pi$, $\mu - \pi$, $e - \mu$ separation respectively (Barrel region only).

Here results are shown for binary classification, although the method can be easily extended to multi-class particle identification. Furthermore, additional discriminating variables such as the ECL pulse shape discrimination [79], newly introduced in Belle II, could be helpful to improve the algorithm performance further. I didn't use any information from the vertex detectors for the particle separation. One could also improve the algorithm performance by exploiting the lower-level information (dE/dx) from SVD.

4. Lepton identification using Boosted Decision Trees (BDT) at Central drift chamber (CDC) of Belle II detector

The following work was done solely by the author, while additional guidance was provided by PID group convenors and Gianluca Inguglia.

4.1. Introduction

In the last chapter, I developed an MVA-based particle identification algorithm for separating leptons (mainly μ and e) from pions (π). The algorithm was based on the crucial information from the ECL and the likelihoods from other sub-detectors to give the full description of the particle identification algorithm (global MVA).

In this chapter, I concentrate only on the CDC of the Belle II detector in an aim to replace the likelihood information used in global MVA-based particle identification by exploring some lower-level information at the CDC level.

4.2. Likelihood ratio-based lepton Identification at Belle II

Before going to the main subject of this chapter, I would like to introduce the likelihood based stable charged particle identification at Belle II. Here, "stable" means the charged particles that live long enough to travel across entire subsections of the detector. I focus solely on electrons, muons, and charged pions in this study. At Belle II, likelihood based charged stable particle identification is based on the combination of measurements from the various sub-detectors [53]:

- Particle energy loss (dE/dx) information from the silicon vertex detector (SVD) and central drift chamber (CDC).

- Measurements of the mass-dependent optical response of time-of-propagation detectors (TOP) and ring-imaging Cherenkov detectors (ARICH).

- Measurements of the energy deposition in the CsI(Tl) scintillation crystals of the electromagnetic calorimeter (ECL).

- Measurements in the muon detector (KLM) to account for the different penetration ranges of scattered muons and hadrons.

In chapter 2, I briefly described the procedure of measuring these properties related to the tracks and the process of constructing likelihood depending on that. Below I describe the method of constructing a global or binary likelihood;

In each sub-detector (d), a likelihood ($\mathcal{L}^d(\mathbf{x}|i)$) is defined for each charged particle hypothesis ($i \in e, \mu, \pi, K, p, d$) as a joint probability density function (PDF) of a given set of observables, $\mathbf{x}$. The PDFs are either predicted from simulation, extracted from data control samples with high purity, or determined analytically. Assuming subdetectors' measurements of each of the identifying observables are independent, a global likelihood for each particle hypothesis i is defined by:

$$\mathcal{L}(\mathbf{x}|i) = \prod_{d}^{d\in\{SVD,CDC,\dots\}} \mathcal{L}^d(\mathbf{x}|i) \quad \text{or equivalently,} \tag{4.1}$$

$$\log\mathcal{L}(\mathbf{x}|i) = \sum_{d}^{d\in\{SVD,CDC,\dots\}} \log\mathcal{L}^d(\mathbf{x}|i). \tag{4.2}$$

Given all possible, mutually exclusive outcomes of identification, $\{A_j\}=\{e, \mu, \dots\}$ for a reconstructed particle candidate, the global likelihood ratio provides the information for identifying such candidate as A_i, using Bayes' theorem one could show that,

$$P(A_i|\mathbf{x}) = \frac{P(\mathbf{x}|A_i)\cdot P(A_i)}{\sum_j P(\mathbf{x}|A_j)P(A_j)} \quad \Rightarrow P(i|\mathbf{x}) = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j}. \tag{4.3}$$

In the latter, we assume that the prior probabilities, $P(A_j)$, are identical for any j, so the likelihood ratio can be interpreted as an actual identification probability. From the individual likelihoods, it is also possible to build binary likelihood ratio discriminators between two hypotheses, i and j:

$$P(i/j|\mathbf{x}) = \frac{\mathcal{L}_i}{\mathcal{L}_i + \mathcal{L}_j}. \tag{4.4}$$

4.3. Accessing likelihood information from CDC

Here I describe the procedure for accessing the likelihood information from the CDC of the Belle II detector.

Ionization current (dE/dx) measurements are obtained for each reconstructed track in the CDC. Signal pulses on each CDC wire are digitized with 31.75 MSPS flash ADCs (see chapter 2 for detector configuration), and values over a nominal threshold are summed to yield the raw ADC readout [80]. Corrections are applied to the track geometry, in particular to the projected $r - z$ path length, based on the polar angle of the track, and the projected $r - \phi$ path length, based on the track geometry in the cell. Wire-to-wire and run-to-run gain variations are also calibrated with high-statistics eeγ events.

Finally, samples of e, μ, π, K, and p tracks from the dedicated control channels are used to determine the gas-gain saturation effects for tracks with varying intrinsic ionization (i.e., saturation relative to electrons) and also to obtain the parameterization of the expected dE/dx as a function of $\beta\gamma = p/m$. Figure 4.1 shows the predicted dE/dx for six long-lived charged-particle hypotheses as a function of momentum.

Then the corrected ADC values from each hit are used to calculate a truncated mean defined by the lowest 5% and highest 25% of measurements on a given track are discarded, and the remainder are averaged. Distributions of this measured truncated mean used to calculate a normalized deviation for each charged particle hypothesis $h \in \{e, \mu, \pi, K, p\}$:

$$\chi_h = \frac{\Delta_{dE/dx}}{\sigma_{\text{prediction}}} = \frac{(dE/dx_{\text{measurement}} - dE/dx_{\text{prediction}})}{\sigma_{\text{prediction}}} \tag{4.5}$$

Where $\sigma_{\text{prediction}}$ predicted resolution on $\Delta_{dE/dx}$. It is parameterized as a function of the track polar angle, the number of hits of the track, and the dE/dx. The χ_h distributions are converted to likelihoods. These likelihoods are then further combined with information from other detectors.

4.4. Boosted Decision Trees (BDT)

I used the same BDT algorithm and hyper-parameters as described in section 3.2. I used BDT to perform this study because our initial plan was to include this measurement in the global MVA, which is BDT-based and available for analysis usage in Belle II software.

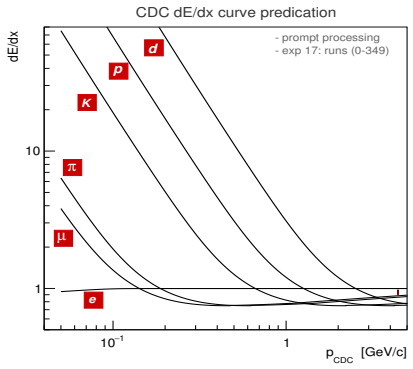


Figure 4.1.: The CDC-dE/dx curve predictions for charged particles.

4.5. Inputs from Central Drift chamber (CDC)

Here I give a complete description of the variables used in the study. In the beginning, I used all the variables listed below; later, I dropped some of them by gaining knowledge about them; see the section 4.7.4 for the final set of variables. Appendices B shows normalized plots of all the variables.

CDCdEdx: Energy deposition of particle tracks at CDC.

CDCdEdxnosat: Energy deposition of particle tracks at CDC without hadron saturation corrections.

CDCdEdx/p_{LAB}: Ratio of particle's energy deposition in CDC over particle's track momentum. Initially, I used this variable for separation inspired by the E/p ratio from ECL; although it doesn't make much sense for CDC, later I dropped it; see the section 4.7.4.

costhCDC: Particle's angular distribution valid in the CDC.

CDCdEdx-lnhits: CDC layer hits for the tracks.

χ_h: χ_h is defined as (CDCdEdx - prediction$_h$)/resolution$_h$, Here, the prediction depends only on $\beta\gamma$ = p/m, but since the mass appears, this is hypothesis-dependent. The resolution is a function of three variables: theta, the number of layers used (after 'truncation'), and the predicted mean. Here h $\in$ e, μ, π.

p^h: Predicted momentum resolution at CDC, h ∈ e, μ, π.

One could see that among the listed variables, most of them are co-related, but BDT is relatively insensitive to that. Because each node splitting of a growing tree is based on the most separable variable at that node, if there have equal importance variable, it will just drop it and use the most important one.

4.6. Algorithm

4.6.1. Event generation and Simulation

- Signal particle samples $e^{\pm}$, $\mu^{\pm}$, and $\pi^{\pm}$ are generated with ParticleGun Module available in the Belle II Analysis Simulation Framework (BASF2) [77].
- 2 Million events per particle generated with uniform momentum (p) ∈ (0.05 - 5.5) Gev/c^2, $\theta \in (0^o, 180^o)$, and $\phi \in (0^o, 360^o)$ distribution.
- GEANT4 [78] is used to reproduce interactions of particles traversing the Belle II detector, considering the varying detector conditions and beam backgrounds. Here simulation is a little bit tricky. We don't have all the above mentioned variables in the mDST (mini-Data Summary Table) level; we have to access the cDST (calibrated Data Summary Table) information for our study.
- Then particle to particle reconstruction has been performed. All generation, simulation, and reconstruction have been performed using BASF2 release-05-02-00.

4.6.2. Reconstruction selection

- Impact parameter cut : | dr |< 2.0 cm and | dz |< 5.0 cm.
- Only particles within CDC acceptance are considered.
- For electrons, Bremsstrahlung correction, has been applied using correctBrems Module in basf2. For that, we are using γ energy(E)< 0.1 GeV.

4.6.3. Momentum and Angular binning

- I generated particle's momentum (p) ∈ (0.05-5.5) Gev/c^2, then divided the total momentum range into 4 sub-momentum ranges, p_{low} ∈ (0.1 - 0.2) Gev/c^2, p_{medium_low} ∈ (0.2 - 0.6) Gev/c^2, p_{medium_high} ∈ (0.6 - 1.0) Gev/c^2 and p_{high} ∈ (1.0 - 5.0) Gev/c^2 depending on the layer hits information from

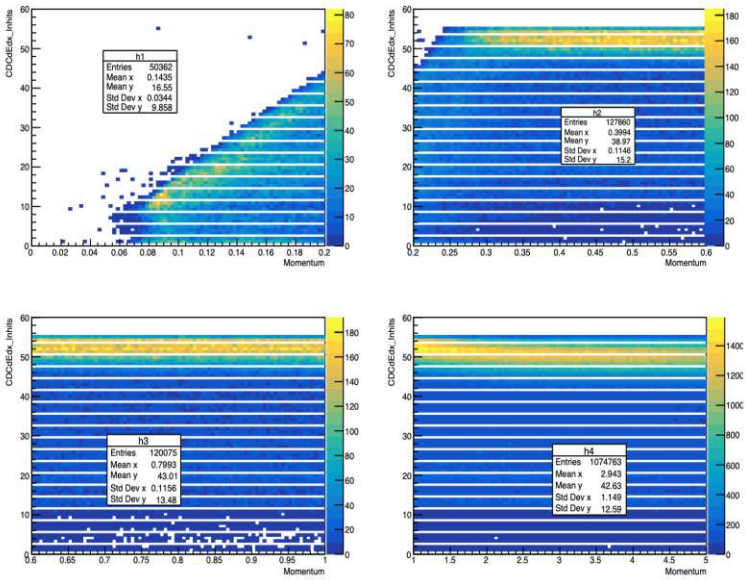


Figure 4.2.: CDC layer hits vs momentum

the CDC (see Figure 4.2). The optimal choice of binning for the low momentum should be up to ~ 0.24 Gev/c^2, but to be consistent with global BDT based Lepton ID [64] I chose the binning up to 0.2 Gev/c^2.

- Apart from replacing the likelihood information used in global MVA, the other crucial importance of doing CDC-based lepton ID studies is that we could extend our global BDT momentum range up to 0.1 GeV/c^2 (Global BDT lepton ID momentum region is up to 0.2 Gev/c^2, see chapter 3). So this study is useful for the low momentum particle < 0.2 Gev/c^2, which are not capable of going to ECL.
- No selection applied on the angular regions for training (unlike section 3.6.3). But of course, one could always get the angular information of different momentum ranges separately, as we trained the whole region at a time.
- So, total 4 BDT's for 4 different momentum regions.

4.6.4. Set of variables for different training regions

Table 4.1 describes the list of variables used for training for the e-π, e-μ, μ-π cases, initially. I didn't use CDCdEdx and χ^h simultaneously for training because they are fully co-related.

Momentum	e-π BDT	μ-π BDT	e-μ BDT
All momemtum regions	CDCdEdxnosat CDCdEdx/p CDCdEdx-lnhits costhCDC χ_e χ_π p^e p^π	CDCdEdxnosat CDCdEdx/p CDCdEdx-lnhits costhCDC χ_μ p^μ	CDCdEdxnosat CDCdEdx/p CDCdEdx-lnhits costhCDC χ_e χ_μ p^e p^μ

Table 4.1.: List of variables according to different separation types.

4.7. Performance of lepton-pion sepeartion

4.7.1. ROC curves

Figure 4.3 shows the comparison of standard Belle II likelihood and BDT based particle separation for all the momentum regions. Table 4.2, Table 4.3, Table 4.4, Table 4.5 gives the variable ranking for different momentum regions. As mentioned, BDTs are quite insensitive to the choice of inputs, variables that have poor separation will be skipped in the tree growing algorithm by construction, but It is useful to check how much each impacts the final classification response. From the ROC curves and variable ranking for different momentum regions, one could conclude that for e-π, e-μ and μ-π cases, separation power coming mostly from CDCdEdxnosat, costhCDC and χ^h. According to experts, one could expect costhCDC and χ^h to contribute to the separation power, but CDCdEdxnosat should not contribute much, as seen from ROC curves, and should behave the same way as CDCdEdx.

4.7.2. Best separable variable

We didn't expect this kind of perfect separation. I did cross-checks to understand where most of the separation power is coming from. I planned to use variable drop and add strategy to understand the best separable variable. The study concluded when I included the CDCdEdxnosat variable in BDT training and observed a large separation. Results are shown only for $0.2 < p_{lab} \leq 0.6$ (Figure 4.4). The conclusion was also the same for other momentum regions. Figure 4.4, left and right, have the variable sequence described in Table 4.6 and Table 4.7 respectively, latter case, we introduced the CDCdEdxnosat randomly.

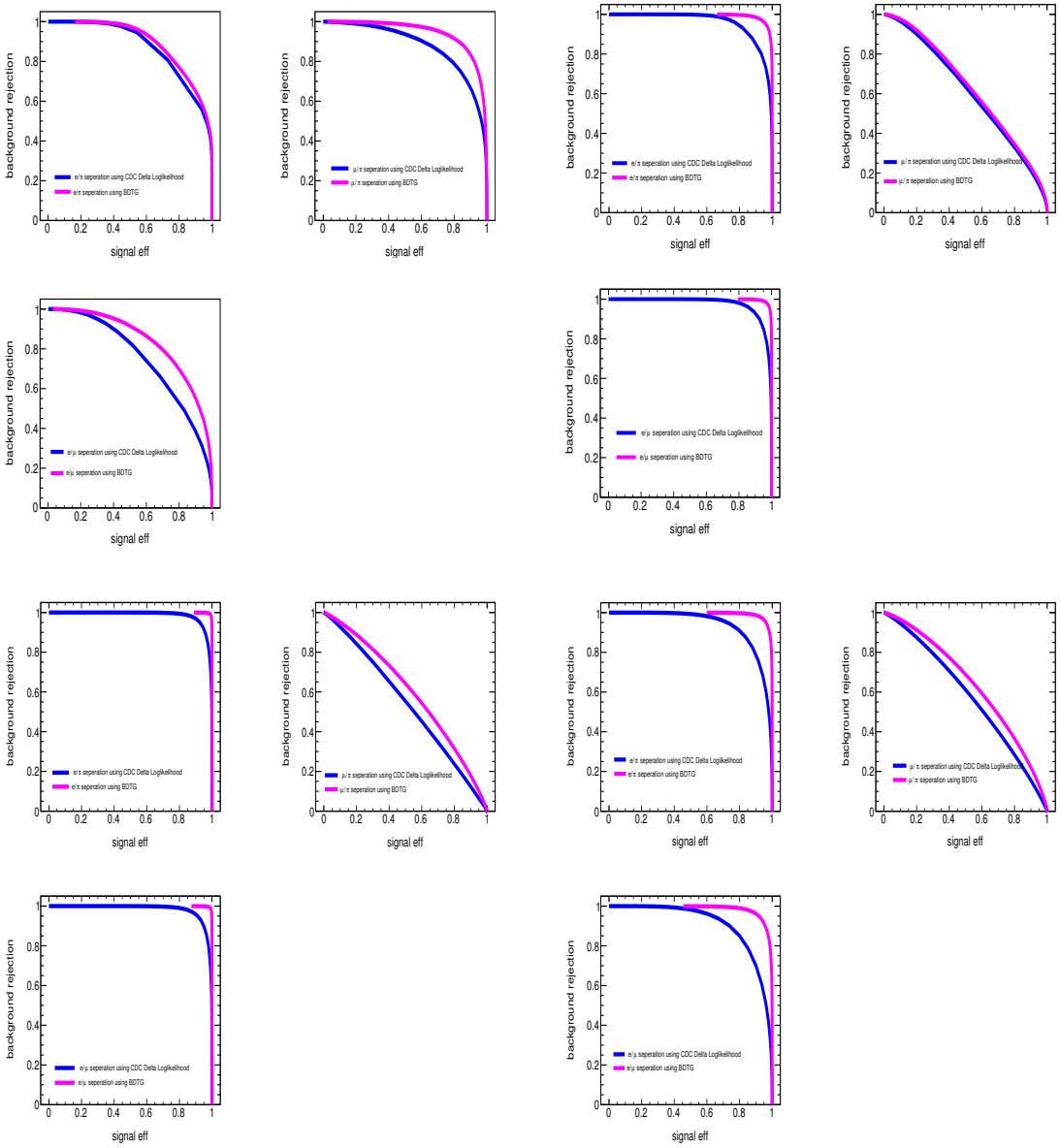


Figure 4.3.: Above left three plots for $0.1 < p_{lab} \leq 0.2$, Above right three plots for $0.2 < p_{lab} \leq 0.6$, below left three plots for $0.6 < p_{lab} \leq 1.0$, below right three plots for $1.0 < p_{lab} \leq 6.0$.

Momentum	e-π BDT	μ-π BDT	e-μ BDT
$0.1 < p_{lab} \leq 0.2$	p^e costhCDC CDCdEdx-lnhits χ_π CDCdEdxnosat χ_e CDCdEdx/p p^π	CDCdEdxnosat costhCDC CDCdEdx-lnhits CDCdEdx/p χ_μ p^μ	CDCdEdxnosat costhCDC p^e CDCdEdx/p CDCdEdx-lnhits χ_μ χ_e p^μ

Table 4.2.: Variable ranking for low momentum region.

Momentum	e-π BDT	μ-π BDT	e-μ BDT
$0.2 < p_{lab} \leq 0.6$	CDCdEdxnosat costhCDC p^e CDCdEdx/p CDCdEdx-lnhits χ_π χ_e p^π	costhCDC χ_μ CDCdEdxnosat CDCdEdx/p CDCdEdx-lnhits p^μ	CDCdEdxnosat costhCDC p^e CDCdEdx-lnhits χ_μ CDCdEdx/p χ_e p^μ

Table 4.3.: Variable ranking for medium low momentum region.

Momentum	e-π BDT	μ-π BDT	e-μ BDT
$0.6 < p_{lab} \leq 1.0$	CDCdEdxnosat costhCDC CDCdEdx-lnhits p^e p^π χ_π χ_e CDCdEdx/p	costhCDC CDCdEdxnosat χ_μ CDCdEdx/p CDCdEdx-lnhits p^μ	CDCdEdxnosat costhCDC CDCdEdx-lnhits p^e p^μ χ_μ χ_e CDCdEdx/p

Table 4.4.: Variable ranking for medium high momentum region.

Momentum	e-π BDT	μ-π BDT	e-μ BDT
$1.0 < p_{lab} \leq 5.0$	CDCdEdxnosat costhCDC p^e CDCdEdx-lnhits χ_π CDCdEdx/p χ_e p^π	costhCDC CDCdEdx-lnhits p^μ CDCdEdxnosat CDCdEdx/p χ_μ	costhCDC CDCdEdxnosat p^e CDCdEdx-lnhits CDCdEdx/p p^μ χ_e χ_μ

Table 4.5.: Variable ranking for high momentum region.

Momentum	1	2	3	4	5	6	7
$0.2 < p_{lab} \leq 0.6$	CDCdEdxnosat CDCdEdx/p CDCdEdx-lnhits costhCDC p^π p^e χ_π χ_e	CDCdEdx/p CDCdEdx-lnhits costhCDC p^π p^e χ_π χ_e	CDCdEdx-lnhits costhCDC p^π p^e χ_π χ_e	costhCDC p^π p^e χ_π χ_e	p^π p^e χ_π χ_e	p^e χ_π χ_e	χ_π χ_e

Table 4.6.: Variable drop and add sequence.

Momentum	1	2	3	4	5	6	7
$0.2 < p_{lab} \leq 0.6$	CDCdEdx-lnhits CDCdEdx/p CDCdEdxnosat costhCDC p^π p^e χ_π χ_e	CDCdEdx/p CDCdEdxnosat costhCDC p^π p^e χ_π χ_e	CDCdEdxnosat costhCDC p^π p^e χ_π χ_e	costhCDC p^π p^e χ_π χ_e	p^π p^e χ_π χ_e	p^e χ_π χ_e	χ_π χ_e

Table 4.7.: Variable drop and add sequence in a random manner for CDCdEdxnosat.

4.7.3. CDCdedx vs CDCdedxnosat

Finally, studies from the above sections concluded that CDCdEdxnosat is the primary variable responsible for the significant separation. But according to the

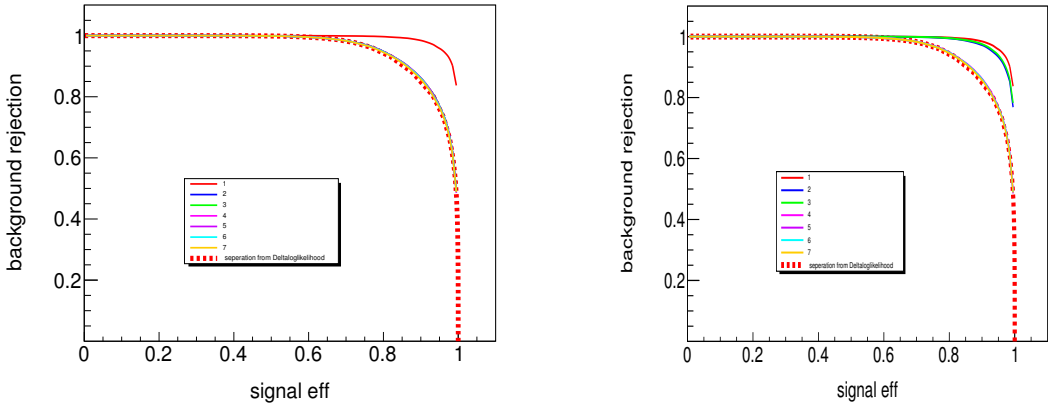


Figure 4.4.: Different ROC curves to know the best separable variable.

expert, CDCdEdxnosat and CDCdEdx should behave similarly. Again, we planned to cross-check by introducing the variable CDCdEdx in training. But we observed some strange behavior from ROC curves and linear correlation shown in figure 4.5 and figure 4.6. I observed CDCdEdxnosat and CDCdEdx behaving complementary to each other at the MC level. I didn't find any correlation between them. I also observed that CDCdEdx and χ_h have a strong correlation, as expected, and planned not to use both simultaneously. After the discussion with the experts, it was concluded that CDCdEdxnosat needs to be better calibrated at the MC level, and we shouldn't use it for analysis purposes.

4.7.4. Final results

After carefully investigating all the variables, I decided table 4.8 as our final set of variables. Figure 4.7 and figure 4.8 describe the signal efficiency vs. background rejection for $0.2 < p_{lab} \leq 0.6$ and $0.6 < p_{lab} \leq 1.0$ for e-π, μ-π, e-μ case. In conclusion, I would like to point out that I didn't observe any vast improvement over likelihood, but some minor improvements have been observed mainly due to angular variable (costhCDC) and hit information from CDC (CDCdEdx-lnhits).

4.8. Conclusions

I did a very preliminary study of lepton-pion identification (binary classification) at the CDC level (using lower-level variables). At first, I observed an improvement over $\Delta log\mathcal{L}_{CDC}$ due to CDCdedxnosat (CDCdedx without saturation correction),

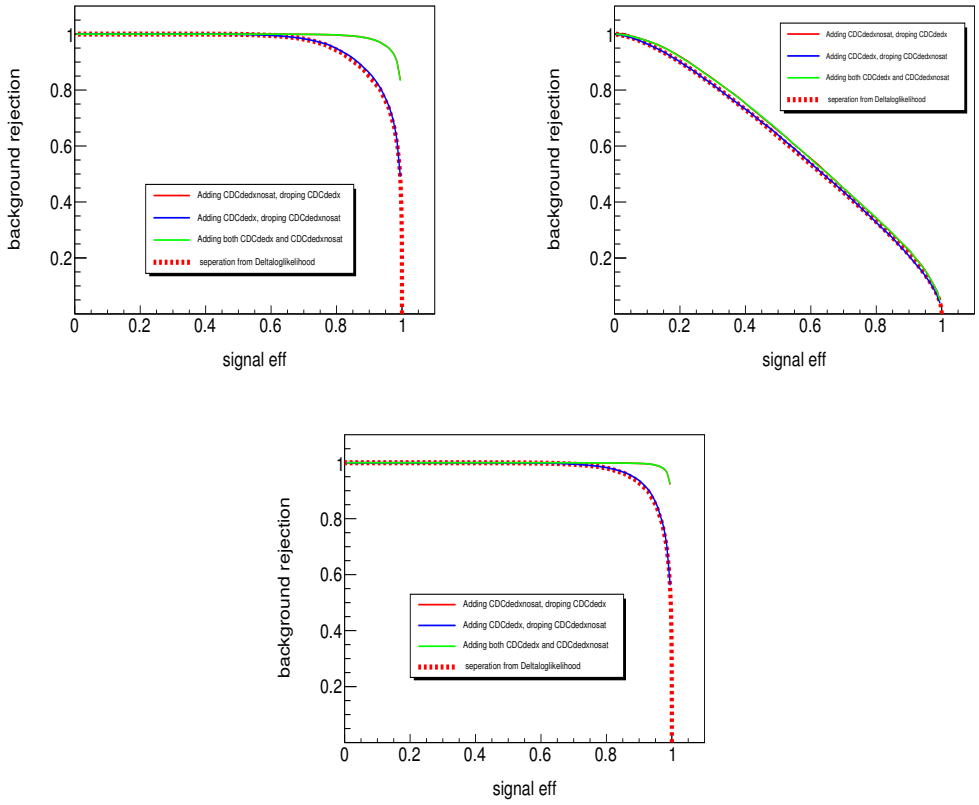


Figure 4.5.: Upper left, upper right, below plots are for e-π, μ-π, e-μ case respectively for $0.2 < p_{lab} \leq 0.6$.

Momentum	e-π BDT	μ-π BDT	e-μ BDT
All momemtum regions	CDCdEdx CDCdEdx-lnhits costhCDC χ_e - χ_π p^e - p^π	CDCdEdx CDCdEdx-lnhits costhCDC χ_μ - χ_π p^μ - p^π	CDCdEdx CDCdEdx-lnhits costhCDC χ_e - χ_μ p^e - p^μ

Table 4.8.: Final list of variables.

but experts suggested to drop it as it was poorly calibrated at the MC level. After dropping it, I didn't observe much improvement over $\Delta log\mathcal{L}_{CDC}$. But still, some improvements mainly come from CDC hit information and angular distribution

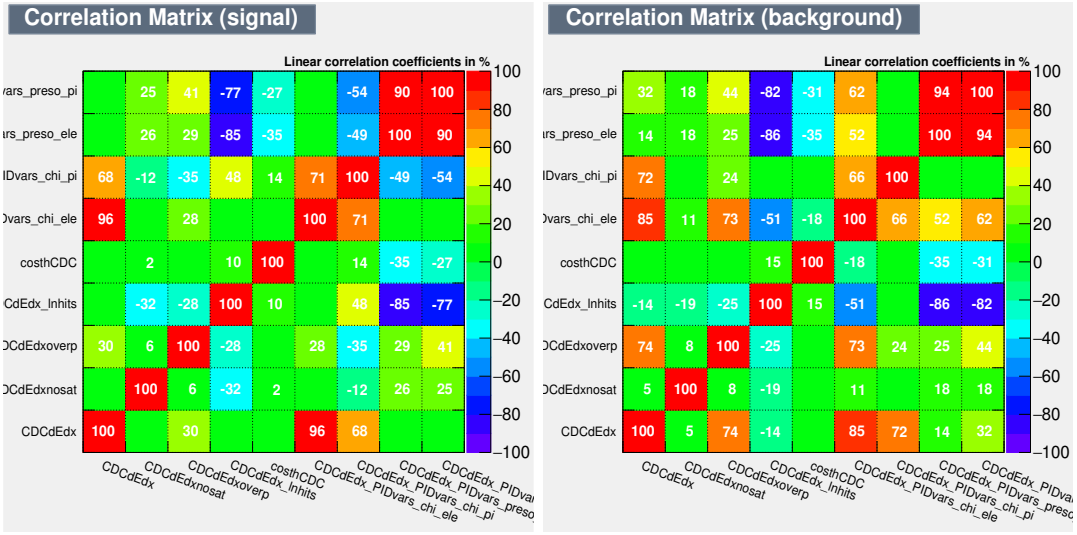


Figure 4.6.: Linear correaltion matrix of e-π case (e is Signal π is background) for $0.2 < p_{lab} \leq 0.6$.

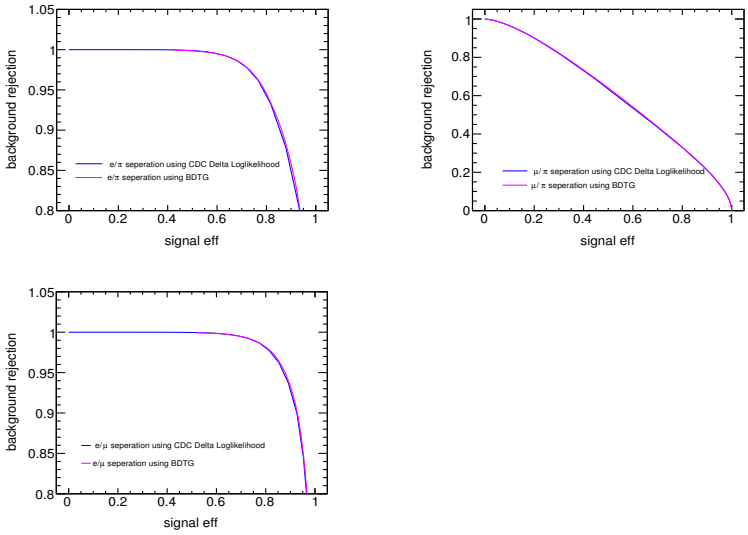


Figure 4.7.: Upper left, upper right, below plots are for e-π, μ-π, e-μ case respectively for $0.2 < p_{lab} \leq 0.6$.

variables. So, if we drop the variable $\Delta log\mathcal{L}_{CDC}$ from global BDT, instead, if we use χ_h, CDCdEdx-lnhits, and costhCDC, or CDC BDT response as an input variable

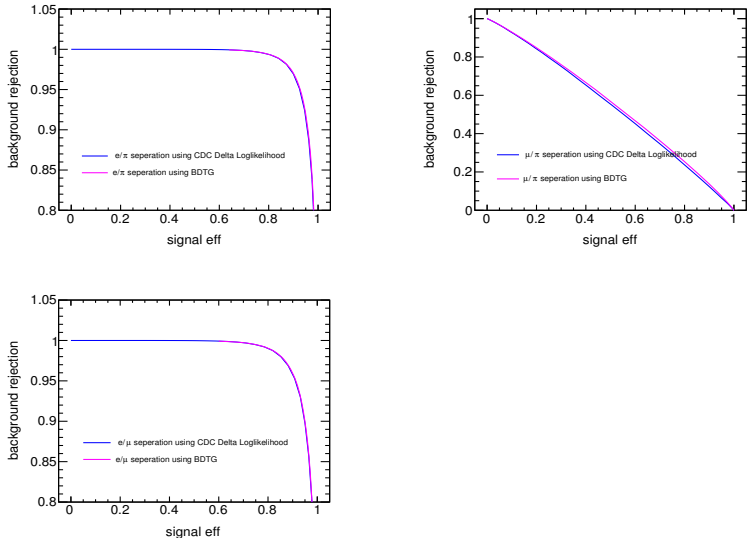


Figure 4.8.: Upper left, upper right, below plots are for e-π, μ-π, e-μ case respectively for $0.6 < p_{lab} \leq 1.0$.

one could expect better performance. This study is also helpful for extending the global BDT training region below 0.2 GeV/c^2 (up to 0.1 GeV/c^2). An internal Belle II note is submitted to the collaboration for review [81].

5. Trigger studies for low multiplicity events at Belle II

The following work was done solely by the author, while additional guidance was provided by Gianluca Inguglia.

5.1. Introduction

Chapter 2 briefly describes the Belle II trigger system. Belle II trigger system is mainly dedicated to the B physics (efficiency $\sim$ 100 %) searches, but to perform analysis related to the dark sector and Low Multiplicity (LM) physics searches, one needs to have an excellent understanding of the trigger lines to get a maximum trigger efficiency. This chapter mainly focused on the ECL-based L1 Tigger for triggering the LM events, such as two muons in the final state mainly dedicated to dark sector searches.

5.2. ECL-based trigger $(ECL_{\mu\mu})$

The ECL-based orthogonal trigger $ECL_{\mu\mu}$ (trigger bit 40), designed to trigger the μ-pair events that are relevant for the dark sector physics searches such as $e^+e^- \rightarrow \mu^+\mu^- Z'(\rightarrow invisible)$ [82], $e^+e^- \rightarrow A'(\rightarrow \mu^+\mu^-)h'(\rightarrow invisible)$ [83], or other searches with two tracks (possibly muons) plus large missing energy in the final state. Other possible CDC-based two-track with an opening angle larger than 90° (ffo)-trigger, a two-track with an opening angle larger than 30° $(ff30)$-trigger, and a KLM-based mu_b2b (muon back to back)-trigger with the same event topology has already been studied [84] [85]. The main aim of this study is to check if the ECL-based trigger could perform better than the others. Triggers are called orthogonal when they are fired by essentially different and independent signatures, such as tracks that can be used by CDC/KLM triggers or energy deposits in the ECL. One event can fire simultaneously multiple orthogonal triggers if it contains the essential information required to fire multiple triggers. For such

a choice of a trigger, the efficiency is defined as,

$$\epsilon_{TRG_{test}} = \frac{N_{TRG_{test} \cap TRG_{reference}}}{N_{TRG_{reference}}} \tag{5.1}$$

where $N_{TRG_{test} \cap TRG_{reference}}$ is the number of events firing both the trigger to be tested and the reference trigger and $N_{TRG_{reference}}$, is the number of events activating the reference trigger.

I present here the performance of $ECL_{\mu\mu}$ using ffo (bit-17) as a reference trigger for $e^+e^- \to \mu^+\mu^-\gamma$ events. The kinematic condition required for $ECL_{\mu\mu}$ to be activated is the difference of phi angles (ϕ) of two tracks in the ECL clusters in the Centre of Mass (CM) frame has to be within 160°, and 200°, and the sum of polar angles (θ) in the CM frame has to be within 165° and 190°.

5.2.1. Event selection

The trigger performance is tested directly on the data. We used the good runs of the early phase 3 data sample available in processing 9 (proc9) and bucket7, which comprises Experiment 7 (Exp7) and Experiment 8 (Exp8), for an estimated offline total integrated luminosity of 5.15 fb^{-1}. Some pre-selections are listed below,

- $e^+e^- \to \mu^+\mu^-\gamma$ events are selected requiring two *good* tracks originating from the vertex with $|dz| < 2.0$ and $dr < 0.5$ and to be identified as muons with a global PID ($\mu_{id} > 0.7$).
- Both μ-tracks are required to be in the ECL barrel region.
- Photons are selected while requiring the hypothesis ID of the associated ECL cluster to be equal to 5, their energy to be greater than 2 GeV, and to be in the CDC acceptance.

5.2.2. Activation curve

The activation curve of a trigger is considered to be a region where it achieves maximum and approximate constant performance. The activation curve for $ECL_{\mu\mu}$ in terms of the maximum cluster energy of the two tracks is shown in Figure. 5.1. Maximum cluster energy is defined as the highest energy between two tracks deposited in the ECL cluster. After a careful investigation plateau region was chosen to be between 0.2-0.5 GeV. The average efficiency of the $ECL_{\mu\mu}$ is $\sim 85\%$ in the plateau region.

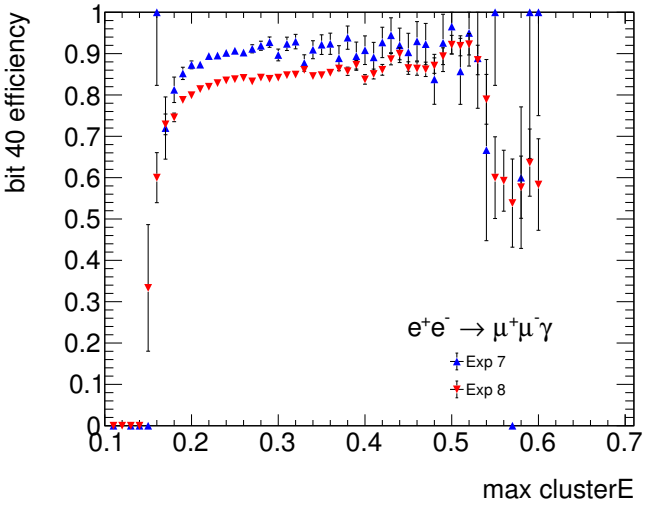


Figure 5.1.: Activation curve of the $ECL_{\mu\mu}$.

5.2.3. Systematic Evaluation

To evaluate the possible source of systematic effects when triggering events with $ECL_{\mu\mu}$ in physics analysis, I checked for dependencies of the efficiency on the kinematic variables such as the azimuthal angle (ϕ) in the laboratory frame, the polar angle (θ) in the laboratory frame, the track(s) quality, and the track(s) transverse momentum. Since the trigger efficiency should not depend on these variables, any dependencies will be considered as a systematic effect.

Dependencies on the run number

The run-by-run trigger efficiency is computed for Exp7 and Exp8 and given in Figure 5.2. For most runs, the calculated efficiency is constant at roughly 85% except for a set of runs with numbers 3800-4000 for exp7 and a few runs for Exp8. The concerned runs may be excluded for analysis purposes and considered as bad runs

Dependencies on the azimuthal angle ϕ in the laboratory frame

Figure 5.3 shows the variation of trigger efficiency with the azimuthal angle (ϕ) for track 1 (the μ^+, left) and 2 (the μ^-, right). A small region of inefficiency is visible in Fig. 5.3 (both left and right) for $\phi \approx 170°$ and $\phi \approx 350°$ for Exp8. All these dependencies are included as a systematic effect.

Figure 5.4 shows the 2D efficiency map between the tracks, where efficiency

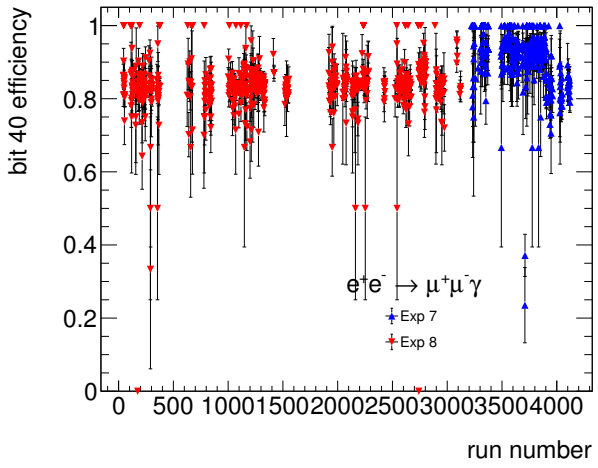


Figure 5.2.: Run by run efficiency of $ECL_{\mu\mu}$.

is calculated in the bins of ϕ for Exp7 (left) and Exp8 (right). In the 2D map, one could see that events are gathered between the angle 160° and 200°, which is consistent with the definition of $ECL_{\mu\mu}$ trigger logic.

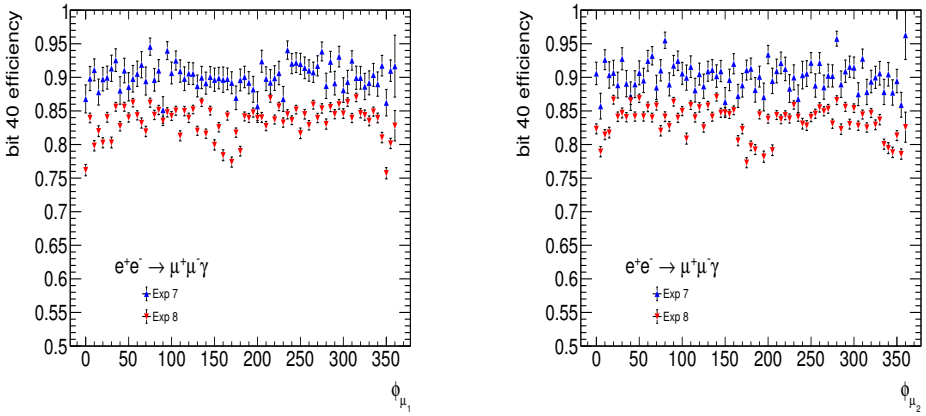


Figure 5.3.: Dependency of the $ECL_{\mu\mu}$ efficiency on ϕ in the laboratory frame, positive track (left) and negative track (right).

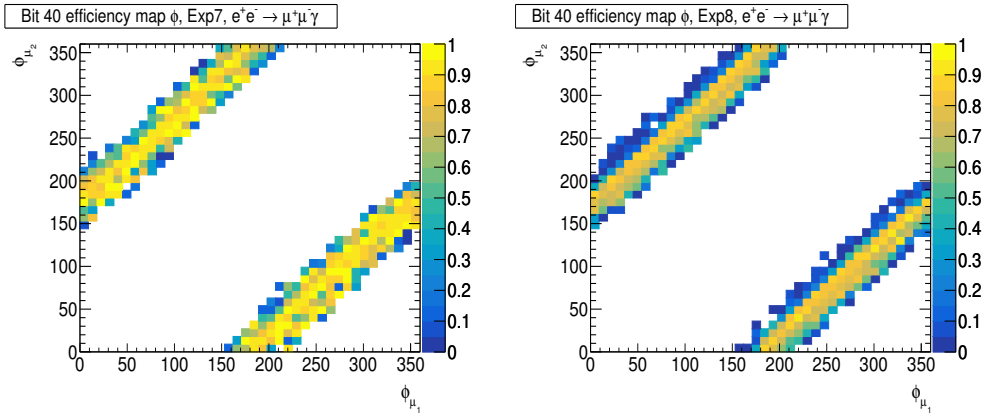


Figure 5.4.: 2D efficiency map of the $ECL_{\mu\mu}$ in terms of ϕ for Exp7 (left) and for Exp8 (right).

Dependencies on the polar angle θ in the laboratory frame

Figure 5.5 shows the efficiency variation for track 1 (the positron, left) and track 2 (the electron, right) as a function of $\cos\theta$. For track 2, some inefficiency is observed in the region between 110° to 130°, and they are considered as a systematic effect.

Figure 5.6 shows the 2D efficiency map, where the efficiency is calculated in bins of the $\cos\theta$ of the two tracks. The observed angular distribution is consistent with the definition of $ECL_{\mu\mu}$ trigger.

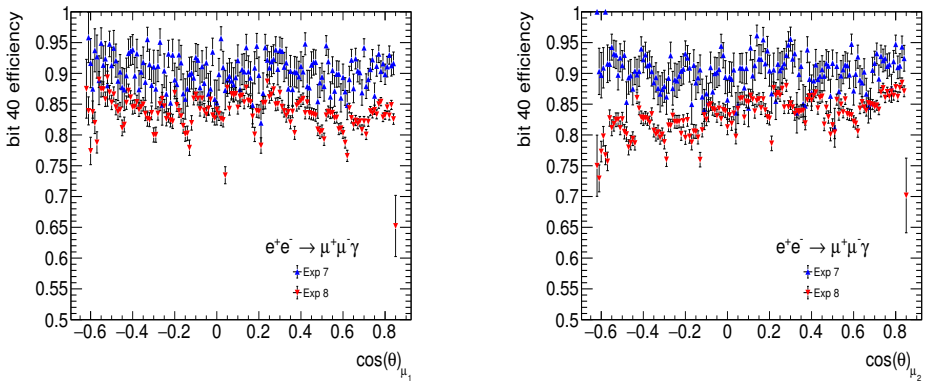


Figure 5.5.: Dependency of the $ECL_{\mu\mu}$ on $\cos\theta$, positive track (left) and negative track (right).

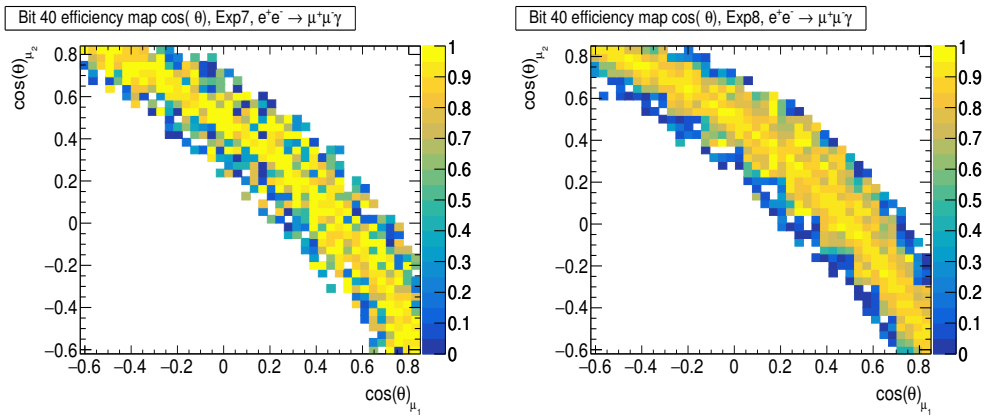


Figure 5.6.: 2D efficiency map of the $ECL_{\mu\mu}$ in terms of $\cos\theta$ of the tracks, for Exp7 (left) and for Exp8 (right) events.

Dependencies on the track quality

The trigger efficiency is also monitored as a function of the number of CDC hits, in particular as a function of the minimum number of CDC hits per track per event, and Figure. 5.7 shows the variation of efficiency for the $e^+e^- \to \mu^+\mu^-\gamma$ events.

From Figure. 5.7, it is visible that there is a slight dependence of the trigger efficiency on the minimum number of CDC hits, and when the minimum number of CDC hits is larger than 35, the efficiency for Exp7 and Exp8 start to differ. We think this effect is due to the CDC TDC discriminator threshold change. But more data will be needed to confirm this effect.

Dependencies on the track transverse momentum

The trigger efficiency is also monitored as a function of the tracks' transverse momentum, in particular as a function of the minimum transverse momentum (p_t) of the tracks per event, and Figure. 5.8 shows the variation of efficiency for the $e^+e^- \to \mu^+\mu^-\gamma$ events. The figure clearly shows a strong dependency of the trigger efficiency on the minimum track transverse momentum, and is one of the most significant sources of systematics. It is also visible that for a minimum transverse momentum larger than ≈ 4.5 GeV, the efficiency for Exp7 and Exp8 starts to merge.

5.2.4. Results

I found a total efficiency of 0.85±0.002 for the combined Exp7 and Exp8 dataset, where the error indicated is statistical. Possible sources of systematic effects have also been studied. Systematic effects are evaluated as half of the variance in the

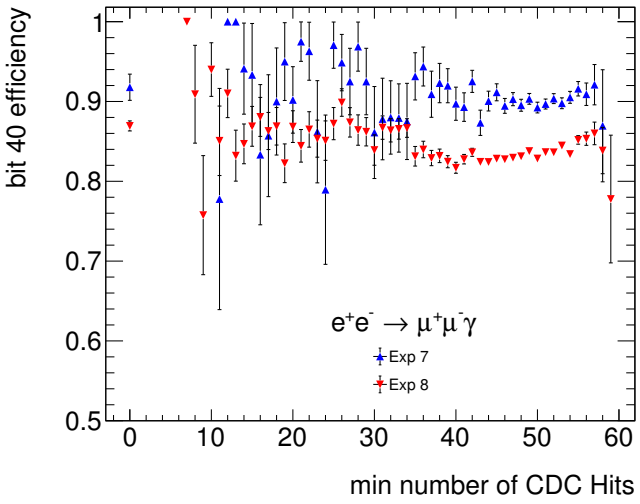


Figure 5.7.: $ECL_{\mu\mu}$ efficiency as a function of the minimum CDC hits per track per event.

plateau efficiency among different bins of the chosen variables divided by the total efficiency. One of the more significant contributions to the systematic effects arises from the dependence on the track transverse momentum, with a value of 16%. Although I am using only Good Runs for this study, some runs still show a small efficiency value (as discussed before, see Figure 5.2) and could be considered Bad Runs. The systematic effect due to the max cluster energy is $\sim 2\%$. Dependencies on the polar angle of the tracks systematically affect the efficiency at a level of 1.6%. The track quality has a systematic effect of 1.5%. In addition, a dependence on the run number is $\sim 3.4\%$. All systematic contributions are then summed in quadrature. A summary of all the systematics is given in Table 5.1.

5.2.5. Conclusions

I did a very preliminary study of ECL-based trigger $ECL_{\mu\mu}$, and try to understand the capability of triggering the events $e^+e^- \to \mu^+\mu^- Z'(\to invisible)$, $e^+e^- \to A'(\to \mu^+\mu^-)h'(\to invisible)$, by looking at the triggering capability of $e^+e^- \to \mu^+\mu^-\gamma$ events, and the estimated efficiency for Exp7+Exp8 with proc9 data is found to be $\epsilon_{ECL_{\mu\mu}}^{e^+e^-\to\mu^+\mu^-\gamma}$=0.85±0.002$_{stat}$ ± 0.167$_{sys}$, which is mainly dominated by the systematic effects. I also found that efficiency for $ECL_{\mu\mu}$ is $\sim 5\%$ lower than other CDC-based triggers such as ffo and $ff30$. Detailed study of $ECL_{\mu\mu}$ and other CDC-based triggers is submitted to collaboration described in the note [84].

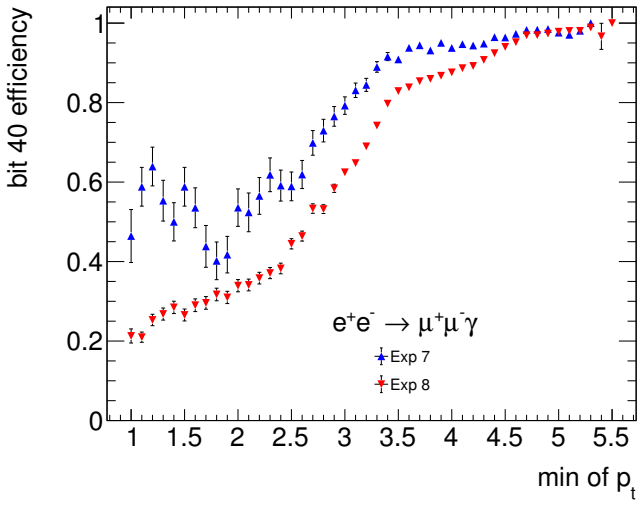


Figure 5.8.: $ECL_{\mu\mu}$ efficiency as function of the minimum p_t.

