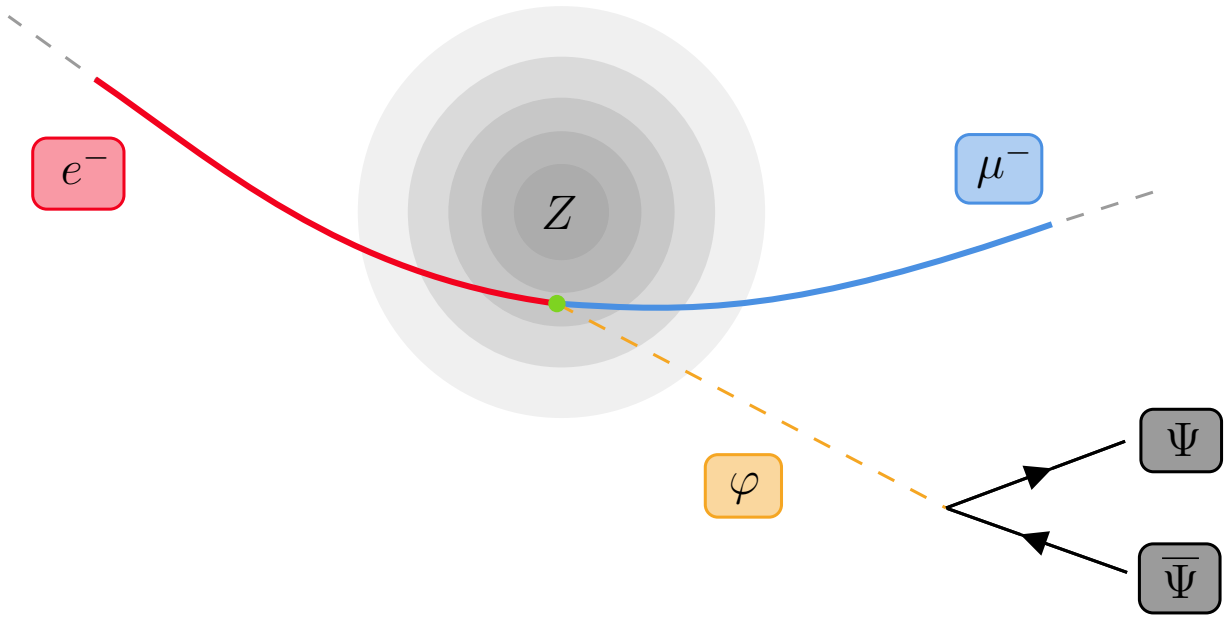
CHALMERS
UNIVERSITY OF TECHNOLOGY



Neural Network-based study on background for the Dark Leptonic Scalar model at NA64

A Deep Learning approach for optimization of dimuon selection with NA64 at CERN

Master's thesis in Physics

EMIL ZAYA

DEPARTMENT OF PHYSICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

Master's thesis 2024

Neural Network-based study on background for the Dark Leptonic Scalar model at NA64

A Deep Learning approach for optimization of dimuon selection with NA64 at CERN

Emil Zaya

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Physics
Division of Subatomic, High Energy and Plasma Physics
Theoretical Subatomic Physics
Chalmers University of Technology
Gothenburg, Sweden 2024

Neural Network-based study on background for the Dark Leptonic Scalar model at NA64
Emil Zaya

Supervisor: Paolo Crivelli, ETH Zürich
Examiner: Martin Cederwall, Chalmers University of Technology

Master's Thesis 2024
Department of Physics
Division of Subatomic, High Energy and Plasma Physics
Theoretical Subatomic Physics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: An illustration of a Brehmstrahlung-like process producing a dark leptonic scalar φ. The electron e^- (red line) approaches the heavy nucleus Z and is scattered. The electron then interacts with a dark scalar φ (yellow line), introducing a change in lepton flavor which leaves a negatively charged muon μ^- (blue line). The mediator particle φ then decays into a pair of fermionic dark matter. More information in section 3.1.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2024

Neural Network-based study on background for the Dark Leptonic Scalar model at NA64

Emil Zaya
Department of Physics
Chalmers University of Technology

Abstract

The search for a particle candidate that could explain the origin of dark matter is a central goal in modern astro-particle physics. Numerous experiments employing various measurement strategies are being developed to try and understand this elusive phenomenon. The NA64 experiment situated at the north area of CERN, utilizing the CERN Super Proton Synchrotron (SPS), is an active target experiment aiming to look for signatures like missing energies with hopes of finding signals that correspond to Dark Matter (DM) particles. These dark particles are modelled to explain the physical process of kinetic mixing between the Standard Model (SM) and the hypothesised corresponding Dark Sector (DS). The main purpose of this project is to study the background for a Dark Leptonic Scalar model (DLS) using a highly accurate Monte Carlo simulation for the NA64 experiment. More precisely, the `GEANT4` particle simulator was used for the NA64 experiment to simulate the results of the experimental setup used in 2023. The results of this was compared with real data taken in 2023, and a first step was benchmarking the simulation which was done by using dimuon ($\mu\mu$) events. Furthermore, the simulation results were used as a means of perfecting the methods of event selection. The main source of background for DLS particle φ are $\mu\mu$ production, kaon κ and pion π decay. The main purpose of this thesis is to produce a trained Neural Network (NN) model that can be used for optimizing the selection of events. The background for the DLS φ was simulated and trained on a NN for selecting $\mu\mu$ events as a means of benchmarking the method. The selection of $\mu\mu$ using a trained NN is compared to traditional methods of selection, where an increase of 36 % of the final state events is seen with the NN selected data. A future study could be to simulate the DLS φ particles and train them on a NN to use for event selection. The hopes are to gain a higher signal-to-background ratio and a larger amount of data for the DLS model.

Keywords: NA64, CERN, particle physics, simulation, GEANT4, Neural Network, Python, C++, beyond the standard model, dark matter

Acknowledgements

First and foremost, I would like to thank my supervisor, Paolo Crivelli, for having faith in me and for giving me the opportunity to work in the field of particle physics at CERN, a goal I've had since I started my journey in physics. I would also like to express my sincere gratitude to Martina Mongillo and Benjamin Oberhauser for all the guidance and support throughout this project. Furthermore, I want to thank the members of the NA64 collaboration, and in particular Laura Bueno, for all the many helpful discussions. And last but not least, I want to thank my family.

Emil Zaya, Gothenburg, July 2024

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ANN	**A**rtificial **N**eural **N**etwork
BCE	**B**inary **C**ross **E**ntropy
BSM	**B**eyond the **S**tandard **M**odel
CDM	**C**old **D**ark **M**atter
Λ-CDM	Lambda **C**old **D**ark **M**atter
DLS	**D**ark **L**eptonic **S**calar
DM	**D**ark **M**atter
DS	**D**ark **S**ector
ECAL	**E**lectromagnetic **CAL**orimeter
EOT	**E**lectron **O**n **T**arget
GEANT4	**GE**ometry **AN**d **T**racking
HCAL	**H**adronic **CAL**orimeter
LDM	**L**ight **D**ark **M**atter
LFC	**L**epton **F**lavour **C**hanging
MC	**M**onte **C**arlo
MIP	**M**inimum **I**onizing **P**article
MM	**M**icro-**M**ega
MOND	**MO**dified **N**ewtonian **D**ynamics
NA64	**N**orth **A**rea **64**
NN	**N**eural **N**etwork
POT	**P**roton **O**n **T**arget
SM	**S**tandard **M**odel
SPS	**S**uper **P**roton **S**ynchrotron
SRD	**S**ynchrotron **R**adiation **D**etector
SGD	**S**tocastic **G**radient **D**ecent
STD	**ST**raw **D**etector

Contents

1 | Introduction

Within the domain of physics, and more broadly in the field of scientific exploration, the seamless merging of theoretical models with experimental observations is paramount. It enables us to grasp the intricacies of nature and its diverse phenomena. Particle physics is a specialized area that investigates the characteristics of subatomic particles, exploring factors like their fundamental properties, interactions governed by fundamental forces, and their behavior as quantum-mechanical particles. The study of particle physics plays a crucial role in addressing inquiries within the realms of astrophysics and cosmology, one of which stands as arguably one of the biggest mysteries in today's science; namely Dark Matter (DM).

The Standard Model of particle physics (SM) is widely considered as one of the greatest scientific achievements of modern physics. Its comprehensive framework provides a structured understanding of the fundamental particles and their interactions and there are many examples that show the successful integration of theoretical predictions of the SM with experimental observations. The UA1 and UA2 experiments at the Large Electron-Positron Collider (LEP) at CERN conducted precision measurements in the electroweak sector of the SM. LEP, which operated from 1989 to 2000, provided data on the Z and $W^{\pm}$ bosons, crucial for explaining the electroweak theory [8, 13, 9, 10]. Other notable experiments, such as those involving the top quark and the Higgs boson, have significantly contributed to validating the SM. The heaviest known elementary particle, the top quark, was first observed at Fermilab in 1995 through experiments conducted by the CDF and D0 experiments [2, 1]. Additionally, the observation of the Higgs boson at the Large Hadron Collider (LHC) by the ATLAS and CMS collaborations in 2012 marked a groundbreaking moment. This discovery not only fulfilled a long-standing prediction within the SM but also provided crucial insights into the mechanism of mass generation for elementary particles [17, 39].

While the SM has demonstrated significant success it does possess limitations and the established consensus amongst most scientist today is that it cannot provide a complete description of nature. The most obvious flaw is the absence of a proper incorporation of gravity. Attempts to include gravity lead to a non-renormalizable theory, indicating a fundamental disconnect between the principles governing quantum field theory and general relativity. This fact alone implies that the SM cannot be the complete description of our universe. However, other issues have emerged after more experimental investigations. In the SM, fermion masses are generated through the Higgs mechanism, which requires a coupling between left- and right-

handed fermion fields [28]. However, it appears that no right-handed neutrinos have been observed, leading to the theoretical assumption that neutrinos should be massless [32]. Despite this, the detection of atmospheric neutrino oscillations by Super-Kamiokande in 1998 provides compelling evidence that neutrinos actually possess a non-zero mass [25]. This suggests the existence of a mechanism beyond the Higgs framework that could be responsible for neutrino mass generation, but such a mechanism has not yet been identified.

In addition, the Standard Model (SM) faces the so-called strong Charge-Parity (CP) problem. CP violation is experimentally verified in weak interactions and is consistently incorporated into the SM. However, Quantum Chromodynamics (QCD), which governs the theory of strong interactions within the SM, also theoretically allows for CP violation. This potential violation can be described through the Peccei–Quinn theory, which introduces a θ-parameter to quantify the strength of CP violation in QCD [34]. Experiments, however, require this θ-parameter to be exceedingly small, as even minimal deviations would result in observable effects inconsistent with experimental data [11]. The necessity for θ to be nearly zero has prompted physicists to search for a more profound explanation for the suppression of CP violation in QCD, leading to proposals such as the existence of the axion [43, 44].

The multitude of questions that have arisen from experiments which are all challenging the validity of the SM, therefore, strongly motivates the existence of physics Beyond the Standard Model (BSM). In this thesis, one possible explanation will be studied in detail: Dark Matter (DM). DM is a hypothesized form of matter that does not emit, absorb, or reflect electromagnetic radiation, making it essentially invisible. It constitutes about 27% of the total mass and energy content of the universe. Its existence is inferred from galaxy rotational curves [40, 24], gravitational lensing [23], Cosmic Microwave Background (CMB) anisotropies [35], or large-scale structure formation [45], suggesting that the DM relic density abundance, $\Omega_{\mathrm{DM}} = 0.27$, exceeds that of atoms by a factor $\Omega_{\mathrm{DM}}/\Omega_{\mathrm{baryon}} = 5$. Despite extensive research, the specific particle or particles composing DM remain unidentified, presenting a major unresolved question in the field of astro-particle physics. The absence of DM in the SM is not only a significant problem in itself, but discovering DM could offer insights into explaining other anomalies within the SM, which is the reason as to why they are so appealing and searched for.

The NA64 experiment at CERN is a fixed-target experiment that uses missing-energy techniques, aiming to identify various types of DM candidates. One of these candidates is the so-called Dark Leptonic Scalar (DLS) model. The DLS model constitutes a Lepton Flavour Changing (LFC) interaction and data selection of this interaction signature is the main focus of this thesis (see section 3.1.3 for more information). The physical processes is simulated in the `GEANT4` package, a highly accurate particle physics simulator. In order to experimentally claim a discovery, careful background analysis is performed on the expected signal signature in simulation and data. One of the main background studied for this model is the dimuon background, coming from the pair production in the electromagnetic shower

$e^- Z \to e^- Z\gamma$, $\gamma \to \mu^+\mu^-$; a SM event. Other background for this model are the pion and kaon decay, $\pi^- \to \mu^-\nu_\mu$ and $K^- \to \mu^-\nu_\mu$, mainly coming from contamination from the initial beam. Various selection cuts are applied to the simulation data to optimize the detection of the expected DM signal against the background signal. The objective is to refine the selection criteria to enhance the DM signal yield. This optimization is achieved by comparing the data obtained from traditional selection cuts with data selected by a trained Neural Network (NN) on the simulated data.

1.1 Outline of the thesis

In essence, the main goal of this thesis is to study the following questions:

- How well can a NN be trained to identify $\mu\mu$ events on real experimental data?
- Could such a model be trusted for identification of the DLS model φ on real experimental data?

The first part of the thesis (Theory) is the theoretical part giving the prerequisites needed to understand the depth of the DLS model as well as the methods used for the implementation of the NN. This includes the motivations of DM, Beyond the Standard Model (BSM) and also a brief theoretical explanation of the NN-techniques used in the project. The second part of the thesis (Methods and Results) explains in detail the experimental setup of NA64. The layout as well as detailed description of the different detectors that are essential for the measurements of the missing energies and hence of crucial importance to the training process of the NN. Simulation techniques for producing different types of interactions and particles, as well as methods in reconstruction and analysis of data is presented. Finally, the results obtained from the simulations, as well as the data analysis on the comparison between the simulation and actual data taking are presented. This thesis is concluded with a discussion emphasising the use of the results and how this method should be used for future use. Limitations in the implementation is presented and further discussions on future implementation is presented.

Part I

Theory

2 | The Dark Matter problem

The DM problem is regarded as one of the biggest unanswered questions in today's astro-particle physics. The concept of DM started with anomaly measurements that astronomers observed centuries ago. Today, this is a fact backed up by many different kinds of observations, all pointing to the same strange phenomenon. In the following chapter, the experimental evidence that currently exists for DM will be addressed to shed light on the main motivations for considering DM as a candidate for new physics BSM. The different families of possible DM candidates that are widely searched for in today's research at particle accelerator experiments, such as NA64 at the CERN SPS, will also be presented. Lastly, experimental anomalies and discrepancies observed when comparing experimental values to theoretical predictions guide us in selecting certain DM candidates. *The relic abundance of dark matter* and $g-2$ *dipole moment discrepancy* (which are presented in sec. 2.3.1 and 2.3.2) are used to constrain the parameters of new DM models, aiding in the development of more accurate theories. These topics are presented in the final sections of this chapter.

2.1 Evidence for dark matter

The notion of unseen mass in the universe has puzzled astronomers for many years. One early and notable case was the discovery of Uranus in 1781, which displayed irregularities in its orbit that led astronomers to predict the existence of another planet influencing its motion. This prediction eventually led to the discovery of Neptune in 1846. However, there have been efforts to explain similar anomalies by proposing the existence of some unseen matter. One example of this is the observed anomalous precession of Mercury's orbit, where the existence of a planet named Vulcan was proposed as a means of explaining these anomalies. Mercury's orbital discrepancies were however later resolved with Einstein's general theory of relativity, which offered a new understanding of gravity without the need for additional unseen planets. This historical context is reflected in today's astrophysics, where unexplained gravitational behaviors prompt two primary approaches. The first approach is to hypothesize the existence of some unseen mass that interacts gravitationally but not electromagnetically, to account for the observed phenomena. The second approach involves modifying our current theories. In 1983, Mordehai Milgrom proposed Modified Newtonian Dynamics (MOND) as an alternative to DM, suggesting that the laws of gravity themselves might need adjustment to explain the behaviors observed in galaxies. Extensive astronomical observations have placed stringent constraints on both approaches. Most physicists today agree that DM offers the most compre-

hensive explanation for a wide range of observations, including the rotation curves of galaxies, the CMB measurements and gravitational lensing effects observed in galaxy clusters, which all fits better to the Λ-CDM model. Nevertheless, the debate continues as researchers strive to refine their models and seek definitive evidence.