Source of systematics	Estimated effect in $e^+e^- \to \mu^+\mu^-\gamma$
Run number dependencies	0.034
$\phi_{LAB}(e^+)$ dependencies	0.005
$\phi_{LAB}(e^-)$ dependencies	0.005
$\theta_{LAB}(e^+)$ dependencies	0.016
$\theta_{LAB}(e^-)$ dependencies	0.016
Max clusterE dependencies	0.022
Min number of CDC hits	0.015
Min p_T of the tracks	0.160
Total	**0.167**

Table 5.1.: Systematic effects evaluated as half of the variance in the plateau efficiency among different bins of the selected variables divided by the total efficiency.

Part III.

Physics Analysis

6. Search for prompt visible decay of Z′ in muonic final state

The following work was done in collaboration with one other Ph.D. student Martina Laurenza; guidance was provided by Gianluca Inguglia and Enrico Graziani. The author provided major contributions to the Section 6.2, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11, 6.13.6 and 6.13.7.

6.1. Analysis overview

I present here the search for the process $e^+e^- \to \mu^+\mu^- Z', Z' \to \mu^+\mu^-$, where Z' is decaying promptly to muons, introduced in Section 1.2.3. This analysis is planned to carry out with data collected on the $\Upsilon(4S)$ resonance in 2020 and 2021a+b runs, amounting to 178.47 fb^{-1} at the Belle II experiment. The signal search strategy consists of finding a peak in the dimuon mass distribution in events with four muons having an invariant mass compatible with $\Upsilon(4S)$ collision energy and nothing else in the final state. The effect of Initial State Radiation (ISR) partially spoils this picture, allowing the presence of radiated photons and moving the four-muon invariant mass away from the collision energy. We selected events with exactly four tracks, where at least three are identified as muons, and the other is not an electron. We require the total four-track invariant mass, M(4-tracks), to be restricted between 10 and 11 GeV/c^2 to minimize the effect of ISR. The ISR contribution is simulated in the signal but not in some of the most important backgrounds (all the four lepton processes) as the used generator doesn't include it. We applied a tight selection criterion to remove events with photons, as they are not expected to be present in the signal events (see section 6.3.4).

The main background components that arise from SM processes mimic our signal topology are $e^+e^- \to \mu^+\mu^-\mu^+\mu^-$, $e^+e^- \to \mu^+\mu^-(\gamma)$, and $e^+e^- \to e^+e^-\mu^+\mu^-$. Where γ converts to electron pair and electron misidentified with muons. A Multivariate Analysis (MVA) technique is implemented to reject the backgrounds depending on the kinematic features of the events. This is a very important distinction concerning the previous searches done by BaBar and Belle (see section 6.5).

We use two trigger lines in logical OR: the CDC trigger line fff (CDC-based three or more than three tracks trigger), replaced by ffy (CDC-based two tack

and other track are identified based on neural network algorithm) for most of the 2021 data, and the logical OR of the CDCKLM lines. For data directly, trigger bits were set, but for MC, efficiency weights are applied; see section 6.4 for a detailed trigger study.

Further, a kinematical fit procedure is applied to the events that pass all the selections, imposing the 4-track invariant mass to coincide with the collision energy. This is found to improve the dimuon mass resolution and hence increase the sensitivity of the analysis (see section 6.6). Then the signal yield extraction is performed through a fit to the dimuon mass distribution. We didn't rely on the absolute background predictions as the lack of ISR contributions in the main background process would make such a prediction extremely unreliable. We planned to keep free all the parameters in the background fitting (see section 6.8). We excluded the dimuon mass interval corresponding to the J/ψ region since this is a peaking background for our search (see section 6.11). For the moment, we limit the last dimuon mass point to 9 GeV/c^2, because we don't have a good sensitivity above (see section 6.11). Then I computed the 90% exclusion limit on the coupling constant and production cross-section. Different levels of flat systematics are tested, and it is found to have a negligible effect on the sensitivity (see section 6.8.1). I also tested the reliability of the fit procedure using Toy MC techniques and found that fitting methods are stable (see section 6.9). The procedure of calculating local significance and converting it to a global prediction is also given in section 6.10.

Then after taking permission from the working group, we look at the ee$\mu\mu$ control data to validate our analysis procedure. Most of the important sources of systematics are estimated from there. After finalizing all the systematics, a mass dependent systematics is computed. A partial Unblinding of 2019 data is also presented in the section, mainly to find the exact data MC discrepancy due to the ISR effect, which is planned to discard from the final result. After all, a final expected sensitivity projection is given in section 6.15.

6.2. Analysis framework and data samples

Signal simulation: I produced 20000 signal events for different Z′ masses at steps of 25 MeV/c^2 from 0.22 GeV/c^2 to 10 GeV/c^2, with a fixed width of 10^{-6} GeV/c^2 (well below the detector resolution) and a coupling constant g′ = 1, using **MadGraph@5-NLO** [86] with the **Lmu_minus_Ltau_UFO** model, which provides the framework for simulating SM and Beyond the Standard Model processes and computing their cross section [87]. The generator includes Initial State Radiation (ISR) effects. For the MLP training and testing, I produced an additional more dense signal samples with generated mass from 220 MeV to 10 GeV

Process	σ [nb]	$\int Ldt$ [fb^{-1}]	MC generator
$e^+e^- \to \mu^+\mu^-\mu^+\mu^-$	0.351×10^{-3}	2000 and 5000	AAFH [88]
$e^+e^- \to \mu^+\mu^-(\gamma)$	1.148	900	KKMC [89]
$e^+e^- \to \tau^+\tau^-(\gamma)$	0.919	200	KKMC [89] + TAUOLA [90]
$e^+e^- \to e^+e^-\mu^+\mu^-$	18.83	200	AAFH [88]
$e^+e^- \to u\bar{u}(\gamma)$	1.605	200	KKMC [91]+PYTHIA+EvtGen
$e^+e^- \to d\bar{d}(\gamma)$	0.401	200	KKMC [91]+PYTHIA+EvtGen
$e^+e^- \to c\bar{c}(\gamma)$	1.329	200	KKMC [91]+PYTHIA+EvtGen
$e^+e^- \to s\bar{s}(\gamma)$	0.383	200	KKMC [91]+PYTHIA+EvtGen
$e^+e^- \to B^0\bar{B}^0$	0.510	200	EvtGen
$e^+e^- \to B^+B^-$	0.540	200	EvtGen
$e^+e^- \to \pi^+\pi^-(\gamma)$	0.167	1000	PHOKARA [91]

Table 6.1.: Early Phase 3 MC14a samples used for background studies with the cross section of the process σ and the equivalent integrated luminosity $\int$Ldt. All the details about the generators are given in the internal note [92]

at 5 MeV steps.

Background samples: We used the official run independent samples so-called **MC14a**. In particular:

- For background and discriminant variable studies we used 2000 fb^{-1} of $\mu^+\mu^-\mu^+\mu^-$ events;
- For the neural network training and testing purpose, I produced an additional 5000 fb^{-1} of $\mu^+\mu^-\mu^+\mu^-$ and 900 fb^{-1} of $\mu^+\mu^-(\gamma)$ events.
- Table 6.1 summarizes all the expected background components for our study and corresponding integrated luminosities.
- For data-MC comparison, we used the $e^+e^-\mu^+\mu^-$ control sample.

Data: We are targeting an integrated luminosity $\int$Ldt$\sim$ 178.47 fb^{-1} of data, collected during the 2020 and 2021a+b runs. It corresponds to the so-called "Moriond dataset", without the 2019 data and without the off-resonance samples.

For the analysis we used **root v6-21** with **TMVA**, **RooFit** and **RooStats** packages [93].

6.3. Event reconstruction and selections

6.3.1. Definitions of Signal and Background

We select events with exactly four tracks coming from the interaction region, with impact parameters concerning the collision point in the longitudinal and transverse direction within $|dr| < 0.5$ cm and $|dz| < 2$ cm. The sum of the four charges is required to be zero.
For each event, there are four possible dimuon pairs of oppositely charged tracks, and we didn't attempt to distinguish one of these pairs as more likely coming from a Z′ decay. Each event will therefore contribute with four possible 4-track different candidates, each with a different candidate Z′ invariant mass, causing unavoidably some combinatorial background.
A 4-track candidate has a pair of tracks, named "candidate tracks" that, in case of signal, are the decay products of the Z′, and two other tracks called "recoil tracks". This condition is checked by requiring the PDG code matching. Candidate mass is the invariant mass of the candidate track pair: it coincides with the Z′ mass in case of a proper signal. "Candidate" tracks and "recoil" tracks exist for the background too, but none of them come from a Z′ decay.
We restrict the total four-track invariant mass, M(4-track), to be limited between 10 and 11 GeV/c^2.

6.3.2. Muon identification

Our decay topology consists of four muons in the final state. To have a good muon identification efficiency, different combinations of μID threshold have been tested, where μID is a likelihood based particle identification probability for muons, which is defined as follows,

$$\mu\text{ID} = \frac{\mathcal{L}_\mu}{\mathcal{L}_\text{e} + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_\text{K} + \mathcal{L}_\text{p} + \mathcal{L}_\text{d}} \tag{6.1}$$

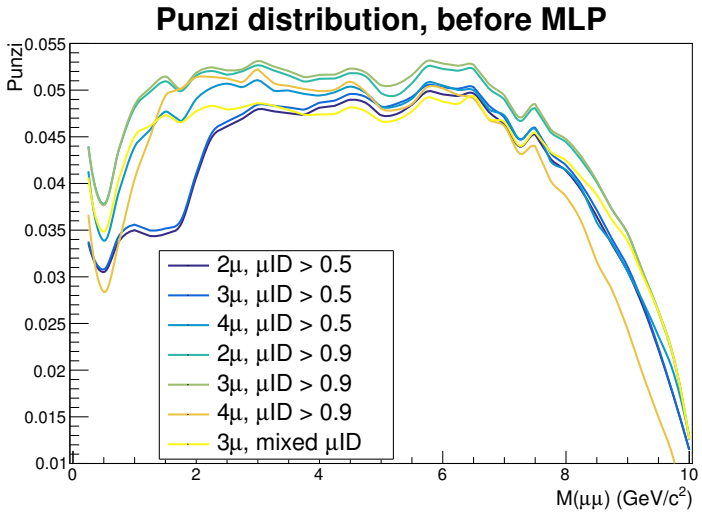
where $\mathcal{L}_\text{X}$, X = e, μ, π, K, p, d, is the likelihood from different sub-detector components for different particle hypotheses. An optimization on muon identification is needed to reduce the fake rate mainly arising from pions in hadronic processes.

The tested combinations are:

- at least 2 same-charge tracks identified as muons with:
  - $\mu\text{ID} > 0.5$;
  - $\mu\text{ID} > 0.9$;

- at least 3 tracks identified as muons with:
  - μID > 0.5;
  - μID > 0.9;
  - μID > 0.9 for the tracks with the same charge and μID > 0.5 for the remaining muon;
- 4 tracks identified as muons with:
  - μID > 0.5;
  - μID > 0.9;

We used the Punzi Figure Of Merit (FOM) [94] to select the best case. Figure 6.1a shows the Punzi FOM for different requirements.



(a) Punzi FOM before MLP.

Figure 6.1.: Punzi FoM as a function of different μID threshold.

The best result requires at least three tracks identified as muons with μID > 0.9. Moreover, we asked for the fourth track to have a low electron ID (eID < 0.5). This cut is applied for low candidate masses up to 1 GeV, only to reject the $e^+e^- \to \mu^+\mu^-(\gamma)$ process contribution to the background in the low dimuon mass region, rising from misidentification of electrons from photon conversion into muons (see section 6.3.3).

The choice of requiring three muons marks a difference concerning the selections used by BaBar and Belle because they both needed the presence of two identified same-charge as muons.

6.3.3. Characterization of background events

The main SM background processes contributing to the analysis's final state are $e^+e^- \to \mu^+\mu^-\mu^+\mu^-$, $e^+e^- \to \mu^+\mu^-(\gamma)$, $e^+e^- \to e^+e^-\mu^+\mu^-$, $e^+e^- \to \tau\tau$, $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$), $e^+e^- \to \pi^+\pi^- J/\psi(J/\psi \to \mu^+\mu^-)$.

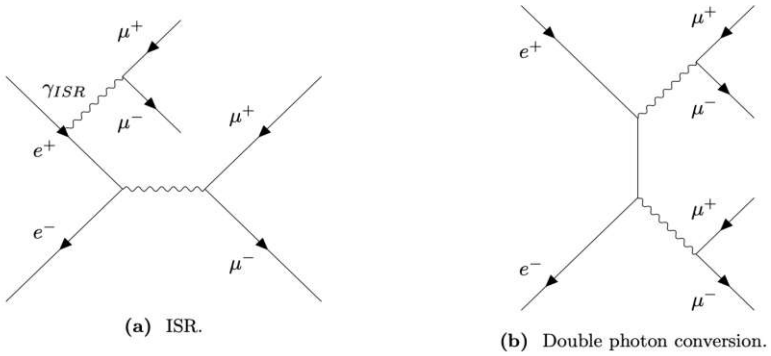
Due to its high charged track multiplicity, the $e^+e^- \to b\bar{b}$ process turns out to be negligible. In Belle II, we don't have a generator for the $e^+e^- \to \pi^+\pi^- J/\psi(J/\psi \to \mu^+\mu^-)$ process, which is expected to contribute, according to the BaBar experience, as we don't have the process generated, we planned to veto the J/ψ mass region. Other processes such as Υ resonances, $e^+e^- \to \pi^+\pi^-\Upsilon(\Upsilon \to \mu^+\mu^-)$ are for the moment outside of our search region. Processes with light hadron resonances (ρ, ω, etc.) are also expected to contribute to low dimuon masses, mainly through the misidentification of pions in muons. They are not included in the used generators.

The dominant background is the SM $\mu^+\mu^-\mu^+\mu^-$, followed by the $e^+e^- \to \mu^+\mu^-(\gamma)$ for the low Z' masses only and partly by the $e^+e^- \to e^+e^-\mu^+\mu^-$ due to misidentification. The $e^+e^- \to \mu^+\mu^-\mu^+\mu^-$ process mainly proceeds through two processes (see Figure 6.2):

- **ISR**, produces a muon pair by converting an off-shell ISR photon. Due to this feature, it typically gives rise to low mass muon pairs in the forward direction recoiling against a high mass muon pair.
- **Double-photon conversion** is the dominant one for most of the mass phase space, except for the low mass region.

The dimuon mass spectrum for the $\mu\mu\mu\mu$ process is shown in Figure 6.3. The contributions from the ISR in the low mass region and double conversion above 1 GeV are visible.

The $e^+e^- \to \mu^+\mu^-(\gamma)$ process contributes to the low dimuon mass region due to the misidentification of electrons from photon conversion into muons. If we require three tracks to be identified as muons (first, we planned to select only three tracks with μID >0.9), to pass the selections, one misidentification will lead to the huge background. The $e^+e^- \to \mu^+\mu^-(\gamma)$ contribution has been studied in detail. The electron to muon misidentification is possible mainly due to two reasons: very low energy electrons having momentum below 500 MeV/c, not reaching the muon detector and therefore undergoing a non-optimal μID and electrons with high momentum, going in the forward or backward gaps between barrel and endcap muon detector. Figure 6.4 shows the distribution of the highest electron ID in the event track, requiring three tracks identified as muons for the entire event topology for the $\mu^+\mu^-$ γ process. One can see that some tracks are misidentified as electrons; as a result, we applied a loose cut of electron ID<0.5 to ensure the



(a) ISR. (b) Double photon conversion.

Figure 6.2.: The Feynman diagram for the ISR and double photon conversion.

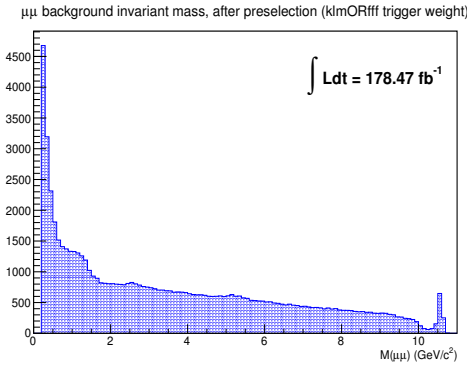


Figure 6.3.: Dimuon mass $M(\mu\mu)$ distribution for the SM four muon background.

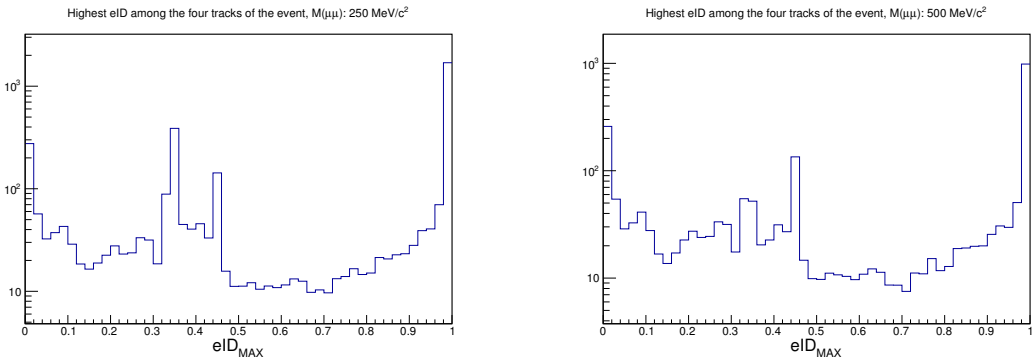
fourth track is not an electron in the low mass region (below 1 GeV dimuon mass system).

6.3.4. ISR cuts

We don't want any neutral depositions, particularly those due to photons in our events. Requiring the absence of photons will reduce the possible background sources. One could expect two kinds of ISR (and partly FSR) in the events,

1. **ISR/FSR with a high energy photon emitted in acceptance and detected by the ECL**: in this case, the ISR/FSR photon is back to back in the center of mass frame (CMS) concerning the system of the four reconstructed tracks.

2. **ISR with undetected photon** emitted at a low angle. This case is charac-



(a) Candidate mass: 250 MeV/c^2. (b) Candidate mass: 500 MeV/c^2

Figure 6.4.: Highest eID among the four tracks in the $\mu^+\mu^-$ γ process, for two low Z′ signal masses.

terized by missing energy when the total momentum of the four track system points in the very forward or backward direction in CMS.

To minimize the effect of ISR, we selected the photon energy in the ECL clusters larger than 100 MeV and defined the *Rest Of the Event* (ROE) as a system composed of all detected photons, and E(ROE) as the energy of that system. As already mentioned, the SM four background generator doesn't include ISR and FSR, and we can't look at the 4μ data directly, so we, therefore, use ee$\mu\mu$ control channel, with two identified electrons replacing two identified muons and directly look at the data.

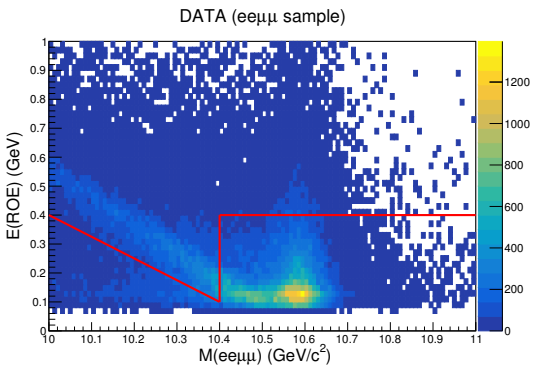
Figure 6.5a shows the total energy of the ROE photons as a function of the total 4-track invariant mass. Based on this two-dimension distribution, we decided to select events that satisfy the relation:

$$E(\text{ROE}) < E^*, \tag{6.2}$$

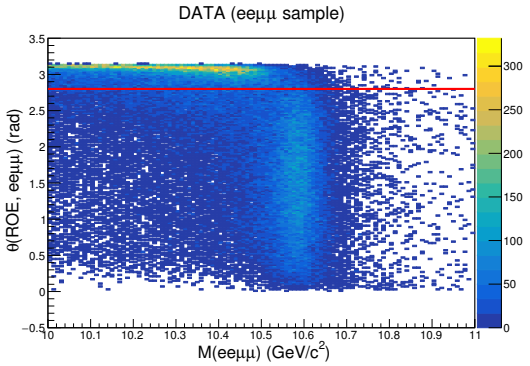
where:

$$E^* = \begin{cases} -0.75 \cdot M(ee\mu\mu) + 7.9\ GeV & \text{if } M(ee\mu\mu) < 10.4\ GeV, \\ 0.4\ GeV & \text{if } M(ee\mu\mu) \geq 10.4\ GeV. \end{cases} \tag{6.3}$$

The selection removes the diagonal stripe visible in Figure 6.5a due to the ISR and FSR process.
Figure 6.5b shows the two-dimensional distribution of the 3-dimensional angle between the ROE and the four-track system momentum directions, as a function of the four-track invariant mass, M(ee$\mu\mu$). The horizontal band in the upper part of the distributions corresponds to the case in which a photon is emitted back-to-back



(a) ROE total energy as a function of the 4-track invariant mass. The region below the red line is selected to apply the cut described in equation 6.2



(b) three-dimensional angle between the ROE and the 4-track system momentum direction as a function of invariant mass. The region below the red line is selected to apply the cut described in equation 6.4.

Figure 6.5.: Two-dimensional variables used to construct the "ISR" cuts, from the ee$\mu\mu$ control samples data-set. The red lines in the pictures represent the "ISR" cuts.

to the momentum direction of the 4-track system.

We applied the following selection:

$$\theta(\mathrm{ROE}, ee\mu\mu) < 2.8, \tag{6.4}$$

$\theta(\mathrm{ROE}, ee\mu\mu)$ is the angle between the ROE and the four-track system direction. These selections are trivially generalized to the $\mu\mu\mu\mu$. For signal, the MadGraph generator includes the ISR effects of point 2., but not those of point 1. Therefore, the selections listed in this section will also impact signal efficiency.

The selection described up to now is named and later referenced as "ISRB".

6.3.5. Summary of event selection

We summarise here the list of event selections before going to the final background suppression.

- exactly four reconstructed charged tracks with a total charge of zero;
- invariant mass of the four-track system restricted between 10 and 11 GeV/c^2;

- at least three tracks identified as muons with μID>0.9 and for masses below 1 GeV/c^2 we require in addition the fourth track has to be an electron ID < 0.5;

- minimal photon activity in the ECL. This is done through the "ISRB" requirement, which sets an upper threshold for the total photon energy;

All these requirements are set at the event level. An event that passes these selections produces four candidates, corresponding to the four possible neutral pairs eligible to be a Z′ candidate. The dimuon mass distribution after all the selections mentioned above is shown in Figure 6.6

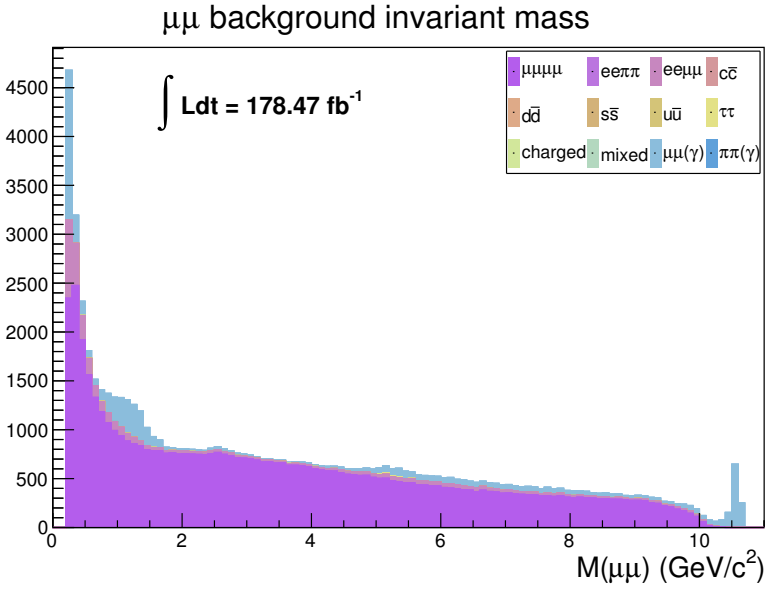


Figure 6.6.: Dimuon mass distribution after all selection summarized in this subsection 6.3.5 and the corresponding contributing background components.

6.4. Trigger study

Before going to the main background suppression, I would like to give a brief description of the triggers used for this analysis, as final background suppression technique using MVA includes the trigger weights. For this measurement, we use two trigger lines in logical OR: the CDC 3-track trigger and the CDCKLM. The CDC 3-track trigger requires the presence of at least 3 tracks with polar angles approximately in the barrel region of the detector: it is the fff trigger

line in 2020 and the beginning of 2021 (up to experiment 16), then replaced by the ffy trigger line for the rest of 2021, where a neural algorithm identifies the y track. The CDCKLM is the logical OR of the CDCKLMn exclusive trigger lines, which require n CDC tracks (n = 4) matched with barrel KLM clusters. The CDCKLM single muon trigger efficiencies have been studied in the internal Belle II note BELLE2-NOTE-TE-2022-008. Table Figure 6.7 summarizes the results for different momentum and angular regions obtained by that note. We compute a CDCKLM event trigger probability by building the logical OR of the single muon efficiencies, depending on the number of identified muons in the event and their momentum and polar angle direction. The CDCKLM trigger efficiency is shown in Figure 6.8 as a function of the Z′ mass. We will refer to this trigger in the following either as "CDCKLM" or simply as "KLM" or "klm".

id	θ [°]	p [GeV/c]	Efficiency	Systematic Uncertainty				
				θ-p	ϕ	nCDCHits	run	Total**
100	51 - 117	>1.2	0.888±0.001	0.058	0.000	0.006	0.007	0.059
0	43 - 51	0.75 - 2.0	0.346±0.004	0.263	0.049	0.009	0.010	0.264
1	43 - 51	2.0 - 6.0	0.554±0.003	0.275	0.040	0.002	0.008	0.275
2	51 - 78	0.75 - 1.4	0.521±0.003	0.053	0.032	0.005	0.016	0.056
3	51 - 78	1.4 - 5.5	0.932±0.001	0.013	0.019	0.002	0.004	0.014
4	78 - 84	0.75 - 1.4	0.442±0.006	0.060	0.062	0.004	0.019	0.063
5	78 - 84	1.4 - 5.0	0.807±0.003	0.068	0.038	0.008	0.019	0.071
6	84 - 117	0.75 - 1.4	0.567±0.003	0.031	0.050	0.009	0.020	0.038
7	84 - 117	1.4 - 5.0	0.865±0.001	0.024	0.027	0.007	0.007	0.026
8	117 - 125	0.75 - 2.0	0.381±0.006	0.299	0.088	0.008	0.023	0.300
9	117 - 125	2.0 - 4.5	0.568±0.007	0.287	0.057	0.011	0.026	0.289
c	112 - 125	0.75 - 1.2	0.175±0.011	0.110	0.076	0.012	0.014	0.112
c	112 - 125	1.2 - 4.5	0.351±0.007	0.079	0.140	0.016	0.023	0.084

Figure 6.7.: CDCKLM trigger efficiency for the signal track muon in different momentum and polar angle intervals. The error in the efficiency column is statistical only. The Table also shows the estimated systematic uncertainties. **Table and description are taken from the internal note BELLE2-NOTE-TE-2020-028**.

As we don't have any publicly available results for the fff and ffy trigger efficiency, we plan to estimate it ourselves. The efficiency is measured using an external and orthogonal reference trigger: the choice was the ECL trigger line *hie*, which is fired when the total energy deposition in the barrel and part of the forward ECL calorimeter exceeds 1 GeV. The efficiency is calculated as follows:

$$\varepsilon(fff) = \frac{N_{fff} \quad \& \quad N_{hie}}{N_{hie}} \tag{6.5}$$

Where N is the number of events, the same expression also holds for the ffy trigger. We studied the fff and ffy efficiencies as a function of the minimum and the second minimum transverse momentum of the 4 tracks (third and fourth in decreasing order of p_T). Being a blinded analysis, we planned to estimate the efficiency from $ee\mu\mu$ control channel data, but we found that trigger efficiency strongly depends on the process. Various final states are used to reduce the dependence on the process: $\mu\mu$ee, $\mu\mu eX$, $\pi\pi$ee, $\pi\pi$eX, where X is any particle. The average efficiency over these final states was assumed. The fff trigger efficiency is shown in Figure 6.8 as a function of the Z′ mass. A similar study is also performed to estimate the ffy trigger efficiency.

We finally combined them event by event basis, the fff/ffy trigger efficiency with the CDCKLM with a logical OR. This number is interpreted as a probability and used as an event weight for MC. In data, we require the OR of the CDC fff/ffy trigger bit with the CDCKLM lines. The signal efficiencies of the CDCKLM trigger, fff/ffy trigger, and the logic OR between them are shown in Figure 6.8, as a function of the Z′ mass. The drop of the trigger efficiency for high masses could be explained by the fact that the heavier the mass, the two muons recoiling against the Z' will be softer (very low momentum) and preferentially emitted in the forward/backward direction due to the boost. CDCKLM and fff/ffy triggers are based on tracks in the barrel and are inefficient at low transverse momenta.

6.5. Final background suppression

This is the crucial part of the analysis and represents a significant difference concerning the approaches adopted by BaBar and Belle. The core idea is to find a set of variables that discriminate between signal and background and then combine them with an MVA technique. The background suppression overcompensates the unavoidable loss of efficiency, and the final sensitivity overall improves. The $\frac{\varepsilon_S}{\sqrt{\varepsilon_B}}$ is a rough performance estimator, where ε_S and ε_B are the relative signal, and background efficiencies are resulting due to the MVA selection only. Observation

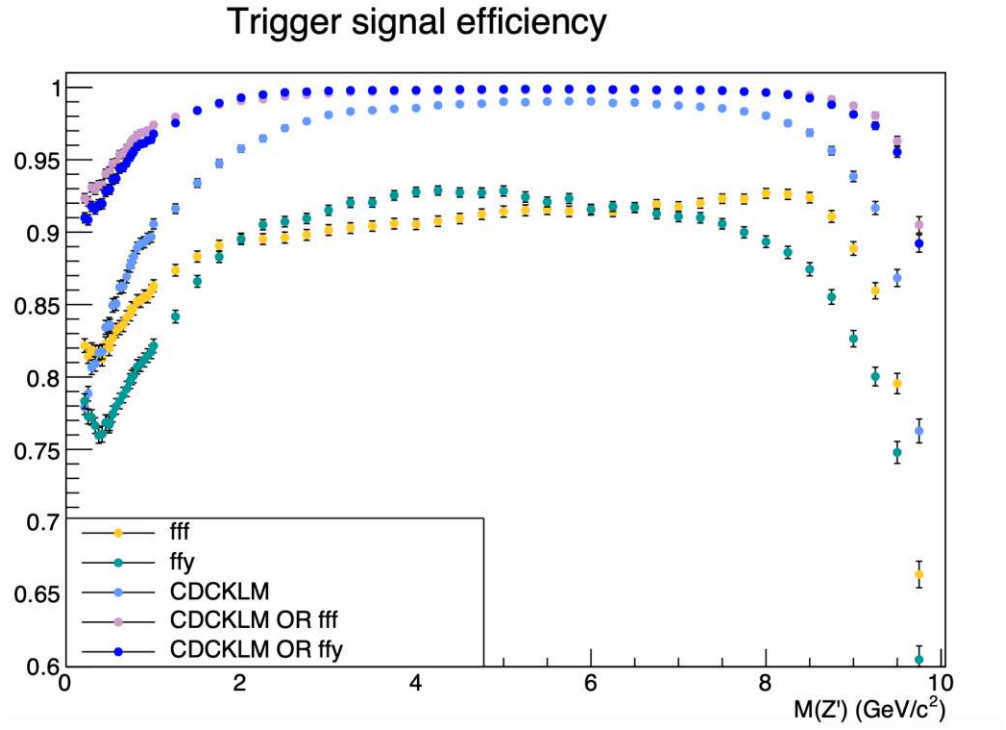


Figure 6.8.: fff, ffy, CDCKLM, CDCKLM OR fff and CDCKLM OR ffy trigger signal efficiency as a function of the Z′ mass.

of $\frac{\varepsilon_S}{\sqrt{\varepsilon_B}} > 1$, is profitable.

6.5.1. Discriminant variables

The signal over background separation relies on a few variables sensitive to a different class of features:

- presence of a dimuon ($\mu^+\mu^-$) resonance in case of a signal both in the candidate dimuon and in the recoil invariant mass system recoiling against the two recoil muons;
- production mechanism: in the case of signal, the Z' is emitted through an FSR process by radiation of a final state muon, and the four muon background arises through the ISR and the double-photon conversion process. When studying the background as a function of the mass (assuming the scanning procedure) for a specific mass hypothesis M, in two out of the four 4-track candidates, one of the two pairs will be automatically close to M. In

contrast, the other pair will be preferentially emitted at the lowest possible value, $2m_\mu$. This feature identifies a *quasi-two body* process with a (non-existing, quasi) particle of mass M recoiling against a *"zero"* (actually $2m_\mu$) mass object. In the center-of-mass (CMS) system, the kinematics of such a process is closed, with a momentum P_0 of the two quasi-particles,

$$P_0 = \frac{\sqrt{(s + M^2 - (2m_\mu)^2)^2 - 4sM^2}}{2\sqrt{s}}, \tag{6.6}$$

where M is the candidate mass, $2m_\mu = 0.210\ GeV/c^2$ is twice the dimuon mass and s=10.58^2 GeV$^2/c^4$ is the total invariant squared mass.

After a long and detailed study, the considered following sets of variables: (all are in the center-of-mass frame):

- $p_{\mu\mu}$: Magnitude of the candidate dimuon pair 3-momentum; In the 4μ background events, we expect the dimuon momentum $p_{\mu\mu}$ to be strongly peaked around P_0, differently from the signal case, at least for two of the four 4-track candidates. Therefore, we expect $p_{\mu\mu}$ to be a very discriminating variable: this is shown in Figure 6.9, where the dimuon momentum is plotted for a 3 GeV$/c^2$ signal and background. The visible peak in the background distribution signifies this quasi-two-body process.

- $|\cos(\varphi_{hel})|$: Absolute value of the cosine of the helicity angle. The helicity angle is defined as the angle in the candidate dimuon rest frame between the momentum direction of the μ^+ and the momentum direction of the CMS system. The cosine of the helicity angle will reflect the decaying particle's quantum-mechanical nature. It is expected to be flat for an unpolarized spin-1 boson with three polarization degrees. On the other hand, it is expected to peak at one (in absolute value) in the case of background, where no resonance is present. This is actually seen in Figure 6.10.

- p_{μ_0} and p_{μ_1}: Magnitude of the candidate muon 3-momentum. In Figure 6.11, the two-dimensional distribution of the candidate muon momenta is shown. Here the difference between the signal and background is former populates a middle-bottom region of the distribution, while the latter concentrates on the extreme upper edges. The linear upper border of the distribution (left plots) is determined by the relation $p_{\mu_0} + p_{\mu_1} = P_0$, which holds for a two-body final state. In the background case, events stay preferentially along the line being a quasi-two-body process. In the signal case, events accumulate far from the upper border because they are not quasi-two-body-like, and the two muons share more democratically the available momentum as they come from a resonance decay.

- $p_{\mu_0}^{rec}$ and $p_{\mu_1}^{rec}$: Magnitude of the recoil muon 3-momentum; Figure 6.12 shows the recoil muon momentum. In this case, the muon pair is opposite to the Z′ candidate, and the same straight line limits the distribution as before but in the bottom part.

- $pt_{\mu_0}^{rec}$ and $pt_{\mu_1}^{rec}$: Transverse momentum of the recoil muons with respect to the z-axis. Figure 6.13 shows the distribution of recoil muon transverse momentum.

- pt_{μ_0} and pt_{μ_1}: Transverse momentum of the candidate muons to the z-axis. A 2D plot of candidate muon's transverse momentum is shown in Figure 6.14 (left plots). We observe that signal and background events cluster around a hyperbolic shape. This is the kinematical consequence of the two muons having a well-defined invariant mass, either because they come from a resonance (Signal) or forced to have that mass due to the scanning procedure (background). The ultimate reason for the hyperbolic relation is that the invariant mass squared is proportional to $p_0 \times p_1 \times (1 - cos\theta)$, where p_0 and p_1 are the candidate muon momenta and θ their opening angle. The signal populates preferentially in the central part while the backgrounds are around the corners. To make use of this feature, we constructed two new variables. The first variable we constructed is asymmetry:

$$A(pt_{\mu_0}, pt_{\mu_1}) = \frac{I_1 - I_2}{I_1 + I_2}, \tag{6.7}$$

  where I_1 and I_2 are curvilinear coordinates along the hyperbole defined by the projection (see Figure 6.15) and measure quantitatively how much a point is placed in a central or in an endpoint position. The second variable, much less discriminating, is the signed distance between the points (x_0, y_0) and (x_1, y_1), so how much a point is far from the hyperbole:

$$D(pt_{\mu_0}, pt_{\mu_1}) = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}. \tag{6.8}$$

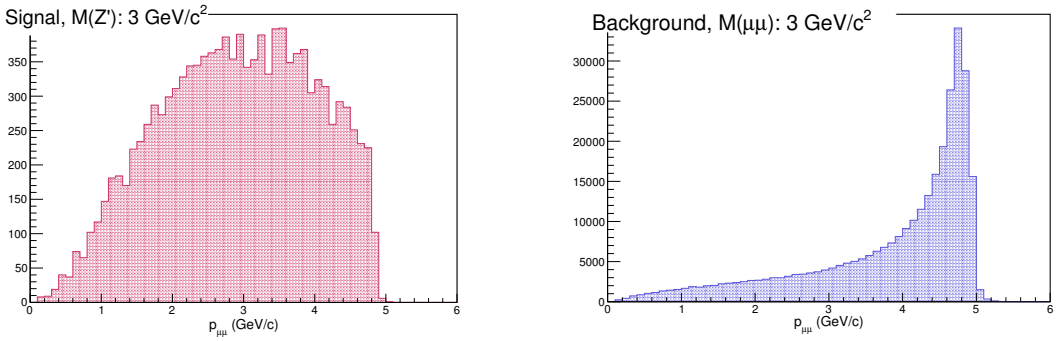
  The plot of $D(pt_{\mu_0}, pt_{\mu_1})$ versus $A(pt_{\mu_0}, pt_{\mu_1})$ is shown in Figure 6.14, right plots. These newly constructed variables are given as a variable for training.

- $P^T(p_{\mu\mu}, p_{min}^{rec})$ and $P^T(p_{\mu\mu}, p_{max}^{rec})$: Transverse projections of the candidate dimuon momentum with respect to the minimum and the maximum recoil muon momentum direction. In the case of signal, these are the transverse projections of the Z' momentum on the recoil muons.

- $pt_{\mu\mu}$ and $P^T(p_{\mu\mu}, p_{min}^{rec})$: Candidate dimuon transverse momentum with respect to the z-axis and transverse projection of the candidate dimuon momentum to the minimum recoil muon momentum direction. The reason

behind this pair of variables is that $pt_{\mu\mu}$ is sensitive to the ISR background, while $P^T(p_{\mu\mu}, p^{rec}_{min})$ is sensitive to the FSR-like nature of the signal;

- $P^T(p^{rec}_{\mu\mu}, p_{min})$ and $P^T(p^{rec}_{\mu\mu}, p_{max})$: Transverse projections of the recoil dimuon momentum on minimum and maximum candidate muon momentum direction.



(a) signal. (b) Background.

Figure 6.9.: Dimuon momentum $p_{\mu\mu}$ for a 3 GeV$/c^2$ signal (Figure 6.9a) and for the 4μ background around dimuon mass of 3 ± 0.25 GeV$/c^2$ (Figure 6.9b).

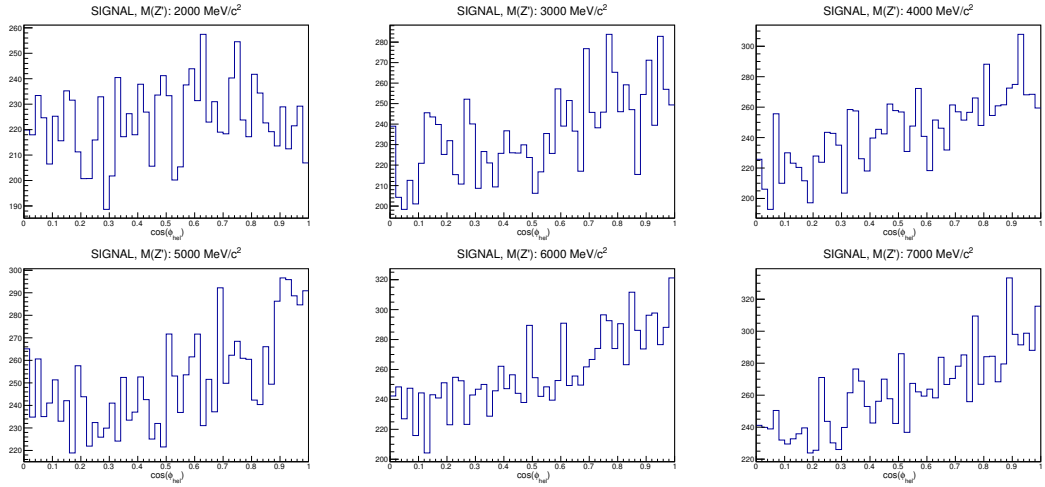
Figures 6.16 and 6.17 shows some of the 2D variables containing projections.

All the variables listed above, but the first two, have been studied based on their two-dimensional relations, but during training, individual variables are given as input. These variables, except $|\cos(\varphi_{hel})|$, underwent careful pre-processing before the actual usage in an MVA algorithm because NN can't handle arbitrary large values. As well as this step played an essential role in making the inputs more homogeneous with mass to drastically reduce the dependence on the Z′ mass.

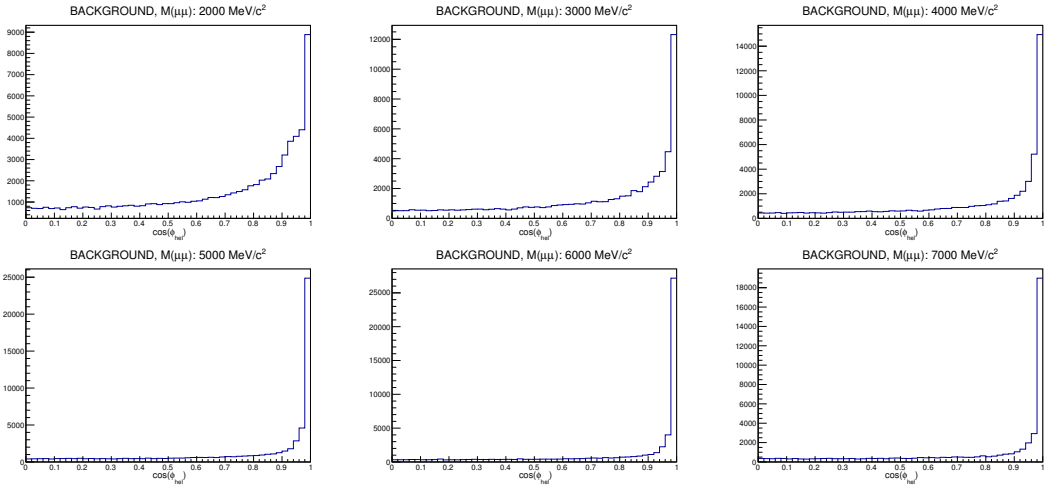
6.5.2. Multi Variate Analysis (MVA)

After carefully studying different MVA algorithms available in TMVA, we plan to use the MLP Neural Network for training, testing, and application due to its high performance. A detailed explanation of how an MLP neural network works is already given in section 3.3. The same list of standard MLP hyper-parameters is also chosen for this study presented in Table 3.2.

As already mentioned, for MLP training, we used dense signal samples and high statistics background samples to predict the shapes accurately. Using a dense signal sample approximates the signal as a continuous mass distribution of backgrounds. After a long detailed study, we ended up with a configuration of five



(a) Signal.



(b) Background.

Figure 6.10.: Absolute value of the cosine of the helicity angle for some signal Z′ masses (a) and background. For background signal mass value ± 0.25 GeV/c^2 region has shown.

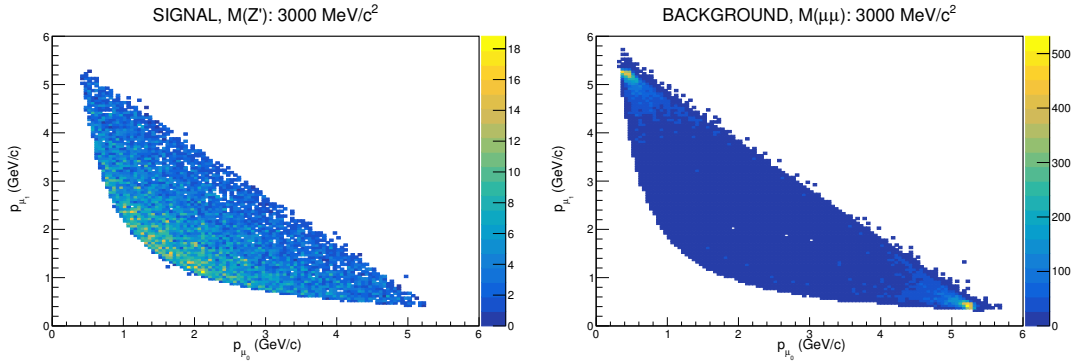


Figure 6.11.: 2D distribution of candidate muon momentum, for 3 GeV/c^2 signal and 3 $\pm$ 0.25 GeV/c^2 background.

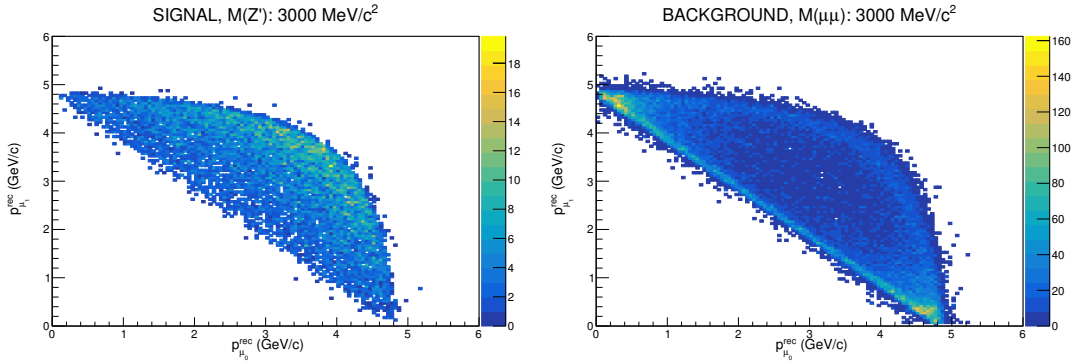


Figure 6.12.: 2D distribution of recoil muon momentum, for 3 GeV/c^2 signal and 3 $\pm$ 0.25 GeV/c^2 background.

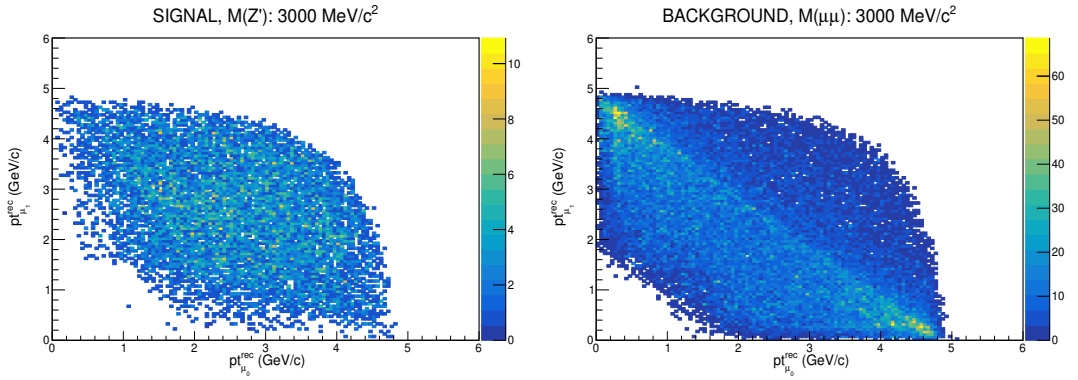


Figure 6.13.: 2D distribution of recoil muon transverse momentum, for a 3 GeV/c^2 signal and a 3 $\pm$ 0.25 GeV/c^2 background.

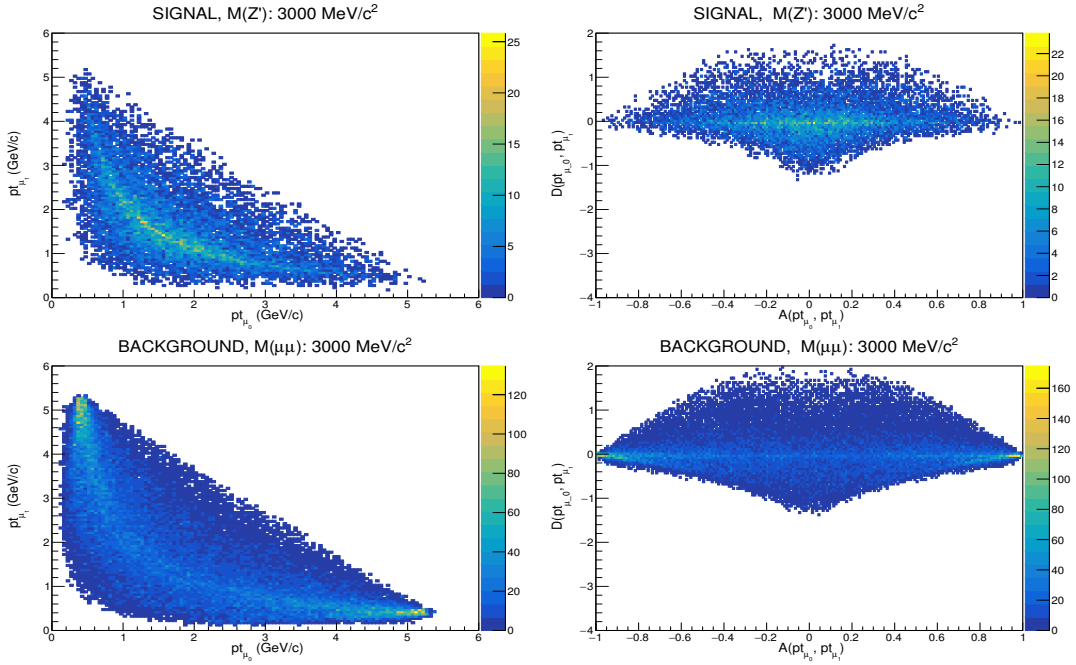


Figure 6.14.: 2D distribution of candidate muon transverse momentum, for 3 GeV/c^2 signal (top) and 3 ± 0.25 GeV/c^2 background (bottom) (left). The right plot shows the 2D distribution of the two newly constructed variables.

different mass intervals for training: 0.0-1.0 GeV/c^2, 1.0-3.75 GeV/c^2, 3.75-6.25 GeV/c^2, 6.25-8.25 GeV/c^2 and 8.25-10 GeV/c^2. The effect of the trigger and luminosity weights are also considered during training.

The ROC curves for each mass range are shown in Figure 6.18. ROC curves represent the background rejection as a function of signal efficiency. The nearer the area of this curve to 1 is better the net performance. The net output is for different MLP ranges are shown in Figure 6.19. The best separation happens in the third and fourth mass ranges. The low mass region is more difficult because the background is higher, and the 4μ ISR diagram contributes heavily. Now we checked the quantity $\frac{\epsilon_S}{\sqrt{\epsilon_B}}$ to quantify the performance as a function of the MLP output in Figure 6.20 and as a function of the signal efficiency in Figure 6.21, when $\frac{\varepsilon_S}{\sqrt{\varepsilon_B}} > 1$ the application of the net is convenient.

Since the training is performed at MC level and important background the SM 4μ don't include ISR effects, we don't expect these performances to be exactly reproducible on real data.

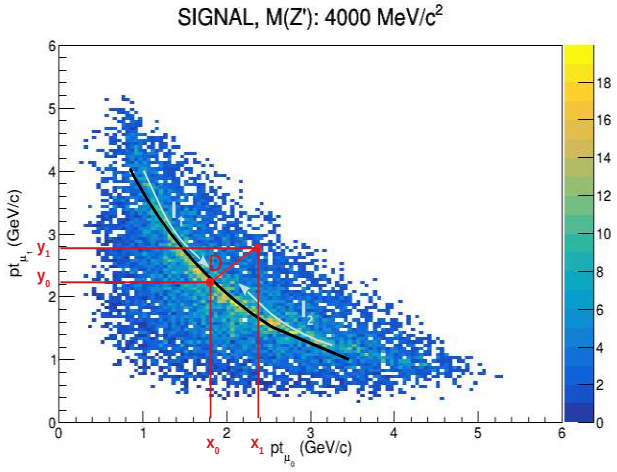


Figure 6.15.: Construction of two discriminant variables starting from candidate muon transverse momenta.

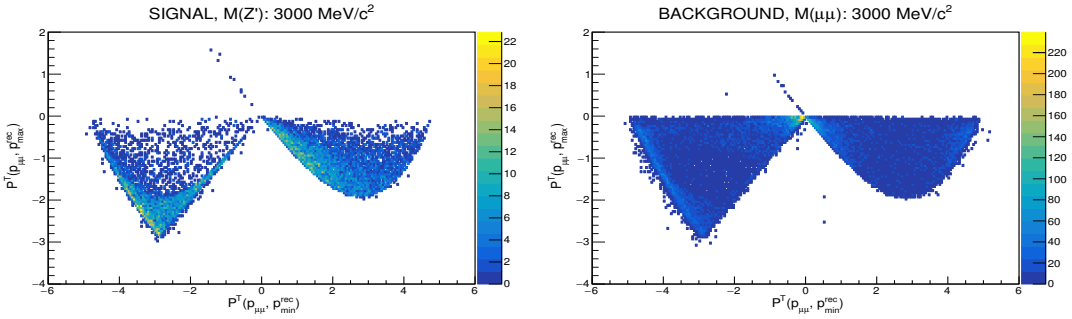


Figure 6.16.: projection of the dimuon transverse momenta to the minimum versus maximum recoil muon directions, for a 3 GeV/c^2 signal and a 3 $\pm$ 0.25 GeV/c^2 background.

6.5.3. Optimization of the MLP selection

Now to determine the analysis cut on the MLP output, we optimize it using Punzi figure of merit (f.o.m) [94], defined by the equation:

$$P(t) = \frac{\varepsilon(t)}{a/2 + \sqrt{B(t)}}, \tag{6.9}$$

where $\varepsilon(t)$ is the absolute signal efficiency after a given cut t, a is the significance level expressed in term of number of σ, in our case a=1.64 (corresponding to one sided 95% CL), and $B(t)$ is the number of background events left after the cut t.

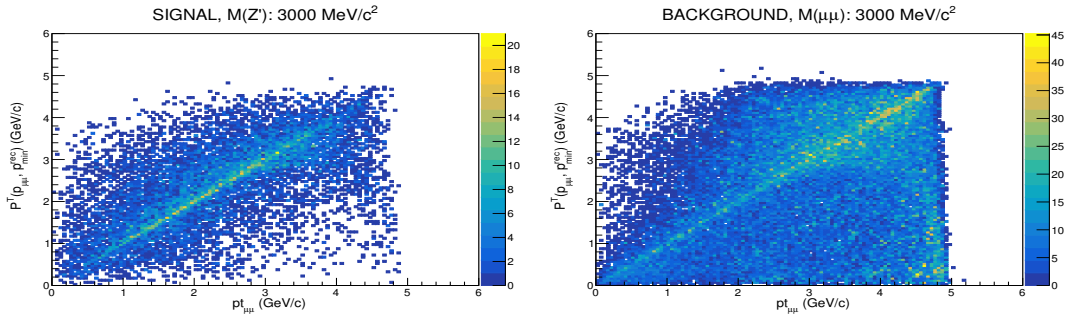


Figure 6.17.: projection of the dimuon transverse momenta to the minimum recoil muon direction versus the transverse dimuon momentum, for a 3 GeV/c^2 signal and a 3 $\pm$ 0.25 GeV/c^2 background.

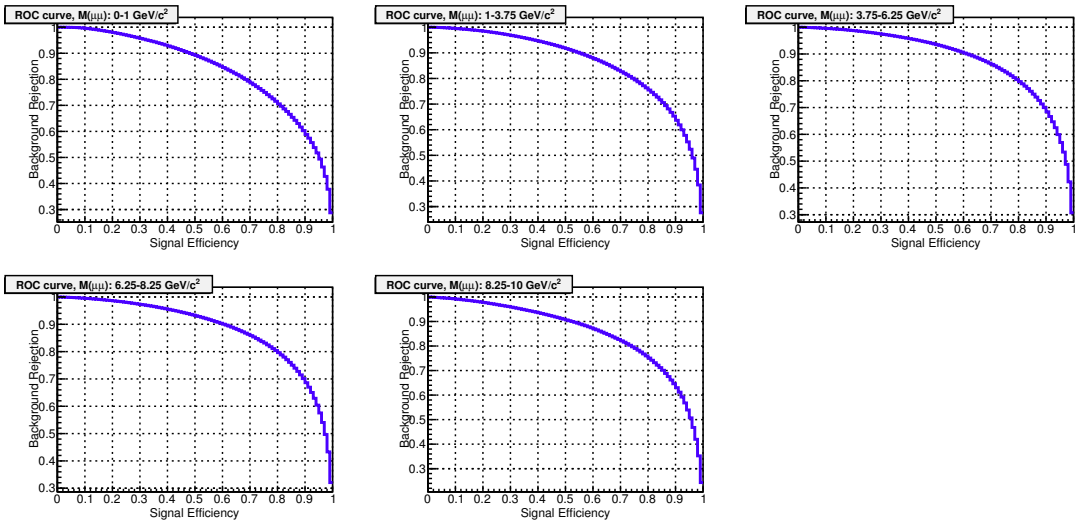


Figure 6.18.: ROC curves for different MLP mass ranges.

To determine the cut on the network output mass by mass, we chose some sample signal mass points between 0.212 GeVc^2 to 10 GeV/c^2 and determined the Punzi f.o.m for each mass as a function of different cut values (t). The cut value on the network output for each mass point is assigned corresponding to the maximum Punzi f.o.m (t(P^{max})). Now the obtained cut values are fitted with different polynomials for different MLP ranges separately to get an analytical function for any mass hypothesis. In Figure 6.22 bottom, an example of the Punzi f.o.m as a function of different cut thresholds is shown for a Z′ mass of 5 GeV. The figure also shows the variation in signal efficiency and the number of surviving background events after the different applied thresholds. Figure 6.23 shows the chosen cut values corresponding to the maximum Punzi f.o.m as a function of

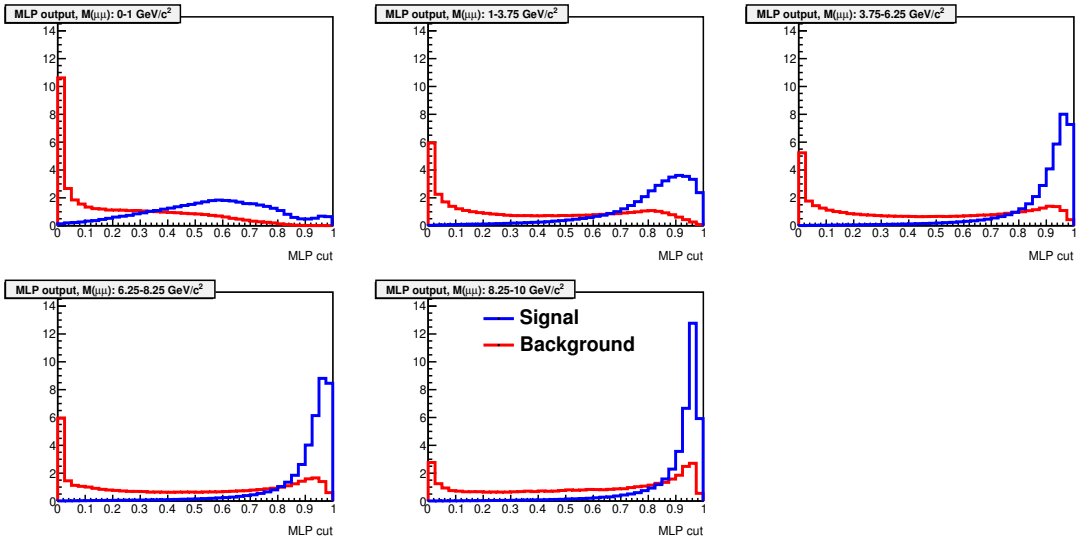


Figure 6.19.: MLP output neurons for different MLP mass ranges. Red is the background, and blue is the signal.

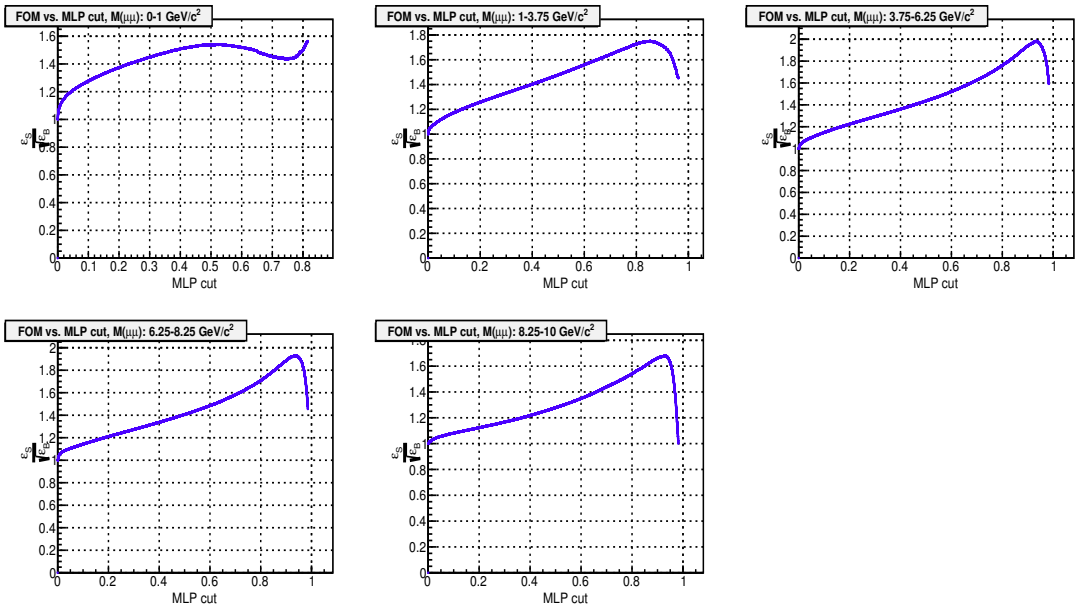


Figure 6.20.: $\frac{\varepsilon_S}{\sqrt{\varepsilon_B}}$ as a function of the network output for different MLP mass ranges.

the candidate dimuon mass along with the fitted polynomials for different MLP ranges.

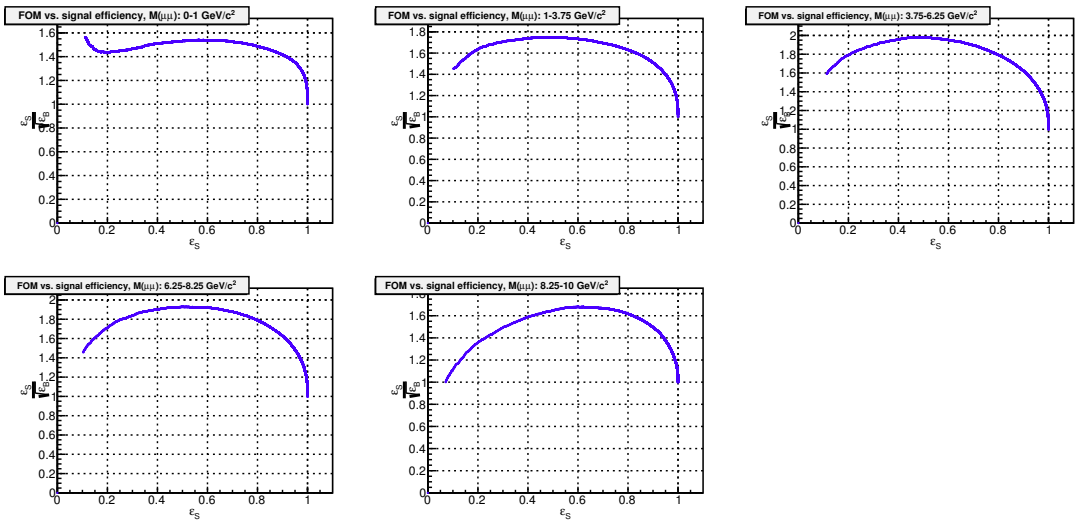


Figure 6.21.: $\frac{\varepsilon_S}{\sqrt{\varepsilon_B}}$ as a function of the relative signal efficiency for different MLP mass ranges.

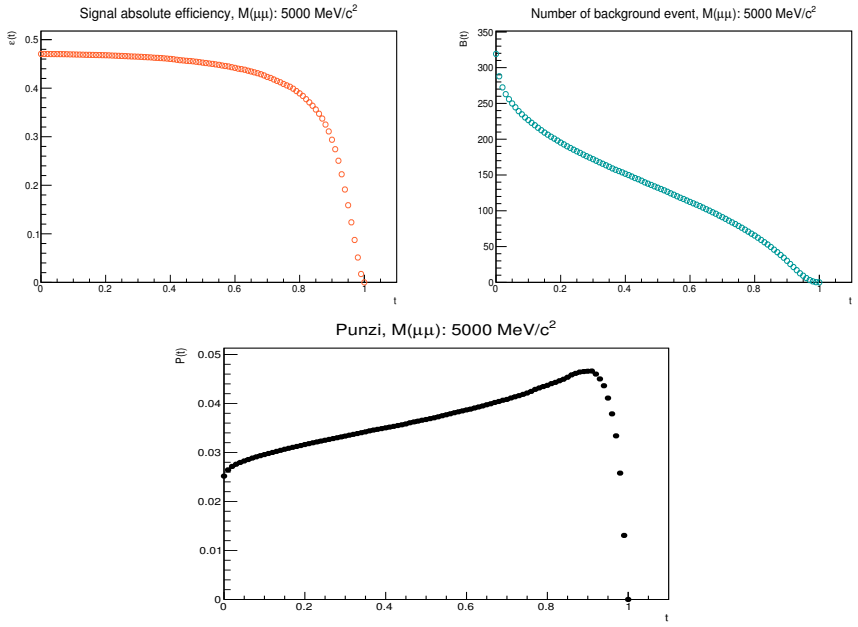


Figure 6.22.: **Top-left**: Signal efficiency as a function of the cut t on the neuron output. **Top-right**: Number of background events left as a function of the cut t on the neuron output. **Bottom**: Punzi f.o.m computed using Equation 6.9 as a function of the cut t on the neuron output.

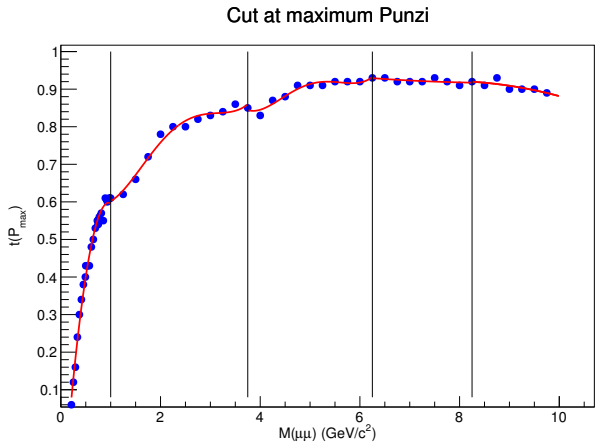


Figure 6.23.: Value of the cut that maximizes the Punzi f.o.m as a function of the mass, also polynomial fits are shown

Figure 6.24 shows the expected signal efficiency as a function of mass after applying all the analysis cuts. It also shows the spline interpolation of the signal efficiency for any mass hypothesis used later in the analysis to estimate the sensitivity. Finally, dimuon invariant mass distribution for the backgrounds after and

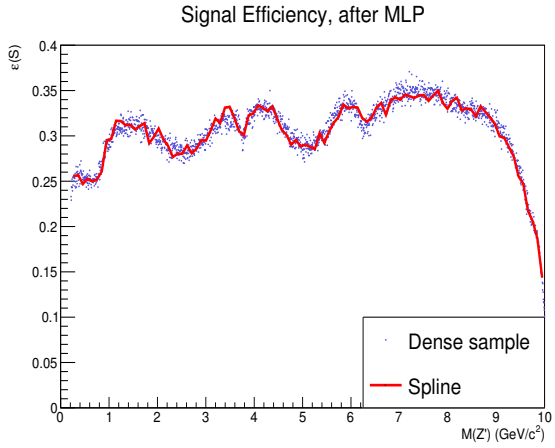
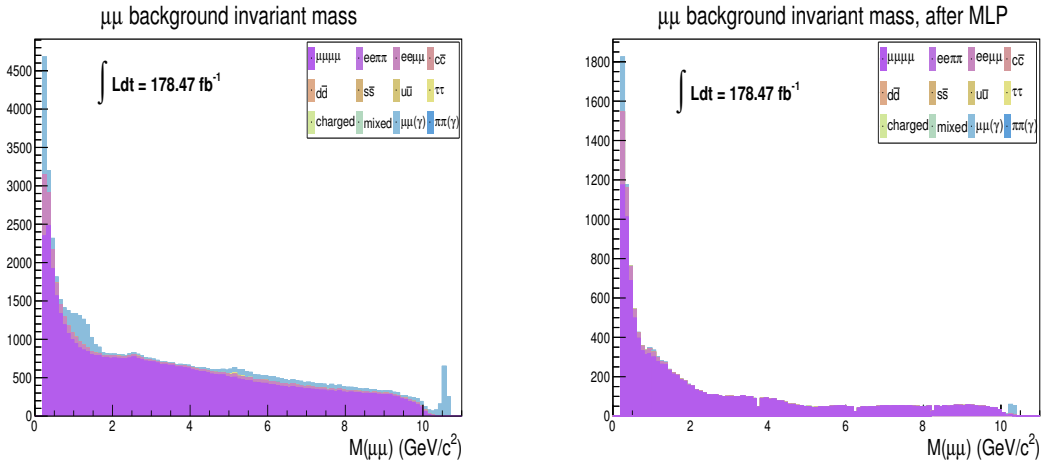


Figure 6.24.: Signal efficiency after all the analysis cut as a function of mass, as well as the spline signal efficiency interpolation, is shown by the red curve.

before the MLP application is shown in Figure 6.25. A background rejection factor as a function of mass is also determined by taking a bin-by-bin ratio between these two plots, as shown in Figure 6.26. A significant reduction is obtained in the entire range, especially for the central masses.



(a) before MLP.

(b) after MLP.

Figure 6.25.: Dimuon invariant mass distribution for backgrounds before (6.25a) and after (6.25b) MLP application. The main contributions come from $\mu^+\mu^-\mu^+\mu^-$ and $\mu^+\mu^-(\gamma)$ processes.

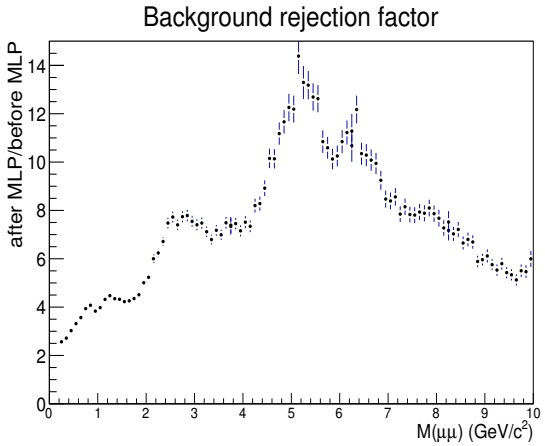


Figure 6.26.: MLP rejection factor as a function of invariant mass.

6.6. 4C kinematic constraint

Events that passed all the selections, including MLP, were subject to a kinematic fit procedure to improve the dimuon mass resolution, which is the key parameter to improve the analysis sensitivity.

Kinematic fitting is a technique that uses an event hypothesis, formulated in terms of kinematic constraints, to improve the measured objects of the event or to predict unmeasured ones. In addition, the procedure usually provides a χ^2 of the

fit that can be used to test the event hypothesis quantitatively.

In this analysis, I perform a so-called "4C" hard momentum constraint fit, imposing that the four-momentum of the system made of the four charged tracks coincides with the initial state four-momentum determined by the beam. This is implemented in `OrcaKinFit` [95] via the set of equations:

$$\begin{aligned}
\sum_i E_i - E_{\text{beam}} &= 0 \ , \\
\sum_i (p_x)_i - (p_x)_{\text{beam}} &= 0 \ , \\
\sum_i (p_y)_i - (p_y)_{\text{beam}} &= 0 \ , \\
\sum_i (p_z)_i - (p_z)_{\text{beam}} &= 0 \ .
\end{aligned}$$

where the index i runs over the four charged tracks. The beam four-momentum values are automatically taken by `basf2` from the conditions database. The usage of the 4C kinematic fitting is found to improve the dimuon mass resolution as shown in Figure 6.27.

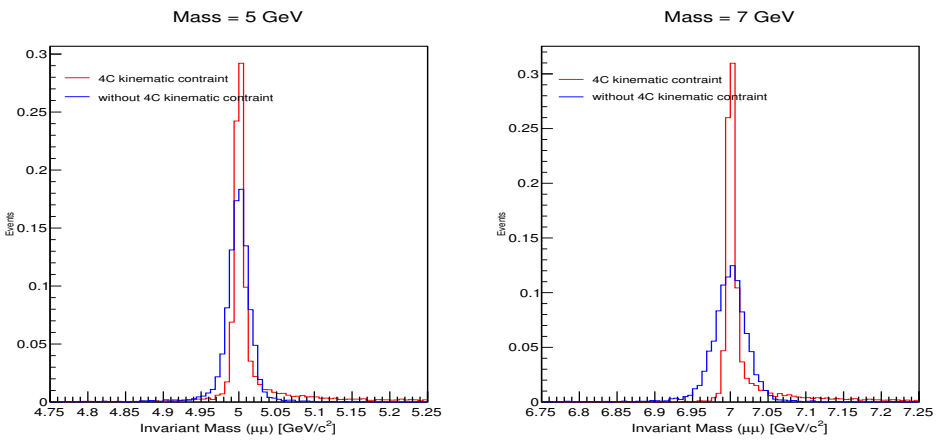


Figure 6.27.: Above plot shows a comparison of two superimposed histograms with and without 4C kinematic constraint for signal mass points 5 and 7 GeV.

6.7. Signal Modeling

I studied the signal shapes by looking at the reduced dimuon invariant mass distributions $M_{\mu\mu}$ after the 4C kinematical fit. The reduced mass is defined as follows

$$M_{\mu\mu}^{\text{reduced}} = \sqrt{M_{\mu^+\mu^-}^2 - 4m_\mu^2}. \tag{6.10}$$

The advantage of considering the reduced mass is that reduced mass distribution is smoother than the invariant mass, especially in the low mass region (see Figure 6.28). Generally, I use the reduced mass for fitting and show the results as a function of the invariant mass (because of the easier interpretation). The reduced dimuon mass distribution is fitted using the sum of two Crystal ball (CB) functions sharing the same mean values μ. The crystal ball function is given by

$$f(x,\alpha,n,\mu,\sigma) = N \cdot \begin{cases} \exp(-\frac{(x-\mu)^2}{2\sigma^2}), & \text{for } \frac{x-\mu}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-\mu}{\sigma})^{-n}, & \text{for } \frac{x-\mu}{\sigma} \leqslant -\alpha \end{cases} \tag{6.11}$$

where,

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right), \tag{6.12}$$

$$B = \frac{n}{|\alpha|} - |\alpha|\,, \tag{6.13}$$

$$N = \frac{1}{\sigma(C+D)}, \tag{6.14}$$

$$C = \frac{n}{|\alpha|} \cdot \frac{1}{n-1} \cdot \exp\left(-\frac{|\alpha|^2}{2}\right), \tag{6.15}$$

$$D = \sqrt{\frac{\pi}{2}}\left(1 + \text{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right). \tag{6.16}$$

N is a normalization factor, while α, n, μ, and σ are parameters that are fitted with the data, and the erf is the error function. The signal model is therefore given by

$$f = N_1 \cdot f_1(x,\alpha_1,n_1,\mu,\sigma_1) + N_2 \cdot f_2(x,\alpha_2,n_2,\mu,\sigma_2). \tag{6.17}$$

where f_1 and f_2 are two CB functions.

I performed an extended 1D unbinned maximum likelihood (1DUML) fit for each mass hypothesis in the range of reduced mass ± 0.05 GeV for masses below 1 GeV and reduced mass ± 0.2 GeV for masses above 1 GeV. This corresponds

to around 10-100 mass resolution, depending on the mass. The weighted mass resolution σ_w is given by

$$\sigma_w = \sqrt{f_{CB1} \cdot \sigma_1^2 + f_{CB2} \cdot \sigma_2^2} \tag{6.18}$$

and

$$f_{CB1} = \frac{N_1}{N_1 + N_2}, f_{CB2} = \frac{N_2}{N_1 + N_2}. \tag{6.19}$$

Figure 6.29 shows the fitting of some reduced mass distributions. Trigger weights and the Belle II recommended μID corrections are considered during the fitting procedure.

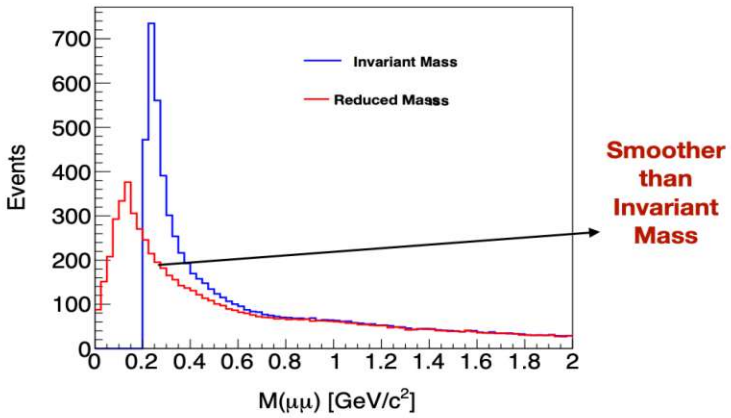


Figure 6.28.: Comparison of reduced mass and invariant mass. Reduced mass is found to be smoother than invariant mass.

6.7.1. Signal shaping

The signal shape at a specific mass is described by seven parameters: f_{CB1}, σ_1, α_1, n_1, σ_2, α_2, n_2. In our final Z' search strategy at data, we plan to look at thousands of signal hypotheses: in principle, it is possible, though uncomfortable and unpractical, to perform fits for each of those hypotheses and store the parameters for later use. We didn't follow that procedure and opted to use analytical functions for each CB parameter to model their dependence as a function of the Z' mass. I fit all the CB parameters separately in three different mass regions: 0.212 GeV - 1 GeV, 1 GeV - 7 GeV, and 7 GeV - 9 GeV. Figure 6.30, 6.31, 6.32 show the modeling of the CB parameters for the three different mass ranges, respectively.

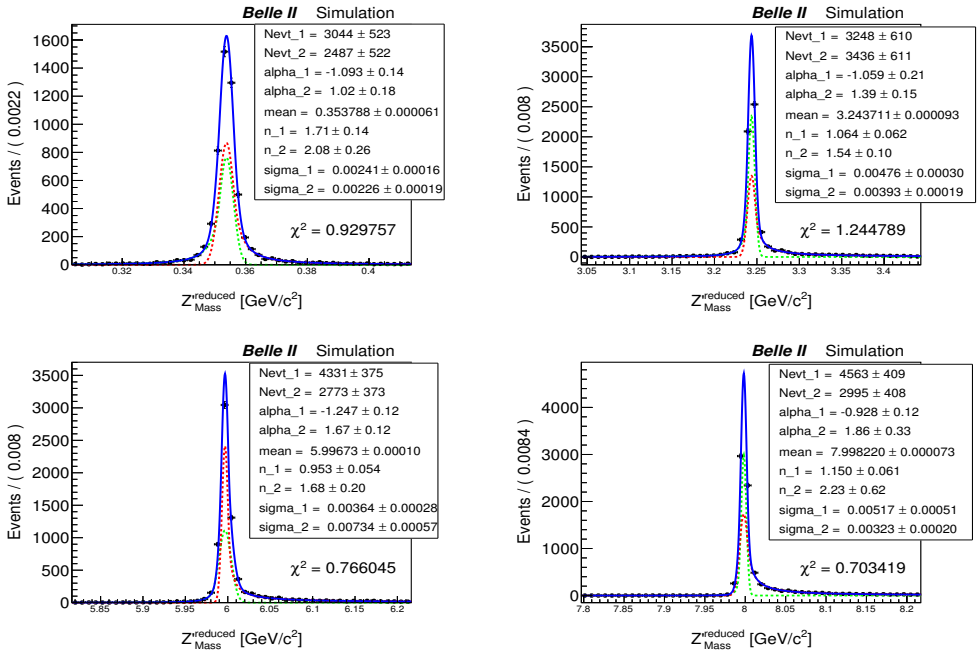


Figure 6.29.: Examples of double Crystall fits for 0.412 GeV (top left), 3.25 GeV (top right), 6.0 GeV (bottom left), and 8.0 GeV (bottom right) masses, respectively.

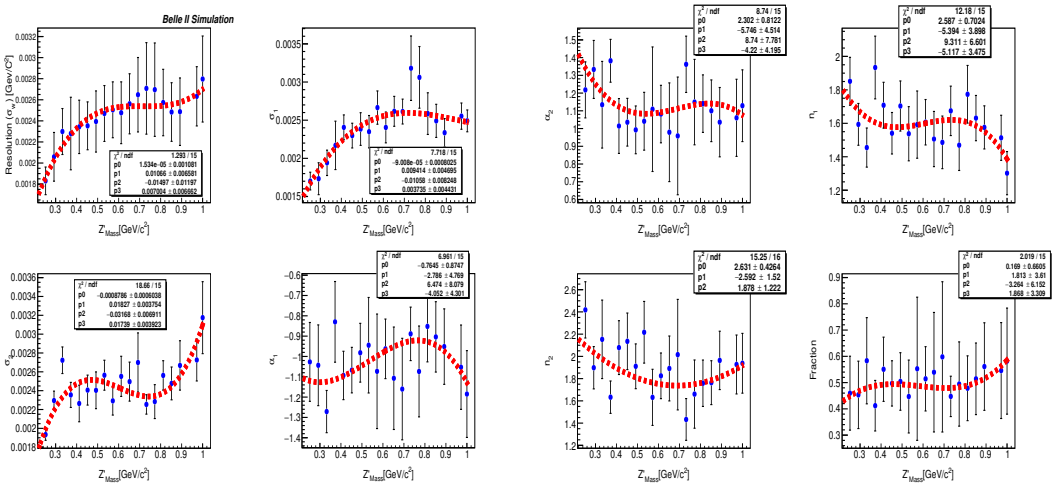


Figure 6.30.: Modeling of double CB parameters for 0.212 GeV - 1 GeV mass range.

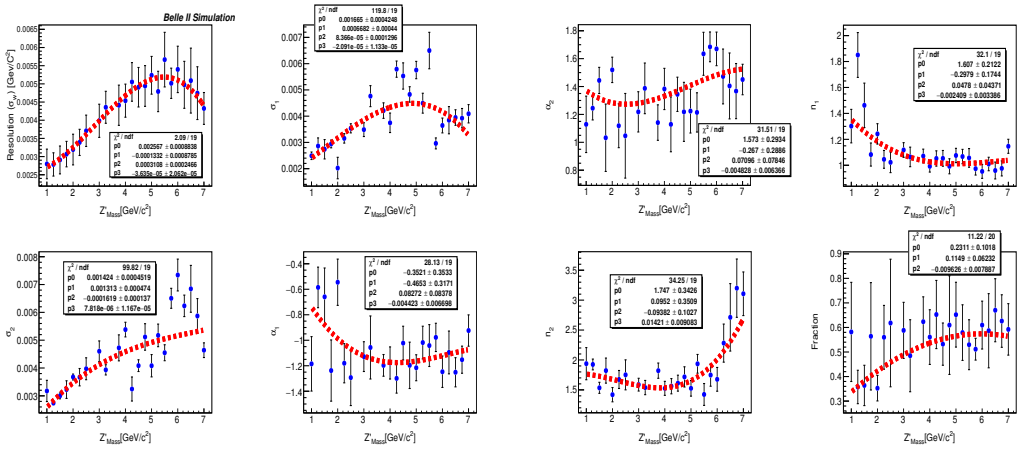


Figure 6.31.: Modeling of double CB parameters for 1 GeV - 7 GeV mass range.

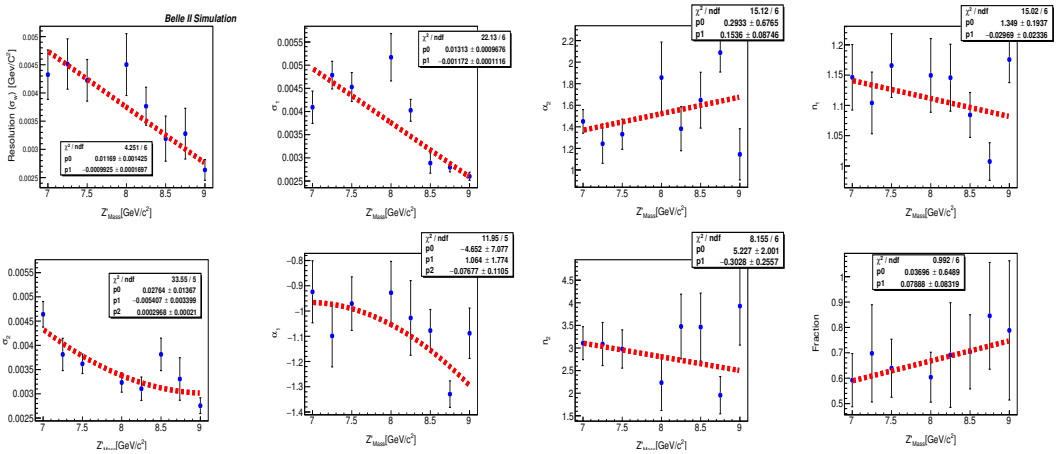


Figure 6.32.: Modeling of double CB parameters for 7 GeV - 9 GeV mass range.

The mass resolution, one of the most important parameters of interest with direct consequences to the sensitivity, ranges approximately between 2 and 5.5 MeV/c^2.

Table 6.2 summarizes all the information about the continuous analytic functions found for the different mass ranges. I also cross-checked the reliability of the obtained analytical functions by comparing the resolution σ_w in the polynomial modeling with the corresponding values obtained from the double CB fits mass by mass and with the values obtained by using the polynomial modeling for f_{CB}, σ_1 and σ_2. Figure 6.33 (left) compares these three cases. Figure 6.33 (right) compares the expected χ^2 value while fixing all the parameters to their polynomially

Mass range (GeV)	polynomial (x = mass
0-1	σ_w = 0.00001534 + 0.01066x - 0.01497x^2 + 0.007004x^3 σ_1 = -0.00009008 + 0.009414x - 0.01058x^2 + 0.003735x^3 σ_2 = -0.0008786 + 0.01827x - 0.03168x^2 + 0.01739x^3 α_1 = -0.7645 - 2.786x + 6.474x^2 - 4.052x^3 α_2 = 2.302 - 5.746x + 8.74x^2 - 4.22x^3 n_1 = 2.587 - 5.394x + 9.311x^2 - 5.117x^3 n_2 = 2.631 - 2.592x + 1.878x^2 f_{CB1} = 0.169 + 1.813x - 3.264x^2 + 1.868x^3
1-7	σ_w = 0.002567 - 0.0001332x + 0.0003108x^2 - 0.00003635x^3 σ_1 = 0.001665 + 0.0006682x + 0.00008366x^2 - 0.00002091x^3 σ_2 = 0.001424 + 0.001313x - 0.0001619x^2 + 0.000007818x^3 α_1 = -0.3521 - 0.4653x + 0.08272x^2 - 0.004423x^3 α_2 = 1.573 - 0.267x + 0.07096x^2 - 0.004828x^3 n_1 = 1.607 - 0.2979x + 0.0478x^2 - 0.002409x^3 n_2 = 1.747 + 0.0952x - 0.0938x^2 + 0.01421x^3 f_{CB1} = 0.2311 + 0.1149x - 0.009626x^2
7-9	σ_w = 0.01169 - 0.0009925x σ_1 = 0.01313 - 0.001172x σ_2 = 0.02764 - 0.005407x + 0.0002968x^2 α_1 = -4.652 + 1.064x - 0.07677x^2 α_2 = 0.2933 + 0.1536x n_1 = 1.349 - 0.02969x n_2 = 5.227 - 0.3028x f_{CB1} = 0.03696 + 0.07888x

Table 6.2.: Continuous analytic functions for all the mass ranges.

modeled values, with the value coming directly from the double CB fit with all the parameters floating.

6.8. Background fit procedure

The main surviving background components after all the analysis selections are shown in Figure 6.34. It is mainly dominated by the $\mu^+\mu^-\mu^+\mu^-$ and, to a lesser extent, $\mu^+\mu^-(\gamma)$ processes in the low mass region. BaBar and Belle's analysis

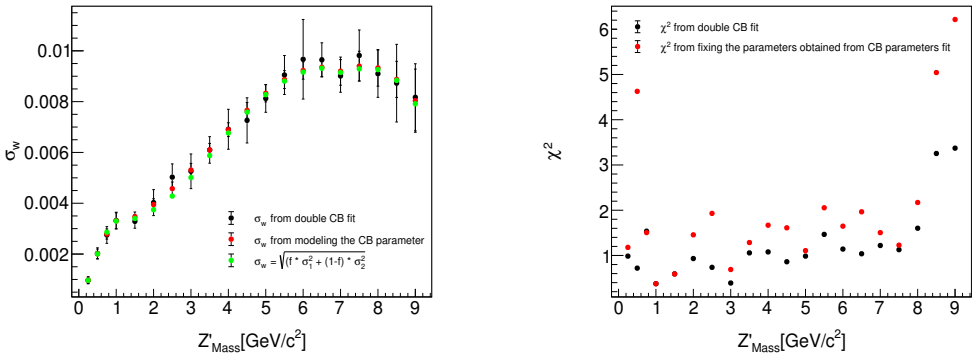


Figure 6.33.: Comparison of the mass resolution σ_w in the polynomial modeling, from individual fits mass by mass and with the values from the polynomial modeling of f_{CB}, σ_1 and σ_2, σ_w (left), and χ^2 (right)

observed that the background due to the SM $\mu^+\mu^-\mu^+\mu^-$ process is lower than the MC prediction at data because of the lack of ISR in the generator. In this section, **I scale the $\mu^+\mu^-\mu^+\mu^-$ component by a factor 0.65**, to be eventually adjusted after the post-unblinding direct measurement of the background. I would like to remind you that BaBar and Belle didn't use any background suppression using MVA techniques. During partial unblinding in section 6.14 we found that a 0.65 scale factor is not required because MLP can reject the unexpected sources of backgrounds (mainly ISR). The final expected sensitivity projection is given in section 6.15.

The fitting procedure proceeds as follows,

- I select fit intervals of $\pm 30\sigma_w$ around each interested mass point.
- If the fitting interval extends to two different MLP ranges, I use only the MLP corresponding to the range where the central mass value is located to avoid discontinuities.
- I parameterize the background using first-order Chebychev polynomials. Using a low-order polynomial provides an accurate description through the entire mass range because the mass resolution σ_w is small and massive background reduction due to MLP. I also checked the effect of a higher-order polynomial. Figure 6.35 shows the fitted second-order coefficient as a function of the mass when a second-order polynomial is used. These are average values obtained using samples created with a bootstrap technique. Values are minimal everywhere, with an average of -0.005, confirming that using

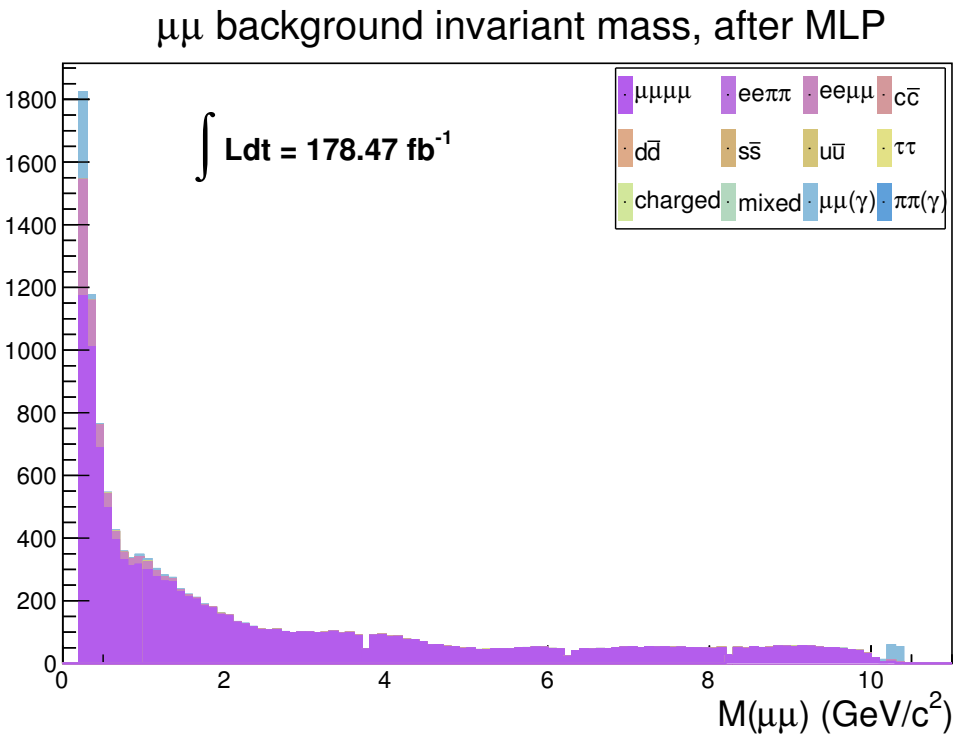


Figure 6.34.: dimuon candidate invariant mass background distribution after MLP application. The main contributions come from $\mu^+\mu^-\mu^+\mu^-$ and $\mu^+\mu^-(\gamma)$ processes.

the first-order polynomial effectively describes the background shape. In addition, the usage of second-order polynomials leads to slightly worse final sensitivities. This is another crucial difference with respect to BaBar and Belle analyses, which use more complicated analytical functions. The usage of the MLP numerically reduces the background in our case hence the lower order polynomials.

- I parameterize the signal shapes using the double CB functions with parameters taken from our analytical modeling and kept fixed to those values

- I fit the MC background distributions with a 1D Unbinned Maximum Likelihood technique, using two hypotheses: a) background-only hypothesis; b) signal + background hypothesis.

For the background-only hypothesis, our model is

$$f^{bkg}(M) = N_{bkg} \cdot (1 + a_1 \cdot M) \tag{6.20}$$

while for the signal + background hypothesis it is

$$f^{sig+bkg}(M) = N_{sig} \cdot (CB_1 + CB_2) + N_{bkg} \cdot (1 + a_1 \cdot M) \tag{6.21}$$

where M is the reduced dimuon mass. The fit returns the number of background events N_{bkg} in the background-only hypothesis and the number of signal events N_{sig}, and the number of background events N_{bkg} in the signal + background hypothesis. In this latter case, we allow negative values for N_{sig} because this is shown to improve the final sensitivity slightly. Figure 6.36 shows some example mass points where both hypotheses are checked on background-only distributions.

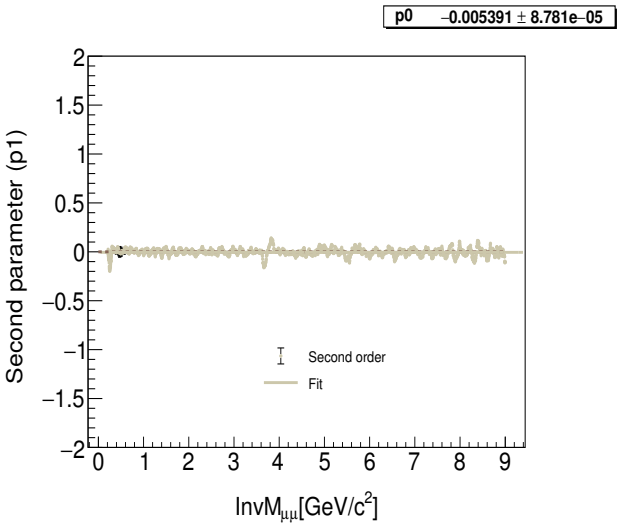


Figure 6.35.: Values of the fitted second order coefficient as a function of the mass, when I parameterize the background using a second order polynomial: one could see that values are very close to zero.

6.8.1. Preliminary sensitivity studies

We want to estimate the sensitivity for the process $e^-e^- \to \mu\mu Z', Z' \to \mu\mu$ both in terms of the process cross section σ and of the coupling constant g'. I use the asymptotic calculator [96] with a one-sided Profile likelihood test statistics to estimate the upper limit on the cross-section in the presence of systematic uncertainties affecting the signal efficiency for each mass point. I also tested other available calculators in the RooStat and traditional toy mc method for upper limit estimation; it was found that all were consistent. I planned to choose the asymptotic calculator as it is very fast in terms of computing time. The model explained

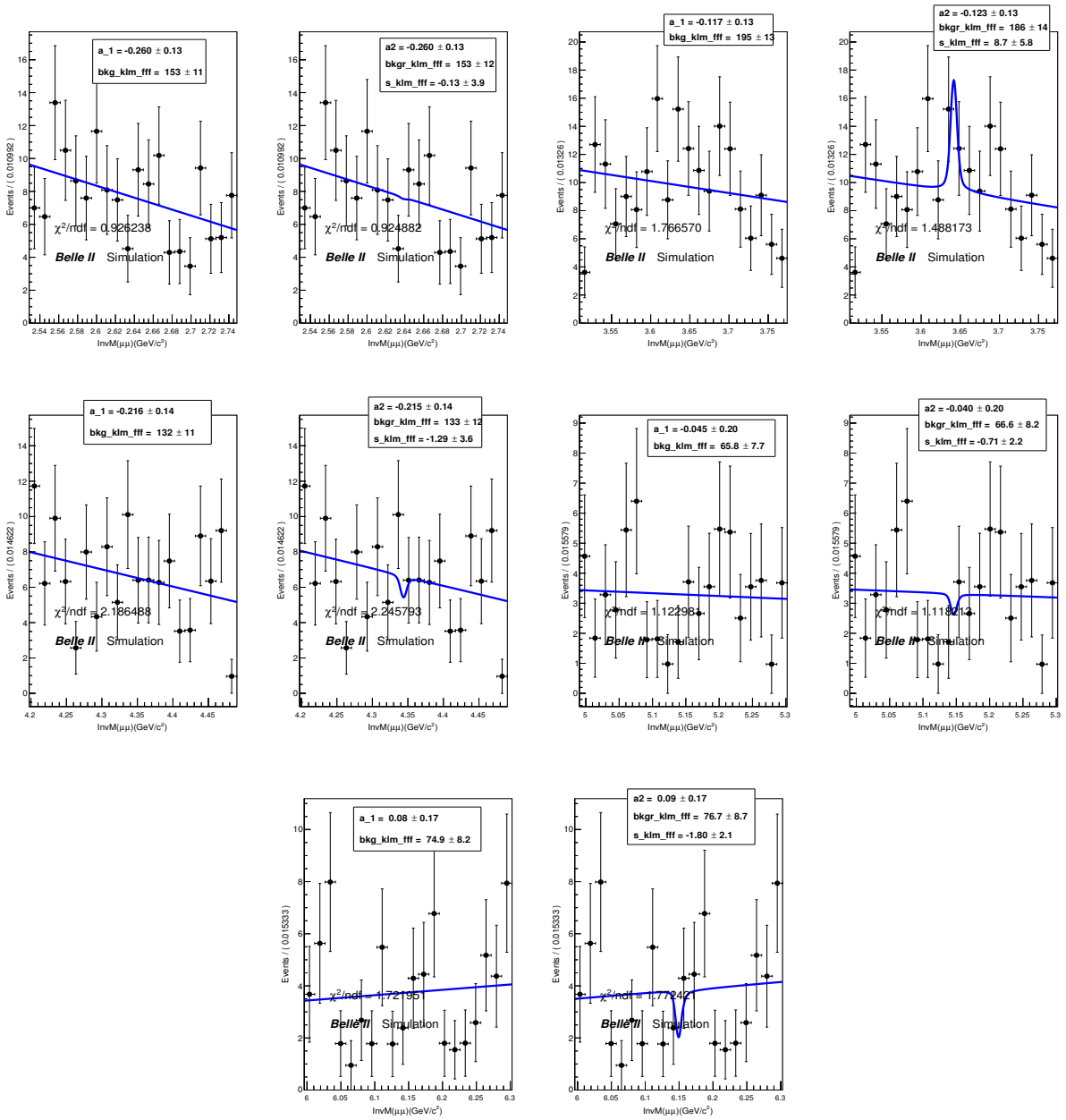


Figure 6.36.: Fit results for mass points 2.65 GeV/c^2, 3.65 GeV/c^2, 4.35 GeV/c^2, 5.25 GeV/c^2, 6.15 GeV/c^2: background-only (left) and signal+background (right) hypotheses.

in the previous section is slightly modified now to incorporate systematics. The number of events N_{sig} is replaced by,

$$N_{sig} = \mathcal{L} \times \sigma \times \epsilon_{sig} \tag{6.22}$$

where $\mathcal{L}$ is the target integrated luminosity, σ is the process cross-section, and ϵ_{sig} is the signal efficiency. The signal + background model is, therefore

$$f^{sig+bkg}(M) = \mathcal{L} \times \sigma \times \epsilon_{sig} \cdot (CB_1 + CB_2) + N_{bkg} \cdot (1 + a_1 \cdot M) \tag{6.23}$$

As the signal efficiency is multiplied by the integrated luminosity, I technically assign systematic uncertainties to the latter, applying Gaussian smearing whose effects propagate on the cross-section. Using the Gaussian smearing on the luminosity, I get the final model PDF to be fitted:

$$f^{sig+bkg}(M) = G(\mathcal{L}) \times \mathcal{L} \times \sigma \times \epsilon_{sig} \cdot (CB_1 + CB_2) + N_{bkg} \cdot (1 + a_1 \cdot M) \tag{6.24}$$

where $G(\mathcal{L})$ is the Gaussian-smeared luminosity. The relevant fitted parameter is the process cross-section σ. The signal efficiency used is already shown in Figure 6.24. Figure 6.37 shows the estimated 90% CL upper limits on the cross-section σ_{UL} for mass points 0.5 GeV/c^2, 1.1 GeV/c^2, 1.95 GeV/c^2, 2.55 GeV/c^2 using the described procedure. The upper limit on the signal yield is given by

$$N_{UL}^{sig} = \mathcal{L} \times \sigma_{UL} \times \epsilon_{sig} \tag{6.25}$$

Figure 6.38 shows the upper limits on the cross-section and the signal yield for our target integrated luminosity of 178.47 fb^{-1} using 1%, 10% and 20% flat systematic uncertainties for all mass hypotheses (at this point we are not finalized with our all systematics, and a constant systematics is given to all mass values, later we evaluated mass dependent systematics see section 6.15). Figure 6.38 also shows the expected sensitivity projection at 500 fb^{-1} with 20% flat systematics. This is done to compare our results with the BaBar easily.