The very first clue of DM dates back to the 1930s, with significant contributions from Fritz Zwicky. Zwicky applied the virial theorem on the Coma Cluster and discovered that the visible mass could not account for the gravitational effects observed, leading to the inference of a substantial amount of unseen mass, which he termed *dark matter* [45]. Upon further investigation, the rotation curves of spiral galaxies provide compelling evidence for DM. According to Newtonian dynamics, the rotational velocity v of stars in a galaxy should decrease with distance r from the galactic center, following the relation

$$v \propto \frac{1}{\sqrt{r}}.$$

However, observations show that the rotational velocity remains approximately constant at large radii, which contradicts this expectation. An example of such observation is the rotation curves of the stellar disks of galaxy NGC 2903 which is shown in figure 2.1. This discrepancy can be explained by the presence of a DM halo that extends well beyond the visible part of the galaxy, providing the necessary gravitational pull to maintain the observed velocities [40].

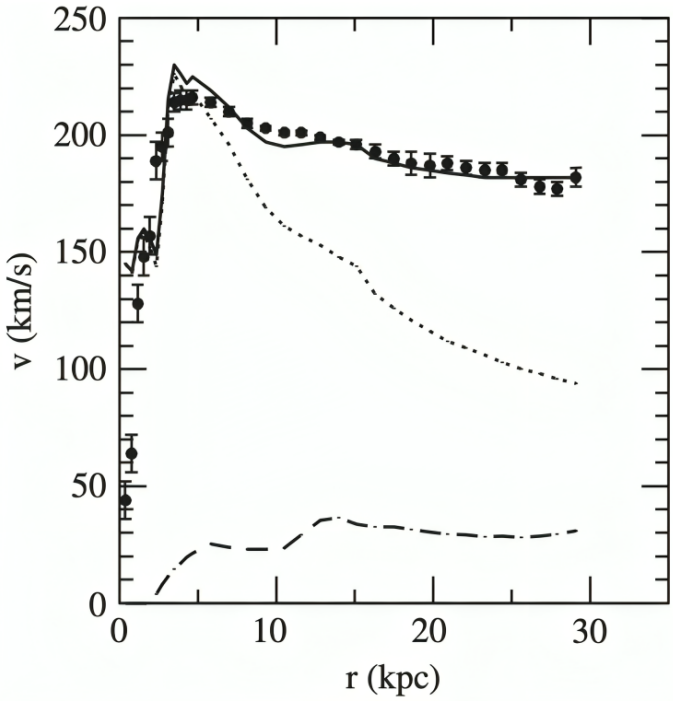


Figure 2.1: Rotation curve of the High Surface Brightness (HSB) galaxy NGC 2903. The dotted and dashed curves represent the Newtonian rotation curves for the stellar and gaseous disks, respectively. The solid curve depicts the rotation curve as predicted by modified Newtonian dynamics [40].

While DM provides a compelling explanation for these observations, alternative theories like MOND have been proposed. MOND modifies Newton's laws to explain the flat rotation curves of galaxies without invoking dark matter. In MOND, the acceleration a is modified as

$$\mu\left(\frac{a}{a_0}\right) a = \frac{GM}{r^2},$$

where $\mu(x)$ is an interpolating function such that $\mu(x) \approx x$ for $x \ll 1$ and $\mu(x) \approx 1$ for $x \gg 1$, and a_0 is a characteristic acceleration scale. This modification yields the observed flat rotation curves but struggles to explain the full range of astrophysical phenomena attributed to DM [15].

Instead, physicists are persuaded by other observations like the measurements of the CMB, which serves as a major contribution for the understanding and validation for DM. This faint glow of microwave radiation is a relic from the early universe, dating back to approximately 380,000 years after the Big Bang when the universe had cooled enough for protons and electrons to combine into hydrogen atoms, allowing photons to travel freely. Small fluctuations in the wavelengths observed as temperature give insights to the state of the early universe. These temperature fluctuations, known as anisotropies, are imprints of the density variations in the early universe. By studying these anisotropies, one can infer the composition and distribution of matter in the universe at that time. The CMB measurements, particularly from missions such as the Wilkinson Microwave Anisotropy Probe (WMAP) and the Planck satellite, have provided detailed maps of these fluctuations. Analyzing the CMB involves solving the Boltzmann equation for DM in the early universe. This mathematical framework allows us to predict how the abundance of DM evolved over time. By comparing these theoretical predictions with the observed properties of the CMB, we can place stringent constraints on DM models. These constraints depend significantly on the assumed interactions and behaviors of DM particles. One widely accepted model is the thermal freeze-out mechanism. In this scenario, DM particles were in thermal equilibrium with ordinary matter in the hot, early universe. As the universe expanded and cooled, the interaction rate of DM particles dropped below the universe's expansion rate, causing them to decouple from ordinary matter and cease interactions. This freeze-out process left behind a relic density of DM, which matches well with observations from the CMB. According to the Λ-CDM model, about 26% of the universe's mass-energy content is dark matter, while ordinary matter constitutes only about 5% [3].

Other notable studies have also revealed that the gravitational effects observed in galaxies, clusters of galaxies, and large-scale structures cannot be accounted for by visible matter alone. The Bullet Cluster, for instance, provides compelling evidence through gravitational lensing. When two clusters of galaxies collided, the hot gas, which makes up the majority of the visible mass, interacted and slowed down, while the DM, inferred from gravitational lensing, passed through unaffected [15]. This separation of DM and visible matter directly indicates the presence of a significant amount of unseen mass. Similarly, the gravitational lensing of light by galaxy

clusters indicates more mass than what is observable. These phenomena collectively point to a substantial amount of DM, which interacts primarily through gravity. The mass discrepancies observed is consistent in many of these astrophysical observations [14]. This persistent anomaly cannot be dismissed as a mere experimental error; instead, it suggests an underlying physical phenomenon. The robust and reproducible evidence significantly strengthens the scientific basis for the DM hypothesis.

2.2 Dark matter detection techniques

The search for DM has been going on for many years, and because of the mysterious nature of this potential sector of particles, different experiments using different techniques have been proposed over the years. The three most common ones are direct detection, indirect detection, and production at particle accelerators. Direct detection techniques assume that the universe contains a relic of DM, through which the Earth travels. The experiments use highly dense materials that are more likely to interact with DM, in the hopes that some faint interaction happens that could be considered a signal event of DM. This would be portrayed as an excitation signal, where a DM particle is assumed to scatter with the nucleus of the given material. The difficulty in this type of experiment is *background*, which is actually similar in all of the experiments looking for these faint signals. Signals considered background are SM events that would give the exact same event as a DM signal, and the quantification of these types of signals is crucial for claiming the discovery of DM. Minimizing and identifying these types of backgrounds is therefore of great importance in these types of experiments. The other type of experiment is indirect detection which is a study on galaxies and large-scale structures where one looks for gamma rays and other types of signals using telescopes to try to model DM production. The premise of indirect experiments is that DM particles could decay or annihilate with each other into detectable SM particles. The aim of this thesis is to study the production of DM using particle accelerators. Production of DM involves high-energy beams at fixed target experiments assuming that DM could exist in a so-called Dark Sector (DS) which mixes with the SM with mediator particles, often referred to as portals. NA64 is an example of an experiment focusing on producing DM particles by looking at events of missing energy (more on NA64 in chapter 5).