The corresponding 90% CL upper limits in terms of the coupling constant g'^{UL} are given by

$$g'^{UL} = g'_{ref}\sqrt{\frac{\sigma^{UL}}{\sigma_{ref}}} \tag{6.26}$$

where g'_{ref} is the reference coupling constant used in the Madgraph generator to obtain the reference cross section (σ_{ref}). Figure 6.39 shows the expected 90% CL sensitivity in terms of g' for 178.47 fb^{-1} and 500 fb^{-1} integrated luminosities and compared with the results from BaBar. The effect of the different assumptions on the systematic uncertainties is shown in Figure 6.40. They do not play a big role, accounting for a ≈2.5% worsening for a 20% change in the size of the systematic

uncertainty. The comparison with BaBar is better seen in Figure 6.41, where ratios of the BaBar results with our expected sensitivities are shown for various assumptions on the systematic uncertainty sizes. Regardless of the systematic uncertainties, which do not play a significant role, our expectations at the 178.47 fb^{-1} target luminosity are worse than BaBar below $\approx$1 GeV/c^2, comparable, but on average better, in the range 1–6.5 GeV/c^2, and worse above $\approx$7.5 GeV/c^2. Our expected sensitivities for 500 fb^{-1} integrated luminosity are better than BaBar by an average $\approx$40%.

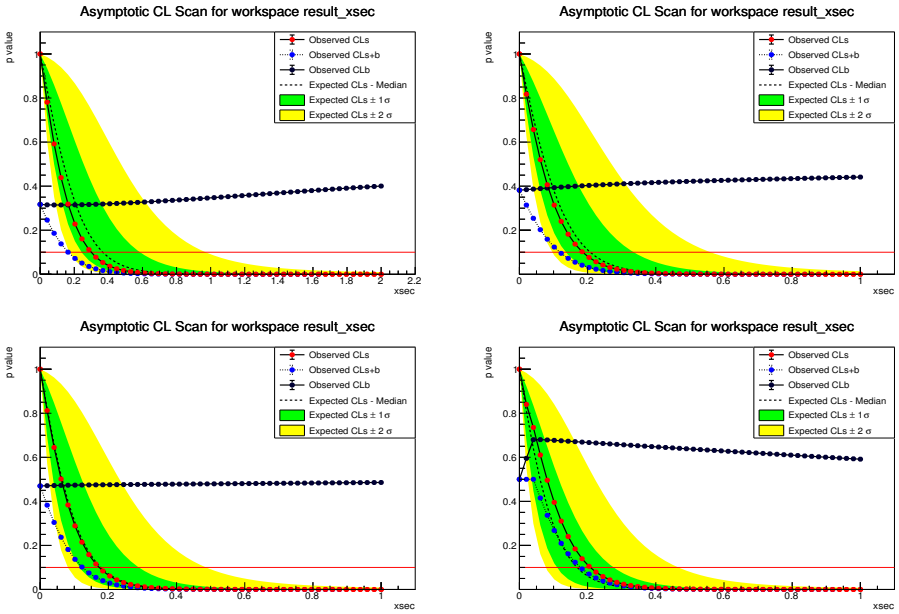


Figure 6.37.: Upper limit estimation on cross section for masses 0.5 GeV/c^2 (top left), 1.1 GeV/c^2 (top right), 1.95 GeV/c^2 (bottom left), and 2.55 GeV/c^2 bottom right.

6.9. Fit stability studies

6.9.1. Background generation using PDF

In this section, I check the stability and self-consistency of the fitting procedure by using a toy MC technique. Both signal and background shapes are generated from PDF distributions (double Crystal ball function for the signal and first-order Chebyshev polynomial for the background), with Poisson fluctuations on the expected yield. The injected signal yields are 5, 10, 15, 20, and 50 events, and the

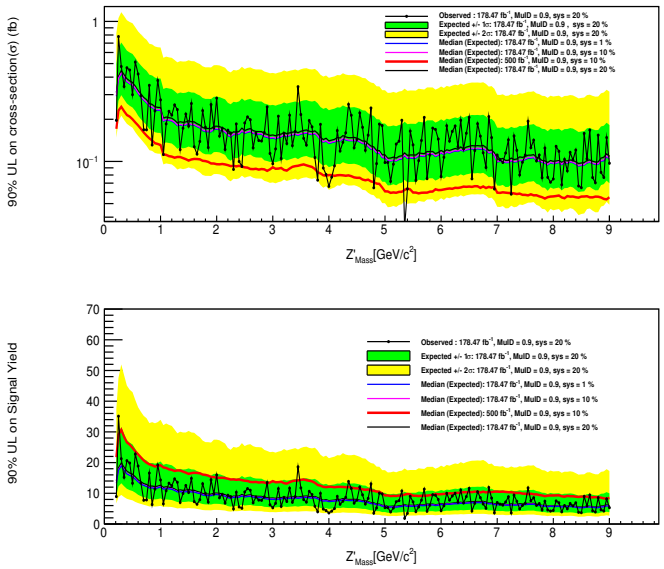


Figure 6.38.: Estimated 90% CL upper limit on the cross-section (top) and on the signal yield (bottom) for different luminosity and systematic scenarios.

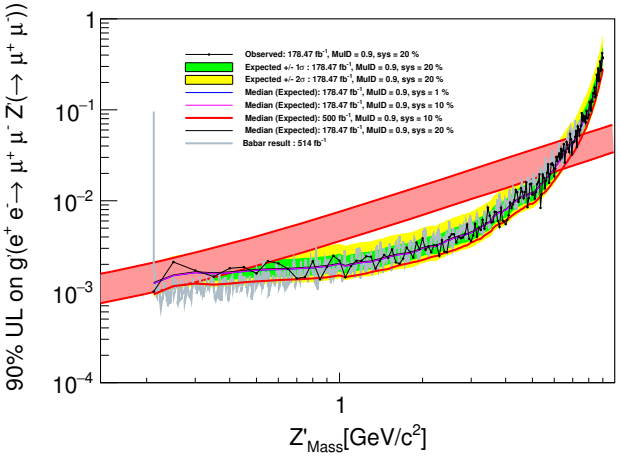


Figure 6.39.: Expected 90% CL on the coupling constant g' with different luminosity and systematics scenarios. Also shown band is that would explain the observed $(g-2)_\mu$.

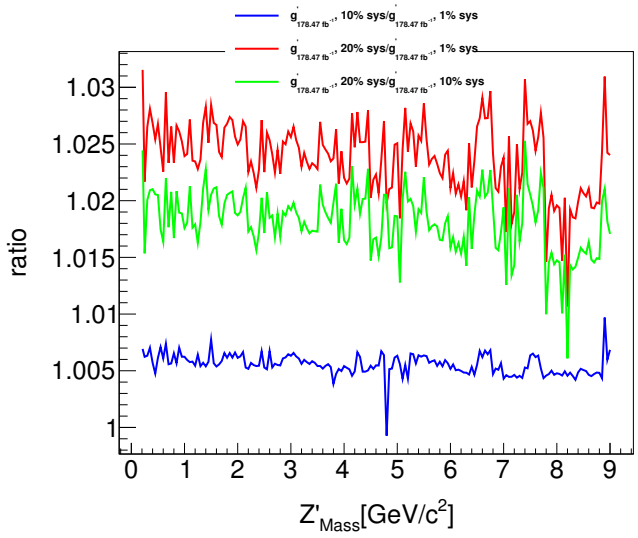


Figure 6.40.: Effect of systematic uncertainties on the sensitivity for our 178.47 fb^{-1} target integrated luminosity. Ratios of expected sensitivities corresponding to 1%, 10%, and 20% assumptions on systematic uncertainties are shown.

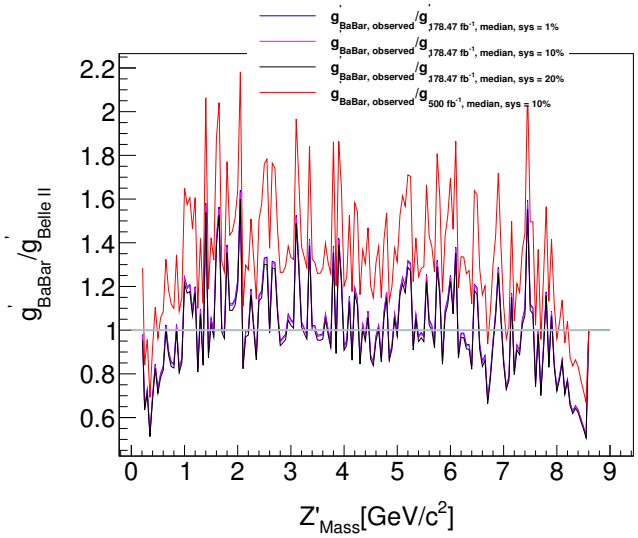


Figure 6.41.: comparison of expected Belle II sensitivities for 178.47 fb^{-1} and 500 fb^{-1} of the BaBar results.

background is expected from MC for 178.47 fb^{-1} integrated luminosity. 175 sample mass points are tested, with an interval of 0.05 GeV/c^2. A total of 500 toys are generated for each mass point. A summary of the results is shown in Figure 6.42, separately for each injected signal yield. Pull distributions of some mass hypothesis are given in Figure 6.44, Figure 6.45, Figure 6.46, Figure 6.47. The pulls are, on average, all slightly negative and the pull sigmas are all slightly above 1. The same tendency shows up in the bias of the fitted yields (bottom-right plot of Figure 6.42), or relative yield variations, defined as (fitted yield - true value)/(true value): it is on the average negative for all injected signal yields, but with a negligible value. Distributions of relative yield variations are shown in Figure 6.43, with rms in the range $\approx$ 2–6%.

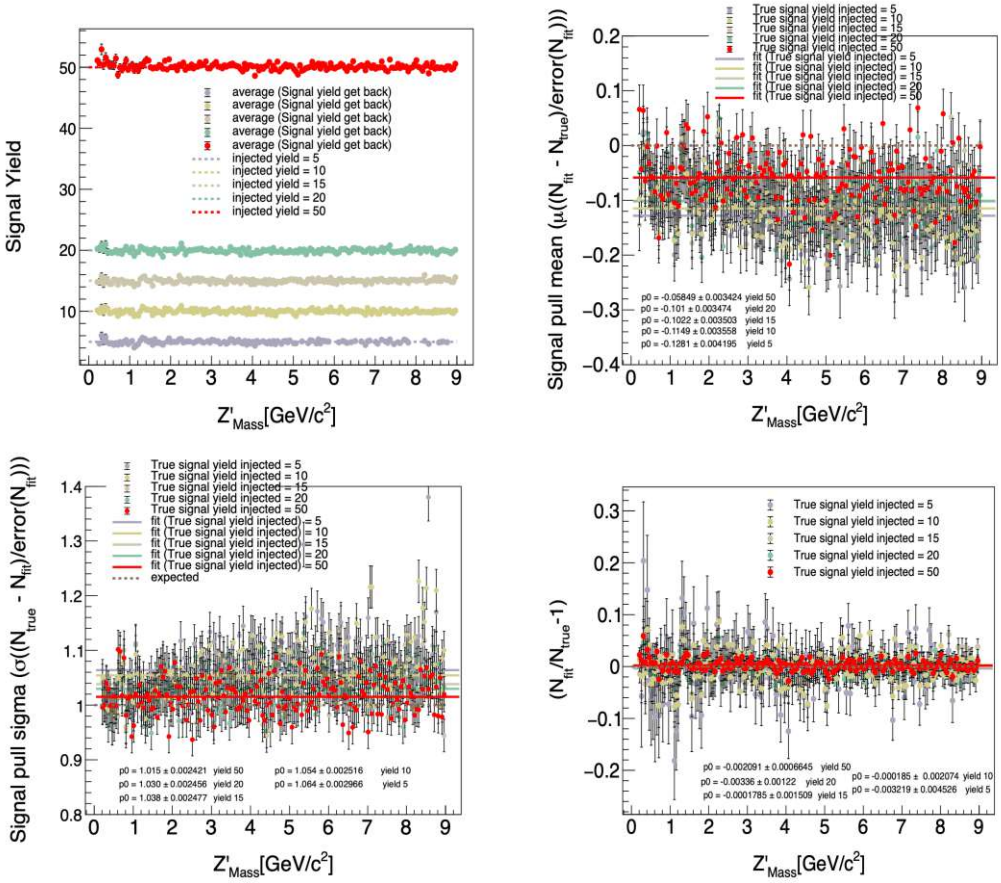


Figure 6.42.: Distribution of signal yield, pull means, pull sigmas and relative yield variations. In the above plots, all the dashed lines are expected values and solid lines are the fitted results.

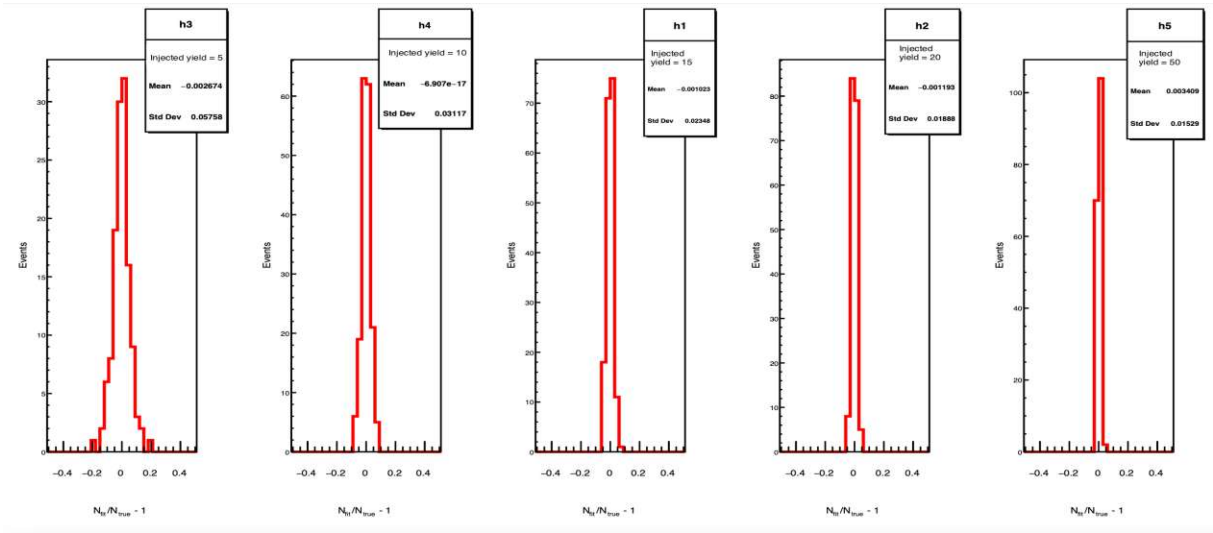


Figure 6.43.: Distributions of relative yield variations for the different injected signal yields .

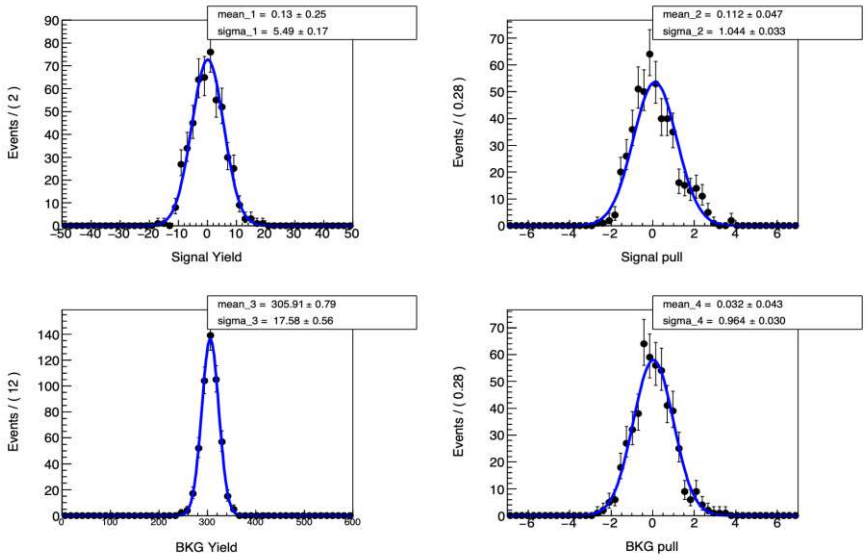


Figure 6.44.: Pull distribution for the mass point 1.5 GeV/c^2 with injected signal yield 0.

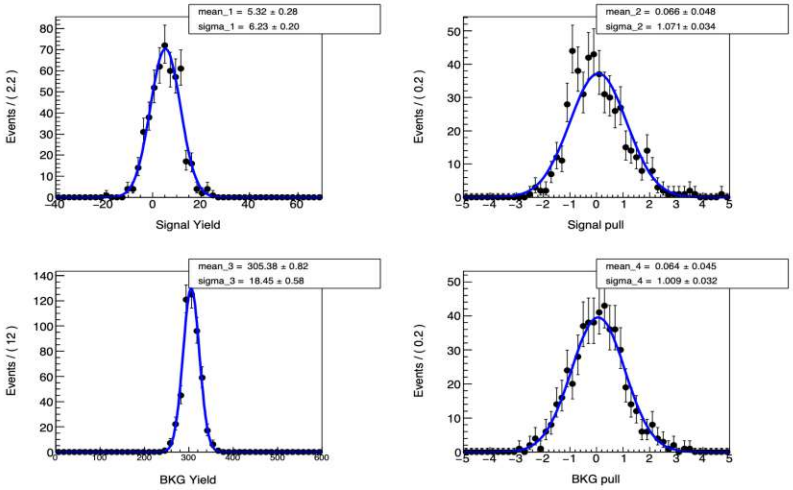


Figure 6.45.: Pull distribution for the mass point 1.5 GeV/c^2 with injected signal yield 5.

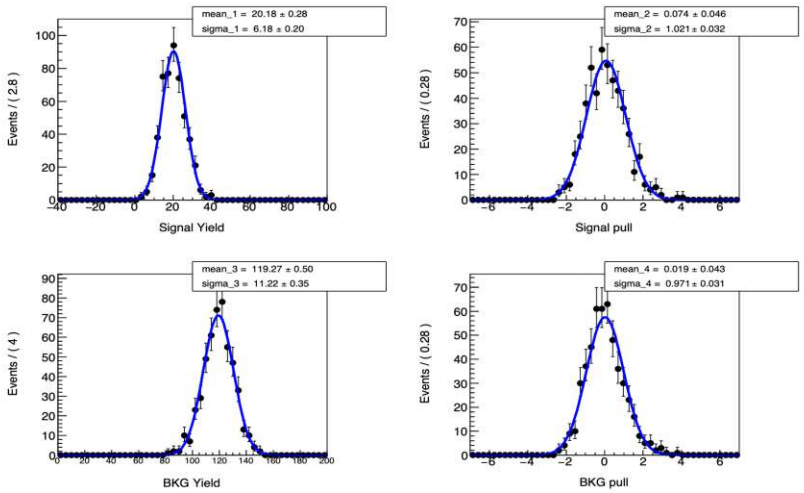


Figure 6.46.: Pull distribution for the mass point 4.5 GeV/c^2 with injected signal yield 20.

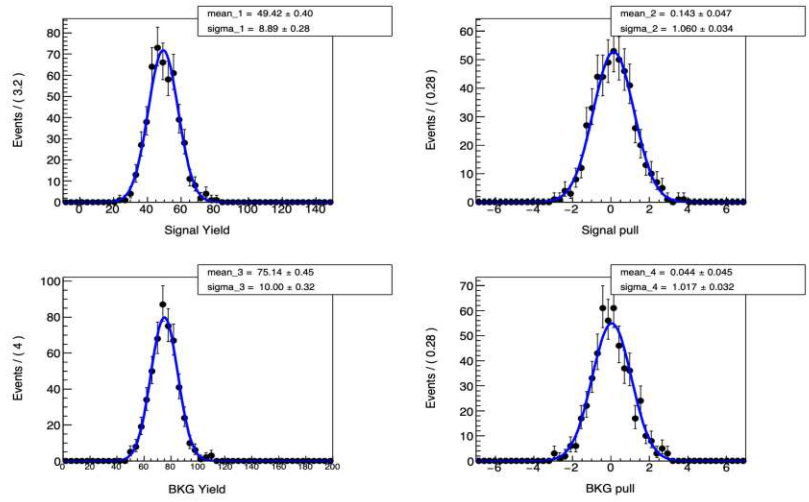


Figure 6.47.: Pull distribution for the mass point 7.5 GeV/c^2 with injected signal yield of 50.

6.9.2. Background generation using MC (bootstrap)

In this section, I check the stability of the fitting procedure in a more realistic scenario by using a bootstrap technique. In this case, signal shapes are generated from PDF distributions, and background shapes are generated from MC distributions, with Poisson fluctuation on the expected yields. A total of 500 toys are generated. The injected signal yields are 0, 5, 10, 15, 20, and 50. The background is expected from MC for 178.47 fb^{-1} integrated luminosity. Sample mass hypothesis are 0.5 GeV/c^2, 1.5 GeV/c^2, 2.5 GeV/c^2, 3.5 GeV/c^2, 4.5 GeV/c^2, 5.5 GeV/c^2, 6.5 GeV/c^2, 7.5 GeV/c^2, 8.5 GeV/c^2. A summary of the results is given in Figure 6.48. Distributions of some mass hypotheses are shown in Figure 6.49, Figure 6.50. The pull mean and pull sigma seem to be consistent with 0 and within 1-2σ and are also consistent with the toys generated from the PDF in the previous section.

6.10. Significance

I describe here the procedure for calculating significance. The significance is computed as the square root of twice the difference between the negative logarithm of the likelihoods (NLL) of the fits on background-only samples with signal+background and background pdf. Describe by the equation given below:

$$\mathrm{S} = \sqrt{2 \cdot (\mathrm{NLL}_{sig+bkg} - \mathrm{NLL}_{bkg})} \tag{6.27}$$

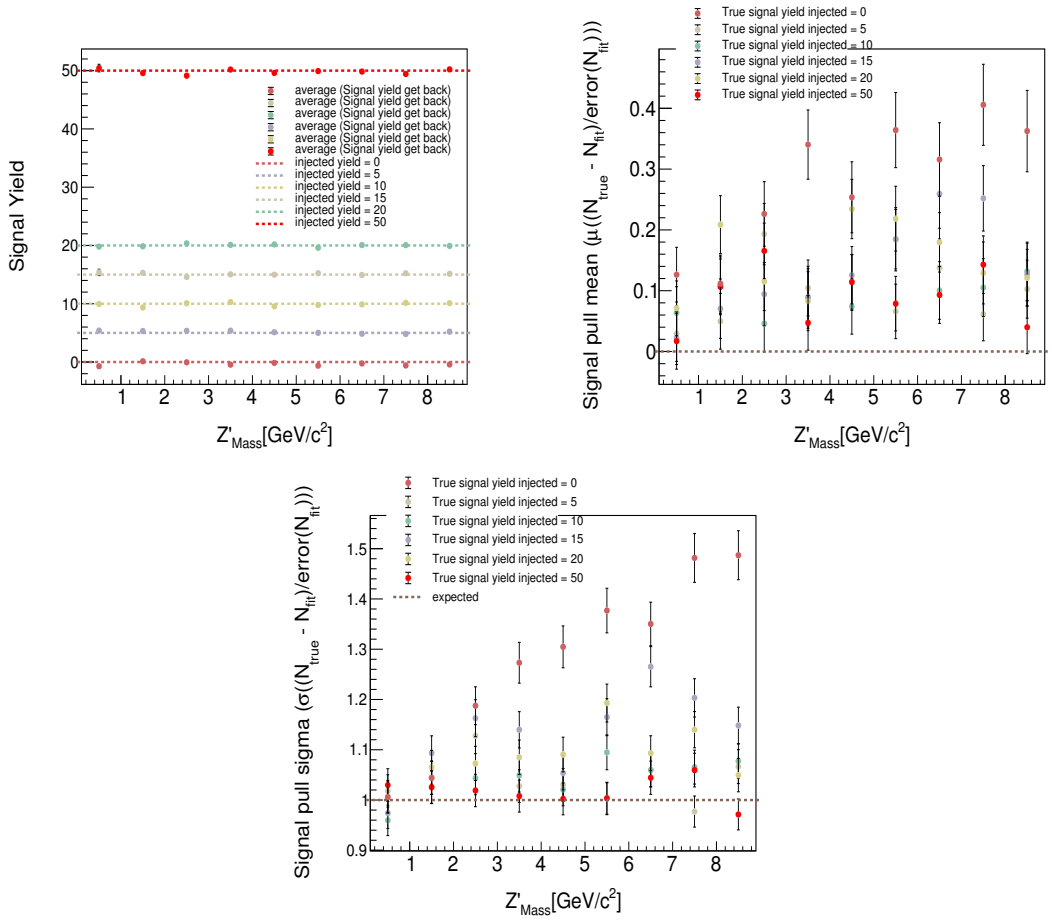


Figure 6.48.: Distribution of signal yield, pull mean, and pull sigma. In the above plots, all the dashed lines are expected and dots are obtained for the average of toys

Significances are computed on the simulated samples and reflect their statistical fluctuations. In more detail, the background MC sample is fitted with signal+background pdf (double crystal ball + first-order Chebyshev polynomial) and $\mathrm{NLL}_{signal+bkg}$ is computed. Then again, it is fitted with background-only pdf (first-order Chebyshev polynomial), and NLL_{bkg} is computed.
Figure 6.51 shows the significance mass by mass. As the signal yield is allowed to be negative in the fit, so for mass points where the obtained yield is negative, I assign a negative significance for those mass points. Negative signal yields do not make sense physically, but they are expected to improve the sensitivity and reduce biases.

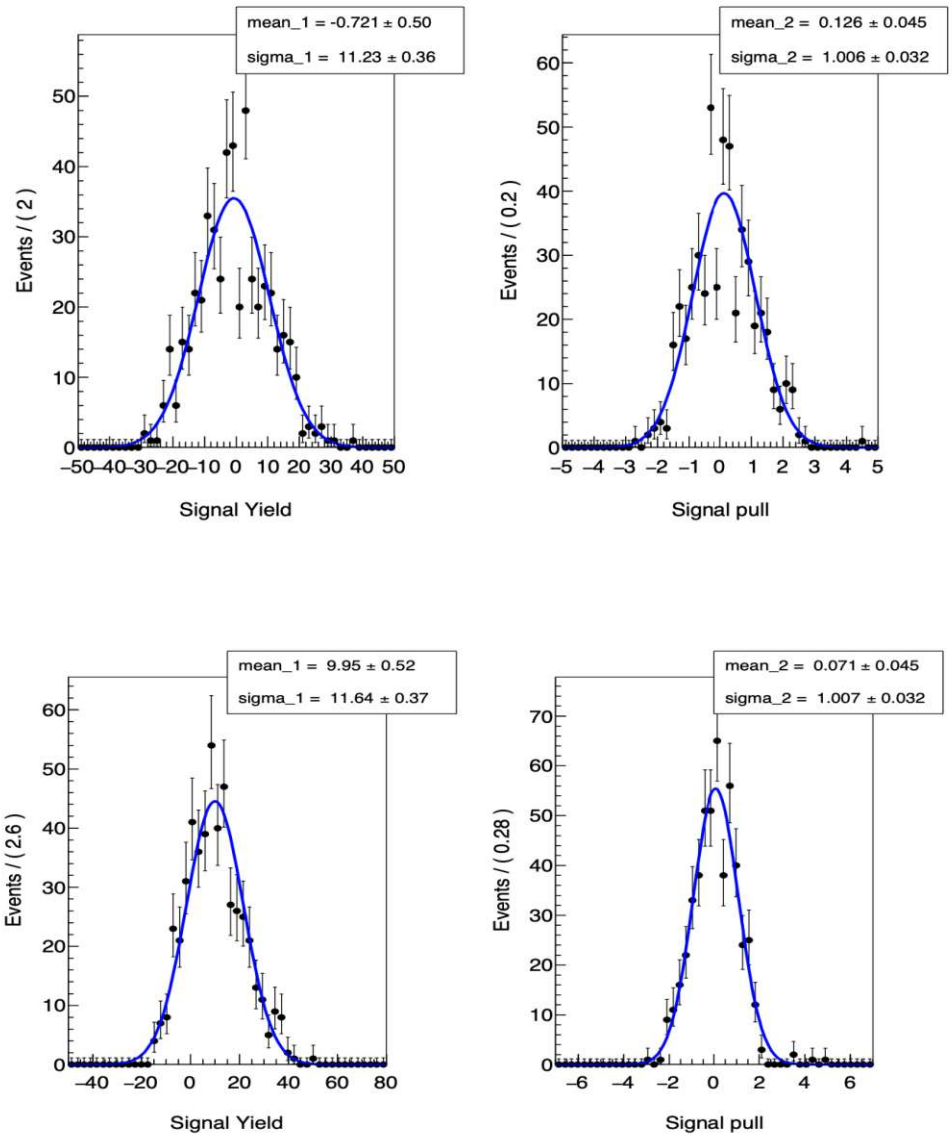


Figure 6.49.: Pull distribution for the mass point 0.5 GeV/c^2 with injected signal yield 0 and 10 respectively.

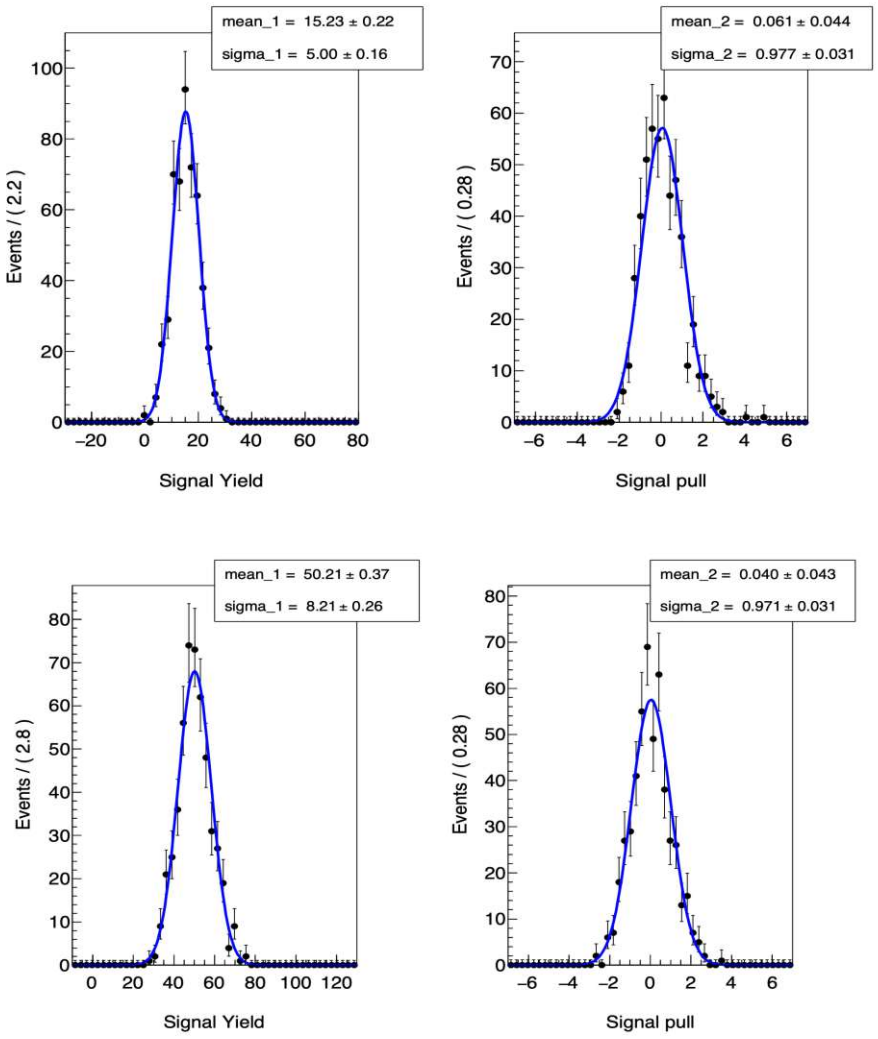


Figure 6.50.: Pull distribution for the mass point 7.5 GeV/c^2 and 8.5 Gev/c^2 with injected signal yield 15 and 50 respectively.

6.11. Scanning strategy and Look-Elsewhere Effect (LEE)

I devise a scanning strategy to a predefine set of mass points over which to run our fitting procedure before looking at the final data. The starting point is set to Z' masses just above the dimuon threshold, at 0.212 GeV/c^2. We limit the last point to 9 GeV/c^2 because we don't have a good sensitivity above. The chosen

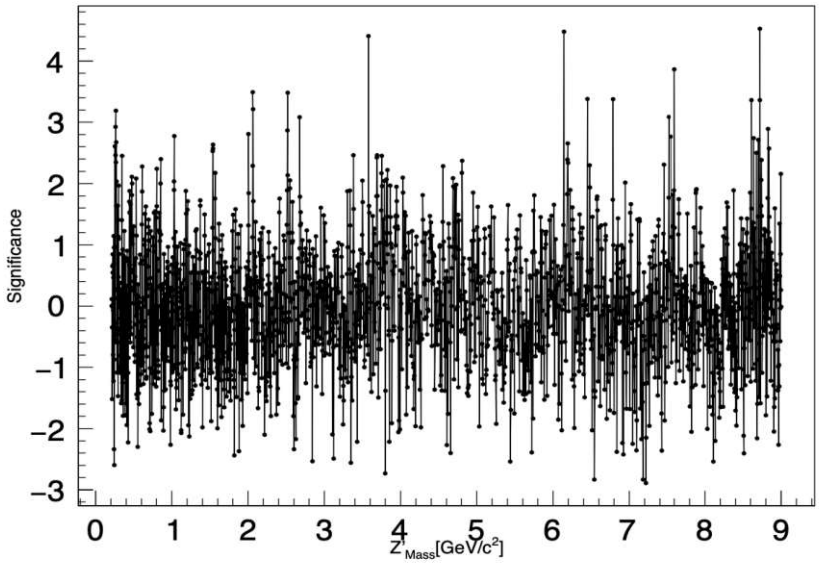


Figure 6.51.: Significance obtained allowing negative signal yields.

scan step is one signal mass resolution. With this choice, I get 2315 mass points in the interval 0.212–8.999 GeV/c^2.

I checked if using one σ scanning step causes some losses if a signal is present at a mass within two predefined masses. I performed some checks to confirm if there some signal present within one σ step and will be able to catch it or not using one σ step; I injected signals halfway between two contiguous masses among 2315 predefined for scanning. Then run our fitting procedure on the three concerned masses, the one in which I inject signal and the two contiguous ones, half σ away on each side, and compare the resulting significances. In all cases, at least one of the fits corresponding to the contiguous masses resulted in a significance negligibly degraded or not degraded at all, compared to the significance from the fit centered on the injection mass. We, therefore, decide to stay with 1 σ steps. In the scanning strategy, I also excluded the region around the J/ψ mass, which is a peaking source of background for our search. We will not give results for the 3.07–3.12 GeV/c^2 dimuon mass interval, determined from the ee$\mu\mu$ control sample study (see section 6.12).

The significance calculated in the last section is the local excess from background fluctuations. Since we search for a signal of Z' with an unknown mass, the significance of observing a local excess from background fluctuations anywhere in the search range must be taken into account to get the global significance. I approx-

imate the global p-value (p_{global}) in the asymptotic limit by introducing the trial factor [97], which is the ratio between the probability of observing the excess at some fixed Z′ mass point to the probability of observing it anywhere in the range.

$$N_{trial} = 1 + \frac{1}{p_{local}} \langle n_{up}\langle Z_{test}\rangle\rangle \; e^{\frac{Z^2_{test}-Z^2_{local}}{2}} \tag{6.28}$$

where $\langle n_{up}\langle Z_{test}\rangle\rangle$ is the average number of up-crossings of the significance Z (Z_{local}) above a reference significance Z_{test}. The local significance is translated to global significance by

$$p_{global} = p_{local} \; + \langle n_{up}\langle Z_{test}\rangle\rangle \; e^{\frac{Z^2_{test}-Z^2_{local}}{2}} \tag{6.29}$$

I implemented the LEE study with a toyMC technique. Using the bootstrap method, I generated 500 different toy background distributions starting from the original one, with random Poisson fluctuations to the background yield. For each toy, I fitted 2315 mass points in 1σ mass resolution step and computed the number of up-crossings for each toy using a significance threshold of $+1\sigma$.

Figure 6.52 shows the distribution of the number of up-crossings. The mean number of up-crossings is $\langle n_{up}\langle Z_{test}\rangle\rangle \sim 58$. For a $p_{local} \approx 3 \cdot 10^{-7}$, corresponding to 5σ, we obtain $N_{trial} = 1189$, $p_{global} = 0.000135$, corresponding to 3.38σ.

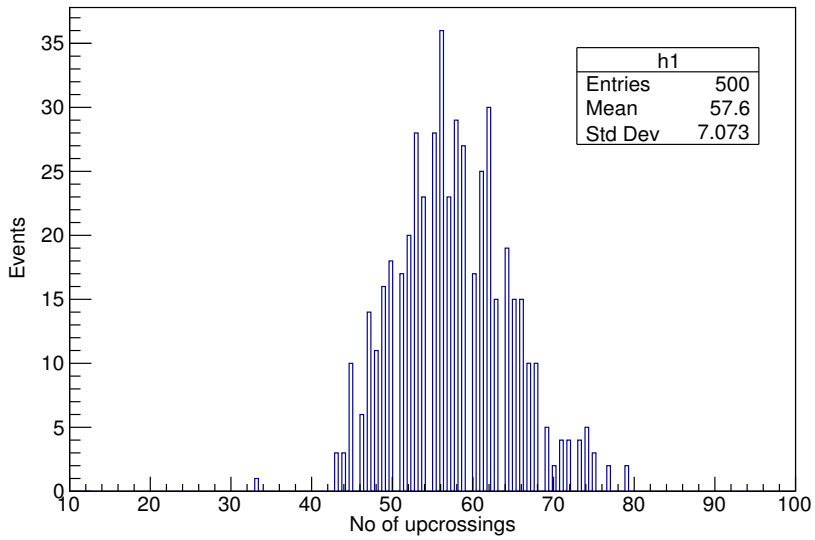


Figure 6.52.: Distribution of the up-crossings considering $Z_{test} = 1$.

6.12. Control sample study

6.12.1. Control sample selection

Being a blinded analysis, we planned to validate all our procedures and evaluate possible systematics sources in the control channel. The Control channel is the transition exactly similar to our signal topology. We used the $ee\mu\mu$ transition as our control channel. The main advantage to using a control channel is that we can look at the data but keep the signal region fully blinded.

We used the same selections as for the main analysis (see section 6.3.5), with the only difference being in particle ID requirement: we required the presence of two identified muons, with μID>0.9, and two electrons with eID>0.5. We will have four possible track pairs with zero net charges again, but in this case, they are not perfectly equivalent: $\mu\mu$, ee, eμ+μe.

Two main physics processes contribute to ee$\mu\mu$ final states: they are the SM ee$\mu\mu$ process and the $\mu\mu(\gamma)$ process.

The SM ee$\mu\mu$ process is generated with AAFH, which does not contain ISR and FSR. We, therefore, expect severe discrepancies between data and MC, even more, prominent than in the 4μ case, because the presence of electrons make it more worse due to FSR contributions. Comparisons of data and MC will suffer from these effects. So we decided to check the MLP relative efficiency defined as the ratio of events before and after the MLP selection, which will cancel the ISR effects but conserve all essential pieces of information. We also check the data MC ratio and MLP relative efficiency very narrow mass region around $\Upsilon(4S)$ to get the minimal ISR effect.

6.12.2. Data MC comparison

The ee$\mu\mu$ invariant mass distribution before the MLP selection is shown in Figure 6.53. The bin-by-bin ratio between data and MC yields is shown at the bottom. We performed a fit on this ratio with a zeroth order degree polynomial. The pattern of the discrepancy follows the expectations, with data above MC for invariant masses below 10 GeV/c^2 (not shown in the plot) and above, data well below MC due to the lack of ISR and FSR processes in the generator. The fitted average value of the data/MC ratio above 10 GeV/c^2 is close to 0.5. The studies of BaBar and Belle on the 4μ final state found such a ratio in the range of 0.65–0.7, which we also expect after unblinding. The difference between ee$\mu\mu$ and $\mu\mu\mu\mu$ final states can be understood in terms of FSR contributions due to electrons.

In Figure 6.54 the same distributions are shown after the MLP selection. We show here the $ee\mu\mu$ invariant mass distribution for those events in which at least one of the four possible 2-track candidates passes the MLP selection. Results for

the five mass intervals, corresponding to five different MLP ranges of application, are presented separately. The discrepancies are similar to the previous case for the same reasons. Figure 6.55 shows comparison of MLP relative efficiencies as a function of the ee$\mu\mu$ invariant mass for different MLP ranges. As expected, we observed better agreement. It is also observable from the plots that data/MC agreement is consistently better in a restricted region around the $\Upsilon(4S)$. This is because there is much less space available for ISR and FSR effects in that region, and the kinematic features of data and MC are much closer. Now with a goal of checking the discrepancies in conditions where data and MC are more comparable, we set the cut: 10.55 GeV $< M(ee\mu\mu) <$ 10.63 GeV. The MLP relative efficiency distributions for data and MC after this cut is shown in Figure 6.56. We finally got a sensible improvement in the low and medium mass ranges, with discrepancies below 10%.

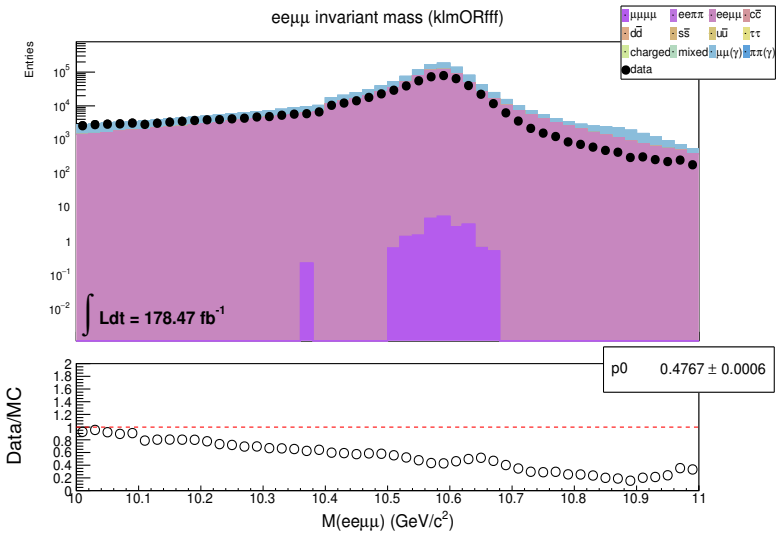


Figure 6.53.: ee$\mu\mu$ invariant mass distribution and data/MC ratio, before MLP.

We also studied data/MC yields and data/MC relative MLP efficiency ratios as a function of the $\mu\mu$ invariant mass. The distribution of the yields before and after the MLP selections are shown in Figure 6.57 and Figure 6.58, respectively.

Figure 6.59 shows the data/MC MLP relative efficiency ratios as a function of the $\mu\mu$ invariant mass, where the ee$\mu\mu$ invariant mass was restricted to be around the $\Upsilon(4S)$ mass, between 10.55 GeV and 10.63 GeV. These last plots demonstrate that the relative effect of the MLP selection on the background is well reproduced in MC, at the level of 10–20%, in conditions in which data and simulation are comparable. Although the background for this search will be measured directly

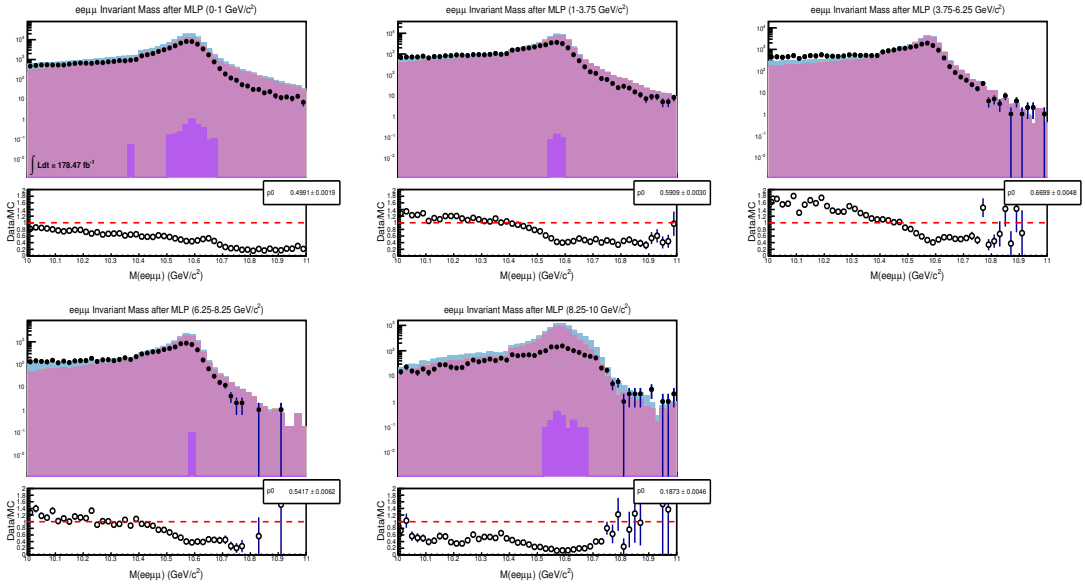


Figure 6.54.: ee$\mu\mu$ invariant mass distribution and data/MC ratio, after MLP application in the five MLP mass ranges.

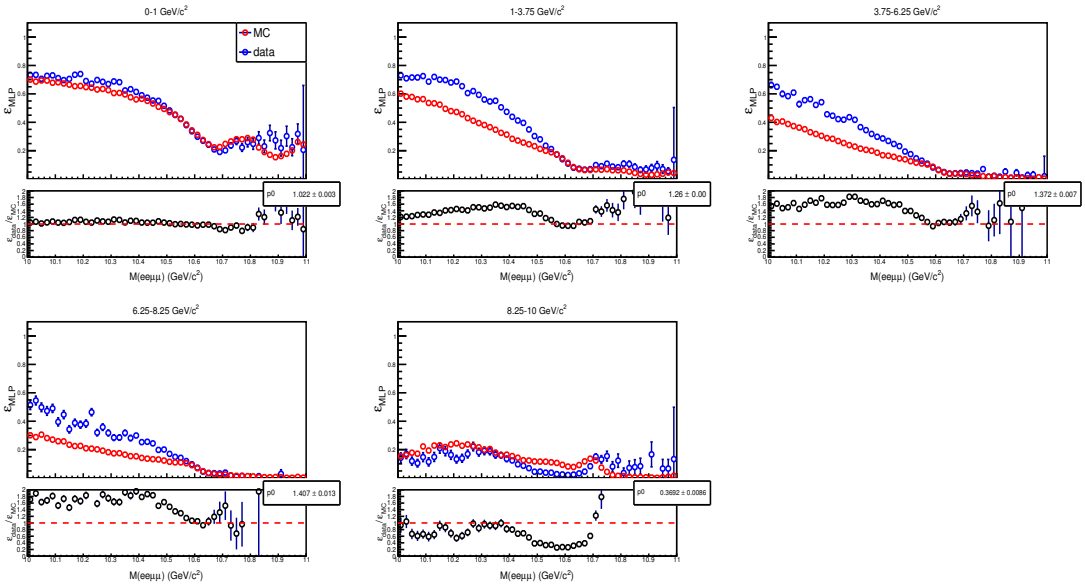


Figure 6.55.: Data and MC MLP relative efficiencies and data/MC MLP relative efficiency ratios as functions of the ee$\mu\mu$ invariant mass in the five MLP mass ranges.

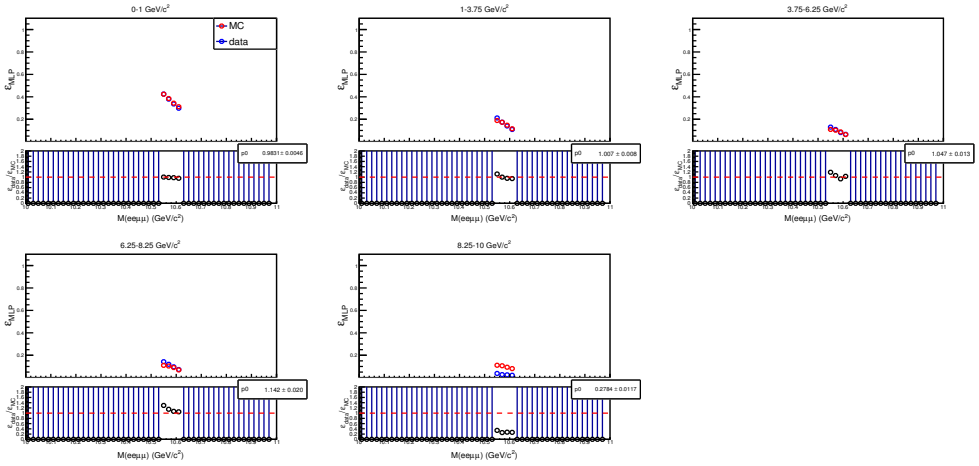


Figure 6.56.: Data and MC MLP relative efficiencies and data/MC MLP relative efficiency ratios as a function of the ee$\mu\mu$ invariant mass in the five MLP mass ranges, when the M(ee$\mu\mu$) is restricted to be around the Υ(4S).

from data using fitting procedure, these checks are highly beneficial for the study of the systematics due to the MLP selection affecting the signal efficiency, as, in that case, the ISR process is taken into account at generator level. Moreover, comparing MLP efficiencies after selection on the MVA output must be considered a check of the agreement in signal-like regions, because network is trained for separating $\mu\mu\mu\mu$ background and signal.

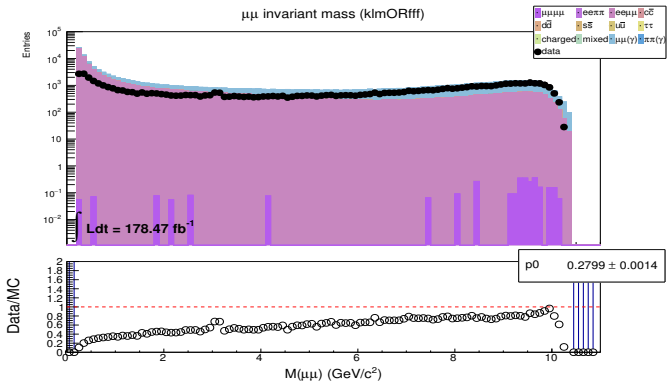


Figure 6.57.: $\mu\mu$ invariant mass distributions, before MLP application.