2.3 Dark matter candidates

The definition of the dark sector is broad, accommodating numerous models, and are intriguing candidates for explaining the origin of DM. However, its physics can be systematically explored using specific portal interactions as a classification scheme. While a mediator acting as a portal is not essential for creating a dark sector, a small interaction with the SM allows for detectable signatures in particle physics experiments and provides a mechanism for calculating the observed relic density.

$$\mathcal{L} = \mathcal{L}_{SM} + \mathcal{L}_{DM} + \mathcal{L}_{Portal} \tag{2.1}$$

The DM particles that make up the DS, marked as $\mathcal{L}_{DM}$, cannot be detected directly and can be assumed to be entirely separate to the SM particles $\mathcal{L}_{SM}$. The mixing, often referred to as *kinetic mixing*, is instead happening through a mediator particle; a portal $\mathcal{L}_{Portal}$. Gauge and Lorentz symmetries limit the ways mediators can couple to SM particles, which can be classified by their spin and parity. The most common and popular portals possibilities are [6, 18]:

$$\mathcal{L} \supset \begin{cases} -\frac{\epsilon}{2\cos\theta_W} B_{\mu\nu} F'^{\mu\nu}, & \text{vector portal} \\ (\mu\varphi + \lambda\varphi^2) H^\dagger H, & \text{Higgs portal} \\ y_n L H N, & \text{neutrino portal} \\ \frac{a}{f_a} F_{\mu\nu} \tilde{F}^{\mu\nu}, & \text{axion portal} \end{cases} \tag{2.2}$$

Here, H is the Higgs doublet, L is a lepton doublet of any generation, $B_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu$ is the hypercharge field strength tensor, $F_{\mu\nu}$ ($\tilde{F}_{\mu\nu}$) is the (dual) field strength tensor of the SM photon field, θ_W is the weak mixing angle, and $F'_{\mu\nu} \equiv \partial_\mu F'_\nu - \partial_\nu F'_\mu$ is the field strength tensor of a dark $U(1)'$ vector boson. More about this can be found in [6] which is a great source for investigating different models that are looked for at fixed target experiments similar to NA64.

The focus of this thesis is on the production of a mediator scalar particle that could produce scalar or fermionic dark matter particles. This particle is referred to as the Dark Leptonic Scalar particle φ, with the leptonic referring to the fact that it contains a lepton number. This property gives rise to a Lepton Flavor Change (LFC), which means that e.g. an incoming e^- would scatter with a nucleus in a Bremsstrahlung-like process, most likely resulting in an outgoing μ^- and the producing DLS φ. This interaction is precicely what is referred to as a kinetic mixing which arises in the Lagrangian; a term that couples with the SM as $\propto e\epsilon$. The DLS φ is expected to be massive, within the region of LDM/sub-GeV mass scale, which lies within the mass scales of detectable DM in experiments.

2.3.1 The relic abundance of dark matter

Models describing the production of DM particles in the early universe are essential in order to account for the DM density observed today. These models not only help explain the universe's dynamics but also provide constraints on DM properties, enabling us to focus the searches on specific classes of models that meet these criteria. The relic abundance of DM is a fundamental concept in cosmology, crucial for understanding the density and distribution of DM from the early universe to the present day. Different types of DM predict varying formation scenarios for cosmic structures and are categorized into cold, warm, and hot DM based on the velocity distribution of the DM particles. Cold Dark Matter (CDM) consists of particles moving at non-relativistic velocities, with examples including Weakly Interacting Massive Particles (WIMPs) and axions. CDM is characterized by its ability to form small-scale structures early in the universe, leading to a bottom-up structure formation process. Hot Dark Matter (HDM) consists of particles moving at relativistic velocities. These high speeds prevent the formation of small structures, leading to

a top-down structure formation process where large structures form first and then fragment into smaller ones. This model struggles to explain the detailed structures observed in the universe today and is largely ruled out. Warm Dark Matter (WDM) particles have velocities intermediate between those of CDM and HDM. Sterile neutrinos are a typical example of WDM. These particles suppress the formation of very small structures due to their higher velocities, leading to a smoother distribution of matter on small scales compared to CDM. Thermal equilibrium is a state where the DM particles and the particles in the primordial plasma (photons, baryons, leptons, etc.) interact frequently enough that they share the same temperature and energy distribution. This occurs in the early universe when the temperature is extremely high, and the interaction rates between particles are much higher than the expansion rate of the universe. As explained before, the freeze-out mechanism leaves a constant of DM left which is used as constraint on the models sought after in today's particle accelerators. The full window of DM particles ranges from less than a zeV ($\sim 10^{-22}$ eV) to hundreds of solar masses $M_\odot$ ($\sim 10^{68}$ eV). An extensive search is now being made for so called Light Dark Matter (LDM), which are candidates in the mass range between keV and GeV and Weakly Interacting Massive Particles (WIMP) that constitute the GeV to $\sim$ 100 TeV mass scale.

2.3.2 The $g-2$ anomalous magnetic dipole moment

The $g-2$ magnetic dipole moment discrepancy is another experimental result that further challenges the legitimacy of the SM. The magnetic moment for a muon is given by

$$\vec{\mu} = g_\mu \left(\frac{q}{2m}\right) \vec{S} \tag{2.3}$$

where g_μ is the gyromagnetic ratio of the muon, q is its charge, and m is its mass. According to the Dirac equation, which describes free spin-$\frac{1}{2}$ particles, the predicted value of g_μ is approximately $g_\mu \approx 2$. This is the result of tree-level Feynman diagram calculation. Higher-order contributions arise from quantum effects involving virtual particles such as leptons, hadrons, photons, Z-bosons, W-bosons, and even the Higgs. And it turns out that the higher-order corrections introduce an anomalous magnetic moment. The anomalous magnetic moment defined as

$$a_\mu = \frac{g_\mu - 2}{2} \tag{2.4}$$

and the most recent measurement of the anomalous magnetic moment was measured by Fermilab's Muon g-2 experiment [5], leading to the following result:

$$\Delta a_\mu = a_\mu^{\text{exp}} - a_\mu^{\text{SM}} = 116,592,059(22) \times 10^{-11}\,(0.19 \text{ ppm}) \tag{2.5}$$

This corresponds to a deviation of 4.2σ from the SM prediction, strongly suggesting the existence of new physics BSM. Consequently, several new BSM theories have been proposed to resolve this issue, and these models help constrain their parameter spaces. This will be demonstrated in detail in section 3.1.1.

3 | Physics Beyond the Standard Model

It is now time to introduce a new model BSM which is of interest for this thesis. The purpose of this chapter is to show how both the $g-2$ anomaly and relic density of dark matter could be explained by a model that mixes a dark scalar φ with the SM and introduces a change of lepton flavor with the interacting leptons. In addition, exact tree level cross section calculations of the DLS production is explained and presented in Appendix A.

3.1 Dark Leptonic Scalar

In order for most experiments that search for dark matter today to discover dark matter, mediator interactions between the dark matter particle and some Standard Model particle is a necessity. One of these models is the Dark Leptonic Scalar model (DLS), it inclundes the DLS φ which serves as a scalar portal boson carrying SM leptonic numbers L_e and L_μ. In its interaction with the SM, it introduces a non-diagonal interaction between the electron and muon, i.e. $e \leftrightarrow \mu$, whilst still conserving both the leptonic- flavor and numbers. The φ predominantly decays invisibly, $\Gamma(\varphi \to \text{invisible})/\Gamma_{tot} \sim 1$, e.g. into dark sector particles, never to be seen. Thus escaping stringent constraints placed today on the visible decay modes of the φ into SM particles from collider, fixed-target, and atomic experiments.
The portal term that mixes the dark matter particle with the SM is given in this model as

$$\mathcal{L}_{\varphi\mu e} = -h_{\mu e}\bar{e}_L\mu_R\varphi + h.c. \tag{3.1}$$

where $e_L = \left(\frac{1-\gamma_5}{2}\right)e$, $\mu_R = \left(\frac{1+\gamma_5}{2}\right)\mu$. The interaction 3.1 is invariant under the L_e, L_μ flavor global transformations $\varphi(x) \to \exp(i\alpha_e + i\alpha_\mu)\varphi(x)$, $\mu(x) \to \exp(-i\alpha_\mu)\mu(x)$, $e(x) \to \exp(-i\alpha_e)e(x)$. The $SU_L(2) \otimes U(1)$ invariant generalization of the interaction 3.1, we get [26]

$$\mathcal{L}_{\varphi\mu e} = -\frac{h_1 h_2}{M}\left(\bar{\nu}_e, \bar{e}\right)_L H\varphi\mu_R + h.c. \tag{3.2}$$

where we define $h_{\mu e} = \frac{h_1 h_2 \langle H\rangle}{M}$, with the vaccum expecation value of the Higgs isodoublet H being $\langle H\rangle = 174$ GeV. Looking at the solution for the simultaneous

broken symmetry, and in the unitary gauge we have $H = \left(0, \frac{h}{\sqrt{2}} + \langle H \rangle\right)$. Note that the complex scalar mediator $\varphi(x)$ is a singlet under the $SU_C(3) \otimes SU_L(2) \otimes U(1)$ SM gauge group. The DLS particle has multiple possible interaction channels, one of which is interacting with the Higgs isodoublet like

$$\mathcal{L}_{\varphi H} = -\lambda_{\varphi H} H^\dagger H \varphi^* \varphi$$

the Higgs boson would then decay invisibly into a pair of ϕ, $h \to \varphi^* \varphi$ with a given decay rate of

$$\Gamma(h \to \varphi^* \varphi) = \frac{\lambda_{\varphi H}^2 v^2}{16\pi m_h} \sqrt{1 - \frac{4m_\varphi^2}{m_h^2}} \tag{3.3}$$

Here m_h is the mass of the Higgs boson and $v = 246$ GeV is Higgs field vacuum expectation value. From the already careful existing bounds on the Higgs boson in the invisible decay mode, one can obtain an upper bound on the coupling constant $\lambda_{\varphi H} \leq 0.01$ [26] [46]. The interaction 3.2 is nonrenormalizable and it conserves both L_e and L_μ flavor numbers in the approximation of massless neutrinos. The renormalizable interaction can be obtained with the use of a vectorlike vector fermion E,

$$\mathcal{L}_{E\mu e} = \left(h_1 \left(\bar{n} u_e, \bar{e}\right)_L H E_R + h_2 \bar{E}_L \mu_R \varphi + h.c.\right) - M \bar{E} E \tag{3.4}$$

The DLS could in principle also interact with other dark matter particles not visible in the SM. Firstly, one could be the case of the $\varphi(x)$ field interacting with two dark matter complex scalars $s_1(x)$ and $s_2(x)$ given by

$$\mathcal{L}_{\varphi s_1 s_2} = g_{\varphi s_1 s_2} \varphi s_1 s_2 + h.c. \tag{3.5}$$

The coupling constant has the dimension of the mass. The interaction is invariant under the global transformation $\phi \to \exp(i\alpha_1 + i\alpha_2)\varphi$ and $s_i \to \exp(-i\alpha_i) s_i$ with $i = 1, 2$. As a consequence, the lepton flavours are conserved. Another model is when the dark scalar interacts with two light dark matter fermions Ψ_1 and Ψ_2 which has the interaction

$$\mathcal{L}_{\varphi \Psi_1 \Psi_2} = g_{\varphi \Psi_1 \Psi_2} \varphi \bar{\Psi}_1 \Psi_2 + h.c. \tag{3.6}$$

Their decay rates are given as

$$\Gamma(\varphi \to s_1 s_2) = \frac{g_{\varphi s_1 s_2}^2}{8\pi} \frac{p_1}{m_\varphi^2} \tag{3.7}$$

$$\Gamma(\varphi \to \Psi_1 \Psi_2) = \frac{g_{\varphi \Psi_1 \Psi_2}^2 p_1}{4\pi} \left(1 - \frac{(m_1 - m_2)^2}{m_\varphi^2}\right) \text{ where} \tag{3.8}$$

$$p_1 = \frac{1}{2m_\varphi} \sqrt{\left(m_\varphi^2 - (m_1 + m_2)^2\right)\left(m_\varphi^2 - (m_1 - m_2)^2\right)} \tag{3.9}$$

where p_1 is the momentum of particle 1 in the rest frame of φ, and m_1 and m_2 are

masses of particle 1 and 2. Here we assume $m_\varphi > m_1 + m_2$.

From the decay width of φ to fermion-fermion paricles, we have the decay with φ to $\mu^+ e^-$ as

$$\Gamma(\varphi \to \mu^+ e^-) = \frac{g^2_{\varphi\mu e} p_e}{4\pi}\left(1 - \frac{(m_e - m_\mu)^2}{m^2_\varphi}\right) \tag{3.10}$$

where p_e is the electron momentum in the center of mass frame. The annihilation corss section of s_1 and s_2 into μe pair in the nonrelativistic approximation in s-wave is

$$\begin{aligned}\sigma(s_1 s_2 \to e^+\mu^-)v_{rel} &= \sigma(s_1 s_2 \to e^-\mu^+)v_{rel} \\ &= |M|^2_1 \frac{p_{ecm}}{16\pi(m_1 + m_2)m_1 m_2}\end{aligned} \tag{3.11}$$

where

$$|M|^2_1 = g^2_{\varphi s_1 s_2} h_{\mu e} \frac{(m_1 + m_2)^2 - m_e^2 - m_\mu^2}{\left(m^2_\varphi - (m_1 + m_2)^2\right)^2} \tag{3.12}$$

If one assumes the simplest cases, with $m_1 = m_2 \gg m_\mu$, the annihilation cross section is then given as

$$\sigma(s_1 s_2 \to e^+\mu^-)v_{rel} = \frac{h^2_{\mu e} g^2_{\varphi s_1 s_2}}{8\pi\left(m^2_\varphi - 4m_1^2\right)^2} \tag{3.13}$$

Using equal masses does not change our main conclusions. For fermions Ψ_1, Ψ_2 in the nonrelativistic approximation, the annihilation cross section $\sigma(\Psi_1\Psi_2 \to e^-\mu^+)v_{rel}$ is given by 3.11 where we now instead use $|M|^2_2$ with

$$|M|^2_2 = \frac{g^2_{\varphi\Psi_1\Psi_2}}{2} h^2_{\mu e} \frac{1}{\left(m^2_\varphi - (m_1 + m_2)^2\right)^2}\left((m_1 + m_2)^2 - m_e^2 - m_\mu^2\right) m_1 m_2 v^2_{rel} \tag{3.14}$$

3.1.1 $g - 2$ anomaly

In the context of our investigation into the anomalous magnetic moment of the muon (a_μ), we delve into the framework of BSM physics, specifically focusing on the DLS model. This model has been proposed as a potential explanation for the observed deviation in the muon's gyromagnetic ratio (g-2) from the predictions of the SM. The deviation in the muon's magnetic moment is illustrated in figure 3.1 and given by:

$$\Delta a_\mu = \frac{h^2 m^2_\mu}{16\pi^2 m^2_\varphi}\int_0^1 dx \frac{x^2(1-x)}{(1-x)\left(1 - \frac{m^2_\mu}{m^2_\varphi}x\right) + \frac{m^2_e}{m^2_\varphi}x},$$

where h represents the coupling constant of the DLS, m_μ and m_e are the masses of the muon and electron respectively, and m_φ denotes the mass of the dark lepton scalar. For the electron's magnetic moment, a straightforward substitution of m_μ with m_e and vice versa yields:

$$\Delta a_{e(\mu)} = \frac{h_{\mu e}^2}{48\pi^2} \frac{m_{e(\mu)}^2}{m_\varphi^2}, \tag{3.15}$$

with $h_{\mu e}$ representing the coupling constant between the muon and electron within the DLS model. In the limit where $m_\varphi \gg m_\mu$, the expressions simplify, leading to the ratio of the electron's to muon's deviation as $\Delta a_e / \Delta a_\mu = (m_e/m_\varphi)^2$. This limit is often considered in theoretical analyses due to its simplifying assumptions.