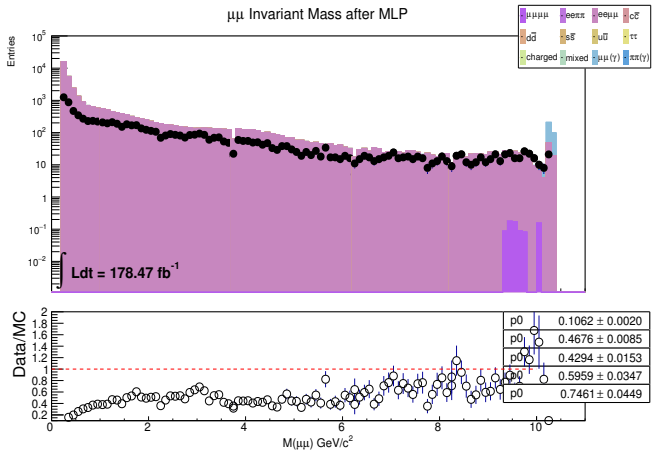


Figure 6.58.: $\mu\mu$ invariant mass distributions, after MLP application.

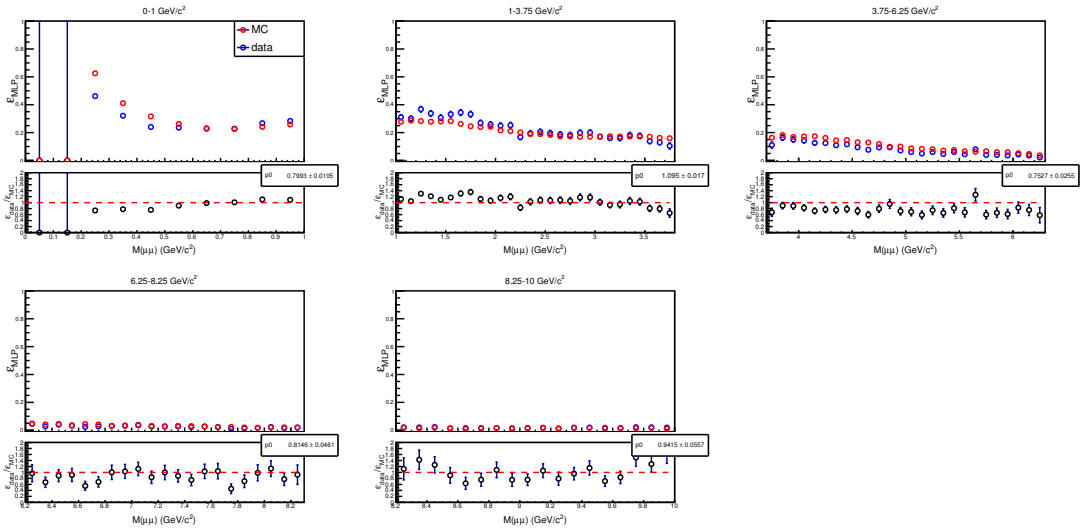


Figure 6.59.: Data and MC MLP relative efficiencies and the relative efficiency ratios in the five MLP mass ranges, as a function of the $\mu\mu$ invariant mass, M($\mu\mu$), for $10.54 < \mathrm{M}(\mathrm{ee}\mu\mu) < 10.62$.

6.13. Systematics

Systematic uncertainties are evaluated in terms of the effect of different procedures and selections on the signal efficiency. Our primary goal is to compute a mass dependent systematics instead of flat or constant systematics for all mass points (see section 6.8.1, where we used flat systematics to evaluate the upper limits). The main sources of systematic uncertainties are tracking efficiency, trigger efficiency, particle identification selections, ISR cuts, MLP selection, fitting procedure for signal extraction, data/MC mass resolution discrepancy, and signal efficiency interpolation. Other sources of systematic uncertainties, such as data and MC discrepancies in momentum resolution and beam energy shift, were found to be negligible due to the 4C kinematical fitting procedure; hence discarded.

6.13.1. Tracking

The tracking efficiency to track a μ track is taken from an existing study within the Belle II collaboration, and the calculated efficiency is 0.13 $\pm$ 0.16 (stat) $\pm$ 0.89 (syst)%. Since exactly four tracks are required in our event topology, the associated systematic uncertainty is propagated in terms of quadrature 4 times, taking into account both statistic and systematic errors (as shown in Equation 6.30).

$$syst = \sqrt{4 \cdot (0.16^2 + 0.89^2)} \tag{6.30}$$

Thus the final systematics is evaluated to be 1.80%.

6.13.2. Trigger

The systematics effect due to the trigger is evaluated independently for CDCKLM and fff/ffy trigger lines. For fff trigger, as mentioned in section 6.4, efficiency was measured using different final states. Figure 6.60 shows the total $klmORfff$ signal efficiency for different final states under consideration as a function of the Z′ mass. We assume the size corresponding to half of this band (variation between maximum and minimum) as a systematic uncertainty due to this source. Figure 6.60 right plot shows the mass dependent systematics variation due to the fff trigger.

Systematic uncertainties related to the CDCKLM trigger lines were evaluated, assuming all the CDCKLM uncertainties on the efficiencies are completely correlated. This was done by moving all the efficiencies in the cells given in Table 6.7 accordingly to a gaussian distribution centered around the klm efficiency value and with a width corresponding to the quadratic sum of the systematic and statistic uncertainties. The procedure was repeated 100 times for each mass hypothesis. The results are shown in Figure 6.61.

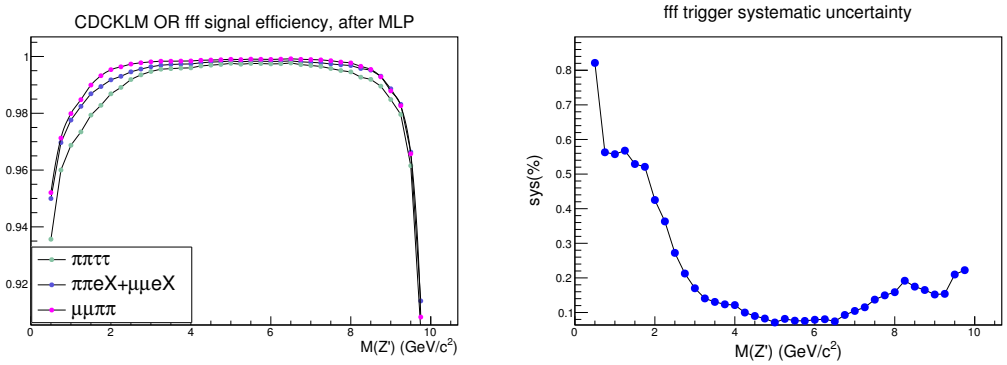


Figure 6.60.: Left: CDCKLM OR fff trigger signal efficiency as a function of the Z′ mass for different configurations of the final states. Systematics is taken as the half width of the maximum and minimum variation. Right: fff trigger systematic uncertainty, corresponding to half of the band shown in the left plot as a function of mass.

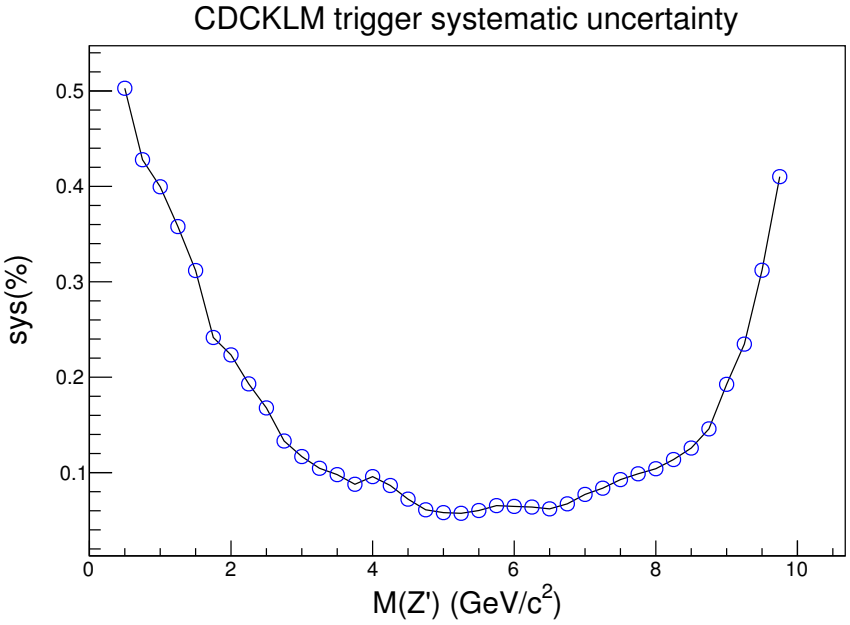


Figure 6.61.: Systematic uncertainties on the CDCKLM OR fff trigger due to uncertainties on the CDCKLM lines as a function of the Z′ mass.

6.13.3. Particle identification

Systematic uncertainties associated with particle identification are evaluated by considering the official Belle II efficiency and fake rate corrections, as provided by the LeptonID and the HadronID working groups. Corrections are charge dependent and given in terms of momentum p and the track's polar angle θ for a specific cut on the PID variable (such as μID>0.9). We select the PID correction and corresponding associated statistic and systematic uncertainties for each track with a given momentum and polar angle (p, θ).

For each track, the procedure is the same as for the trigger systematics evaluation (see the last section), with statistical and systematic uncertainties summed in quadrature. Also, in this case, we evaluated the systematics in the correlated scenario. The evaluated systematics as a function of the Z′ mass is shown in Figure 6.62.

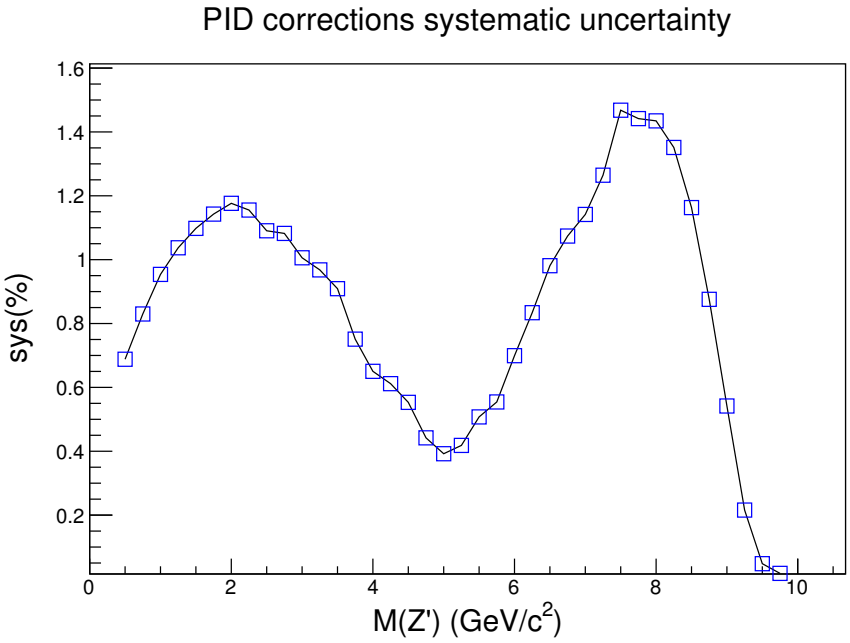


Figure 6.62.: PID systematic uncertainties as a function of the Z′ mass, for μID > 0.9.

6.13.4. ISR cut

Systematic uncertainties due to the ISR selections (see section 6.3.4) can also affect signal efficiency. The signal generator includes the ISR process but not the large-angle hard-radiation component (hard means high energetic photons), which can

produce photons in acceptance. The signal efficiency needs to be corrected for this effect. Events can also be rejected due to the presence of photons from the beam backgrounds. All the samples used in this analysis contain beam backgrounds.

We studied the effect of the lack of the hard ISR component in the signal generator using the $\mu\mu(\gamma)$ process (because the generator used to generate $\mu\mu(\gamma)$ events take into account ISR). Events are simulated without beam background and identical selections are applied as for the $Z' \to \mu^+\mu^-$ search. In these conditions, the ISR selections rejected 2.8% of the events. We then checked the effect of changing the energy cut used in the ISR selections (see section 6.3.4) by ±5%. It turned out that the effect is negligible for samples without beam background, and for samples with beam background, contribution at the simulation level was found to be ±0.26%.

The same ±5% ISR cut variation is also applied to the signal simulation. The result, as a function of the Z′ mass, is shown in Figure 6.63. This confirms, for the beam background only, what is also found for the $\mu\mu(\gamma)$ process.

We also cross-checked these effects on the ee$\mu\mu$ control sample data. Due to the presence of electrons, FSR processes are also expected to contribute. The impact of changing by ±5% the energy cut in the ISR selection is shown in Figure 6.64. Once again, it confirms the MC-based estimation.

We assumed the largest among the effects in Figure 6.63 and 6.64, as a function of the mass, as the systematic uncertainty due to this source.

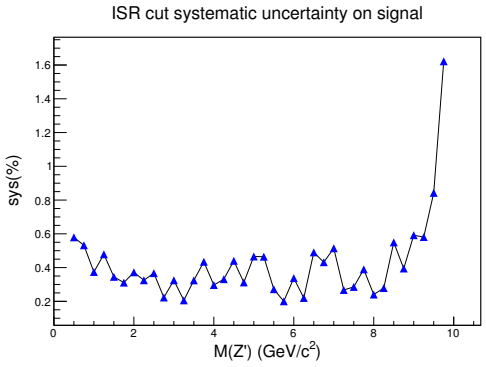


Figure 6.63.: ISR cuts systematic uncertainties on MC signal events as a function of Z′ mass.

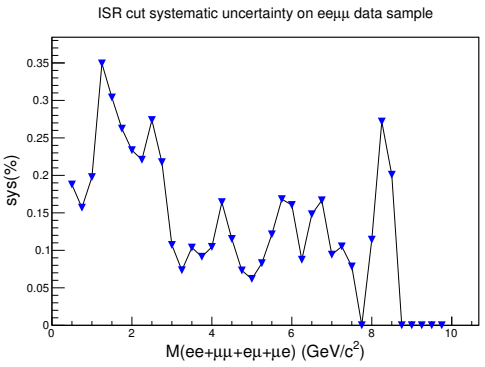


Figure 6.64.: ISR cuts systematic uncertainties on ee$\mu\mu$ data sample.

6.13.5. MLP selection efficiency

Systematics due to the MLP selection is determined from the ee$\mu\mu$ control sample, particularly studies from data and MC comparisons of MLP relative efficiencies

(see section 6.12.2). Although they are related to the background, the assumption is that the uncertainties estimated in those conditions also represent the signal. Cutting the MLP output will automatically select a background signal-like. This is a conservative approach, in a sense, because MLPs are designed to choose a good fraction of signal events and suppress a massive amount of background. We also checked the possibility of using samples with some resonances (J/ψ, $\Upsilon(1,2,3S)$, etc.), but it turned out that, because of the different production mechanism compared to the signal, they are heavily suppressed by the MLP selection.

We use the results found in section 6.12.2 on the relative MLP efficiencies following a tight cut on the ee$\mu\mu$ invariant mass around the $\Upsilon(4S)$. The basic assumption is that, in this situation, data and MC are more directly comparable because ISR and FSR effects (absent in MC) are less relevant. The further assumption is that the uncertainties in these conditions hold in the entire mass interval 10-11 GeV, as anticipated, for the signal generated with ISR too. The results are summarized in Table 6.3, displayed numbers are coming from p_0 polynomial fit. The discrepancies we found in the relative MLP efficiencies in section 6.12.2 and Table 6.3 are of 20%, a value we assume for the whole mass range as the systematic uncertainty due to this source. However, our final systematics, due to the MLP selection efficiency, will be evaluated on the final $\mu\mu\mu\mu$ sample after unblinding. We will measure the data/MC ratio of MLP efficiencies mass by mass, excluding a 5σ region around the mass under study. We expect pretty much better results compared to the ee$\mu\mu$ case illustrated here.

Invariant mass	μID	0-1 (GeV/c^2)	1-3.75 (GeV/c^2)	3.75-6.25 (GeV/c^2)	6.25-8.25 (GeV/c^2)	8.25-10 (GeV/c^2)
ee$\mu\mu$	0.9	0.98	1.0	1.05	1.14	0.27
$\mu\mu$	0.9	0.8	1.09	0.75	0.81	0.94

Table 6.3.: p_0 parameters from data MC MLP efficiencies ratio polynomial fit in the five MLP mass ranges.

6.13.6. Fit bias

In section 6.9, I checked the stability and self-consistency of the fitting procedure with a PDF-based and MC-based toyMC technique. I want to find whether floating the background component induces a bias on the extracted signal yield or not. Systematic uncertainties associated with the fit are evaluated in this section using many mass points using the bootstrap technique and taking the average as a source of systematics. A series of signal + background fits is performed with background statistics equivalent to that expected for the target luminosity of 178.47 fb^{-1} and signal events are injected for each Z' mass according to a Poissonian distribution with the expected value set to the 90%CL excluded yield, evaluated using

a background-only sample without taking into account systematic effects. Fitted signal yields are compared with the correct number of injected events, and a pull distribution is built for each Z' mass. Figure 6.65 top plot shows pull mean values and widths. The bottom plots show relative yield variations and rms of yield variation for the test masses. The results are compatible with those presented in the section 6.9, showing a negligible average yield bias and an rms of $\approx 4\%$. We, therefore, do not apply corrections to the fit results and assign a 4% systematic uncertainty due to the fit stability.

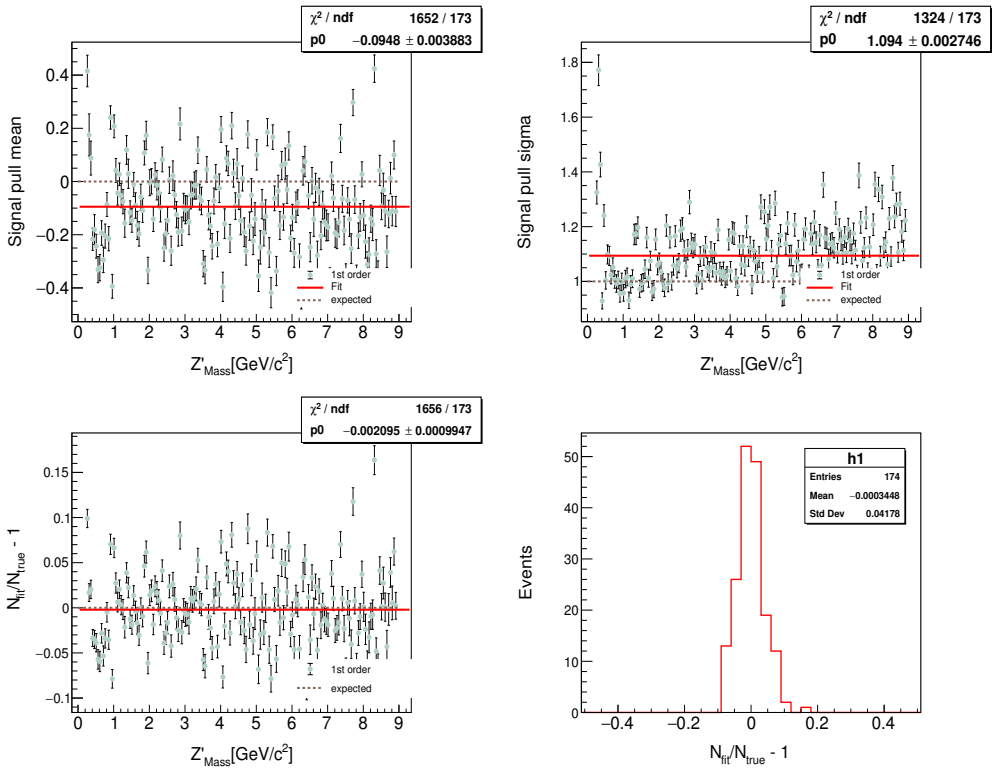


Figure 6.65.: Distribution of the means μ and σ returned by the Gaussian fit of the pulls for different Z' mass hypothesis.

6.13.7. Mass resolution

A data/MC difference of the dimuon mass resolution will affect the fit procedure by changing the signal shapes. I checked the dimuon mass resolution comparing the ee$\mu\mu$ control sample data (see section 6.12) before the MLP application in the region of the J/ψ with the MC predicted signal shapes at the same mass. I

used control sample data before the MLP application because the MVA procedure heavily suppresses the J/ψ due to the production mechanism; see Figure 6.57 and 6.58. I performed a full signal+background fit in an interval centered around the J/ψ using the 2CB signal shape with the parameters fixed as explained in section 6.7.1, and a second fit in which I convoluted the double Crystal ball function with a Gaussian with a coincident mean and a floating width. The last fit represents the dimuon mass resolution data/MC discrepancy at the J/ψ mass. Fit results are shown in fig. 6.66 (left): the Gaussian width is estimated to be 3.08 MeV, with a 25% increase concerning the expected signal mass resolution at that mass. The evaluated estimate is conservative because the J/ψ will be consistently more boosted than a Z' signal at the same mass, with a worse momentum resolution for the two muons. The 25% worsening of the mass resolution at the J/ψ is propagated to all the masses, assuming the same fractional effect. The signal is injected and smeared for eleven sample mass points with a Gaussian extra width corresponding to 25% of the expected resolution. The standard signal+background fit procedure (with a 2CB shape for the signal) is then performed, and the difference in signal yield with respect to the result obtained without the Gaussian smearing is assumed as the systematics due to the mass resolution. Results are shown in fig. 6.66 (right). The average effect is estimated to be -7%. This number is considered as a systematic uncertainty due to this source.

6.13.8. Luminosity

We consider a 1% systematic uncertainty on the integrated luminosity, according to the Belle II official recommendations.

6.13.9. Signal efficiency interpolation

We consider a systematic uncertainty due to the interpolation of signal efficiency (see section 6.5.2) in mass points for which we don't have generated samples. We use the signal samples produced in 5 MeV steps for the training of MLP (see section 6.5.3). We make two interpolations based on even and odd mass points only, and they are shown in Fig. 6.67. The relative difference as a function of the Z' mass and its one-dimensional distribution is shown in Figure 6.68. There are no evident mass-dependent effects. We, therefore, took a 3% flat systematic uncertainty due to this source.

6.13.10. Summary of systematic uncertainties

Table 6.4 summarizes all the systematic uncertainties evaluated in the previous sections. We summed them in quadrature and presented the result in Figure 6.69

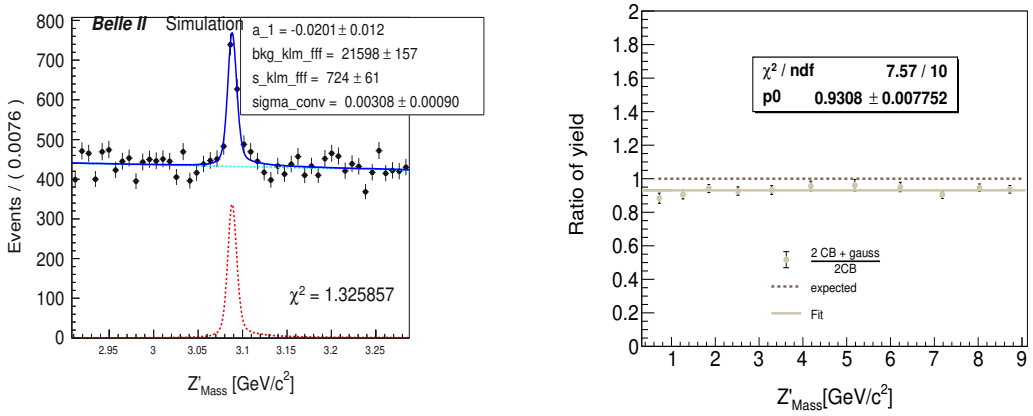


Figure 6.66.: The left plot shows the fit results obtained with a double CB function convoluted with a Gaussian in the ee$\mu\mu$ control sample data in the J/ψ region. The estimated width of the Gaussian is 3.08 MeV, corresponding to a 25% increase in the mass resolution. The right plot shows the propagated effect as a function of the mass, considering the 25% increase in mass resolution for each mass point. The average impact is -7%.

as a function of the Z′ mass. The total uncertainty is mostly dominated by the contribution from the MLP selection and ranges from 21.9% to 22.2%. However, the final MLP systematics will be estimated on the final four-muon sample and is expected to decrease substantially concerning the present evaluation.

6.14. Unblinding of 2019 data

We didn't include 2019 data in our targeted data sample due to inhomogeneous trigger conditions since the CDCKLM lines were unavailable. We planned to study these data, using only fff trigger, with two main goals:

- To verify the systematic uncertainties due to the MLP selection efficiencies are effectively reduced from those evaluated with the ee$\mu\mu$ control sample or not, as mentioned in section 6.13.5.
- Verify the data/MC ratio after the MLP selection, which was supposed to be around 0.65 factor less for the SM $\mu\mu\mu\mu$ process in the MC from BaBar and Belle experience (see section 6.8.1, where I estimated sensitivity based on that factor).

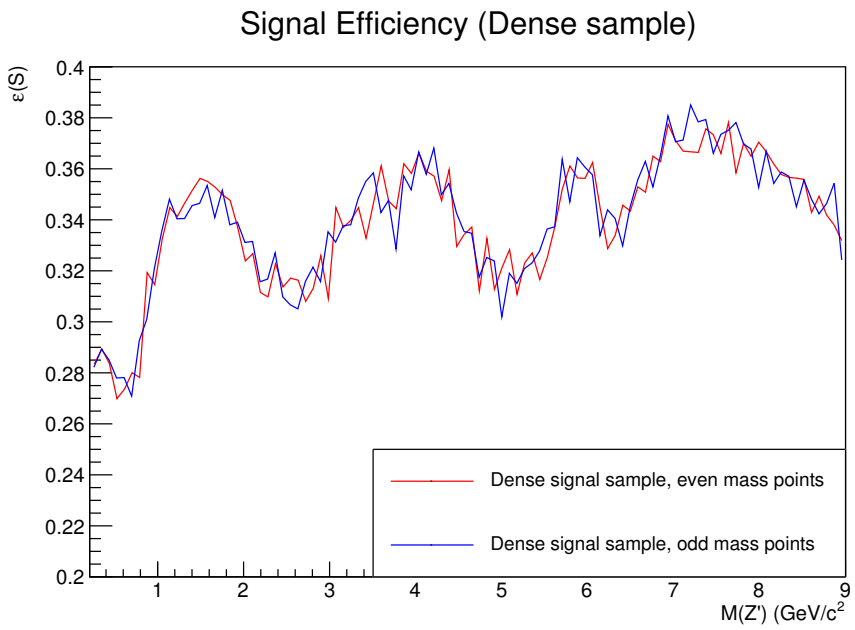


Figure 6.67.: Signal efficiency interpolations based on the even and the odd signal mass points.

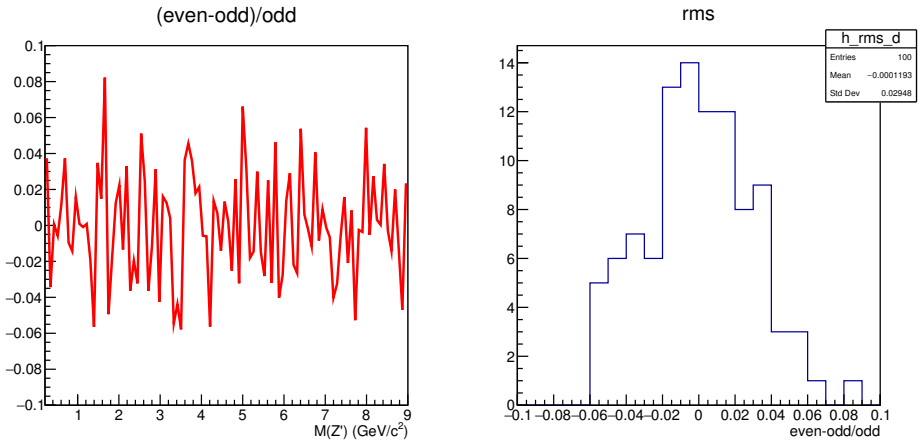


Figure 6.68.: On the left, the difference between the even and odd mass point based signal efficiencies interpolations, and the right plot gives the rms of this deviation about 3%.

Table 6.4.: Systematic uncertainties.

Source	syst(%)
Tracking	1.8
fff trigger	0.1-0.8
CDCKLM trigger	0.1-0.5
Particle ID	0.7-3
ISR cut	0.5-2
MLP selection	20
Fit bias	4
Mass resolution	7
Interpolation	3
Luminosity	1
Total	21.9-22.1

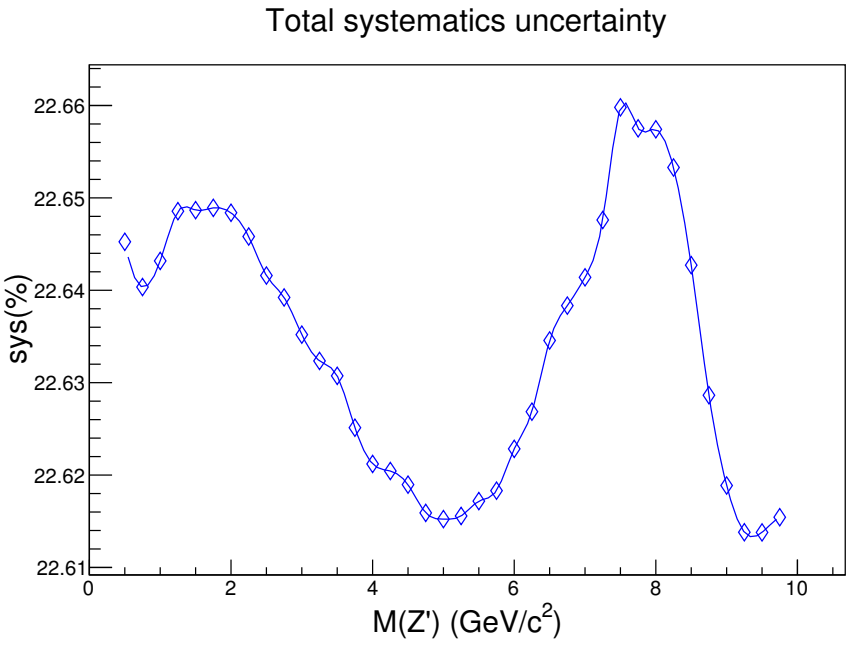


Figure 6.69.: Total systematic uncertainties obtained as the quadratic sum of all the individual source contributions.

After the working group approval, we looked at the 8.8 fb^{-1} of 2019 data. There is, of course, no risk of unblinding because the BaBar and Belle already searched the same phase space with much higher datasets.

Figure 6.70 shows the data and MC MLP relative efficiencies and the relative efficiency ratios when the total invariant mass is restricted to be around the Υ(4S) in the range 10.54–10.62 GeV/c^2. Table 6.5 reports the p_0 parameters, fitted on the data/MC relative efficiency ratios, in the five MLP ranges. These results confirm that the systematic uncertainties due to the MLP selection on the $\mu\mu\mu\mu$ sample are lower than those evaluated on the ee$\mu\mu$ control sample (see section 6.13.5 and particularly Table 6.3 for comparison). However, final uncertainties will be evaluated on the target 178.47 fb^{-1} sample, with much better statistical precision.

Figure 6.71 show the $\mu\mu$ invariant mass distribution and bin-by-bin data/MC ratio after the MLP selection, with the total invariant mass restricted between $10< M(4-track) <11$ GeV/c^2 around the Υ(4S) (analysis cut). The data/MC ratio turns out to be of order 1. This is because the MLP selection effect in these conditions is no longer the same for data and MC due to the lack of the ISR effects in the AAFH generator. However, our MLP procedure is able to compensate for the ISR effect. It is not a problem for the signal which is generated with ISR effects and thus the systematic uncertainty estimate on the signal efficiency will not be changed. The only consequence is that, in the sensitivity evaluation, scaling the $\mu\mu\mu\mu$ background by a factor of 0.65 is not a correct option. In the next section 6.15 I didn't consider this scaling factor and used the nominal background value as it comes from MC, and gave the final sensitivity projection.

Table 6.5.: p_0 parameters from data/MC MLP efficiencies ratio polynomial fit in the five MLP mass ranges, for 2019 $\mu\mu\mu\mu$ sample.

MLP range	0-1 (GeV/c^2)	1-3.75 (GeV/c^2)	3.75-6.25 (GeV/c^2)	6.25-8.25 (GeV/c^2)	8.25-10 (GeV/c^2)
p_0	1.01±0.07	0.94±0.08	1.1±0.2	1.2 ± 0.3	1.1 ± 0.2

6.15. Results

I summarize here the final results I obtained with the signal region still blinded. Figure 6.72 shows the expected background after all the selections. Unlike in section 6.8, we didn't scale the $\mu\mu\mu\mu$ background here. However, after fully unblinding, the signal region in the final sample will be measured from the background directly from the fitting procedure and does not need any additional hypothesis. Figure 6.24 shows the signal efficiency after all the selections. The expected upper limit on cross-section and signal yield assuming the mass-dependent final systematic uncertainties (see Figure 6.69) is shown in Figure 6.73. The expected upper

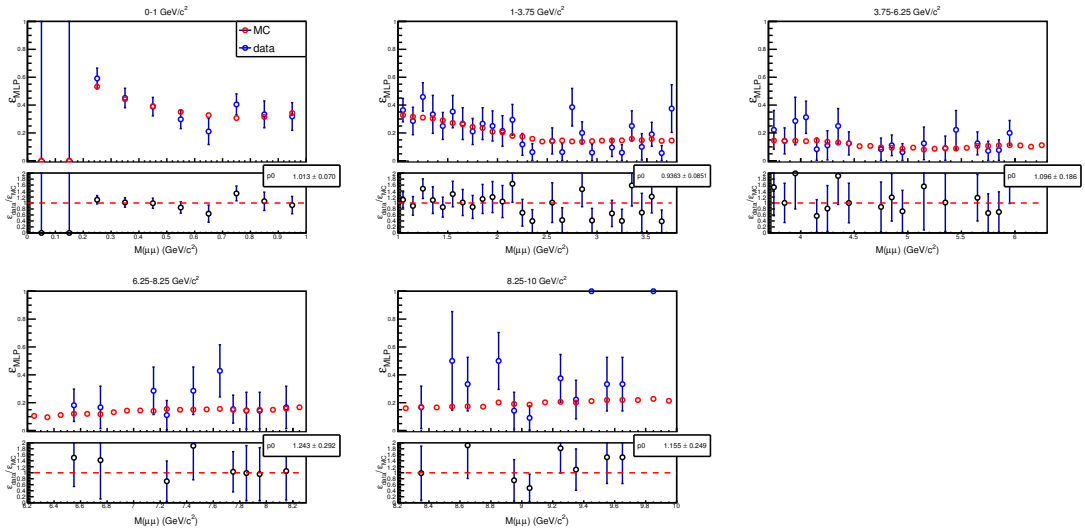


Figure 6.70.: Data and MC MLP relative efficiencies and data/MC relative efficiency ratios as a function of the $\mu\mu$ invariant mass in the five MLP mass ranges, when the 4-track mass is restricted to be around the $\Upsilon(4S)$.

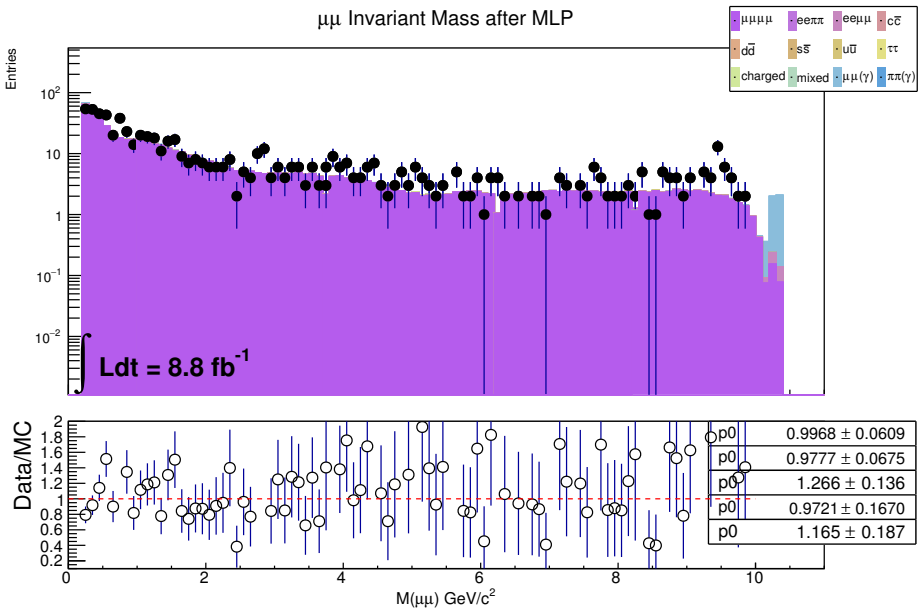


Figure 6.71.: $\mu\mu$ invariant mass distribution and data/MC ratio, after MLP application in the five MLP mass ranges, in the full analysis mass range, 10 GeV/c^2 < M(4-tracks) < 11 GeV/c^2.

limit on the coupling constant g' is also given in Figure 6.74.

Figure 6.75 shows the comparison with BaBar, our expectations at the 178.47 fb^{-1} target luminosity are worse below ≈1 GeV/c^2, comparable, but on average better, in the range 1–6.5 GeV/c^2, and worse above ≈7.5 GeV/c^2.

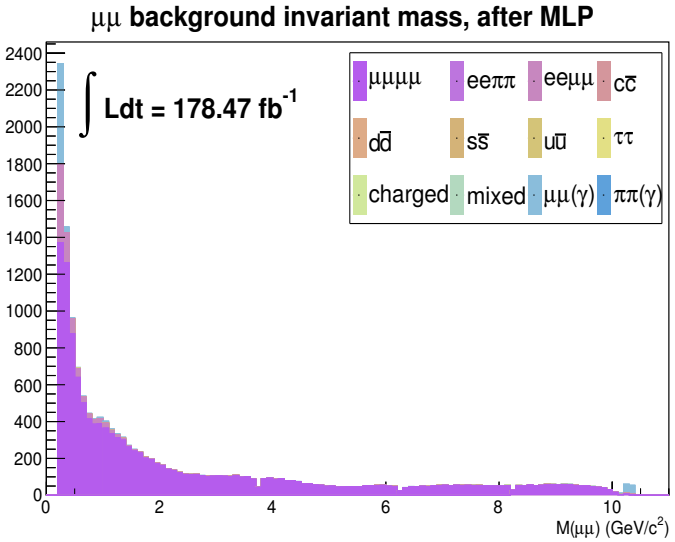


Figure 6.72.: Background after all the selections and dropping the scale factor as a function of the candidate dimuon mass.

6.16. Conclusions and future outlook

We performed a very preliminary search for the process $e^+e^- \to \mu^+\mu^- Z'$, $Z' \to \mu^+\mu^-$ in the $\mu^+\mu^-\mu^+\mu^-$ final state at the Belle II experiment. Our targeted luminosity is 178.47 fb^{-1}, which is almost 2-3 times less data than BaBar and Belle. With this targeted data sample, we expect a similar performance of BaBar except for the low mass region due to the aggressive background suppression. I summarized all the results in section 6.15 and also showed a comparison with BaBar. However, our targeted data sample is less at the moment, but we will definitely obtain a better result compared to the existing ones. We submitted all our findings to the collaboration, which is currently under review, and hope to publish our results soon (BELLE2-NOTE-PH-2022-022 [98]).

Belle II has already gathered a data sample of ∼400 fb^{-1} before its first prolonged shutdown and will resume data taking soon. We are also planning to update this result with higher luminosity collected up to now and aim for another publication. In that measurement, we will definitely get benefit from the higher luminosity, but

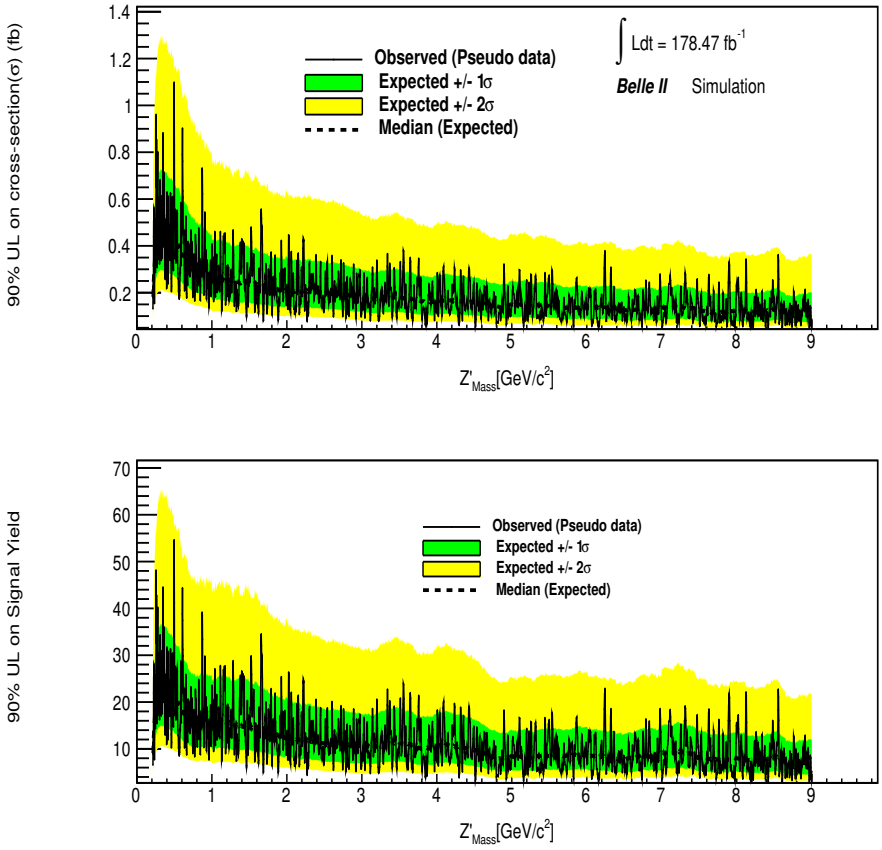


Figure 6.73.: Estimated 90% CL upper limit on the cross section (top) and on the signal yield (bottom) for the process $e^-e^- \to \mu\mu Z'(Z' \to \mu\mu)$ for 178.47 fb^{-1}.

one could also explore the newly developed tool within Belle II collaboration for training the neural network by maximizing the Punzi figure of merit [99] and could get potential improvements in analysis search strategy further.

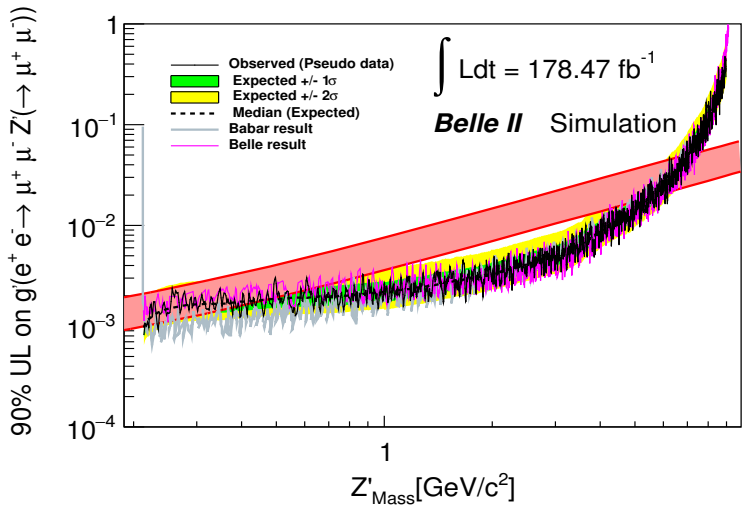


Figure 6.74.: Expected 90% CL sensitivity on the coupling constant g' for $e^-e^- \to \mu\mu Z'(Z' \to \mu\mu)$ for 178.47 fb^{-1}. Also shown is the band that would explain the observed $(g-2)_\mu$.

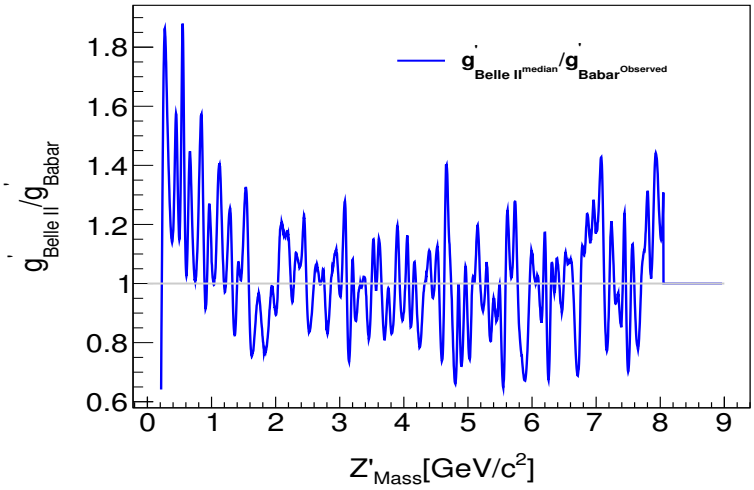


Figure 6.75.: Comparison with BaBar for 178.47 fb^{-1} in terms of the g' sensitivity. Values below the gray line are better in Belle II than in BaBar

7. Search for displaced visible decay of Z' in muonic final state

The following work was done solely by the author, while additional guidance was provided by Gianluca Inguglia.

7.1. Introduction

In this chapter, I tried to do a very preliminary $Z' \to \mu^+\mu^-$ search in the same event topology as described in chapter 6 but in a displaced vertex scenario. A displaced vertex refers to a Z' having a finite lifetime. I tried to do this search as much as possible in a model-independent way. Section 1.2.3 gives the brief theoretical motivation behind the study. The following sections provide a very preliminary Monte Carlo study of $e^+e^- \to \mu^+\mu^- Z'(\to \mu^+\mu^-)$, where Z' has some finite lifetime, at the Belle II experiment.

7.2. Software and data samples

Signal simulation and Background Samples

`EvtGen` [100] was an obvious choice for generating signals in a model-independent way. I tried and somehow generated particles were assigned the wrong angular distribution for different spin configurations. I discussed with an expert, but no solution came out, so I planned to switch back to the madgraph [86]. Madgraph don't have any displaced vertex model which can generate signal with the displaced vertex signatures. Therefore, I generated the prompt $e^+e^- \to \mu^+\mu^- Z'(\to \mu^+\mu^-)$ signal with narrow width approximation (see section 6.2) and then used the Vertex Displacer module available in basf2 to displace the prompt vertexes.

20000 signal events for each Z' mass in an interval of 250 MeV/c^2 from 0.25 GeV/c^2 to 10 GeV/c^2, with a fixed width of 10^{-6} GeV/c^2 (well below the detector resolution) and a coupling constant $g' = 1$, having prompt vertexes (no lifetime)

are generated using **MadGraph@5-NLO** with the **Lmu_minus_Ltau_UFO** model. As I used **Lmu_minus_Ltau_UFO** model, the generated Z' follows a spin 1 statistics, and the Initial State Radiation (ISR) effects are also taken into account during generation. Now the generated prompt signals have given a finite lifetime ($c\tau_{Z'}$) of 0.1, 1, and 10 cm. GEANT4 [101] is used to reproduce interactions of particles traversing the Belle II detector, considering the varying detector conditions and beam backgrounds.

For background, I used the official run independent samples (called MC14a). The contributing background components mimicking our signal topology are mostly arising from the SM QED processes, $e^+e^- \to \mu^+\mu^-\mu^+\mu^-$, $e^+e^- \to \mu^+\mu^-(\gamma)$, $e^+e^- \to e^+e^-\mu^+\mu^-$ and other processes summarizes in Table 7.1. The $e^+e^- \to \mu^+\mu^-\mu^+\mu^-$ processes are easily reducible due to the prompt nature, for $e^+e^- \to \mu^+\mu^-(\gamma)$, γ can travel and convert to e or μ pair and e can misidentified as μ, similarly for $e^+e^- \to e^+e^-\mu^+\mu^-$ process e can misidentified as μ but easily reducible due to the prompt signature.

All the detector simulated signal and background samples are reconstructed back. I used basf2-**release-05-02-00** for generation, simulation and reconstruction purpose. For the analysis I used **root v6-21** with **TMVA**, **RooFit** [93] and **RooStats** packages.

Data samples

For the thesis, I planned to stick with the MC study only.

7.3. Analysis Strategy

I want to search for the process $e^+e^- \to \mu^+\mu^- Z', Z' \to \mu^+\mu^-$, where Z' has $c\tau_{Z'}$ of 0.1, 1, 10 cm. The signal search strategy is same as the prompt case, the presence of a peak in the dimuon mass distribution in events with four muons having an invariant mass compatible with $\Upsilon(4S)$ CM energy and nothing else in the final state (see chapter 6). The displaced signal topology consists of two prompt muons from the interaction point, and the other two tracks are displaced. The effect of ISR also spoils this picture and moves the four-muon invariant mass away from the collision energy. I select the events with exactly four tracks, each identified as muons, and the total four-track invariant mass, M(4-tracks), to be restricted between 10 and 11 GeV. Further track level selections are given in more detail in section 7.4.

As our final state is same as the prompt search, I expect the same background sources. The main expected background component, the SM $\mu\mu\mu\mu$, is easily reducible due to its prompt nature but intends to contribute to the low lifetime Z's.

Process	$\int Ldt$ [fb^{-1}]	MC generator
$e^+e^- \to \mu^+\mu^-\mu^+\mu^-$	2000	AAFH [88]
$e^+e^- \to \mu^+\mu^-(\gamma)$	900	KKMC [89]
$e^+e^- \to \tau^+\tau^-$	200	KKMC [89] + TAUOLA [90]
$e^+e^- \to e^+e^-\mu^+\mu^-$	200	AAFH [88]
$e^+e^- \to u\bar{u}$	200	KKMC [91]+PYTHIA+EvtGen
$e^+e^- \to d\bar{d}$	200	KKMC [91]+PYTHIA+EvtGen
$e^+e^- \to c\bar{c}$	200	KKMC [91]+PYTHIA+EvtGen
$e^+e^- \to s\bar{s}$	200	KKMC [91]+PYTHIA+EvtGen
$e^+e^- \to B^0\bar{B}^0$	200	EvtGen
$e^+e^- \to B^+B^-$	200	EvtGen
$e^+e^- \to \pi^+\pi^-(\gamma)$	1000	PHOKARA [91]

Table 7.1.: Early Phase 3 MC14a samples used for background studies with the equivalent integrated luminosity $\int$Ldt. All the details about the generators are given in the internal note [92]

Other contributing backgrounds are mainly $\mu^+\mu^-(\gamma)$ and $ee\mu\mu$ due to the photons pair production and muon mis-identification could also be easily reducible. I planned to stick with preliminary cut based background selections, given in more detail in section 7.4.3.

Signal events are generated with ISR, but all SM four-lepton processes are generated without ISR. To minimize the effect of ISR, I use the ISR selections given in section 6.3.4.

Belle II trigger system is not fully developed for triggering the displaced signatures like us, some ongoing work is there, but they are not available for analysis use. I planned to trigger our signal topology by using the ffo trigger (two tracks having an opening angle of 90^o) on the prompt muon tracks. I did a very preliminary TSIM study to estimate our efficiency (see section 7.5). The ffo trigger bit is reliable in TSIM for the used basf2 release and background MC samples, but a careful trigger efficiency estimation is required on data. The work on the $\mu\mu\gamma$ control channel is under study.

Further, a kinematical fit procedure is also applied to events that pass all the selections to improve the signal resolution. The signal yield extraction is performed by fitting the dimuon mass distribution. I rely on the MC prediction for the signal shape as the signal is generated using ISR (expected to model perfectly). For the background, I don't want to rely on the absolute prediction of the MC due to lack of ISR and keep the parameters free (see section 7.6 and 7.7). I exclude the dimuon mass interval corresponding to the J/ψ mass, and I limit our search up

to 9 GeV/c^2 as we don't have a good sensitivity above from the prompt analysis experience.