Assuming that the deviation in a_μ is entirely attributed to the DLS model, we can constrain the coupling constant $h_{\varphi\mu}$:

$$h_{\varphi\mu} = (1.1 \pm 0.1) \times 10^{-3} \left(\frac{m_\varphi}{m_\mu} \right). \tag{3.16}$$

However, to ensure the validity of the DLS model, we must impose constraints to avoid undesirable processes such as $\mu \to e\varphi$. Experimental data strongly restricts the coupling constant $h_{\mu e}$ to prevent such decays. This choice of mass hierarchy ensures consistency with experimental constraints while maintaining the viability of the DLS model in explaining the observed deviation in a_μ. The great thing about this constraint is that if experiments are sensitive to this region and no signal is found, then this would exclude the DLS as the solution to the $g_\mu - 2$ problem.

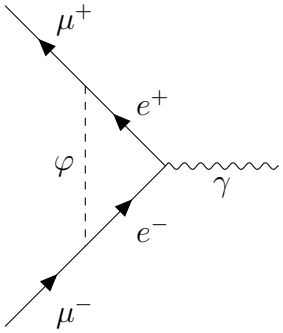


Figure 3.1: One-loop contribution of the dark leptonic scalar φ to Δa_μ, contributing to the muon $g - 2$. First order correction Feynman diagram of the process $\mu^+\mu^- \to \gamma$, with the QED vertex together with the dark scalar φ.

3.1.2 Relic abundance of dark matter

From this we obtain the total cross section

$$\begin{aligned}\sigma_{tot} v_{rel} &= \sigma(\Psi_1\Psi_2 \to e^-\mu^+) + \sigma(\Psi_1\Psi_2 \to e^+\mu^-) \\ &= \frac{h_{\mu e}^2 g_{\varphi\Psi_1\Psi_2}^2 m_1^2 v_{rel}^2}{8\pi\left(m_\varphi^2 - 4m_1^2\right)^2}\end{aligned} \tag{3.17}$$

Thus we see that the nonrelativistic limit model with scalar DM particles has s-wave behavior that contradicts to the Planck data.

For the model with fermionic DM, we have p-wave behavior for the annihilation cross section behavior that allows us to escape the Planck restrictions. If we now assume that the early Universe light DM is in equilibrium with ordinary matter. From the requirement taht the relic density of DM is explained by the model, we can estimat the coupling constant $g_{\varphi\Psi_1\Psi_2}$ using standard formulae for calculation of the DM density. We assume that $\langle\sigma v_{rel}\rangle = \mathcal{O}(1)$ pb, and the average relative velocity of annihilating DM particle is $\langle v_{rel}\rangle \sim c/3$ which corresponds to the observed DM density in the Universe [38]. Considering the simplest case $m_1 = m_2 \gg m_\mu$, gives us

$$\frac{h_{\mu e}^2 g_{\varphi\Psi_1\Psi_2}^2 m_1^2 v_{rel}^2}{8\pi\left(m_\varphi^2 - (m_1+m_2)^2\right)^2} = \mathcal{O}(10)\text{ pb}. \tag{3.18}$$

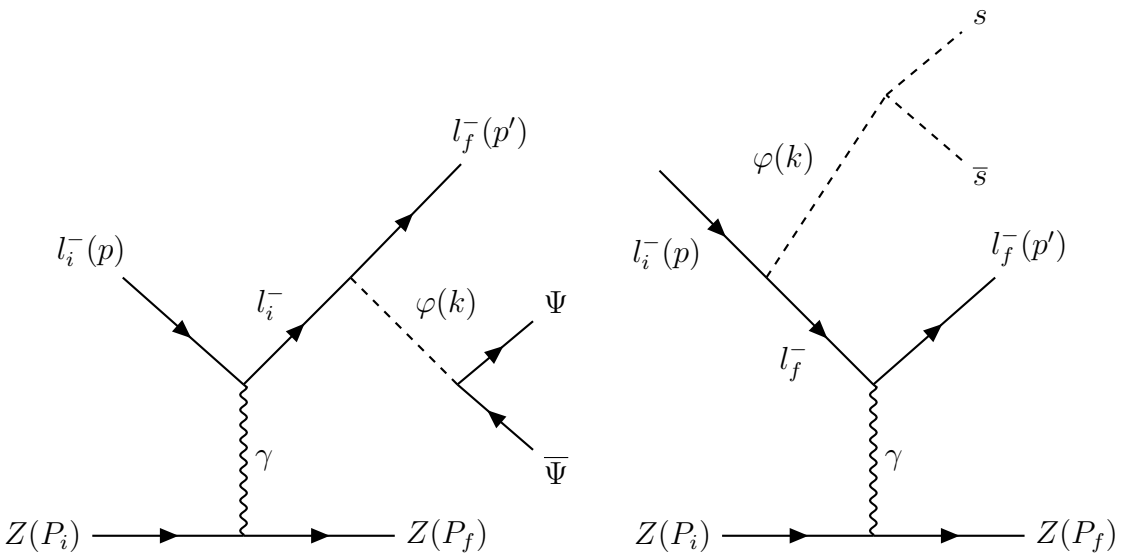
For the case of $m_\varphi = 3m_1$ we find

$$h_{\mu e} g_{\varphi\Psi_1\Psi_2} \sim 10^{-3}\left(\frac{m_\varphi}{\text{GeV}}\right). \tag{3.19}$$

In the assumption that the model explains muon $g-2$ we find that $g_{\varphi\Psi_1\Psi_2} \sim 0.1$ and it depends rather weakly on the φ mass. As a consequence we obtain that $g_{\varphi\Psi_1\Psi_2} \geq h_{\mu e}$ for $m_\varphi \leq 10$ GeV and the meadiator φ decays mainly invisibly into DM particles. So we find that both the $(g-2)_\mu$ anomaly and the dark matter relic abundance can be explained by our model.

3.1.3 Dark Leptonic Scalar φ at NA64

The mediator DLS φ is produced after a charged lepton $l_i^-(p)$ undergoes a brehmsstrahlung-like process as $e^-Z \to \mu^- Z\varphi$; $\varphi \to$ invisible, as shown in the Feynman diagram 3.2.



(a) DLS production after bremsstrahlung process. This example is also showing a possible decay to a pair of fermionic dark matter $\Psi\overline{\Psi}$.

(b) DLS production before bremsstrahlung process. This example is also showing a possible decay to a pair of scalar dark matter $s\overline{s}$.

Figure 3.2: Feynman diagrams depicting two possible creations of the dark leptonic scalar φ in lepton-nucleus scattering [26].

A detailed cross section calculation is shown in Appendix A

4 | Machine Learning in Particle Physics

Machine Learning (ML) is a branch of artificial intelligence that focuses on the development of algorithms and models that enable computers to learn from and make predictions based on data. In recent years, ML techniques have gained immense popularity and have become essential tools in various fields. The applications of Neural Networks (NN) in particle physics has proven to be successful in the handling of data. Efficient data analysis methods are crucial when dealing with large datasets. Traditional data selection methods rely on simple cuts based on expected detector behavior. These simple cuts become much more complex when working with more complicated signal patterns, due to the range of various cut combinations that could be used. This is where NN's might be useful. Deep learning models have been used for jet substructure classification at the LHC, helping distinguish between quark- and gluon-initiated jets, thereby improving signal detection in complex datasets [27].

The primary objective of this thesis is to investigate how a NN can be trained on simulated data by minimizing the loss function for a binary classification. This trained model will be compared to traditional selection methods, focusing on evaluating the signal-to-background ratio on $\mu\mu$ events, to determine if any improvements are achieved. More on this in section 5.2 and 5.4. Additionally, since we are using labeled data, only supervised methods will be taken into account.

4.1 Fundamentals of Neural Networks

NN are computational models inspired by the structure and function of the human brain. At its core, a NN consists of interconnected neurons organized in layers. Each neuron performs a computation on its input data and passes the result to the neurons in the next layer. The strength of the connections between neurons, known as weights, determines the impact of each neuron's output on the subsequent layers.