I planned to use $\mu\mu\gamma$ as our control channel to validate all our procedures, which are currently under study.
Official Belle II recommended PID corrections called "Moriond 2022" PID corrections have been applied to the MC samples.

7.4. Event selection

7.4.1. Signal and background definitions

I select events with exactly two tracks originating from the interaction region having impact parameters $|dr| < 0.5$ cm and $|dz| < 2$ cm in the longitudinal and transverse plane, and the other two tracks with dr > 0.001 cm (for all cτ) are called the "V0 tracks." The sum of the four charges is required to be zero.

Signal and background definitions are exactly same as in section 6.3.1. For each event, there are four possible pairs of oppositely charged tracks; like the prompt case, I also didn't attempt to select one of these pairs more likely coming from a Z′ decay. Although I have two tracks with dr > 0.001 cm in the event, still, I will not be able to understand which tracks are satisfying this condition for the data, and tracks having impact parameter cuts can also fulfill dr > 0.001 cm. A 4-track candidate has a pair of tracks named "candidate tracks" that, in case of signal, are the decay products of the Z′ and are the "displaced tracks" (satisfying the condition of V0 tracks and have a mean lifetime cτ ($\in$ 0.1, 1, 10 cm)). The other two tracks are the "prompt tracks." The signal is defined as the only 2-track candidate (out of four) with both candidate tracks coming from the Z′ decay: this condition is checked by using the MCtruthmatch module available in basf2. Candidate mass is the invariant mass of the candidate track pair: it coincides with the Z′ mass, within the experimental resolution, in case of a proper signal.

In the case of background, all the tracks are prompt, and none of them comes from a Z′ decay.

Unlike the prompt case, all the tracks are identified as muons (for the prompt case, three tracks are identified as muons see section 6.3.2) with a loose μID > 0.5 (this is a global μID depending on different detector likelihoods). Identification of two or three tracks as muons are under study.

Further conditions imposed on V0 tracks are constrained to come from a common vertex (Z′) using a TreeFitter module available in basf2. I restrict the total four-track invariant mass, M(4-track), to be limited between 10 and 11 GeV.

7.4.2. Characterization of background events

The expected contributing backgrounds are similar to the prompt case. The $e^+e^- \to \mu^+\mu^-\mu^+\mu^-$ background process mainly proceeds through ISR and double photon conversion (detailed information in section 6.3.3), which could be easily reducible due to the prompt nature, using the V0 vertex information and di-muon momentum distribution, which expect to differ for signal and background.

The $e^+e^- \to \mu^+\mu^-\gamma$ process expects to contribute to the low dimuon mass region by misidentifying electrons from ISR photon conversion into muons which pass the selections. The same mis-identification can also happen in the $e^+e^- \to e^+e^-\mu^+\mu^-$ process. But the transitions are prompt and easily suppressible from the V0 vertex information.

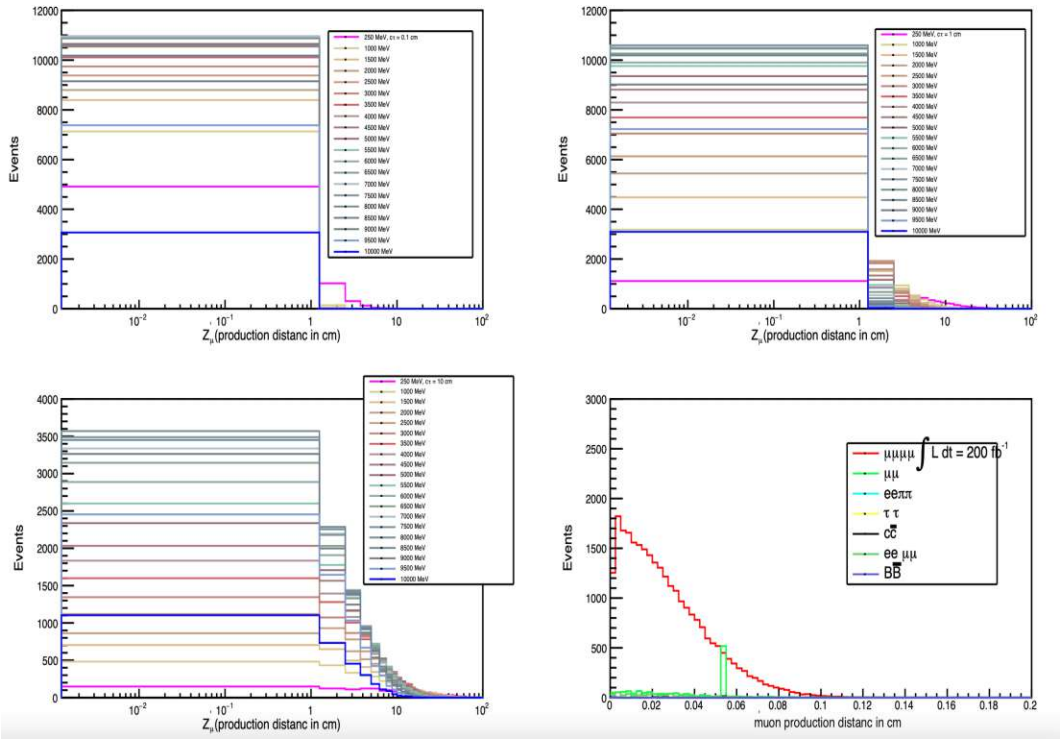
The other backgrounds $e^+e^- \to \tau^+\tau^-$, $e^+e^- \to q\bar{q}$ $(q = u, d, c, s)$, and $e^+e^- \to b\bar{b}$ are found to be negligible. For the $e^+e^- \to \pi^+\pi^- J/\psi(\to \mu^+\mu^-)$ process, the J/ψ mass region is vetoed from the search region, and the $e^+e^- \to \pi^+\pi^-\Upsilon(\to \mu^+\mu^-)$ transitions are outside our search region.

7.4.3. Background suppression using V0 vertex information

Before investigating the necessary background suppressing V0 vertex variables, I looked at the 3D production distance of the V0 tracks (in the case of signal they are μ's originate from Z′) to check how well V0's are reconstructed, which should be distributed exponentially ($N = N_0 e^{-\frac{t}{c\tau}}$) according to the given mean $c\tau$. Figure 7.1 (above four plots) shows the distribution of 3D production distance of μ's (V0 tracks) for $c\tau = 0.1$, 1, 10 cm in case of signal and backgrounds. They are distributed accordingly to the given mean lifetime. Still, as one can see, different mass points having the same generated $c\tau$ travel different 3D distances, and low mass points travel the maximum length. It starts to decrease and gets minimum at specific mass values, then starts to increase again. So, the equation $N = N_0 e^{-\frac{t}{c\tau}}$ with the same $c\tau$ but different N_0 (different efficiency) should give a parallel set of distributions, but this is not the case. It could be due to the resolution effect. Further resolution-related studies are shown in more detail in the section 7.6. Figure 7.1 also shows the comparison of 3D distance for same mass point with different $c\tau$.

V0 vertex distance in the horizontal plane

The V0 vertex distance in the transverse plane (dr) is the most promising variable for rejecting backgrounds over signal. Backgrounds are prompt, and signals are displaced depending on the different $c\tau$. Figure 7.2 (above four plots) shows the distribution of V0 vertex distance dr for signal and backgrounds. The applied



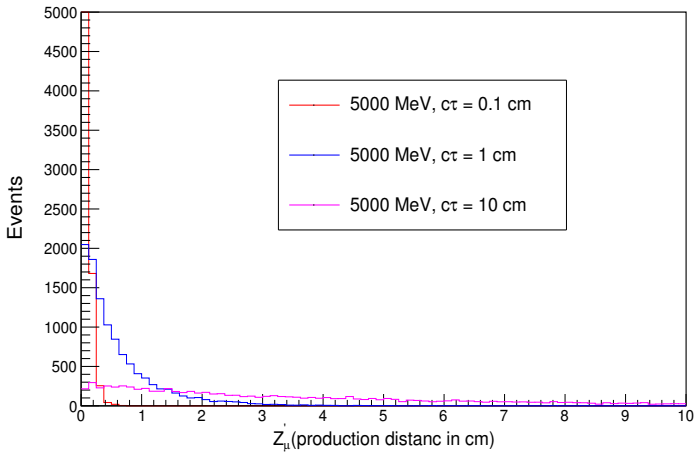


Figure 7.1.: The above four plots show the 3D production distance of muon for $c\tau$ = 0.1, 1.0, 10 cm and of backgrounds, respectively. The below plot shows the 3D production distance of muon for 5 GeV mass with different given $c\tau$'s.

cut on the V0 vertex distance is optimized using the exact Punzi Figure of Merit (FoM) given in equation 7.1.

$$FoM = \frac{\epsilon_s}{\frac{a^2}{8} + \frac{9b^2}{13} + a\sqrt{B} + \frac{b}{2}\sqrt{b^2 + 4a\sqrt{B} + 4B}} \tag{7.1}$$

Where a = 5 (for 5σ discovery), b=1.28 (for 90% CL), B = No of Background events, ϵ_s is signal efficiency.

It is possible to optimize each mass for different $c\tau$'s and get the cut values, but this is quite tedious; I will end up with many different cut values and different background distributions for each mass point. So, I optimize our cuts using the dense signal samples (2000 mass points). The idea is our background dimuon invariant mass is continuously distributed from the dimuon threshold to 10 GeV, and signals are discrete, so I generated a lot of signal mass points to make the combined signal mass distribution continuous. Now one can take the combined signal mass points to describe variables and obtain a single cut value for all the mass points. Figure 7.2 (below) shows the optimized cut values obtained from the Punzi optimization for different $c\tau$'s. As one can see, the chosen cut value is slightly lower than the optimized values, and a common cut of 0.02 cm is chosen for all $c\tau$'s. The cut is chosen where the signal efficiency is close to 99%, as I want to keep the signal efficiency as high as possible, which will cost us in the signal purity, but I could increase that by looking at the other discriminating variables.

applied cut (in cm)	signal efficiency			background rejection (%)
	$c\tau$ = 0.1 cm	$c\tau$ = 1 cm	$c\tau$ = 10 cm	
0.002500	0.999928	0.999952	0.999994	18
0.005000	0.999548	0.999639	0.999933	48
0.007500	0.998667	0.999095	0.999841	60
0.010000	0.997280	0.998282	0.999689	67
0.012500	0.995561	0.997175	0.999469	70
0.015000	0.993413	0.995970	0.999151	73
0.017500	0.990581	0.994244	0.998789	74
0.020000	0.987300	0.992221	0.998427	76
0.022500	0.983214	0.990041	0.997969	77
0.025000	0.978389	0.987408	0.997445	78
0.027500	0.972704	0.984357	0.996863	79
0.030000	0.965976	0.981065	0.996224	79.5
0.032500	0.957955	0.977417	0.995452	80.2
0.035000	0.948235	0.973098	0.994702	80.5
0.037500	0.936586	0.968017	0.993695	81
0.040000	0.922305	0.962311	0.992649	81.4
0.042500	0.904154	0.955791	0.991524	81.7

Table 7.2.: Variation of signal efficiency and back rejection depending on different applied cut thresholds on V0 vertex distance in the horizontal plane. The cut value of 0.02 cm is chosen, keeping the signal efficiency high and acceptable background rejection.

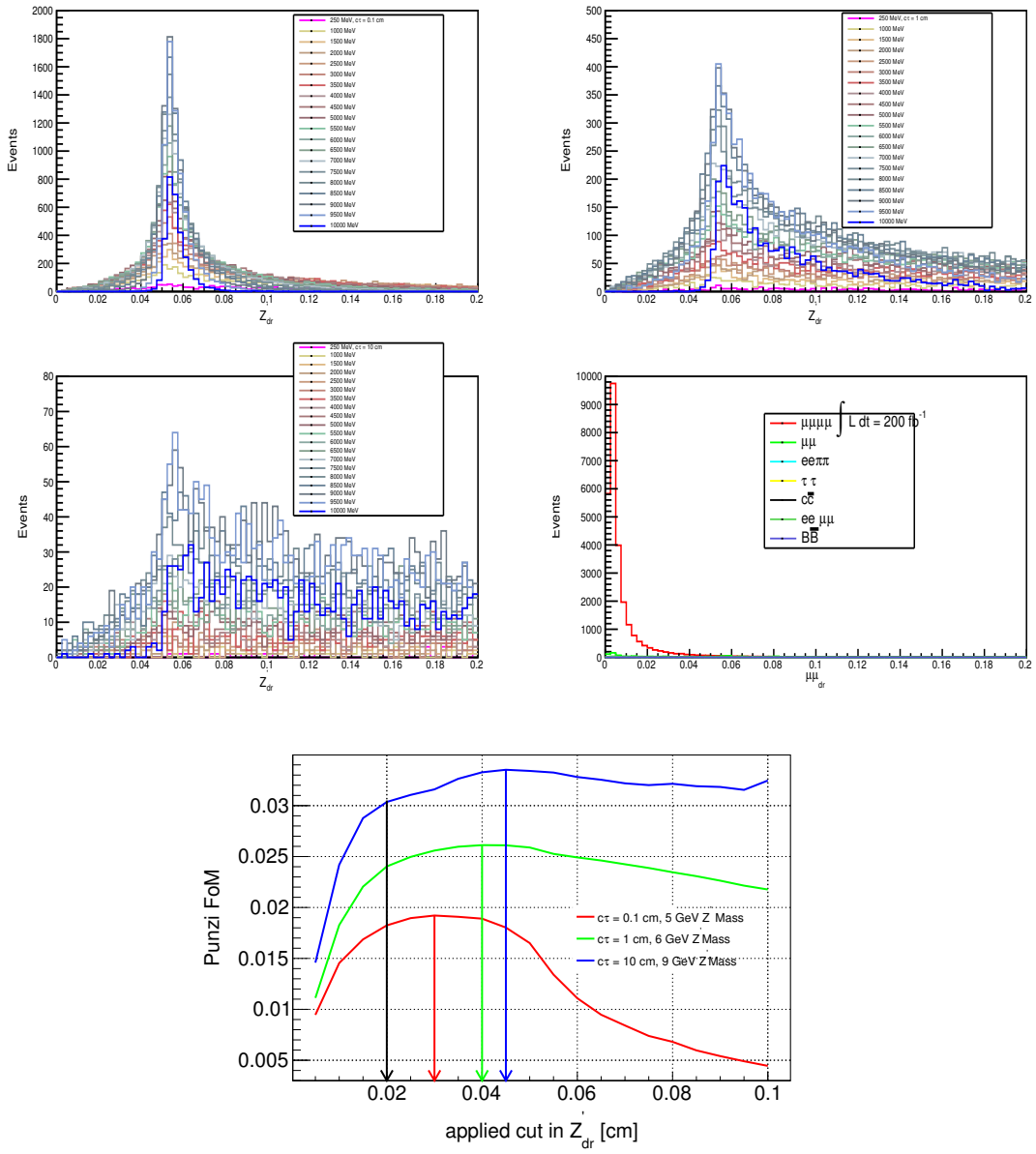


Figure 7.2.: The above four plots show the V0 vertex distance in the horizontal plane (dr) for $c\tau$ = 0.1, 1.0, 10 cm and for backgrounds, respectively. The below plot shows the Punzi FoM depending on different cut thresholds and the black arrow line shows the chosen cut value for all $c\tau$'s.

V0 transverse momentum in CM frame

The momentum projection of the dimuon system in the transverse plane (pt) is a crucial variable for separating events involving photons, mainly the $\mu\mu\gamma$ transition. If one looks at the signal topology in the CM frame, two muons will be back to back, and one muon will radiate a Z′ that decays to the muon pair. So, the Z′ with higher masses are more likely to the beam pipe direction; hence the muon pairs, and have very low pt. Similarly to the $\mu\mu\gamma$ events, two muons will be back to back, and the emitted ISR photon most likely goes to the beam pipe direction and gives a very low momentum projection in the transverse plane.

Figure 7.3 (above four) shows the dimuon transverse momentum distribution for signal and backgrounds, and the below plot shows the optimized cut obtained from the Punzi FoM. As one can see, the given optimized cut from Punzi FoM for three cases is 0.3 GeV, but 0.3 GeV is too tight, so the optimal cut of 0.1 GeV is chosen based on the maximum signal efficiency. A summary of different cut efficiencies is given in the table 7.3.

applied cut on pt	signal efficiency			background rejection (%)
	$c\tau = 0.1$ cm	$c\tau = 1$ cm	$c\tau = 10$ cm	
0.050000	0.999003	0.998954	0.998810	44
0.075000	0.997609	0.997568	0.997301	52.5
0.100000	0.995868	0.995620	0.995127	56
0.125000	0.993595	0.993180	0.992373	58
0.150000	0.990716	0.990276	0.989193	60
0.175000	0.987379	0.986889	0.985364	61
0.200000	0.983443	0.983194	0.980968	63
0.225000	0.979117	0.979258	0.976337	64
0.250000	0.974335	0.974530	0.970967	66
0.275000	0.969283	0.969762	0.965383	67
0.300000	0.963746	0.964542	0.959211	68
0.325000	0.957861	0.958449	0.952616	69

Table 7.3.: Variation of signal efficiency and back rejection depending on different applied cut thresholds on V0 transverse momentum ($\mu\mu_{pt}$). The cut value of 0.1 GeV is chosen, keeping the signal efficiency high and acceptable background rejection.

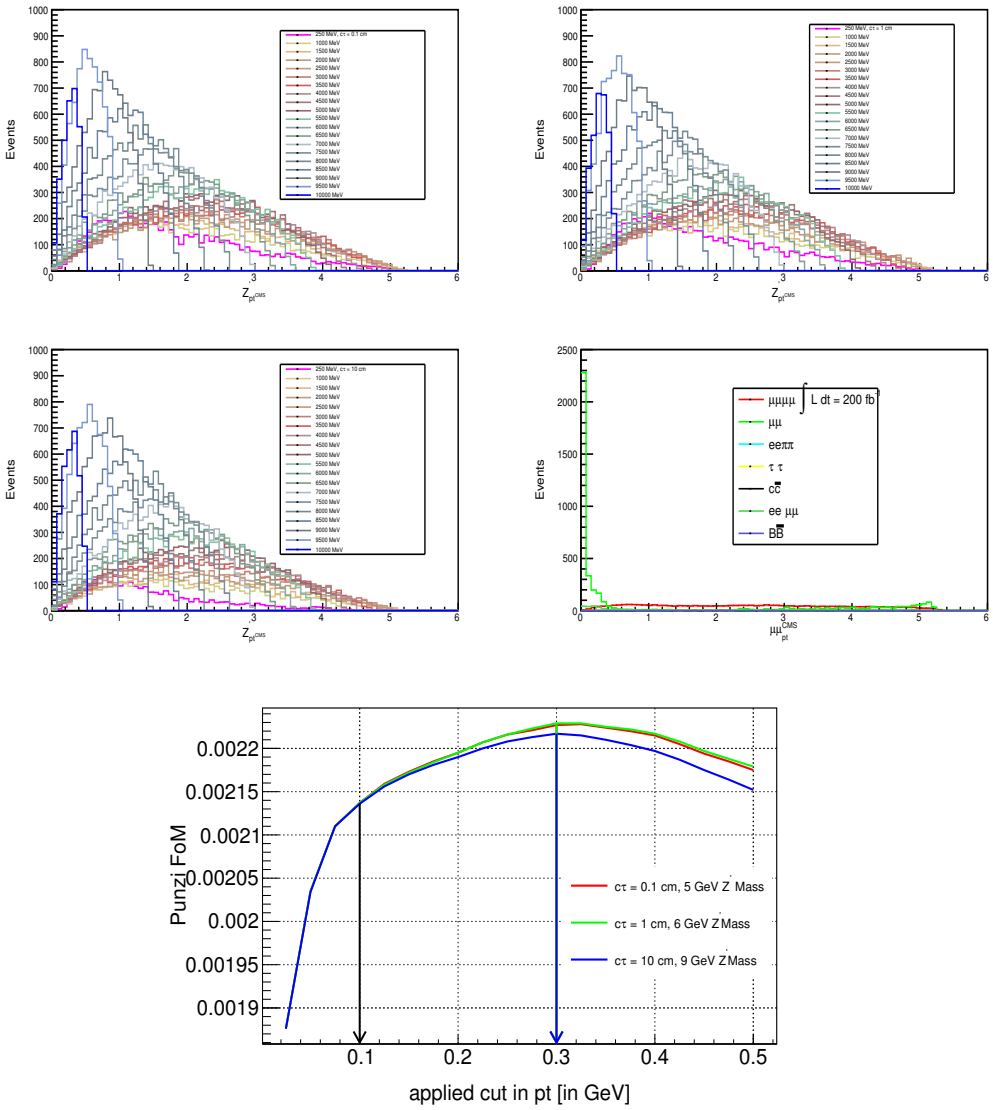


Figure 7.3.: The above four plots show the V0 transverse momentum ($\mu\mu_{pt}$) for $c\tau$ = 0.1, 1.0, 10 cm and for backgrounds, respectively. The below plot shows the Punzi FoM depending on different cut thresholds and the black arrow line shows the chosen cut value for all $c\tau$'s.

V0 momentum in CM frame

The $\mu\mu\mu\mu$ process mainly arises through the ISR and double photon conversation (see section 6.3.3). The double photon conversion can be described as a quasi-

two-body process. In the CM system, the kinematics of such a process is closed, and one can expect the dimuon CM momentum ($p_{\mu\mu}$) to peak around P_0 (see equation 6.6). Figure 7.4 shows the distribution of $p_{\mu\mu}$ in the CM system for signal and backgrounds. A common optimized cut of $p_{\mu\mu} < 5$ GeV is chosen to remove the backgrounds for all $c\tau$'s (cut is optimized using the Punzi FoM).

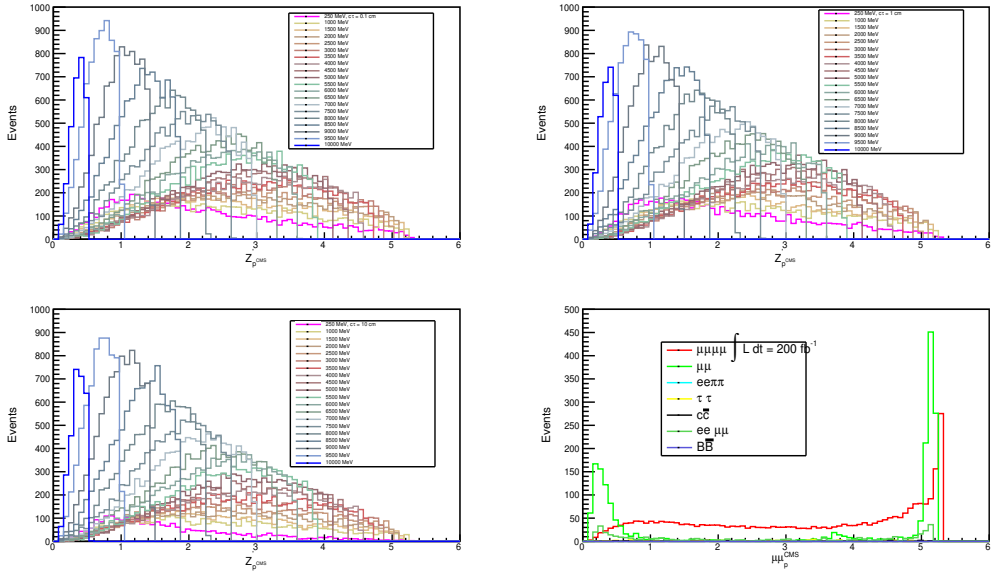


Figure 7.4.: The above four plots show the V0 momentum ($\mu\mu_p$) for $c\tau$ = 0.1, 1.0, 10 cm and for backgrounds, respectively. An optimized cut value of $p_{\mu\mu} < 5$ GeV is chosen to remove the $\mu\mu\mu\mu$ and $\mu\mu\gamma$ events.

V0 angular variables

The V0 angular variables, such as the cosine of V0 theta vertex ($\mathrm{Cos}(\theta_{vertex})$) and the cosine of the pointing angle ($\mathrm{Cos}(\alpha_{PA})$) also helps to separate signal from backgrounds. The θ_{vertex} is defined as the polar angle of the V0 Vertex with respect to IP, and the α_{PA} is the angle between the V0 momentum and the vertex vector (vector connecting IP and fitted vertex).

Figure 7.5 shows the distribution of $\mathrm{Cos}(\theta_{vertex})$ and an optimized cut of >(-0.8) choose for all the $c\tau$'s. Signal has a non-zero vertex polar angle, while the $\mu\mu\mu\mu$ background tends to be either zero or 180° (because they don't form a vertex), but for $\mu\mu\gamma$ some strange peaks are observed at 90°and 60° degrees, and currently I am trying to understand that.

Figure 7.6 (above four) shows the distribution of cosine of α_{PA} for signal and backgrounds. As the signal originates from a 2-body decay, it tends to result in a non-pointing V0 relative to the V0 selected from the $\mu\mu\mu\mu$ and $\mu\mu\gamma$ backgrounds. In figure 7.6 below four plots shows the quantity $-ln(1 - Cos(\alpha_{PA}))$ for the V0 selected from the background and signal. This is done to visualize the structure of the distribution Cos(α_{PA}) more nicely. An optimize cut of >(-0.9) is chosen for all the $c\tau$'s.

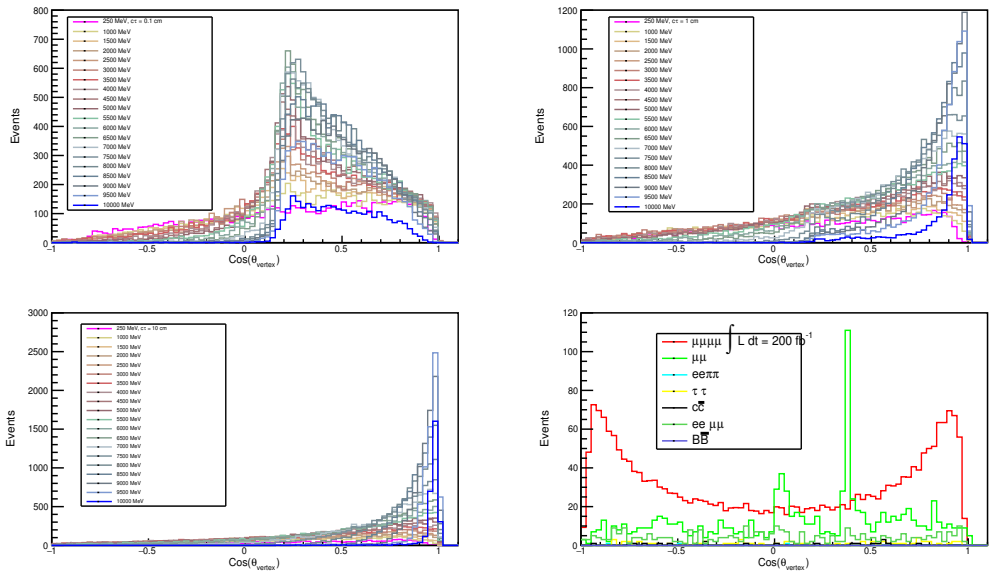


Figure 7.5.: The above four plots show the V0 Cos(θ_{vertex}) for $c\tau = 0.1,\ 1.0,\ 10$ cm and for backgrounds, respectively. An optimized cut value of Cos(θ_{vertex}) >(-0.8) is chosen to remove the $\mu\mu\mu\mu$ and $\mu\mu\gamma$ events.

V0 impact parameters

Like the impact parameter cuts (dr and dz) of the entire event topology, I also selected the V0 tracks satisfying the criteria of the V0 impact parameters.

The difference between the $z0$ impact parameters, (z coordinate of the Point Of Closest Approach (POCA)) $|V0dDeltaz0|$, and the $d0$ impact parameter, (distance to the POCA in the $r - \phi$ plane) $|d_0^{V0}|$, of the V0's daughter w.r.to the V0 vertex point as a pivot for the track is chosen to be within 1 and 0.5 cm respectively. Figure 7.7 shows the distribution of $|V0dDeltaz0|$ and $|d_0^{V0}|$ (for μ^+ track only) for signal and background events.

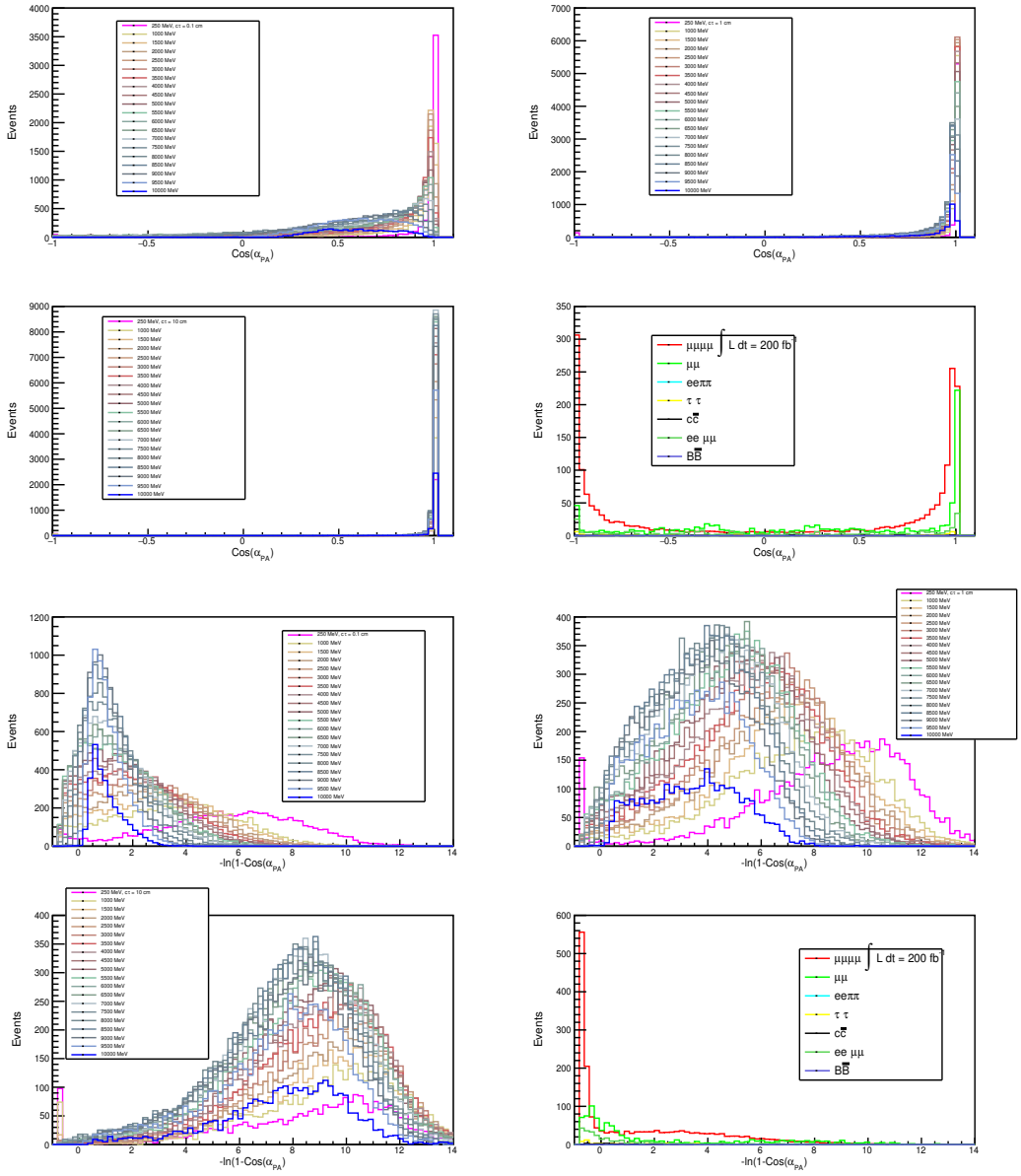


Figure 7.6.: The above four plots show the V0 Cos(α_{PA}) for $c\tau = 0.1,\ 1.0,\ 10$ cm and for backgrounds, respectively. An optimized cut value of Cos(α_{PA}) >(-0.9) is chosen to remove the $\mu\mu\mu\mu$ and $\mu\mu\gamma$ events. The below four plots show the applied transformation on Cos(α_{PA}) for better visualization.

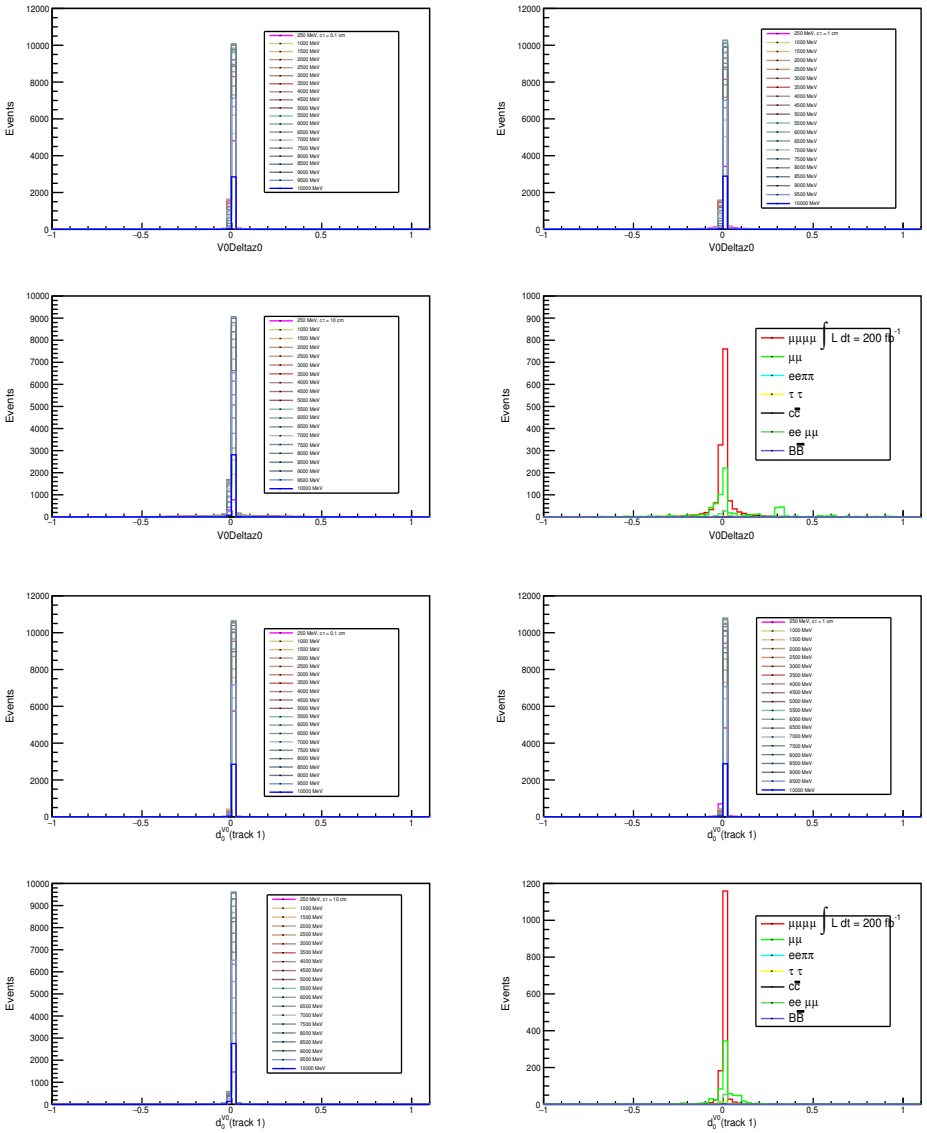


Figure 7.7.: The above four plots show the difference between $z0$ impact parameters w.r.to the V0 vertex point as a pivot of the V0's daughter for $c\tau = 0.1,\ 1.0,\ 10$ cm and for backgrounds, respectively. The below four plots show the $d0$ impact parameter w.r.to the V0 vertex point as a pivot of the one V0 daughter (μ^+) for $c\tau = 0.1,\ 1.0,\ 10$ cm and for backgrounds, respectively.

7.4.4. Summary of event selections

Here I summarise the list of event selections considered for our study.

V0 track selections

- Exactly four reconstructed charged tracks with a total charge of zero. Two of them come from the interaction region having impact parameter cut $|dr| <$ 0.5 cm, $|dz| < 2$ cm and the other two tracks with dr >0.001 cm, called the V0 tracks.
- Each muons have μID>0.5.
- A "TreeFitter" is applied to the V0 tracks and constrained them to come from a common vertex.
- Invariant mass of the four-track system restricted between 10 and 11 GeV.

V0 vertex selections

- Further selections are applied on the V0 vertex to suppress the backgrounds. The V0 vertex distance ($\mu\mu_{dr}$) in the horizontal plane has to be >0.02 cm.
- The V0 transverse momentum ($\mu\mu_{pt}$) in the CM frame has to be >0.1 GeV.
- The V0 momentum ($\mu\mu_{p}$) in the CM frame is < 5 GeV.
- The polar angle of the V0 vertex (Cosθ_{vertex}) is > -0.8.
- The pointing angle of V0 (Cosα_{PA}) is > -0.9.
- V0 impact parameter in the transverse plane ($|V0Deltaz0|$) is <1 cm.
- V0 impact parameter in the horizontal plane ($|d_0^{V0}|$) is <0.5 cm.

ISR cut

For now, I stick with the ISR cut obtained from the prompt Z′ study (see section 6.3.4). I didn't look at the control sample data yet for the displaced Z′ case, but I expect the same sort of photons activity in ECL as observed for the prompt case through the $ee\mu\mu$ control channel data. So I stick with the "ISRB" requirement, which sets an upper threshold for the total photon energy.

4C kinematic constrain

Now, all the events that passed the above selection criteria go through a kinematic 4C fit described in section 6.6. 4C kinematic constrain is useful for improving the dimuon mass resolution.
All these requirements are set at event level. An event that passes these selections produces four candidates, corresponding to the four possible neutral pairs eligible to be a Z' candidate.

7.5. Trigger Simulation (TSIM) study

Triggering the signal events having displaced signature is the most crucial part of this measurement. In Belle II, we don't have any displaced trigger available to trigger such processes. Some studies are ongoing, but they are not available for analysis use. Fortunately, our signal topology has two muons originating from the interaction points. I planned to use the ffo trigger (two tracks with an opening angle of 90 degrees) to trigger our entire signal topology by applying that to the prompt muons (muons without impact parameter cut).

For the moment, I am entirely relying on the trigger simulation to estimate the trigger efficiency. Efficiency is defined by,

$$\text{efficiency} = \frac{\text{From the truth-matched signal events how many prompt muons fired the ffo trigger bit}}{\text{No of truth-matched signal}}. \tag{7.2}$$

The ffo trigger bit is reliable in the TSIM for the used release. However, a complete and reliable efficiency estimate is required from the control sample data. Although, I checked the trigger efficiency in the $\mu\mu\gamma$ control channel at the simulation level, where γ converts to e pair, by applying a very preliminary reconstruction selection, summarized below,

- All the V0 track selections given in section 7.4.4.
- e has eID >0.5 and μ has μID >0.5.

The estimated efficiency from the control channel and the signal MC for different $c\tau$'s is given in figure 7.8. One can see that estimated efficiency from the signal MC depends on the Z′ mass as well as the given $c\tau$'s. The control channel's efficiency is shown in the limited mass region because the invariant di-electron mass phase space constrains it. Figure 7.9 shows the 2D distribution of invariant di-electron mass vs. the 3D distance of γ, clearly indicating the reason behind it. Mass by mass trigger efficiency for the control channel is estimated by taking a mass value $\pm 6\sigma$ region around it (where sigma is taken from signal MC). The estimated efficiency from the control channel and signal MC for $c\tau = 10$ cm agrees with each

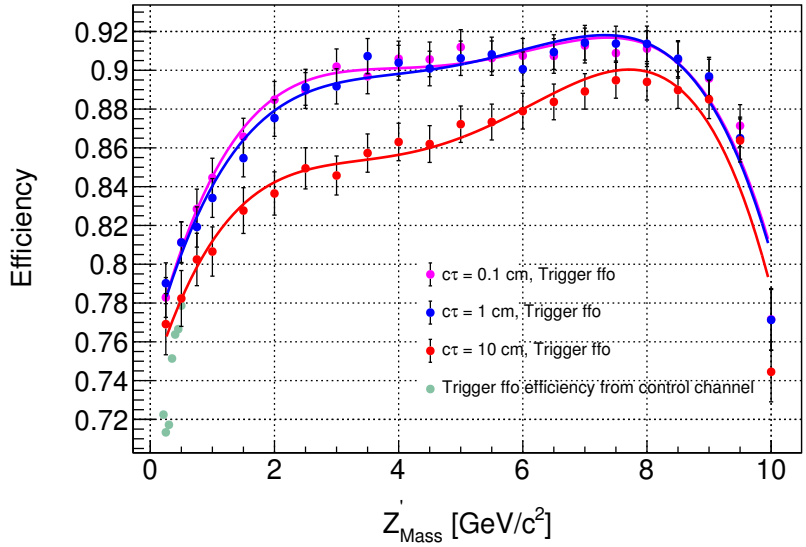


Figure 7.8.: Trigger efficiency estimated from TSIM using ffo trigger bit to fire the entire event topology for different $c\tau$'s. The plot also shows the trigger efficiency calculated from the control channel at the simulation level at the lower mass region.

other in the low mass region. This is due to the fact that the distribution of the 3D distance of the γ vertex and the signal sample with $c\tau = 10$ cm is almost in agreement. However, I need a careful investigation of the $\mu\mu\gamma$ control channel and some other control channels to probe the entire mass region and different $c\tau$'s.

However, for the thesis, I entirely relied on the efficiency estimated from the signal MC. Efficiency is estimated for some sample mass hypotheses, and they are fitted with 4th order polynomial to describe the efficiency for any mass points. The continuous analytic function describing the trigger efficiencies for different $c\tau$'s is given below,

$$\text{Trigger efficiency}_{c\tau=0.1\text{cm}} = 0.7532 + 0.1274 \cdot M - 0.04077 \cdot M^2 + 0.005661 \cdot M^3 - 0.0002804 \cdot M^4 \tag{7.3}$$

$$\text{Trigger efficiency}_{c\tau=1\text{cm}} = 0.7552 + 0.1173 \cdot M - 0.03698 \cdot M^2 + 0.0052 \cdot M^3 - 0.0002624 \cdot M^4 \tag{7.4}$$

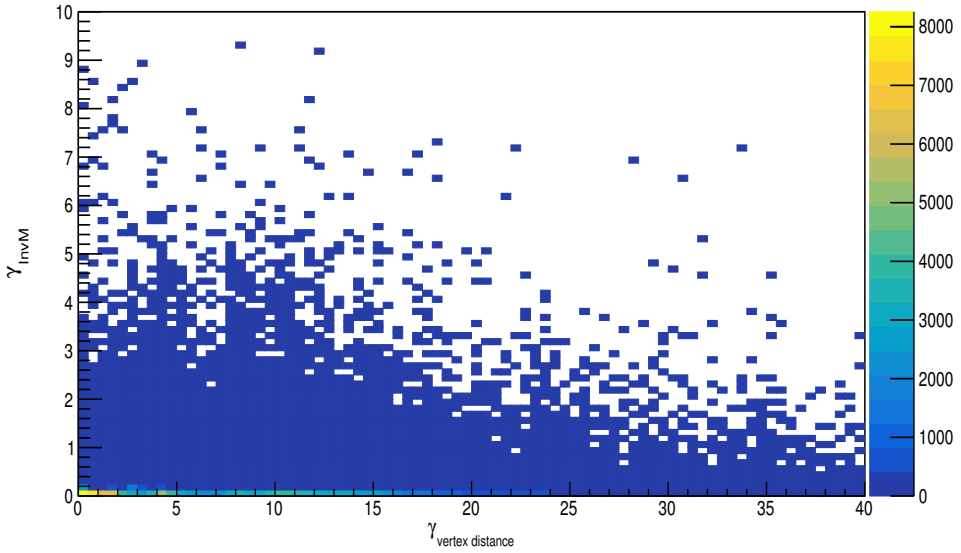


Figure 7.9.: The 2D plot of the photon invariant mass vs. the vertex distance. As one can see, the average photon vertex distance is ∼10cm.

$$\text{Trigger efficiency}_{c\tau=10\text{cm}} = 0.7376 + 0.1058 \cdot M - 0.03682 \cdot M^2 + 0.005629 \cdot M^3 - 0.0002956 \cdot M^4 \tag{7.5}$$

where M is the dimuon invariant mass.

7.6. Signal Modeling

Different signal mass hypothesis are modeled by using the sum of two crystal ball functions. I followed the same strategy as given in section 6.7. The reduced dimuon mass distribution has been fitted for each signal mass hypothesis and the signal shaping technique mentioned in section 6.7.1 is then performed to obtain a continuous analytic function for any mass point. Fittings of some the sample mass points are given in figure 7.10. Figure 7.11 shows the modeling of the CB parameters for the three different $c\tau$'s.

Unlike section 6.7.1, I didn't do the spline fitting for three different regions instead I fitted the whole mass range using different order polynomials summarized in table 7.4. A comparison of signal mass resolution and efficiency is shown also in figure 7.12, As one can see, resolution varies from a minimum value at some low

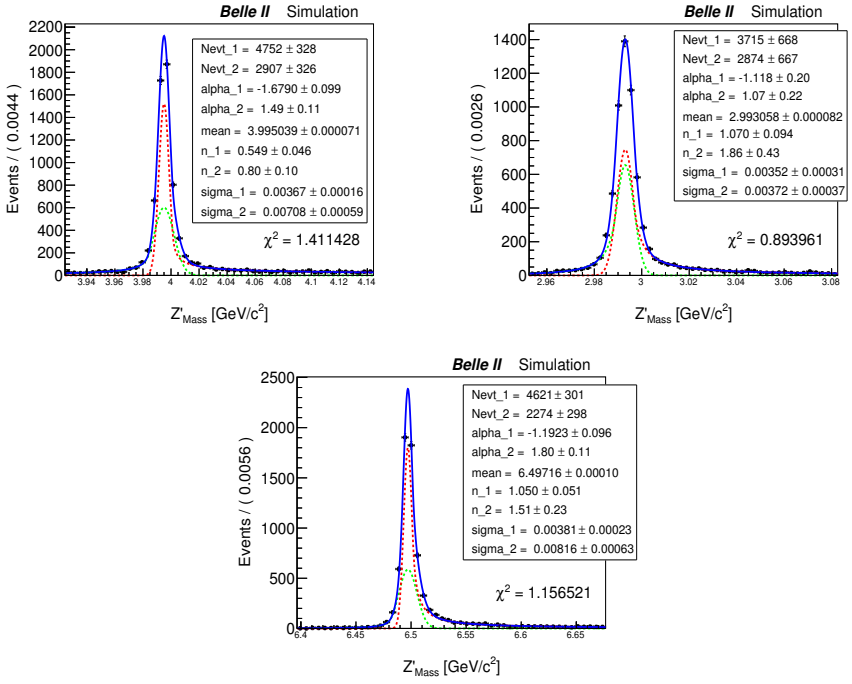


Figure 7.10.: Double crystal ball fitting to the 4.0, 3.0, 6.5 GeV Z' invariant mass having $c\tau$ of 0.1, 1, 10 cm respectively.

mass and then gets maximum at a certain mass point and then starts to decrease again. This behaviour is also same for different cτ's, but the maximum resolution mass point get shifted towards higher mass, this resolution effect is also observed in section 7.4.3.

7.7. Fit procedure

The fit procedure for a very preliminary sensitivity estimate is done for a targeted integrated luminosity of $\int \mathrm{L}\ \mathrm{dt} = 200\ \mathrm{fb}^{-1}$. Figure 7.13 shows the reduced dimuon mass distribution of surviving background components for different cτ's. I followed the same fitting and sensitivity estimate procedure described in section 6.8 and section 6.8.1. Background is fitted with first order Chebychev polynomials (experience from prompt case second or higher order coefficients will be very small and led to worse final sensitivities.) and signal shapes are taken from the double CB functions with parameters taken from the analytical modeling and kept fixed to those values. Then MC background distributions are fitted with a 1D Unbinned Maximum Likelihood technique, using two hypotheses: a) background-only

$c\tau$ (in cm)	polynomial
0.1	σ_w = 0.001755 + 0.001063x - 0.0000437x^2 - 0.000007044x^3
	σ_1 = 0.001099 + 0.002338x - 0.00068x^2 + 0.00007916x^3 - 0.000003406x^4
	σ_2 = 0.001433 + 0.000445x + 0.0002413x^2 - 0.0000317x^3 + 0.000000221x^4
	α_1 = -1.483 + 0.3773x - 0.2469x^2 + 0.04689x^3 - 0.002622x^4
	α_2 = 1.295 + 0.1341x - 0.1185x^2 + 0.03346x^3 - 0.002397x^4
	n_1 = 1.241 + 0.02749x - 0.1534x^2 + 0.03511x^3 - 0.002054x^4
	n_2 = 1.458 - 0.01323x - 0.084x^2 + 0.01142x^3
	f_{CB1} = 0.5643 - 0.1319x + 0.08844x^2 - 0.01819x^3 + 0.001153x^4
	efficiency = 0.1545 + 0.07865x - 0.004959x^2 - 0.000249x^3
1.0	σ_w = 0.001558 + 0.0008463x + 0.00007244x^2 - 0.00001743x^3
	σ_1 = 0.001896 - 0.000685x + 0.0008332x^2 - 0.0001391x^3 + 0.00000606x^4
	σ_2 = -0.0001118 + 0.004061x - 0.001555x^2 + 0.0002384x^3 - 0.00001238x^4
	α_1 = -0.9595 - 0.4076x + 0.1389x^2 - 0.02108x^3 + 0.001196x^4
	α_2 = 0.932 - 0.06156x + 0.1406x^2 - 0.02414x^3 + 0.001195x^4
	n_1 = 1.839 - 0.6517x + 0.1168x^2 - 0.005718x^3
	n_2 = 2.267 - 0.5546x - 0.005729x^2 + 0.008933x^3
	f_{CB1} = 0.6118 - 0.14x + 0.06978x^2 - 0.0104x^3 + 0.0004761x^4
	efficiency = 0.1394 + 0.08859x - 0.006224x^2 - 0.0002033x^3
10	σ_w = 0.002108 + 0.0003329x + 0.0001846x^2 - 0.00002405x^3
	σ_1 = 0.001224 + 0.001536x - 0.00041x^2 + 0.00005935x^3 - 0.000003405x^4
	σ_2 = 0.00395 - 0.003303x + 0.002086x^2 - 0.000351x^3 + 0.00001765x^4
	α_1 = -1.088 + 0.5463x - 0.3633x^2 + 0.06626x^3 - 0.003614x^4
	α_2 = 1.598 - 1.149x + 0.5423x^2 - 0.08282x^3 + 0.004116x^4
	n_1 = 2.177 - 0.7527x + 0.1331x^2 - 0.00664x^3
	n_2 = 1.808 + 0.3716x - 0.2145x^2 + 0.02198x^3
	f_{CB1} = 0.4377 + 0.1399x - 0.03734x^2 + 0.003826x^3 - 0.0001216x^4
	efficiency = 0.04095 + 0.0839x - 0.003534x^2 - 0.0003159x^3

Table 7.4.: Continuous analytic functions for all the mass ranges.

hypothesis; b) signal + back-ground hypothesis as described in section 6.8.