4.1.1 Neurons and Activation Functions

A single neuron in a NN can be represented mathematically as follows

$$z = \sum_{i=1}^{n} w_i x_i + b \tag{4.1}$$

where z is the weighted sum of the inputs x_i with corresponding weights w_i, and b is the bias term. The weighted sum is then passed through an activation function $f(z)$ to introduce nonlinearity and enable the network to model complex relationships

$$y = f(z)$$

Common activation functions that are used in this project are the sigmoid function, ($f(z) = \frac{1}{1+e^{-z}}$) and the rectified linear unit (ReLU) function, ($f(z) = \max(0, z)$).

4.1.2 Neural Network Architecture

A NN consists of an input layer, one or more hidden layers, and an output layer. Each layer is composed of multiple neurons, with connections between neurons in adjacent layers. The number of neurons in the input layer is determined by the dimensionality of the input data, while the number of neurons in the output layer corresponds to the number of output classes or regression targets.

4.1.3 Backpropagation and Gradient-Based Optimization

Backpropagation is a key algorithm used to train NNs by iteratively adjusting the weights and biases to minimize a predefined loss function. The process involves two main steps:

1. *forward propagation*, where input data is passed through the network to compute predictions,
2. and *backward propagation*, where the error between the predictions and the true labels is propagated back through the network to update the model parameters.

Gradient-based optimization methods, such as Stochastic Gradient Descent (SGD) and its variants, are commonly used to minimize the loss function. These methods compute the gradient of the loss function with respect to the model parameters and update the parameters in the direction that minimizes the loss. The learning rate, which determines the size of the parameter updates, is a crucial hyperparameter that affects the convergence and stability of the training process. The Adam algorith will be used in the training of the NN. Adam, short for Adaptive Moment Estimation, is an advanced gradient-based optimization method that combines the advantages of two other popular algorithms: Momentum and RMSProp. Adam adapts the learning rate for each parameter by computing adaptive estimates of lower-order moments, which allows for faster convergence and improved stability compared to standard SGD [30, 20].

By iteratively applying backpropagation and gradient-based optimization, NNs can learn complex patterns and relationships within the data, leading to improved performance on various tasks such as classification and feature extraction. In the context of dimuon selection, NN offer a powerful framework for building predictive models that can effectively discriminate between signal and background events. By leverag-

ing the hierarchical and nonlinear nature of NN architectures, one can capture subtle features and correlations in the data, potentially leading to improved performance in particle identification tasks.

4.1.4 The Binary Cross Entropy loss function

The Binary Cross Entropy (BCE) Loss function, also known as log loss, is widely used in binary classification problems within NN. It quantifies the difference between the predicted probabilities and the actual class labels. Given a binary classification problem, let y_i be the true label of the i-th sample, where $y_i \in \{0, 1\}$, and $\hat{y}_i$ be the predicted probability that $y_i = 1$ [19]. The BCE loss for a single sample is defined as

$$\mathrm{BCE}(y_i, \hat{y}_i) = -\left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)\right) \tag{4.2}$$

For a dataset with N samples, the average BCE loss is computed as

$$\mathrm{BCE}_{\mathrm{avg}} = -\frac{1}{N} \sum_{i=1}^{N} \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)\right) \tag{4.3}$$

The BCE Loss function is derived from the likelihood of the Bernoulli distribution. For a single sample, the likelihood is

$$P(y_i|\hat{y}_i) = \hat{y}_i^{y_i}(1 - \hat{y}_i)^{(1-y_i)} \tag{4.4}$$

Taking the logarithm of the likelihood function gives the log-likelihood

$$\log P(y_i|\hat{y}_i) = y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \tag{4.5}$$

The BCE loss is the negative log-likelihood, averaged over all samples

$$\mathrm{BCE}_{\mathrm{avg}} = -\frac{1}{N} \sum_{i=1}^{N} \log P(y_i|\hat{y}_i) \tag{4.6}$$

In training neural networks, it is essential to compute the gradient of the loss function with respect to the predicted probabilities $\hat{y}_i$. The gradient of the BCE loss for a single sample is given by

$$\frac{\partial \mathrm{BCE}}{\partial \hat{y}_i} = -\left(\frac{y_i}{\hat{y}_i} - \frac{1 - y_i}{1 - \hat{y}_i}\right) \tag{4.7}$$

For the entire dataset, the gradient can be written as

$$\frac{\partial \mathrm{BCE}_{\mathrm{avg}}}{\partial \hat{y}_i} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \mathrm{BCE}(y_i, \hat{y}_i)}{\partial \hat{y}_i} \tag{4.8}$$

The BCE loss function is used as the objective function to minimize during training. The predictions $\hat{y}_i$ are typically obtained by applying an activation function, in my

case, the sigmoid activation function $\sigma(z_i)$ to the output layer of the network

$$\hat{y}_i = \sigma(z_i) = \frac{1}{1 + \exp(-z_i)} \tag{4.9}$$

Here, z_i represents the output of the NN before the activation function. During back-propagation, the gradient of the BCE Loss with respect to the network parameters is used to update the weights via gradient descent.

Part II

Method and Results

5 | Dark matter searches at NA64

Some theoretical models of DM are more compelling than others, but the most appealing models are those that can be tested experimentally. The scientific approach employed at the NA64 experiment relies heavily on creating highly detailed simulations used for modeling our desired physical system. This involves an in-depth analysis of both simulated and experimental data. In order to benchmark the simulation, a comprehensive analysis of dimuon production in simulation and experiment is conducted. In addition to this, proper signal and background detection techniques for the DLS model are employed. This chapter focuses on introducing the goals of NA64 along with the experimental setup, the selection criteria for dimuon and DLS production, outlining the anticipated background and signal characteristics for the DLS model, and describing how these factors are incorporated into the simulation process and compared to real data.

5.1 The NA64 experiment

The NA64 experiment, situated at the North Area of the CERN Super Proton Synchrotron (SPS), is a fixed target experiment that uses an optimized and essentially pure beam from the H4 beam line. The main goal for NA64 is to probe both the visible and invisible signals of a potential new dark force carrier, providing complementary sensitivity to the sub-GeV hidden DM sector. Like many other particle physics experiments, NA64 is a counting experiment. One count is defined as the observed event that passes through all the necessary detectors and meets the conditions necessary for an expected DM signal event. One of the main signal signatures observed by the detectors is missing energy, where the missing energy is assumed to be transferred to DM particles. Depending on the anticipated background-to-signal ratio, which might vary from model to model, one expects a certain number of counts in order to be able to claim a discovery. This is based on the statistical hypothesis testing that is performed in the analysis, where one aims to reject the background-only hypothesis. The most common and accepted measure for a discovery in the High Energy Physics community is the 5σ criteria, i.e. a confidence interval of 5 standard deviations. Simulations of the experiment are crucial in the analysis stage and it is therefore important for the simulation to be accurate in order to draw such conclusions.

5.1.1 General detection techniques at NA64

The type of dark matter mainly searched for at NA64 is light dark matter (LDM), with models compatible with theories aimed at explaining various measured discrepancies, such as the muon $g-2$ anomaly or relic dark matter (see sections 2.3.2 and 2.3.1). Different signal signatures and models follow different strategies for detection. However, some detection techniques are the same for all the models, and some fundamental requirements are

1. Initial energy of the particle needs to be known

2. The initial particle needs to be identified

3. A fully hermetic detector setup; *all* the energy needs to be contained in the detection

These requirements enable the possibility to identify the production of a dark mediator through missing energy. Some models are assumed to be produced entirely invisible, with missing energy as the only signature. Others are visible, meaning that they decay into a SM particle particle again, giving other more specific signatures in the detectors. Another type of signature is the semi-visible signature, where the production is assumed to be decaying to a SM particle and then back to a DM mediator again. Dark force carriers such as the dark photon, A', is an example of a model that is searched for at NA64 in invisible, semi-visible and visible mode [33]. It is produced via kinetic mixing with brehmsstrahlung photons; this is the dominant production process from an electromagnetic shower produced by the beam of electron hitting a dense active target. And the different modes give different signatures, which are all probed at NA64. So far there have been no evidence of LDM production at NA64, however there are still regions in mass-coupling space of these models which require more statistics to be excluded.