7.7.1. Preliminary sensitivity studies

In this section, I estimate the sensitivity for our search $e^+e- \to \mu^+\mu^- Z'(\to \mu^+\mu^-)$ where Z' having $c\tau$ of 0.1, 1, 10 cm in terms of the process cross section σ. As we want to keep our search as much as possible model independent, I will not provide any sensitivity in terms of the coupling constant g'. The asymptotic calculator with a one-sided Profile likelihood test statistics is used to estimate the upper limit on the cross section in presence of 20% flat systematic uncertainties affecting the signal efficiency for each mass points (detailed procedure of sensitivity estimation is given in section 6.8.1). As systematic uncertainties are not studied yet carefully but the experience from the prompt studies systematics uncertainties would not effect much and I expect systematics uncertainties will be below 20%. Figure 7.13

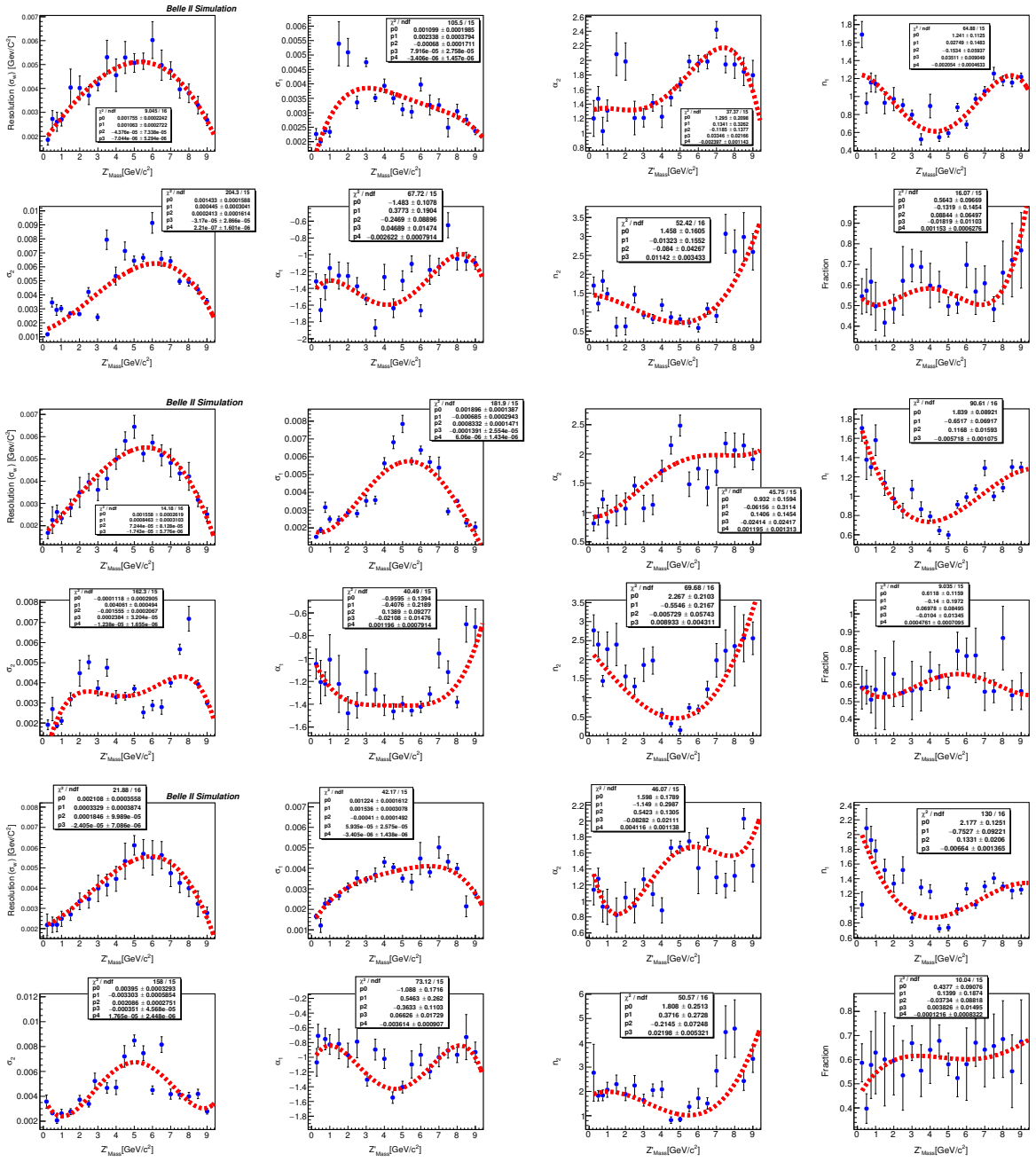


Figure 7.11.: Modeling of the double crystal ball parameters to obtain continuous analytic function, which will describe the parameter for any mass point.

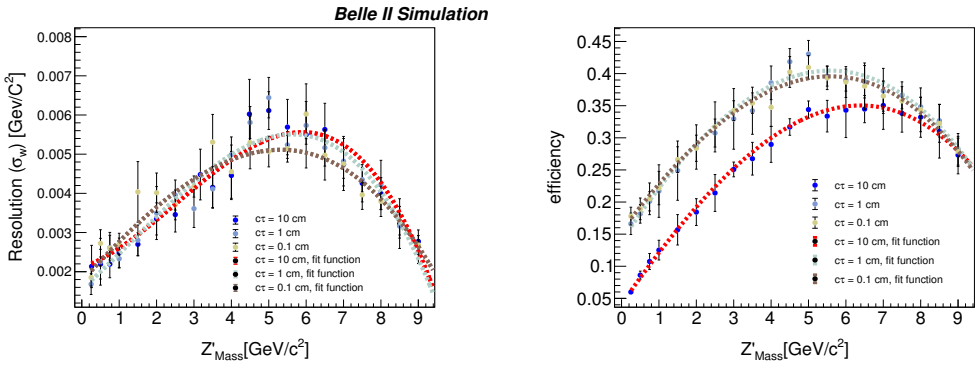


Figure 7.12.: The left plot shows the comparison of mass resolution and right plot shows the comparison of efficiency for different mass hypothesis.

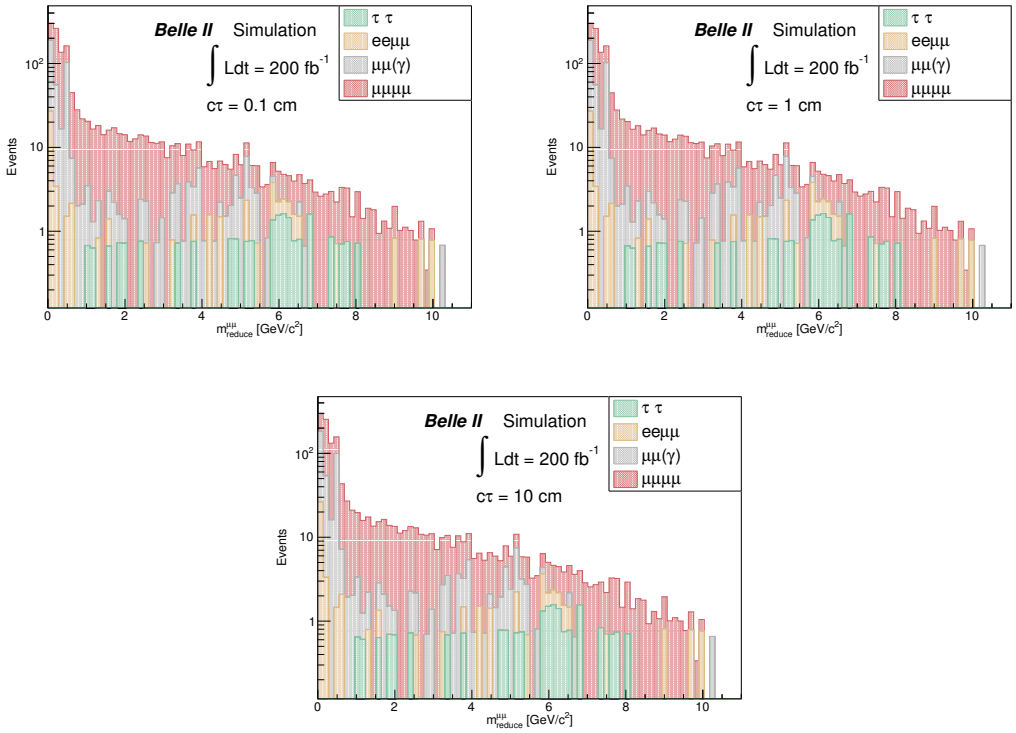


Figure 7.13.: Reduced mass distribution of surviving backgrounds for different cτ's

shows the estimated 90% CL upper limits on the cross-section σ_{UL} for the mass point 0.25 GeV/c^2 for $c\tau$ = 0.1, 1, 10 cm respectively. Figure 7.14 shows the

upper limits on the cross-section (σ_{UL}) and the signal yield (N_{UL}) for our target integrated luminosity of 200 fb^{-1} using 20% flat systematic uncertainties for all mass hypotheses. As one can see from the cross-section plot, increasing the lifetime of Z' causes a loss in the sensitivity due to the loss in signal and trigger efficiency. Also, as expected, the observed upper limit on cross-section (below 0.1 fb on average) seems to improve more from the prompt case (which is at a level of ~ 0.2 fb for cross-section on average) due to more background reduction.

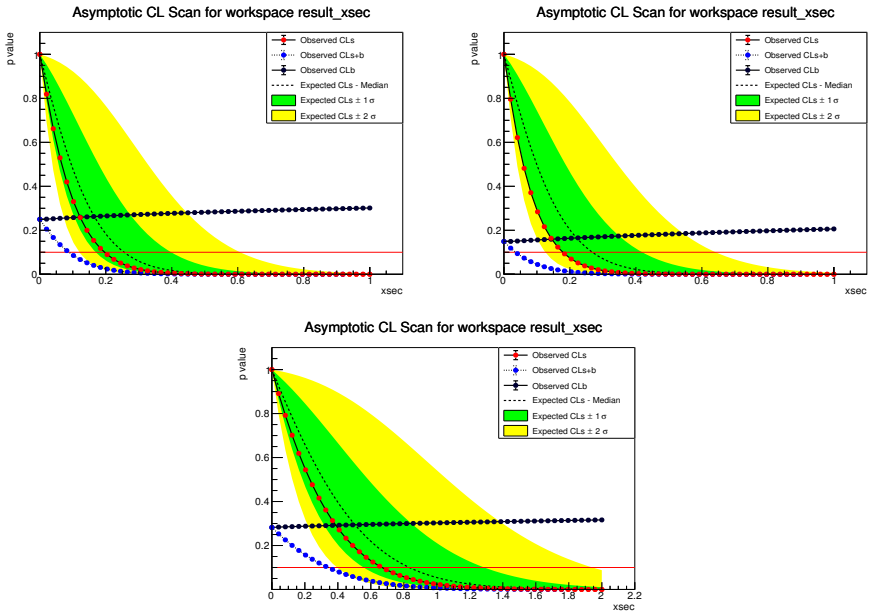


Figure 7.14.: Estimated 90% exclusion limit on the cross-section of 0.5 GeV mass, obtained using asymptotic calculator for $c\tau$'s = 0.1, 1, 10 cm respectively.

7.8. Summary, conclusions and future outlook

In this chapter, first time, I developed some preliminary search strategies to find for the process $e^+e^- \to \mu^+\mu^- Z'(\to \mu^+\mu^-)$, where Z' is long-lived and having $c\tau$ of 0.1, 1, 10 cm, model independently. My primary aim was to use EVTGEN as a generator to generate the signal MC samples, but I found some issues and generated them by Madgraph. Then, using the "Displaced vertex" module gave them a finite lifetime. I selected exactly four tracks in the event; two of them come from the interaction region, and the other two have dr $>$ 0.001 cm, called

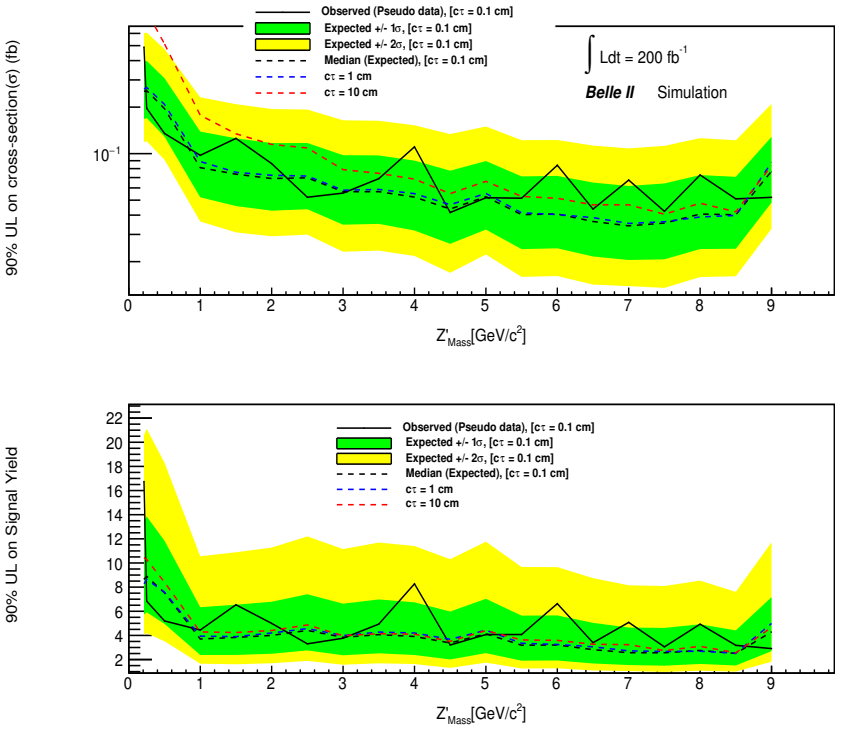


Figure 7.15.: The above plot shows the 90% CL upper limit on the cross-section for different $c\tau$'s and the below plot shows 90% CL on the signal yield.

the V0 tracks. V0 tracks are constrained to come from a common vertex, and depending on the momentum, angular, and distance information of the V0 vertex, background suppression techniques are developed. Then ISR cut is applied to reject the backgrounds arising from photons, and a 4C kinematic constraint is used to improve the signal mass resolution. I also performed a preliminary trigger simulation study to trigger the signal events. My main aim was to trigger the prompt muons originating from the interaction region using the ffo trigger; on average, for $c\tau$ 0.1, 1 cm trigger efficiency is close to 90%, while for $c\tau$ 10 cm, it is 85%. Although the used trigger bit is reliable in the used Belle II software release, but a precise efficiency estimate on data is required. I verified the trigger efficiency using the $\mu\mu\gamma$ control channel at simulation level in the limited phase space region (low mass region) and found to be agreed with the estimated efficiency. However, verification in the entire phase space is still required in data, which is currently under study. After finalizing all the cuts, I estimated the sensitivity of the process cross-section using a dedicated fitting technique. Signal mass points are described by the double crystal ball function, while a first-order Chebyshev polynomial is

used to describe the background. A 90% CL on the production cross-section is set for 200 fb^{-1} integrated luminosity, using a 20% flat systematics uncertainty, assuming no signal will be present at the data. The observed results are very interesting. The upper limit on the cross-section is found to be below 0.1 fb on average, and sensitivity worsened while increasing the lifetime of Z'. The obtained sensitivity on the cross-section for different $c\tau$'s is improved more than the prompt search due to the lower background level.

For the thesis, this study is finalized up to this point, but a dedicated control channel study to validate all the procedures is underway. As mentioned earlier, I applied muonID on each track to identify muons; the possibility of choosing 3 track muonID, like the prompt case, is also under study. There is a possibility to use an MVA technique to suppress the background level further, exploiting all the variables studied for the cut-based analysis to improve the sensitivity. The tools of doing fit stability studies, calculating significance, and LEE are already developed (see section 6.9, 6.10, 6.11); as the analysis cut is not finalized, I didn't present them here. As we didn't have a dedicated control channel that could probe all the phase space of our interest, maybe we could look at 10% of our data, finalize all the systematics sources, verify all the procedures, and discard them from the final result. Which are all currently under discussion, and hopefully, we will converge and submit all our findings to the collaboration soon.

Conclusion

In the thesis, I presented three different works: first, a search for a new dark gauge boson Z' in its decay to the muonic final state in prompt and displaced scenarios, second an MVA-based particle identification algorithm for assisting new physics searches in the critical low momentum region, third trigger performance study for triggering low multiplicity tracks, such as two tracks in the final states, intending to obtain better trigger performance than the currently used triggers, at the Belle II experiment.

The search for the Z' boson in its decay to muons, in the four muons final state: $e^+e^- \to \mu^+\mu^- Z',\ Z' \to \mu^+\mu^-$, has been investigated model dependently $(L_\mu - L_\tau)$ in the prompt and model independently in the displaced cases. In both cases mass phase space $0.212\ \text{GeV}/c^2 < Z'_{\text{mass}} < 9\ \text{GeV}/c^2$ has been explored. The signal search strategy in both cases is to find a peak in the dimuon mass distribution.

The targeted luminosity for prompt search is $178.47\ \text{fb}^{-1}$. BaBar and Belle experiments have already explored the prompt search in the same mass phase region with a larger data set. However, our aim is to obtain better or similar performance despite a lower dataset, which is possible through aggressive background suppression. The main sources of backgrounds in our search are SM $\mu\mu\mu\mu$ and $\mu\mu\gamma$ mainly in the low mass region. We developed an MVA-based background suppression technique depending on several kinematic features of the tracks and helicity angle. After achieving the desired performance, I developed a dedicated fitting technique for signal extraction. The analysis methods were validated, tested, and the possible sources of systematics have been estimated using the $ee\mu\mu$ control channel. It was found that systematics is mainly dominated due to MVA-based background suppression at a level of $\sim 20\%$. The data/MC comparison at the control channel suggests a discrepancy due to the ISR process being absent at the MC level, as expected. Because the generator used to generate four lepton processes doesn't include ISR and FSR effects. To confirm the ISR effect further, we look at the 2019 data and conclude that our MVA-based background suppression technique is able to compensate for the ISR discrepancy for data and MC. This unblinded dataset will not be used to give the final sensitivity projection. After all, at the targeted luminosity, which is 2-3 times less than BaBar and Belle, I estimated a 90% CL on the coupling constant g'. The obtained result is very promising; we got comparable sensitivity in most of the mass spectrum apart from the low mass region, mainly due to the aggressive background suppression. The

estimated results are based on the MC study only. The analysis is currently under review, and we expect to unblind soon.

In the other search, I explored the displaced decay of Z' model independently in the same event topology as mentioned above. Displaced decay means Z' has some finite lifetime, three different lifetime scenarios $c\tau = 0.1,\ 1,\ 10$ cm has been explored. This is the first time I have performed this search. The targeted data luminosity is not finalized yet for this measurement. However, I presented some results depending on 200 fb^{-1} integrated luminosity, based on the MC study only. The expected backgrounds for this search are also SM $\mu\mu\mu\mu$ and $\mu\mu\gamma$. My primary aim was to do a cut-based analysis rather than MVA-based aggressive background suppression because the displaced nature of the signal will reduce the backgrounds drastically. In the signal topology, we have two tracks coming from the interaction region called prompt tracks, and the other two tracks are displaced, called V0 tracks. At first, depending on V0 track selections, a V0 reconstruction is performed; then, depending on V0 vertex information, most of the background suppression technique is developed. The V0 vertex's, displaced nature, kinematic and angular properties helped to reduce the background massively. After achieving the desired background suppression, I proceeded to do a trigger study, which is very crucial for this search as we don't have a displaced trigger available in Belle II. However, fortunately, in our signal topology, we have two muon tracks coming from the interaction region. I planned to trigger these two tracks, hence the entire event topology, based on a CDC-based two track trigger (ffo). The measured efficiency varied from 85% to 90% for higher $c\tau$ to low $c\tau$, respectively. Then I performed a dedicated fitting technique to extract the signal and estimate a 90% CL on the production cross-section, which is found to be below 0.1 fb on average for all the $c\tau$. In comparison with the prompt search, I observe better sensitivity in the production cross-section due to more background rejection. For the thesis, the study is finalized up to this point. Detailed control channel study, systematic evaluation, and trigger study in control channel data are in progress. I will hopefully converge soon and submit our findings to the collaboration for internal review. This study will not only intend us to do this measurement but also help us to understand the capability of Belle's displaced vertex searches in the near future.

The other studies presented in the thesis are developing an MVA-based particle identification algorithm. Semileptonic B decays, especially $b \to s, b \to c$ transitions, providing us testing ground to investigate new physics effects. These decays involve leptons in the final state, and to separate them from hadronic backgrounds, especially in the low momentum region, a better particle identification is required. Because low momentum muons can't reach the KLM, and hadrons easily mimic low momentum electrons due to bremsstrahlung loss. One must need to rely

on the ECL and other sub-detectors to identify them. Likelihood based particle identification worked very well in Belle and Belle II. However, still, in those scenarios, an MVA-based particle identification could provide potential improvements. A BDT-based particle identification algorithm is already developed by the Belle II collaboration and observed potential improvements than the likelihood-based method. My aim was to exploit the neural network to achieve better performance than BDT, especially in the critical low momentum regions. I studied several cluster energy, shower shapes related variables from the ECL, and different likelihoods information from other sub-detectors and gave a global presentation of particle identification in different momentum and angular bins. I observed 1-2% and 2-3% better performance for $e-\pi$ and $\mu-\pi$ separation than BDT in the low momentum regions. This is due to the neural network's better handling of non-linear features. The above mentioned particle identification procedure uses likelihood information from other sub-detectors to give the global interpretation. However, I tried to replace the CDC likelihood by exploring some lower level information from CDC. I studied energy loss information, layer-hits, and angular related feature of tracks within CDC, combined them with BDT and developed a particle identification algorithm concentrating only on CDC. Although I didn't observe huge improvements compared to the likelihood, but we have some potential improvements for $\mu-\pi$ separation in the low momentum region, mainly coming from hit and angular variables. In short, the conclusion of this study is if we replace the likelihood information of CDC in global MVA and use the lower level features such as CDCdEdx, CDCdEdx-layer hits and particle's angular distribution in CDC ("costhCDC"), we could observe potential improvements than likelihood. Other sub-detector based studies could also be useful to improve the global BDT performance.

I also presented some studies related to the trigger in the thesis, mainly aiming at some specific dark sector searches such as $e^+e^- \to \mu^+\mu^- Z',\ Z' \to invisible$ or other dark sector searches especially involving two muons or two tracks in the final state. The CDC-based two track triggers such ffo (two tracks with an opening angle 90°) and $ff30$ (two tracks with an opening angle 30°) and KLM-based triggers such as different CDCKLM lines have been already investigated and observe the trigger performance at a level of ∼90% or above. I aimed to study the ECL-based dimuon trigger $ECL_{\mu\mu}$ if we could observe better performance than the existing ones. I estimated the efficiency in the $e^+e^- \to \mu\mu\gamma$ transition for Exp7+Exp8 with proc9 data and observed 5% less efficiency than the other triggers.

Bibliography

[1] Branko Novakovic et al. "APLICATION OF CONTROL TERMS P1 AND P2 TO ESTIMATION OF ROTATION CURVE OF STARS". In: Sept. 2014. DOI: 10.2014/bran.001 (cit. on p. 9).

[2] N. Aghanim et al. "Planck 2018 results. VI. Cosmological parameters". In: *Astron. Astrophys.* 641 (2020). [Erratum: Astron.Astrophys. 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO] (cit. on p. 10).

[3] Mark Srednicki, Richard Watkins, and Keith A. Olive. "Calculations of relic densities in the early universe". In: *Nuclear Physics B* 310.3 (1988), pp. 693–713. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(88)90099-5. URL: https://www.sciencedirect.com/science/article/pii/0550321388900995 (cit. on p. 11).

[4] John McDonald. "Generation of WIMP Miracle-like Densities of Baryons and Dark Matter". In: (Jan. 2012). arXiv: 1201.3124 [hep-ph] (cit. on p. 11).

[5] Jonathan L. Feng. "Dark Matter Candidates from Particle Physics and Methods of Detection". In: *Ann. Rev. Astron. Astrophys.* 48 (2010), pp. 495–545. DOI: 10.1146/annurev-astro-082708-101659. arXiv: 1003.0904 [astro-ph.CO] (cit. on p. 12).

[6] Benjamin W. Lee and Steven Weinberg. "Cosmological Lower Bound on Heavy-Neutrino Masses". In: *Phys. Rev. Lett.* 39 (4 July 1977), pp. 165–168. DOI: 10.1103/PhysRevLett.39.165. URL: https://link.aps.org/doi/10.1103/PhysRevLett.39.165 (cit. on p. 11).

[7] E. Aprile et al. "Dark Matter Search Results from a One Ton-Year Exposure of XENON1T". In: *Phys. Rev. Lett.* 121 (11 Sept. 2018), p. 111302. DOI: 10.1103/PhysRevLett.121.111302. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.111302 (cit. on p. 12).

[8] Miguel Escudero et al. "Toward (Finally!) Ruling Out Z and Higgs Mediated Dark Matter Models". In: *JCAP* 12 (2016), p. 029. DOI: 10.1088/1475-7516/2016/12/029. arXiv: 1609.09079 [hep-ph] (cit. on p. 12).

[9] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on p. 13).

[10] Henri Ruegg and Marti Ruiz-Altaba. "The Stueckelberg field". In: *Int. J. Mod. Phys. A* 19 (2004), pp. 3265–3348. DOI: 10.1142/S0217751X04019755. arXiv: hep-th/0304245 (cit. on p. 13).

[11] R. D. Peccei and Helen R. Quinn. "Constraints imposed by CP conservation in the presence of pseudoparticles". In: *Phys. Rev. D* 16 (6 Sept. 1977), pp. 1791–1797. DOI: 10.1103/PhysRevD.16.1791. URL: https://link.aps.org/doi/10.1103/PhysRevD.16.1791 (cit. on p. 13).

[12] Paul Langacker. "The physics of heavy Z' gauge bosons". In: *Rev. Mod. Phys.* 81 (3 Aug. 2009), pp. 1199–1228. DOI: 10.1103/RevModPhys.81.1199. URL: https://link.aps.org/doi/10.1103/RevModPhys.81.1199 (cit. on p. 14).

[13] Julian Heeck and Werner Rodejohann. "Gauged $L_\mu - L_\tau$ symmetry at the electroweak scale". In: *Phys. Rev. D* 84 (7 Oct. 2011), p. 075007. DOI: 10.1103/PhysRevD.84.075007. URL: https://link.aps.org/doi/10.1103/PhysRevD.84.075007 (cit. on p. 14).

[14] David Curtin et al. "Illuminating Dark Photons with High-Energy Colliders". In: *JHEP* 02 (2015), p. 157. DOI: 10.1007/JHEP02(2015)157. arXiv: 1412.0018 [hep-ph] (cit. on p. 14).

[15] T. Aoyama et al. "The anomalous magnetic moment of the muon in the Standard Model". In: *Physics Reports* 887 (2020). The anomalous magnetic moment of the muon in the Standard Model, pp. 1–166. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2020.07.006. URL: https://www.sciencedirect.com/science/article/pii/S0370157320302556 (cit. on p. 15).

[16] B. Abi et al. "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm". In: *Phys. Rev. Lett.* 126 (14 Apr. 2021), p. 141801. DOI: 10.1103/PhysRevLett.126.141801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.126.141801 (cit. on p. 16).

[17] Takeshi Araki et al. "Detecting the $L_\mu - L_\tau$ gauge boson at Belle II". In: *Phys. Rev. D* 95 (5 Mar. 2017), p. 055006. DOI: 10.1103/PhysRevD.95.055006. URL: https://link.aps.org/doi/10.1103/PhysRevD.95.055006 (cit. on p. 16).

[18] Sz. Borsanyi et al. "Leading hadronic contribution to the muon magnetic moment from lattice QCD". In: *Nature* 593.7857 (2021), pp. 51–55. DOI: 10.1038/s41586-021-03418-1. arXiv: 2002.12347 [hep-lat] (cit. on p. 16).

[19] R. Aaij et al. "Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$". In: *Phys. Rev. Lett.* 111 (19 Nov. 2013), p. 191801. DOI: 10.1103/PhysRevLett.111.191801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.111.191801 (cit. on p. 16).

[20] Wolfgang Altmannshofer et al. "Quark flavor transitions in $L_\mu - L_\tau$ models". In: *Phys. Rev. D* 89 (9 May 2014), p. 095033. DOI: 10.1103/PhysRevD.89.095033. URL: https://link.aps.org/doi/10.1103/PhysRevD.89.095033 (cit. on p. 16).

[21] Brian Shuve and Itay Yavin. "Dark matter progenitor: Light vector boson decay into sterile neutrinos". In: *Phys. Rev. D* 89 (11 June 2014), p. 113004. DOI: 10.1103/PhysRevD.89.113004. URL: https://link.aps.org/doi/10.1103/PhysRevD.89.113004 (cit. on p. 16).

[22] T. Czank et al. "Search for $Z' \to \mu^+\mu^-$ in the $L_\mu - L_\tau$ gauge-symmetric model at Belle". In: *Phys. Rev. D* 106 (1 July 2022), p. 012003. DOI: 10.1103/PhysRevD.106.012003. URL: https://link.aps.org/doi/10.1103/PhysRevD.106.012003 (cit. on p. 16).

[23] J. P. Lees et al. "Search for a muonic dark force at BaBar". In: *Phys. Rev. D* 94 (1 July 2016), p. 011102. DOI: 10.1103/PhysRevD.94.011102. URL: https://link.aps.org/doi/10.1103/PhysRevD.94.011102 (cit. on p. 16).

[24] J. P. Lees et al. "Search for a Dark Leptophilic Scalar in e^+e^- Collisions". In: *Phys. Rev. Lett.* 125 (18 Oct. 2020), p. 181801. DOI: 10.1103/PhysRevLett.125.181801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.181801 (cit. on pp. 16, 17).

[25] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Phys. Rept.* 516 (2012), pp. 1–102. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph] (cit. on p. 16).

[26] Brian Batell et al. "Muon anomalous magnetic moment through the leptonic Higgs portal". In: *Phys. Rev. D* 95 (7 Apr. 2017), p. 075003. DOI: 10.1103/PhysRevD.95.075003. URL: https://link.aps.org/doi/10.1103/PhysRevD.95.075003 (cit. on p. 16).

[27] J Beacham et al. “Physics beyond colliders at CERN: beyond the Standard Model working group report”. In: *Journal of Physics G: Nuclear and Particle Physics* 47.1 (Dec. 2019), p. 010501. DOI: 10.1088/1361-6471/ab4cd2. URL: https://dx.doi.org/10.1088/1361-6471/ab4cd2 (cit. on p. 16).

[28] R. Bernabei et al. “DAMA results”. In: *10th International Workshop on Neutrino Telescopes*. May 2003, pp. 403–423. arXiv: astro-ph/0305542 (cit. on p. 18).

[29] M. Ackermann et al. “Search for gamma-ray spectral lines with the Fermi Large Area Telescope and dark matter implications”. In: *Phys. Rev. D* 88 (8 Oct. 2013), p. 082002. DOI: 10.1103/PhysRevD.88.082002. URL: https://link.aps.org/doi/10.1103/PhysRevD.88.082002 (cit. on p. 19).

[30] Joerg Jaeckel and Andreas Ringwald. “The Low-Energy Frontier of Particle Physics”. In: *Annual Review of Nuclear and Particle Science* 60.1 (2010), pp. 405–437. DOI: 10.1146/annurev.nucl.012809.104433. eprint: https://doi.org/10.1146/annurev.nucl.012809.104433. URL: https://doi.org/10.1146/annurev.nucl.012809.104433 (cit. on p. 19).

[31] F. Abudinén et al. “Search for Axionlike Particles Produced in e^+e^- Collisions at Belle II”. In: *Phys. Rev. Lett.* 125 (16 Oct. 2020), p. 161806. DOI: 10.1103/PhysRevLett.125.161806. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.161806 (cit. on pp. 19, 23).

[32] J. P. Lees et al. “Search for Invisible Decays of a Dark Photon Produced in e^+e^- Collisions at BaBar”. In: *Phys. Rev. Lett.* 119 (13 Sept. 2017), p. 131804. DOI: 10.1103/PhysRevLett.119.131804. URL: https://link.aps.org/doi/10.1103/PhysRevLett.119.131804 (cit. on p. 19).

[33] J. P. Lees et al. “Search for a Dark Photon in e^+e^- Collisions at BaBar”. In: *Phys. Rev. Lett.* 113 (20 Nov. 2014), p. 201801. DOI: 10.1103/PhysRevLett.113.201801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.113.201801 (cit. on p. 19).

[34] F. Abudinén et al. “Search for a dark photon and an invisible dark Higgs boson in $\mu^+\mu^-$ and missing energy final states with the Belle II experiment”. In: (July 2022). arXiv: 2207.00509 [hep-ex] (cit. on p. 20).

[35] I. Adachi et al. “Search for an Invisibly Decaying Z' Boson at Belle II in $e^+e^- \to \mu^+\mu^-(e^\pm\mu^\mp)$ Plus Missing Energy Final States”. In: *Phys. Rev. Lett.* 124 (14 Apr. 2020), p. 141801. DOI: 10.1103/PhysRevLett.124.141801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.141801 (cit. on pp. 20, 23).

[36] T. Abe et al. “Belle II Technical Design Report”. In: (Nov. 2010). arXiv: 1011.0352 [physics.ins-det] (cit. on p. 21).

[37] T. Aushev et al. “Physics at Super B Factory”. In: (2010). arXiv: 1002.5012 [hep-ex] (cit. on p. 21).

[38] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso. “SuperKEKB Collider”. In: *Nucl. Instrum. Meth. A* 907 (2018), pp. 188–199. DOI: 10.1016/j.nima.2018.08.017. arXiv: 1809.01958 [physics.acc-ph] (cit. on p. 21).

[39] S. Kurokawa and Eiji Kikutani. “Overview of the KEKB accelerators”. In: *Nucl. Instrum. Meth. A* 499 (2003), pp. 1–7. DOI: 10.1016/S0168-9002(02)01771-0 (cit. on p. 21).

[40] M. Bona et al. “SuperB: A High-Luminosity Asymmetric e+ e- Super Flavor Factory. Conceptual Design Report”. In: (May 2007). arXiv: 0709.0451 [hep-ex] (cit. on p. 22).

[41] A. Piwinski. “The Touschek effect in strong focusing storage rings”. In: (Nov. 1998). arXiv: physics/9903034 (cit. on p. 22).

[42] P. M. Lewis et al. “First Measurements of Beam Backgrounds at SuperKEKB”. In: *Nucl. Instrum. Meth. A* 914 (2019), pp. 69–144. DOI: 10.1016/j.nima.2018.05.071. arXiv: 1802.01366 [physics.ins-det] (cit. on p. 22).

[43] K. Adamczyk et al. “The Belle II silicon vertex detector assembly and mechanics”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 845 (2017). Proceedings of the Vienna Conference on Instrumentation 2016, pp. 38–42. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.03.100. URL: https://www.sciencedirect.com/science/article/pii/S016890021630136X (cit. on p. 24).

[44] J. Kemmer and G. Lutz. “New detector concepts”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 253.3 (1987), pp. 365–377. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90518-3. URL: https://www.sciencedirect.com/science/article/pii/0168900287905183 (cit. on p. 24).

[45] P. Wieduwilt et al. “Performance of production modules of the Belle II pixel detector in a high-energy particle beam”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 991 (2021), p. 164978. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164978. URL: https://www.sciencedirect.com/science/article/pii/S0168900220313759 (cit. on p. 25).

[46] Christoph Schwanda et al. “Performance of the Belle II Silicon Vertex Detector”. In: *PoS* Vertex2019 (2020), p. 014. DOI: 10.22323/1.373.0014 (cit. on p. 26).

[47] K. Adamczyk et al. “The Design, Construction, Operation and Performance of the Belle II Silicon Vertex Detector”. In: (Jan. 2022). arXiv: 2201.09824 [physics.ins-det] (cit. on p. 26).

[48] “Letter of intent for KEK Super *B* Factory”. In: (June 2004). Ed. by S. Hashimoto et al. (cit. on p. 27).

[49] K. Inami. “Development of a TOP counter for the Super B factory”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 595.1 (2008). RICH 2007, pp. 96–99. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.07.045. URL: https://www.sciencedirect.com/science/article/pii/S0168900208009510 (cit. on p. 29).

[50] T. Iijima et al. “A novel type of proximity focusing RICH counter with multiple refractive index aerogel radiator”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 548.3 (Aug. 2005), pp. 383–390. DOI: 10.1016/j.nima.2005.05.030. URL: https://doi.org/10.1016%2Fj.nima.2005.05.030 (cit. on pp. 29, 30).

[51] Kodai Matsuoka et al. “Performance of the MCP-PMT for the Belle II TOP counter”. In: *PoS* PhotoDet2015 (2016), p. 028. DOI: 10.22323/1.252.0028 (cit. on p. 29).

[52] M. Starič et al. “Likelihood analysis of patterns in a time-of-propagation (TOP) counter”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 595.1 (2008). RICH 2007, pp. 252–255. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.07.018. URL: https://www.sciencedirect.com/science/article/pii/S0168900208009893 (cit. on pp. 29, 30).

[53] Belle II Lepton ID Group. “Muon and electron identification efficiencies and hadron-lepton mis-identification probabilities with Belle II Phase III data”. In: (Apr. 2021) (cit. on pp. 29, 57).

[54] Peter Križan, Samo Korpar, and Toru Iijima. “Study of a nonhomogeneous aerogel radiator in a proximity focusing RICH detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 565.2 (2006), pp. 457–462. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2006.05.

233. URL: https://www.sciencedirect.com/science/article/pii/S0168900206010229 (cit. on p. 30).

[55] I. Adachi et al. "Study of highly transparent silica aerogel as a RICH radiator". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 553.1 (2005). Proceedings of the fifth International Workshop on Ring Imaging Detectors, pp. 146–151. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2005.08.022. URL: https://www.sciencedirect.com/science/article/pii/S0168900205015895 (cit. on p. 30).

[56] T Matsumoto et al. "Studies of proximity focusing RICH with an aerogel radiator using flat-panel multi-anode PMTs (Hamamatsu H8500)". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 521.2 (2004), pp. 367–377. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2003.11.384. URL: https://www.sciencedirect.com/science/article/pii/S0168900203031243 (cit. on p. 30).

[57] T. Abe. "Belle II Technical Design Report". In: (2010). arXiv: 1011.0352 [physics.ins-det] (cit. on pp. 31, 33, 35).

[58] Wouter Verkerke and David Kirkby. "The RooFit Toolkit for Data Modeling". In: 2003 (cit. on p. 32).

[59] A. Abashian et al. "The Belle Detector". In: *Nucl. Instrum. Meth. A* 479 (2002), pp. 117–232. DOI: 10.1016/S0168-9002(01)02013-7 (cit. on pp. 32, 34).

[60] A Abashian et al. "Muon identification in the Belle experiment at KEKB". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 491.1 (2002), pp. 69–82. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(02)01164-6. URL: https://www.sciencedirect.com/science/article/pii/S0168900202011646 (cit. on p. 33).

[61] Yoshihito Iwasaki et al. "Level 1 trigger system for the Belle II experiment". In: *IEEE Trans. Nucl. Sci.* 58 (2011). Ed. by Sascha Marc Schmeling, pp. 1807–1815. DOI: 10.1109/TNS.2011.2119329 (cit. on p. 34).

[62] S Lee et al. "Development of High Level Trigger Software for Belle II at SuperKEKB". In: *Journal of Physics: Conference Series* 331.2 (Dec. 2011), p. 022015. DOI: 10.1088/1742-6596/331/2/022015. URL: https://doi.org/10.1088/1742-6596/331/2/022015 (cit. on p. 34).

[63] W. Altmannshofer et al. “The Belle II Physics Book”. In: *PTEP* 2019.12 (2019). Ed. by E. Kou and P. Urquijo. [Erratum: PTEP 2020, 029201 (2020)], p. 123C01. DOI: 10.1093/ptep/ptz106. arXiv: 1808.10567 [hep-ex] (cit. on p. 39).

[64] Marco Milesi, Daniel Ferlewicz, and Phillip Urquijo. “Lepton identification in Belle II based on Boosted Decision Trees.” In: (Sept. 2019) (cit. on pp. 39, 62).

[65] Coadou, Yann. “Boosted Decision Trees and Applications”. In: *EPJ Web of Conferences* 55 (2013), p. 02004. DOI: 10.1051/epjconf/20135502004. URL: https://doi.org/10.1051/epjconf/20135502004 (cit. on p. 40).

[66] Jerome H. Friedman. “Greedy function approximation: A gradient boosting machine.” In: *The Annals of Statistics* 29.5 (2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451. URL: https://doi.org/10.1214/aos/1013203451 (cit. on p. 40).

[67] Jan Therhaag. “TMVA Toolkit for multivariate data analysis in ROOT”. In: *PoS* ICHEP2010 (2010). Ed. by Bernard Pire et al., p. 510. DOI: 10.22323/1.120.0510 (cit. on pp. 40, 42).

[68] N.E. Cotter. “The Stone-Weierstrass theorem and its application to neural networks”. In: *IEEE Transactions on Neural Networks* 1.4 (1990), pp. 290–295. DOI: 10.1109/72.80265 (cit. on p. 44).

[69] A. J. Bevan et al. “The Physics of the B Factories”. In: *Eur. Phys. J. C* 74 (2014), p. 3026. DOI: 10.1140/epjc/s10052-014-3026-9. arXiv: 1406.6311 [hep-ex] (cit. on pp. 46, 47).

[70] von F. Zernike. “Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode”. In: *Physica* 1.7 (1934), pp. 689–704. ISSN: 0031-8914. DOI: https://doi.org/10.1016/S0031-8914(34)80259-5. URL: https://www.sciencedirect.com/science/article/pii/S0031891434802595 (cit. on p. 47).

[71] A. B. Bhatia and E. Wolf. “On the circle polynomials of Zernike and related orthogonal sets”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 50.1 (1954), pp. 40–48. DOI: 10.1017/S0305004100029066 (cit. on p. 47).

[72] Alireza Khotanzad and Yaw Hua Hong. “Invariant Image Recognition by Zernike Moments”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (1990), pp. 489–497 (cit. on p. 47).

[73] Torben Ferber, Chris Hearty, and Mike Roney. “Design of the ECL Software for Belle II”. In: (Jan. 2016) (cit. on p. 47).

[74] Torben Ferber. "Measurement of the ECL trigger efficiency for Experiment 3". In: (Dec. 2018) (cit. on p. 48).

[75] DN Brown, J Ilic, and GB Mohanty. "Extracting longitudinal shower development information from crystal calorimetry plus tracking". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 592.3 (2008), pp. 254–260 (cit. on p. 48).

[76] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 50).

[77] T. Kuhr et al. "The Belle II Core Software". In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 1. DOI: 10.1007/s41781-018-0017-9. arXiv: 1809.04299 [physics.comp-ph] (cit. on pp. 50, 61).

[78] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl.Instrum.Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on pp. 50, 61).

[79] S. Longo et al. "CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 982 (2020), p. 164562. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164562. URL: https://www.sciencedirect.com/science/article/pii/S0168900220309591 (cit. on p. 56).

[80] N. Taniguchi. "Central Drift Chamber for Belle-II". In: *Journal of Instrumentation* 12.06 (June 2017), p. C06014. DOI: 10.1088/1748-0221/12/06/C06014. URL: https://dx.doi.org/10.1088/1748-0221/12/06/C06014 (cit. on p. 59).

[81] Gianluca Inguglia Rajesh Kumar Maiti. "Lepton - Pion Identification (binary classification) using Boosted Decision Trees (BDT) at Central Drift Chamber (CDC) of Belle II". In: (Sept. 2021) (cit. on p. 70).

[82] I. Adachi et al. "Search for an Invisibly Decaying Z' Boson at Belle II in $e^+e^- \to \mu^+\mu^-(e^\pm\mu^\mp)$ Plus Missing Energy Final States". In: *Phys. Rev. Lett.* 124.14 (2020), p. 141801. DOI: 10.1103/PhysRevLett.124.141801. arXiv: 1912.11276 [hep-ex] (cit. on p. 71).

[83] Michel Bertemes et al. "Search for Dark Higgsstrahlung in $e^+e^- \to \mu^+\mu^-$ and missing energy final states with the Belle II experiment". In: (June 2020) (cit. on p. 71).

[84] Michel Bertemes, Gianluca Inguglia, and Ilya Komarov. "Monitor of CDC trigger performance for low multiplicity events in Phase 3 data". In: (Aug. 2019) (cit. on pp. 71, 77).

[85] Marcello Campajola and Enrico Graziani. "Measurement of the cdcklm1 trigger performances with Experiment 12 Data". In: (Oct. 2020) (cit. on p. 71).

[86] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (2014), p. 79. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph] (cit. on pp. 82, 149).

[87] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (July 2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079. URL: http://dx.doi.org/10.1007/JHEP07(2014)079 (cit. on p. 82).

[88] Frits A. Berends, P. H. Daverveldt, and R. Kleiss. "Complete lowest-order calculations for four-lepton final states in electron-positron collisions". In: *Nuclear Physics B* 253 (1985), pp. 441–463. DOI: 10.1016/0550-3213(85)90541-3 (cit. on pp. 83, 151).

[89] S. Jadach, B. F. L. Ward, and Z. Wąs. "The precision Monte Carlo event generator KK for two-fermion final states in e^+e^- collisions". In: *Computer Physics Communications* 130.3 (2000), pp. 260–325. DOI: 10.1016/S0010-4655(00)00048-5. arXiv: hep-ph/9912214 [hep-ph] (cit. on pp. 83, 151).

[90] N. Davidson et al. "Universal interface of TAUOLA: Technical and physics documentation". In: *Computer Physics Communications* 183.3 (2012), pp. 821–843. DOI: 10.1016/j.cpc.2011.12.009. arXiv: 1002.0543 [hep-ph] (cit. on pp. 83, 151).

[91] H. Czyż, M. Gunia, and J. H. Kühn. "Simulation of electron-positron annihilation into hadrons with the event generator PHOKHARA". In: *Journal of High Energy Physics* 2013.8 (2013), p. 110. DOI: 10.1007/JHEP08(2013)110. arXiv: 1306.1985 [hep-ph] (cit. on pp. 83, 151).

[92] Torben Ferber and Phillip Urquijo. *Overview of the Belle II Physics Generators*. BELLE2-NOTE-PH-2015-006. 2015 (cit. on pp. 83, 151).

[93] Wouter Verkerke and David P. Kirkby. "The RooFit toolkit for data modeling". In: *eConf* C0303241 (2003), MOLT007. arXiv: physics/0306116 [physics] (cit. on pp. 83, 150).

[94] Giovanni Punzi. "Sensitivity of searches for new signals and its optimization". In: *eConf* C030908 (2003). Ed. by L. Lyons, R. P. Mount, and R. Reitmeyer, MODT002. arXiv: physics/0308063 (cit. on pp. 85, 100).

[95] Torben Ferber. `OrcaKinFit`*: Kinematic Fitting for Belle II*. BELLE2-NOTE-PH-2017-002. 2017 (cit. on p. 106).

[96] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an] (cit. on p. 114).

[97] Eilam Gross and Ofer Vitells. "Trial factors for the look elsewhere effect in high energy physics". In: *Eur. Phys. J. C* 70 (2010), pp. 525–530. DOI: 10.1140/epjc/s10052-010-1470-8. arXiv: 1005.1891 [physics.data-an] (cit. on p. 128).

[98] Rajesh Kumar Maiti et al. "Search for a visibly decaying Z' dark boson at in $e^+e^- \to \mu^+\mu^-\mu^+\mu^-$ events". In: (May 2022) (cit. on p. 146).

[99] P. Feichtinger et al. "Punzi-loss:: a non-differentiable metric approximation for sensitivity optimisation in the search for new particles". In: *Eur. Phys. J. C* 82.2 (2022), p. 121. DOI: 10.1140/epjc/s10052-022-10070-0. arXiv: 2110.00810 [hep-ex] (cit. on p. 147).

[100] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4 (cit. on p. 149).

[101] J. Allison et al. "Recent developments in Geant4". In: *Nucl. Instrum. Meth. A* 835 (2016), pp. 186–225. DOI: 10.1016/j.nima.2016.06.125 (cit. on p. 150).

Acknowledgements

Firstly, I would like to express my deepest gratitude and sincerest thanks to my thesis supervisors, Dr. Gianluca Inguglia and Dr. Christoph Schwanda, for their invaluable guidance, unfailing encouragement, taking pains in clearing many things, and for their full cooperation. It's a wonderful journey with them. They are very nice persons. I never heard a word from Gianluca you can't do it; instead "may be", "might be" you could do this, you could try this... :-) Thanks for believing me.

Thanks to all the Belle II HEPHY group members: Abdul, Gelandine, Huw, Paul, Peter, Nadia, Daniel, and Phillip. The dedication and passion you guys put into everything really motivate me to give the best out of myself. You guys are amazing.

Special thanks to Michel Bertemes, Suman Chatterjee, and Souvik Maity for being a part of this journey and providing valuable suggestions in every aspect. You guys are really an inspiration to go further. Thanks for everything.

Thanks to Dr. Enrico Graziani for sharing your efforts, support, and sound expertise and giving a perfect shape to the analysis.

I sincerely thank Belle II Collaboration for providing me necessary computing tools and resources to carry out this analysis. I would also like to thank the Dark Sector working group and the LeptonID group for the fruitful discussions about the analysis.

I am very thankful to the FWF for supporting my Ph.D. work under grant no. P31361-N36. I am also grateful to HEPHY Vienna for providing me necessary office spaces, resources, and infrastructure to carry out this project.

Finally, how could I forget my loved ones..., especially my mama, my dad, my brother, and Soumi, thanks for believing in me and letting me fulfill my dream. I am extremely grateful for your love and support. No word in the dictionary could describe your appreciation.

every ending is a start of new beginning,
thank you everyone.

A. Appendix

A.1. Normalized variables

Go back to the section 3.5.

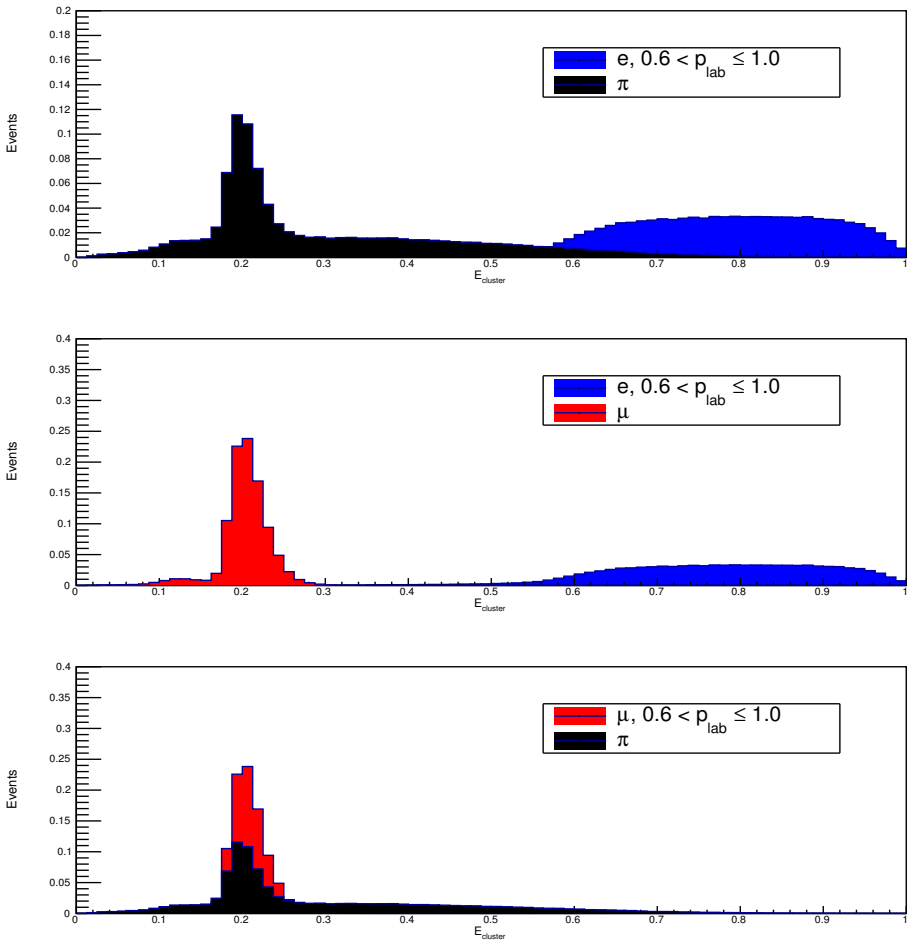


Figure A.1.: .

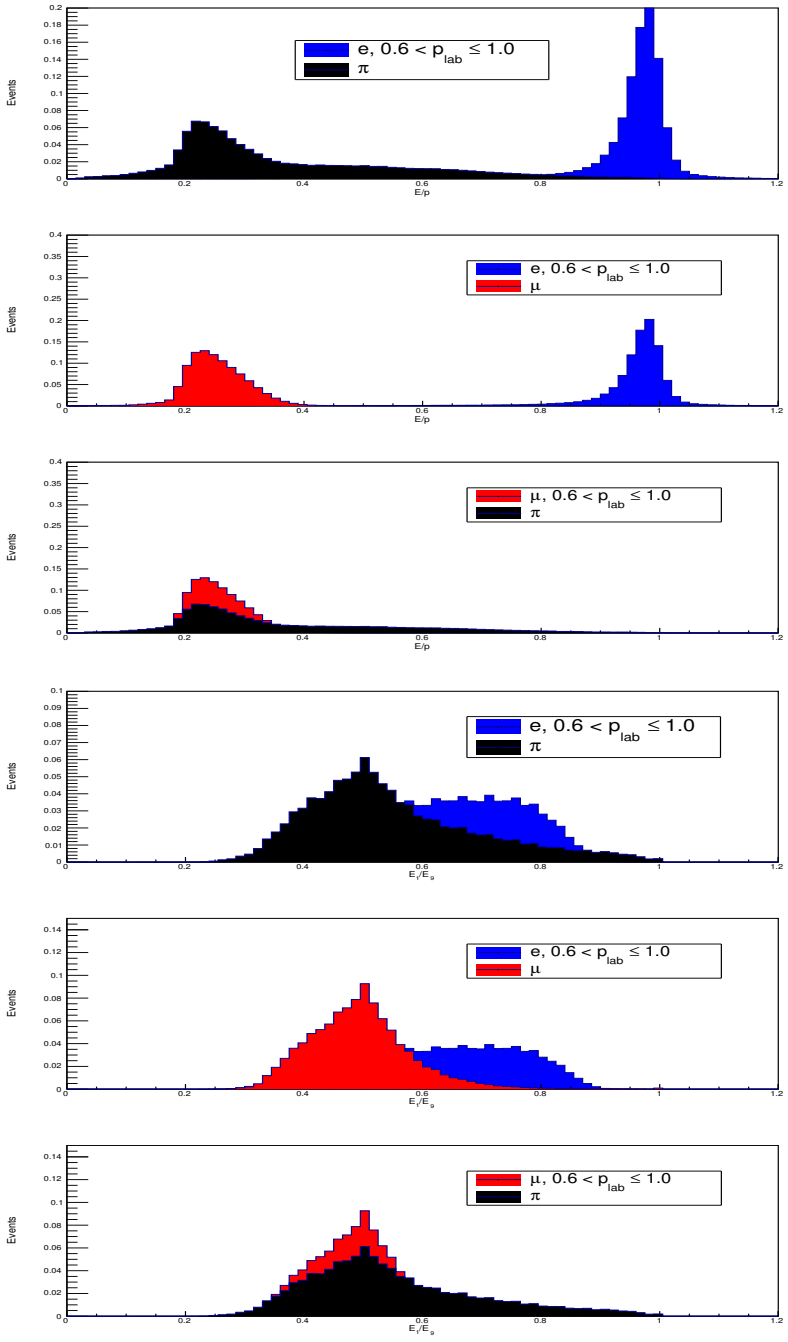


Figure A.2.: .

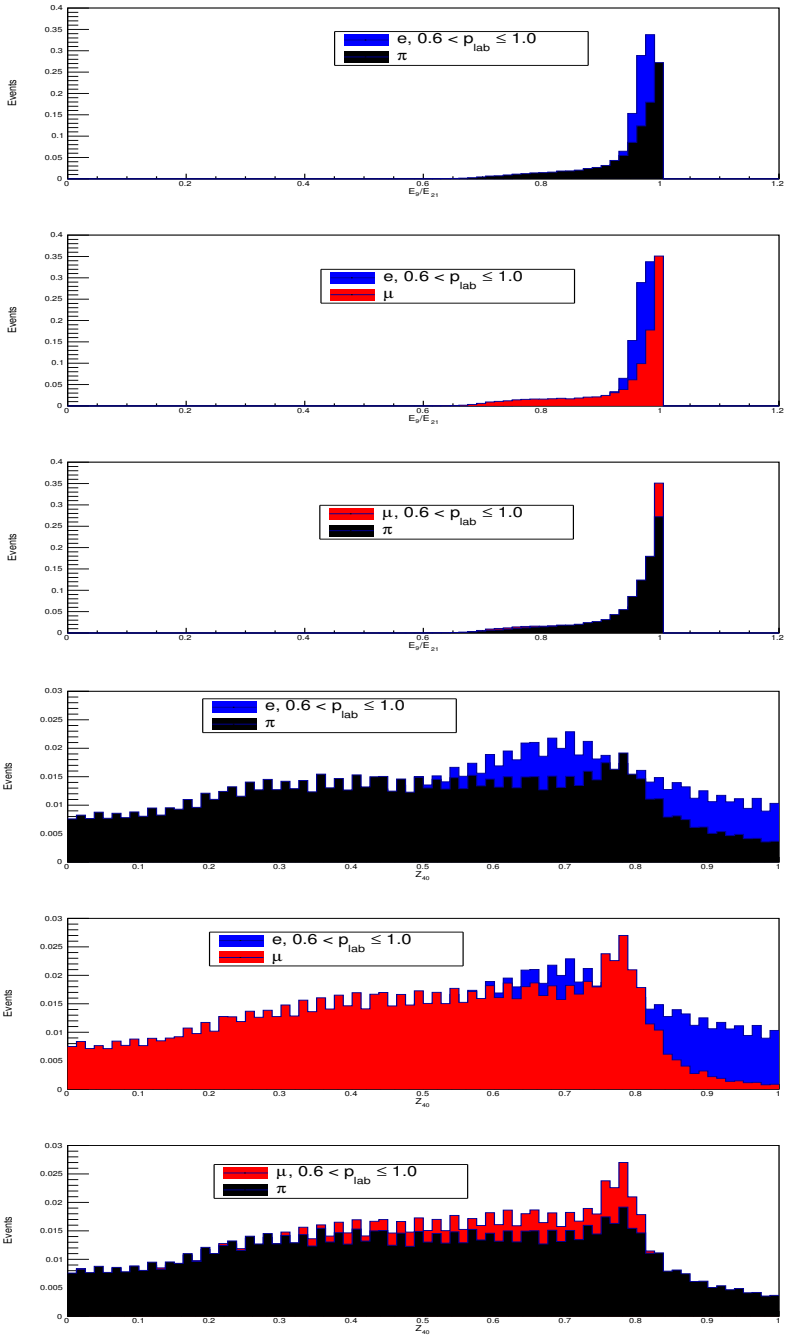


Figure A.3.: .

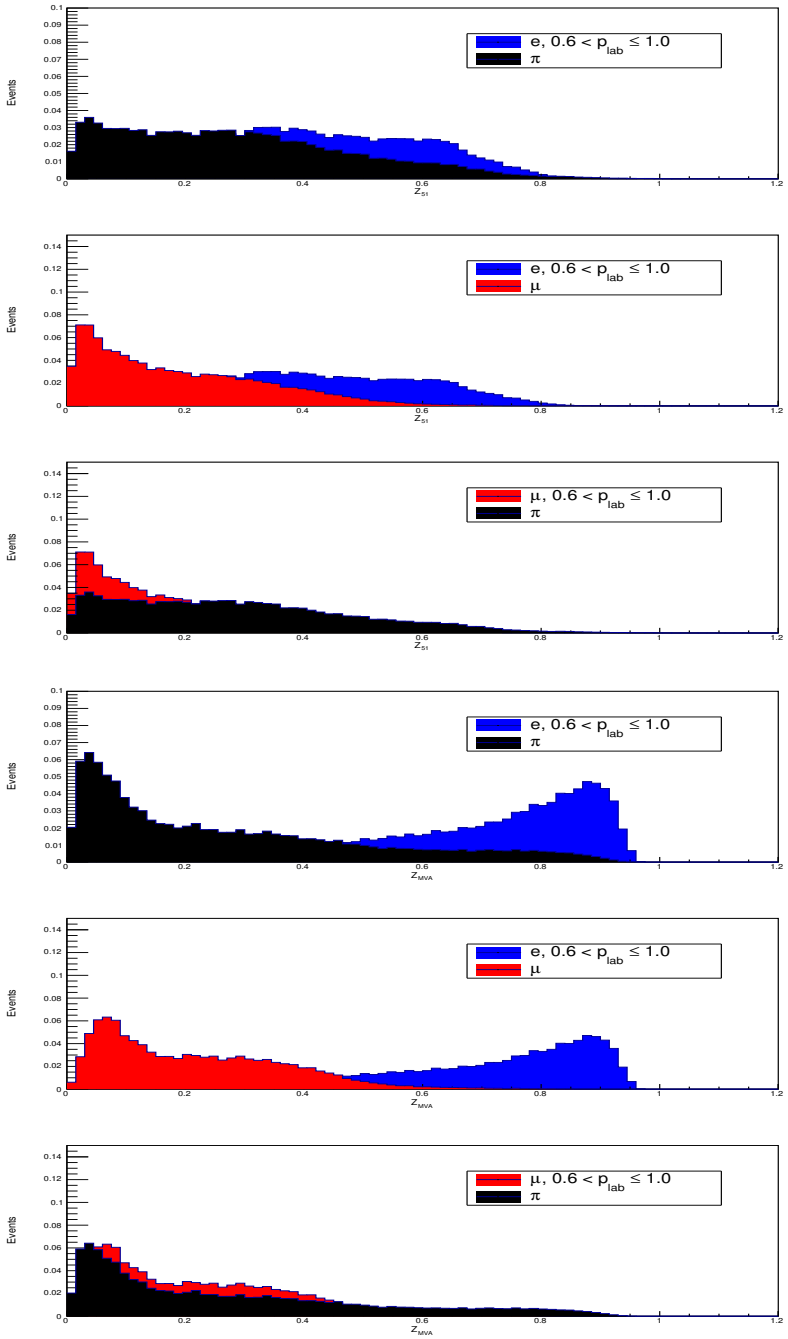


Figure A.4.: .

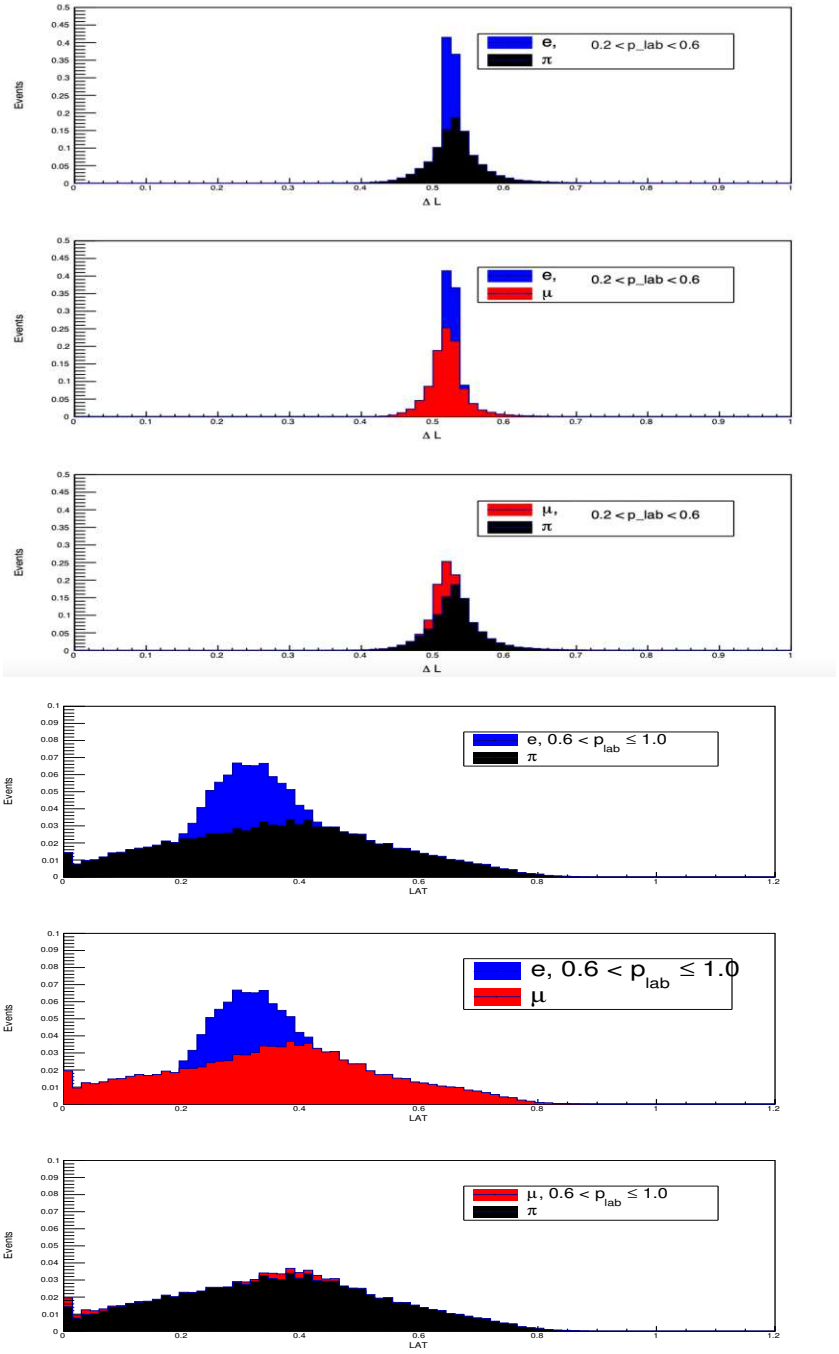


Figure A.5.: .

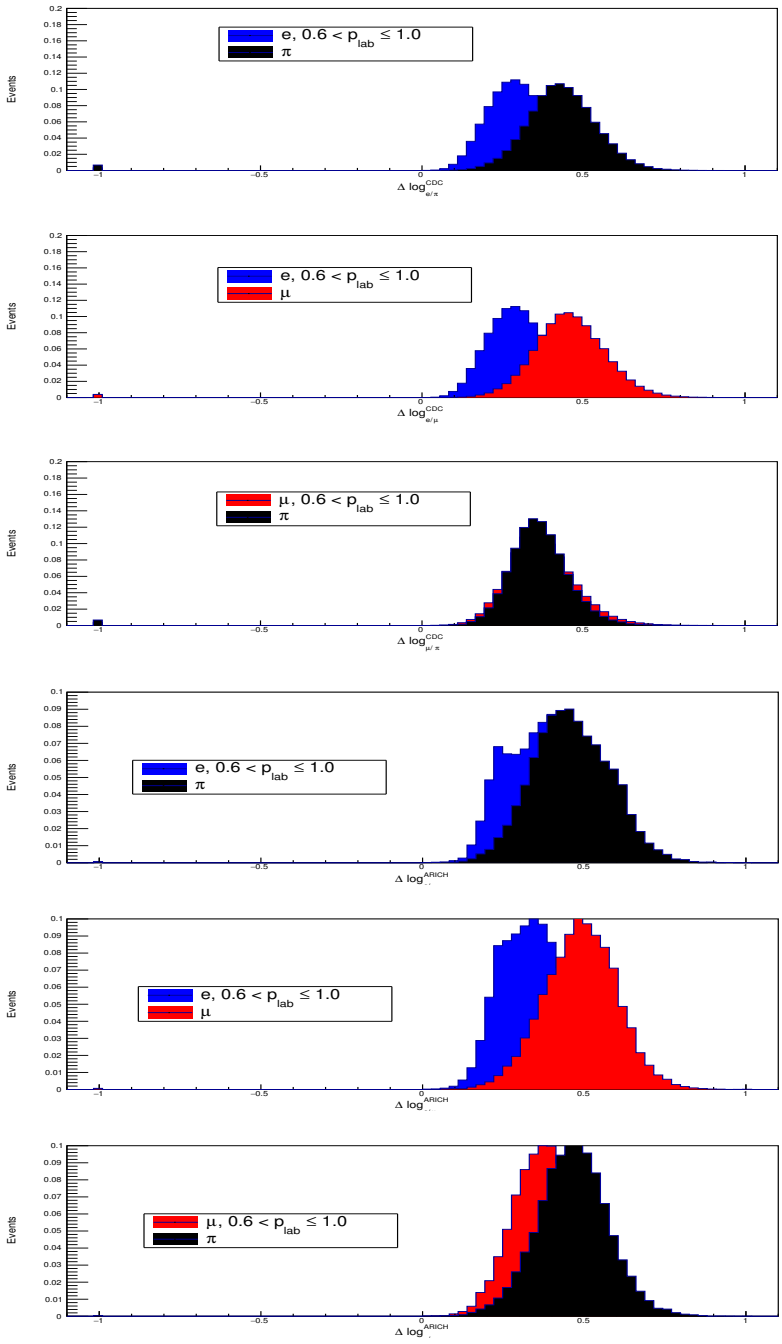


Figure A.6.: .

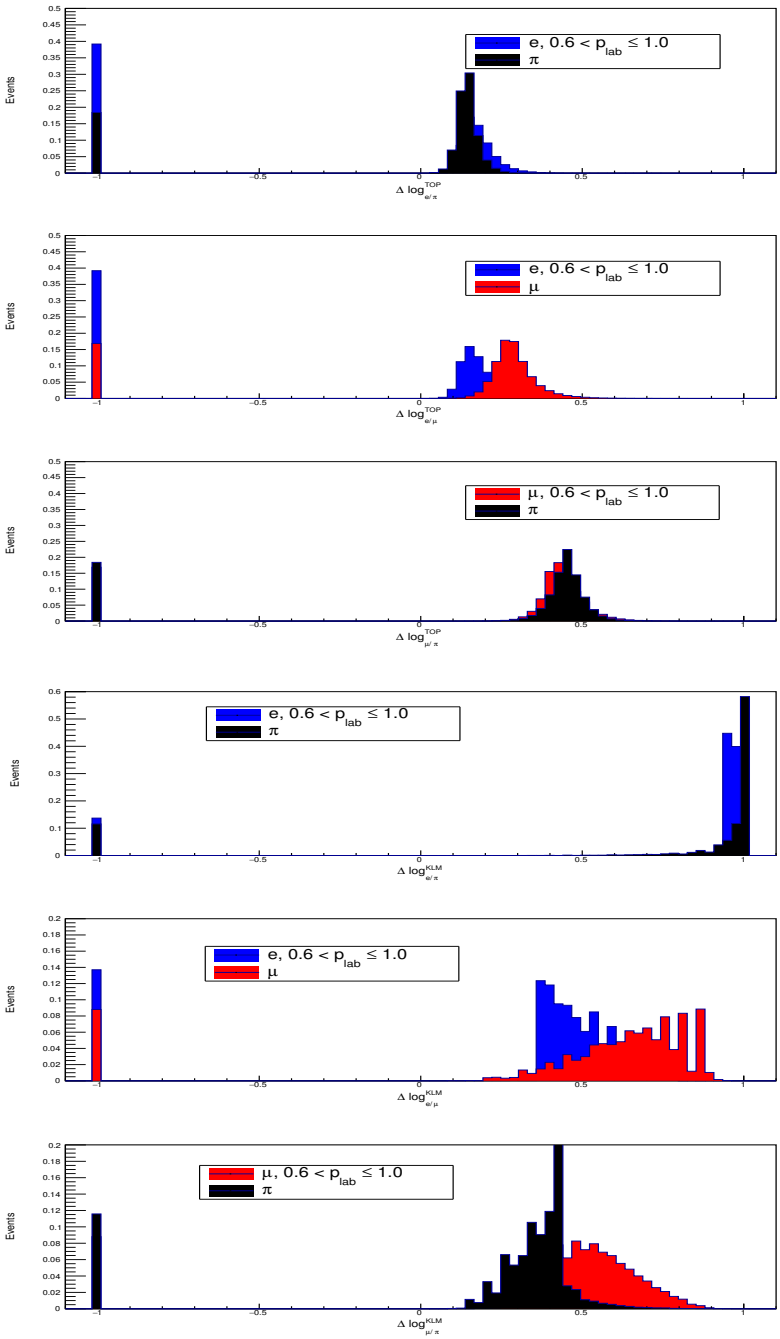


Figure A.7.: .

A.2. Variable ranking

Go back to the section 3.7

Momentum	$\theta_{cluster}$	$e-\pi$ MLP	$\mu-\pi$ MLP	$e-\mu$ MLP	$e-\pi$ BDT	$\mu-\pi$ BDT	$e-\mu$ BDT
‘ $0.2 < p_{lab} < 0.6$	ECL FWD	E/p $E_{cluster}$ $\Delta log\mathcal{L}_{ARICH}$ E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ Δ L $\mid Z_{MVA} \mid$ LAT $\mid Z_{40} \mid$ $\mid Z_{51} \mid$ E_1/E_9	$E_{cluster}$ $\Delta log\mathcal{L}_{CDC}$ E/p E_9/E_{21} $\Delta log\mathcal{L}_{ARICH}$ $\mid Z_{MVA} \mid$ Δ L LAT $\mid Z_{40} \mid$ E_1/E_9 $\mid Z_{51} \mid$	$E_{cluster}$ E/p E_9/E_{21} $\Delta log\mathcal{L}_{ARICH}$ $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{MVA} \mid$ LAT Δ L E_1/E_9 $\mid Z_{40} \mid$ $\mid Z_{51} \mid$	$\Delta log\mathcal{L}_{CDC}$ $E_{cluster}$ $\Delta log\mathcal{L}_{ARICH}$ $\mid Z_{40} \mid$ E/p LAT $\mid Z_{MVA} \mid$ Δ L E_1/E_9 $\mid Z_{51} \mid$ E_9/E_{21}		
$0.2 < p_{lab} < 0.6$	ECL Barrel	$E_{cluster}$ E/p $\Delta log\mathcal{L}_{TOP}$ E_9/E_{21} $\mid Z_{MVA} \mid$ $\Delta log\mathcal{L}_{CDC}$ Δ L E_1/E_9 $\mid Z_{40} \mid$ LAT $\mid Z_{51} \mid$	$E_{cluster}$ E/p $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{MVA} \mid$ $\Delta log\mathcal{L}_{TOP}$ E_9/E_{21} $\mid Z_{40} \mid$ LAT $\mid Z_{51} \mid$ Δ L E_1/E_9	E/p $E_{cluster}$ $\mid Z_{MVA} \mid$ $\Delta log\mathcal{L}_{TOP}$ E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{51} \mid$ $\mid Z_{40} \mid$ E_1/E_9 LAT Δ L	$\Delta log\mathcal{L}_{CDC}$ $E_{cluster}$ Δ L E/p $\mid Z_{MVA} \mid$ $\Delta log\mathcal{L}_{TOP}$ $\mid Z_{40} \mid$ $\mid Z_{51} \mid$ E_1/E_9 LAT E_9/E_{21}		
$0.2 < p_{lab} < 0.6$	ECL BWD	$E_{cluster}$ E/p $\Delta log\mathcal{L}_{CDC}$ E_9/E_{21} E_1/E_9 $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ LAT $\mid Z_{51} \mid$ Δ L	$E_{cluster}$ E/p $\Delta log\mathcal{L}_{CDC}$ E_9/E_{21} $\mid Z_{MVA} \mid$ Δ L LAT $\mid Z_{40} \mid$ E_1/E_9 $\mid Z_{51} \mid$	$E_{cluster}$ E/p $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{MVA} \mid$ E_9/E_{21} $\mid Z_{51} \mid$ E_1/E_9 $\mid Z_{40} \mid$ Δ L LAT	E/p $\Delta log\mathcal{L}_{CDC}$ LAT $E_{cluster}$ $\mid Z_{51} \mid$ $\mid Z_{MVA} \mid$ E_1/E_9 Δ L $\mid Z_{40} \mid$ E_9/E_{21}		

Table A.1.: Low momentum features ranking.

Momentum	$\theta_{cluster}$	$e-\pi$ MLP	$\mu-\pi$ MLP	$e-\mu$ MLP	$e-\pi$ BDT	$\mu-\pi$ BDT	$e-\mu$ BDT
$0.6 < p_{lab} < 1.0$	ECL FWD	E/p $E_{cluster}$ $\Delta log\mathcal{L}_{ARICH}$ E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ LAT $\mid Z_{MVA} \mid$ $\mid Z_{51} \mid$ $\mid Z_{40} \mid$ Δ L E_1/E_9	E/p $E_{cluster}$ $\Delta log\mathcal{L}_{ARICH}$ E_9/E_{21} LAT Δ L $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ $\mid Z_{51} \mid$ $\Delta log\mathcal{L}_{CDC}$ E_1/E_9	E/p $E_{cluster}$ E_9/E_{21} $\Delta log\mathcal{L}_{ARICH}$ LAT $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{MVA} \mid$ Δ L $\mid Z_{40} \mid$ E_1/E_9 $\mid Z_{51} \mid$	$E_{cluster}$ E/p $\Delta log\mathcal{L}_{CDC}$ LAT Δ L $\Delta log\mathcal{L}_{ARICH}$ $\mid Z_{51} \mid$ E_1/E_9 $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ E_9/E_{21}		
$0.6 < p_{lab} < 1.0$	ECL Barrel	E/p $E_{cluster}$ $\Delta log\mathcal{L}_{TOP}$ E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{MVA} \mid$ LAT $\mid Z_{40} \mid$ Δ L $\Delta log\mathcal{L}_{KLM}$ E_1/E_9 $\mid Z_{51} \mid$	$\Delta log\mathcal{L}_{KLM}$ $E_{cluster}$ E/p E_9/E_{21} Δ L $\mid Z_{51} \mid$ $\Delta log\mathcal{L}_{TOP}$ $\mid Z_{MVA} \mid$ $\Delta log\mathcal{L}_{CDC}$ E_1/E_9 $\mid Z_{40} \mid$ LAT	$E_{cluster}$ E/p $\Delta log\mathcal{L}_{KLM}$ E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ Δ L $\Delta log\mathcal{L}_{TOP}$ $\mid Z_{51} \mid$ $\mid Z_{MVA} \mid$ E_1/E_9 $\mid Z_{40} \mid$ LAT	$E_{cluster}$ E/p $\Delta log\mathcal{L}_{CDC}$ Δ L $\Delta log\mathcal{L}_{TOP}$ LAT $\Delta log\mathcal{L}_{KLM}$ $\mid Z_{MVA} \mid$ E_1/E_9 $\mid Z_{51} \mid$ $\mid Z_{40} \mid$ E_9/E_{21}		
$0.6 < p_{lab} < 1.0$	ECL BWD	Δ L E_9/E_{21} E/p $\Delta log\mathcal{L}_{KLM}$ $E_{cluster}$ $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ $\mid Z_{51} \mid$ LAT E_1/E_9	E_9/E_{21} $\Delta log\mathcal{L}_{KLM}$ $E_{cluster}$ Δ L E/p $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ $\Delta log\mathcal{L}_{CDC}$ LAT $\mid Z_{51} \mid$ E_1/E_9	$E_{cluster}$ E/p $\Delta log\mathcal{L}_{KLM}$ Δ L E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{MVA} \mid$ LAT E_1/E_9 $\mid Z_{40} \mid$ $\mid Z_{51} \mid$	Δ L $\Delta log\mathcal{L}_{CDC}$ E/p LAT E_9/E_{21} $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ $E_{cluster}$ $\Delta log\mathcal{L}_{KLM}$ $\mid Z_{51} \mid$ E_1/E_9		

Table A.2.: Medium momentum features ranking.

Momentum	$\theta_{cluster}$	$e-\pi$ MLP	$\mu-\pi$ MLP	$e-\mu$ MLP	$e-\pi$ BDT	$\mu-\pi$ BDT	$e-\mu$ BDT
$p_{lab} > 1.0$	ECL FWD	E/p $E_{cluster}$ $\Delta log\mathcal{L}_{KLM}$ $\Delta log\mathcal{L}_{ARICH}$ E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ LAT $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ E_1/E_9 $\mid Z_{51} \mid$ Δ L	$\Delta log\mathcal{L}_{KLM}$ $E_{cluster}$ E/p $\Delta log\mathcal{L}_{ARICH}$ E_9/E_{21} LAT $\Delta log\mathcal{L}_{CDC}$ E_1/E_9 $\mid Z_{40} \mid$ $\mid Z_{MVA} \mid$ Δ L $\mid Z_{51} \mid$	E/p E_9/E_{21} $\Delta log\mathcal{L}_{KLM}$ $\Delta log\mathcal{L}_{ARICH}$ LAT $E_{cluster}$ E_1/E_9 $\mid Z_{51} \mid$ $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ Δ L	E/p $\Delta log\mathcal{L}_{CDC}$ $E_{cluster}$ $\Delta log\mathcal{L}_{KLM}$ $\Delta log\mathcal{L}_{ARICH}$ E_9/E_{21} Δ L E_1/E_9 $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ $\mid Z_{51} \mid$ LAT		
$p_{lab} > 1.0$	ECL Barrel	E/p $\Delta log\mathcal{L}_{TOP}$ $\Delta log\mathcal{L}_{CDC}$ $\Delta log\mathcal{L}_{KLM}$ $E_{cluster}$ E_9/E_{21} LAT $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ $\mid Z_{51} \mid$ E_1/E_9 Δ L	$\Delta log\mathcal{L}_{KLM}$ $\Delta log\mathcal{L}_{TOP}$ E/p $E_{cluster}$ E_9/E_{21} Δ L $\Delta log\mathcal{L}_{CDC}$ $\mid Z_{MVA} \mid$ LAT E_1/E_9 $\mid Z_{40} \mid$ $\mid Z_{51} \mid$	E/p $E_{cluster}$ $\Delta log\mathcal{L}_{KLM}$ $\Delta log\mathcal{L}_{CDC}$ E_9/E_{21} $\Delta log\mathcal{L}_{TOP}$ LAT $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ $\mid Z_{51} \mid$ E_1/E_9 Δ L	E/p $\Delta log\mathcal{L}_{CDC}$ $E_{cluster}$ $\Delta log\mathcal{L}_{KLM}$ $\Delta log\mathcal{L}_{TOP}$ $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ $\mid Z_{51} \mid$ E_1/E_9 LAT Δ L E_9/E_{21}		
$p_{lab} > 1.0$	ECL BWD	E/p $E_{cluster}$ E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ E_1/E_9 LAT $\mid Z_{MVA} \mid$ $\mid Z_{40} \mid$ Δ L $\Delta log\mathcal{L}_{KLM}$ $\mid Z_{51} \mid$	$\Delta log\mathcal{L}_{KLM}$ E/p $E_{cluster}$ $\Delta log\mathcal{L}_{CDC}$ Δ L LAT $\mid Z_{40} \mid$ E_1/E_9 $\mid Z_{51} \mid$ E_9/E_{21} $\mid Z_{MVA} \mid$	E/p E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ $\Delta log\mathcal{L}_{KLM}$ $E_{cluster}$ Δ L E_1/E_9 LAT $\mid Z_{40} \mid$ $\mid Z_{51} \mid$ $\mid Z_{MVA} \mid$	E/p LAT Δ L E_9/E_{21} $\Delta log\mathcal{L}_{CDC}$ E_1/E_9 LAT $\mid Z_{MVA} \mid$ $\Delta log\mathcal{L}_{KLM}$ $\mid Z_{51} \mid$ $\mid Z_{40} \mid$		

Table A.3.: high momentum features ranking.

B. Appendix

Go back to the section 4.5

B.1. Normalized variables

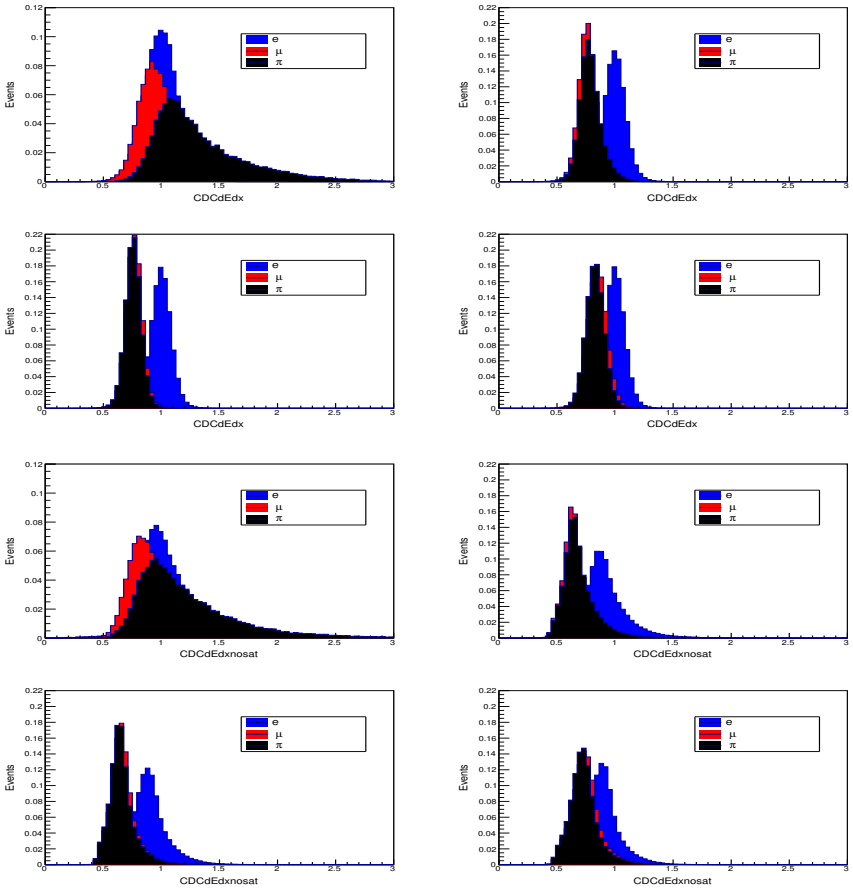


Figure B.1.: Upper and below four plots (upper left ⇒ upper right ⇒ below left ⇒ below right) are for $0.1 < p_{lab} \leq 0.2$, $0.2 < p_{lab} \leq 0.6$, $0.6 < p_{lab} \leq 1.0$ and $1.0 < p_{lab} \leq 5.0$.

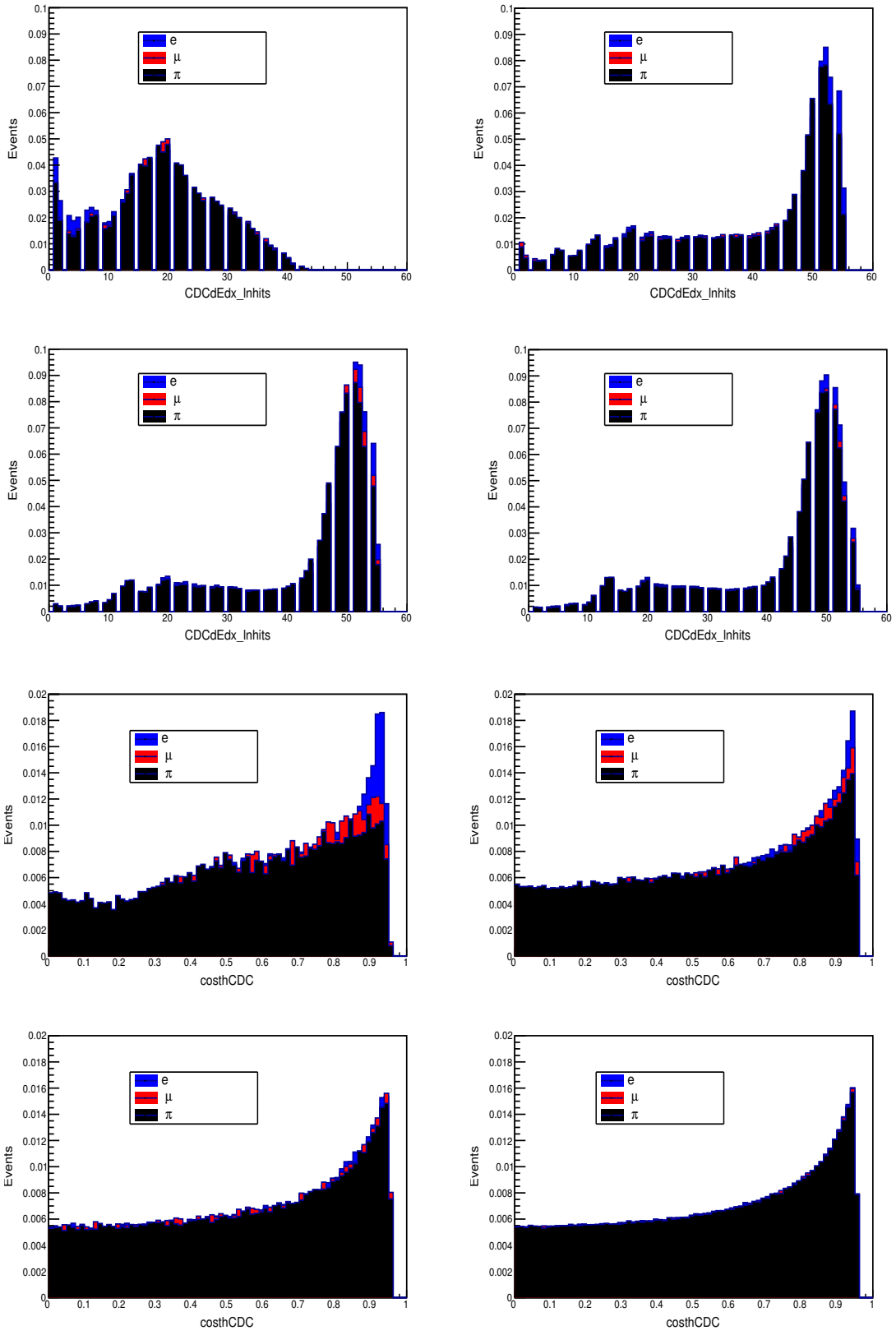


Figure B.2.: Upper and below four plots (upper left ⇒ upper right ⇒ below left ⇒ below right) are for $0.1 < p_{lab} \leq 0.2$, $0.2 < p_{lab} \leq 0.6$, $0.6 < p_{lab} \leq 1.0$ and $1.0 < p_{lab} \leq 5.0$.

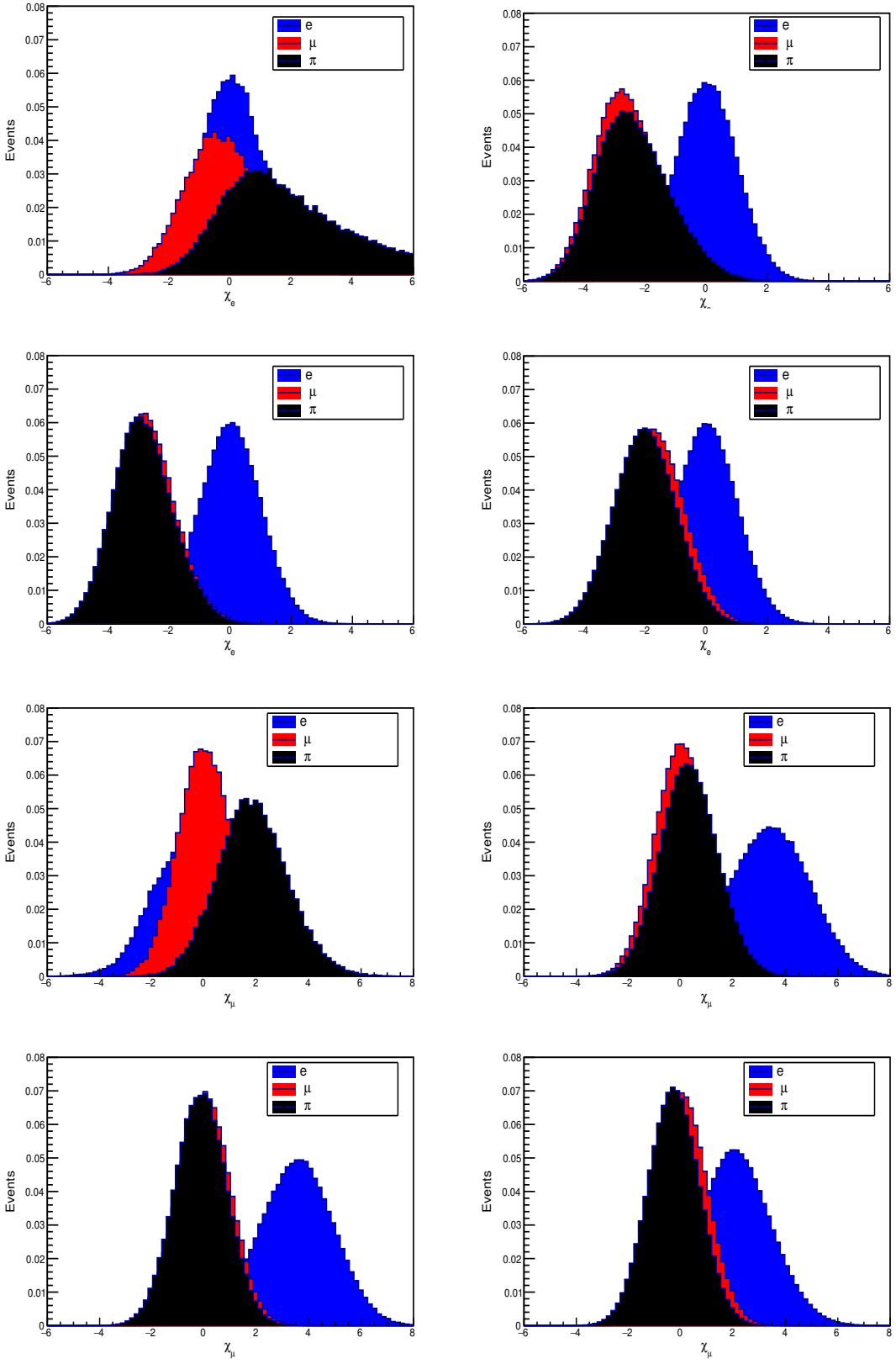


Figure B.3.: Upper and below four plots (upper left ⇒ upper right ⇒ below left ⇒ below right) are for $0.1 < p_{lab} \leq 0.2$, $0.2 < p_{lab} \leq 0.6$, $0.6 < p_{lab} \leq 1.0$ and $1.0 < p_{lab} \leq 5.0$.

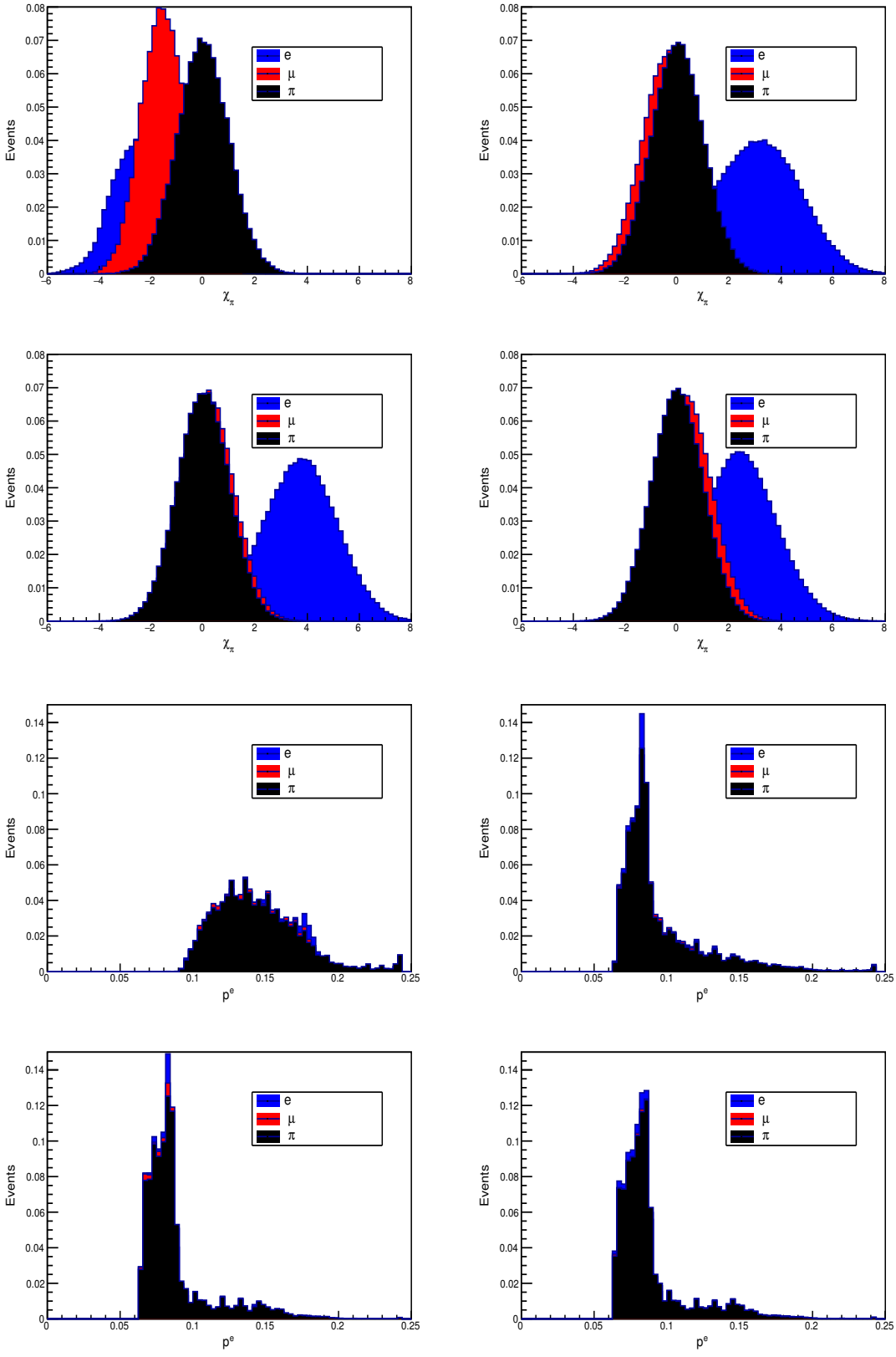


Figure B.4.: Upper and below four plots (upper left ⇒ upper right ⇒ below left ⇒ below right) are for $0.1 < p_{lab} \leq 0.2$, $0.2 < p_{lab} \leq 0.6$, $0.6 < p_{lab} \leq 1.0$ and $1.0 < p_{lab} \leq 5.0$.

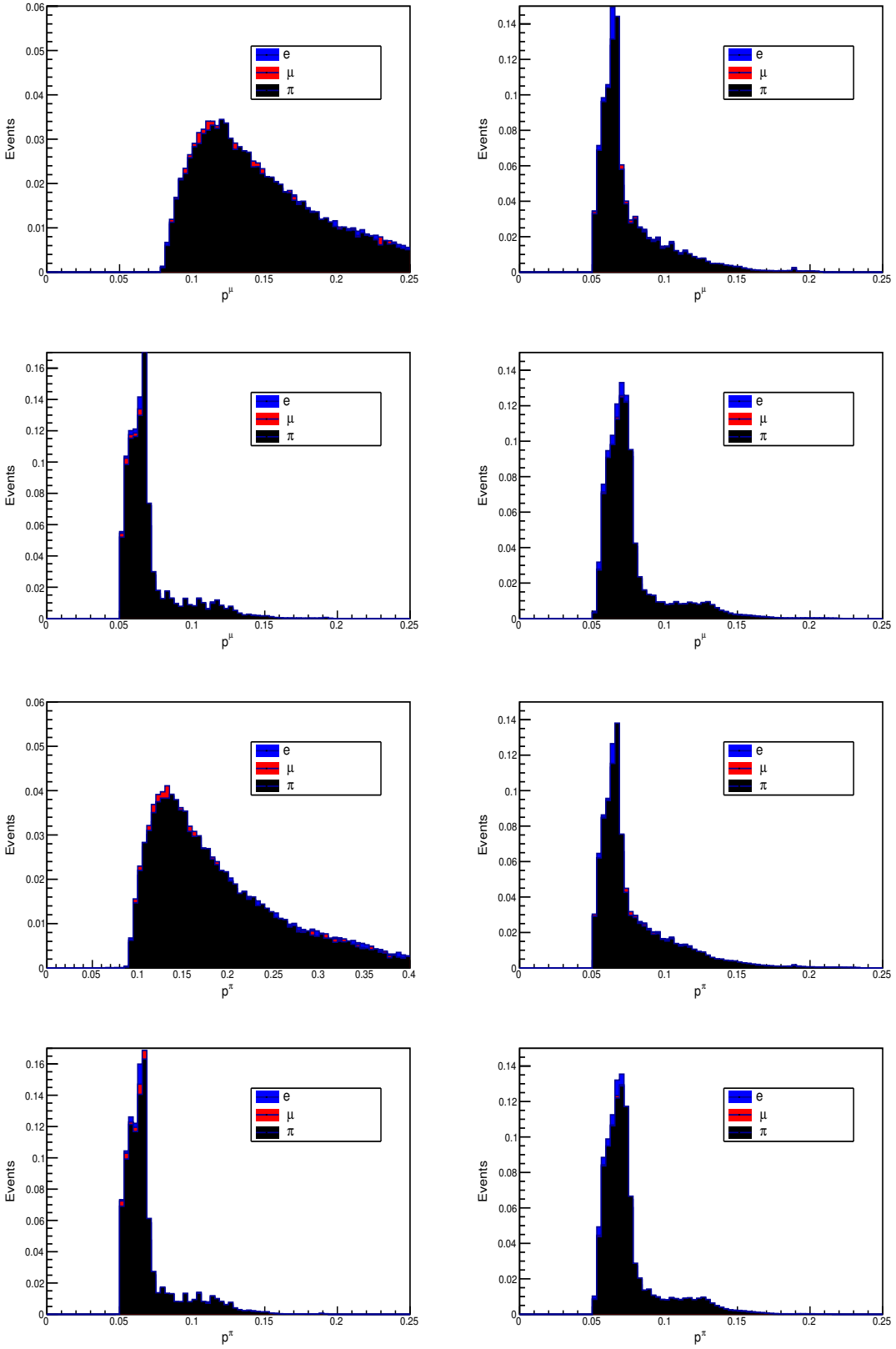


Figure B.5.: Upper and below four plots (upper left ⇒ upper right ⇒ below left ⇒ below right) are for $0.1 < p_{lab} \leq 0.2$, $0.2 < p_{lab} \leq 0.6$, $0.6 < p_{lab} \leq 1.0$ and $1.0 < p_{lab} \leq 5.0$.