Another rather new model that is searched for at NA64 is the DLS model coming from the LFC particle φ, which has a completely different signal signature. Here one expects missing energy, as well as a clear signal signature from a μ^- in the downstream detectors (as illustrated in 5.5). There are also e^-, e^+, and μ modes at NA64, all capable of probing different dark matter models. While the DLS model in principle is feasible in all these modes, the emphasis for this thesis lies in the electron mode, NA64e.

5.1.2 Experimental setup

The NA64 experiment uses the H4 beam line coming from the CERN SPS. The produced particles from the H4 beam are divided and sent to different experiments, one of which is NA64. The H4 beam line provides a very pure 100 GeV e^- beam with a maximal intensity of $\sim 10^7$ Electrons On Target (EOT) per SPS spill of 4.8 s. Originating from a primary proton beam of 400 GeV/c onto a beryllium

target, with intensities reaching up to $5-7\times10^{12}$ Protons On Target (POT)[1]. The electron beam primarily derives from π^0 production, followed by the decay process $\pi^0 \to \gamma\gamma \to e^-e^+$. The degree of hadron contamination in the electron beam is $\pi/e \lesssim 10^{-2}$ [12, 7]. Detailed specifications of the CERN SPS are described in Ref. [41].

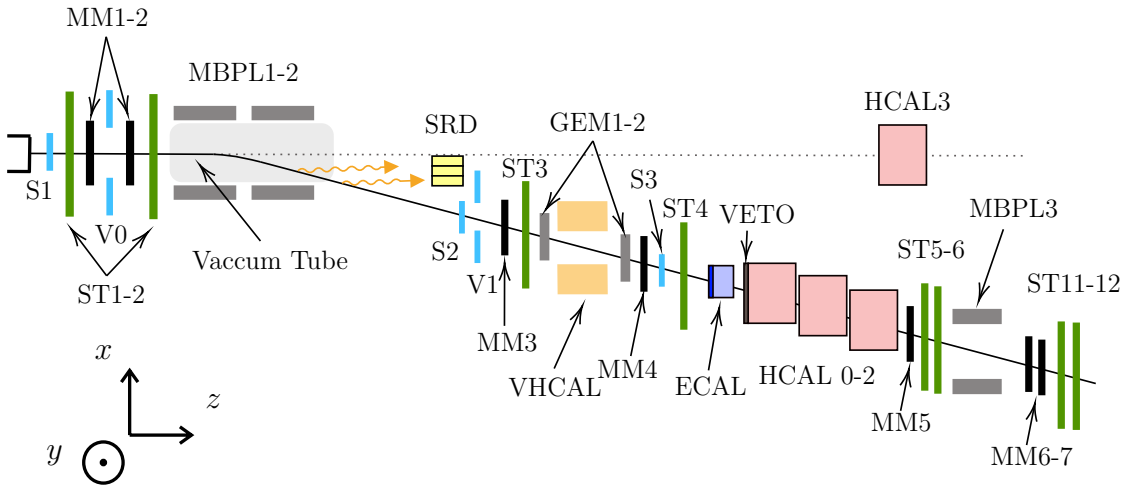


Figure 5.1: Schematic view from the top of the NA64e experimental setup in 2023 [33]. The 100 GeV electron beam (black line) arrives from the left and passes through a couple of tracking detectors before being bent and sent to the active target.

Like mentioned before, the main objective of the experiments at NA64 is to look for missing energies within the signal regions that could be considered as DM. For this to be possible, careful particle identification, tracking, energy determination and a setup that is fully hermetic are crucial. The NA64 experiment therefore consists of many different detectors, each serving a specific purpose. The experimental setup for the electron run used in 2023 is shown in figure 5.1.

Only triggered events are considered and compared with in the analysis procedure. This way pileup and out-of-time events are minimized and rejected. Pileup events are defined as events where two or more electrons pass the selection, instead of a single electron. To do this, the incoming beam is first and foremost defined by two scintillators, S_0 and S_1, a veto counter, V_1, and two straw detectors STD_1 and STD_2 are placed along the beam axis. To determine the incoming electron momentum, a magnetic spectrometer composed of two dipole magnets ($MBPL_1$, $MBPL_2$) providing a magnetic field of 7 Tm and six Micromegas (MM) trackers are used: two upstream (MM_1, MM_2) and four downstream (MM_3-MM_6). Moreover, the precision of the momentum reconstruction is further improved by four Gas Electron Multipliers (GEM_1-GEM_4) and two straw tubes (STD_3 and STD_4) positioned before the target. After being magnetically deflected, the electrons are transported through a 10.2-meter-long vacuum vessel to minimize particle interactions. The reason of the bending is to obtain higher precision for determining the energy of the incoming electrons. The other reason is to filter out unwanted particles. Since a pure electron

[1]Information as of 2024.

beam is sought for, by bending the beam, positively charged and neutral particles get deflected at different directions, not to be detected.

The Synchrotron Radiation Detector (SRD) measures the energy of photons emitted by the bent beam and is crucial for maintaining high electron identification. The trigger system relies on scintillators S_0 and S_1 and a second set of counters placed after the evacuated tube (S_2, S_3, and S_4). The core of the experiment is the active target, the Electromagnetic CALorimeter (ECAL), composed of 36 modules arranged in a 6x6 matrix with a total dimension of 229×229×471 mm^3. The schematic of the ECAL is showed in figure 5.2. This transversal segmentation allows exploiting the shape of the electromagnetic shower to further reject background. Each module of the calorimeter consists of 150 alternating layers of absorber (Pb) and scintillator material to stop incoming particles and measure their energy deposit. The ECAL is subdivided into two parts, the first called the preshower (PS), which is approximately 5 cm thick and helps identify electrons that start to shower immediately in the dense material.

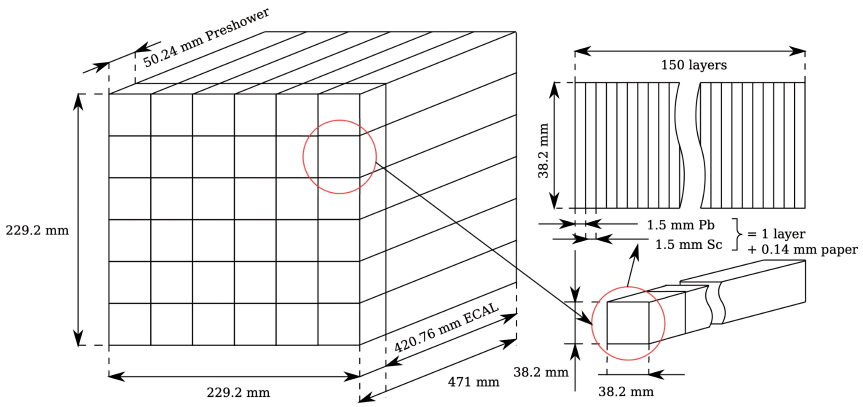


Figure 5.2: Schematic of the 6×6×2 ECAL module, a Shashlik type calorimeter. First part depicts the preshower, used for particle identification, with each cell consisting of alternating layers of absorber (Pb) and scintillator material [21].

To measure particles overcoming the main target and ensure complete hermeticity needed to detect missing energy, a veto detector (V_2) and three large Hadronic CALorimeter (HCAL) modules are installed after the ECAL. The HCAL module is displayed in figure 5.3. The HCALs can detect various types of hadronic particles resulting from the cascade of interactions within the active target, including hadrons and muons, and serve as a tool for signal identification. The HCAL module is arranged in a 3x3 matrix with a total length of approximately 160 cm and consists of a sandwich of 48 stacked absorber-scintillator plates, with iron as the absorber material. Additionally, a fourth module, often called the zero-degree HCAL, is aligned with the unbent beam axis to reject events where upstream interactions generate hard neutral particles. The veto counter, measuring 5 cm in thickness, comprises three longitudinally arranged scintillator stripes. Selections are made based on hits detected on the central scintillator to enhance alignment accuracy.

The detection setup that has been presented thus far remains mostly the same for most DM searches. The reason is because these detectors and the selection that are applied on these are mainly used for selecting a pure electron beam hitting the ECAL.

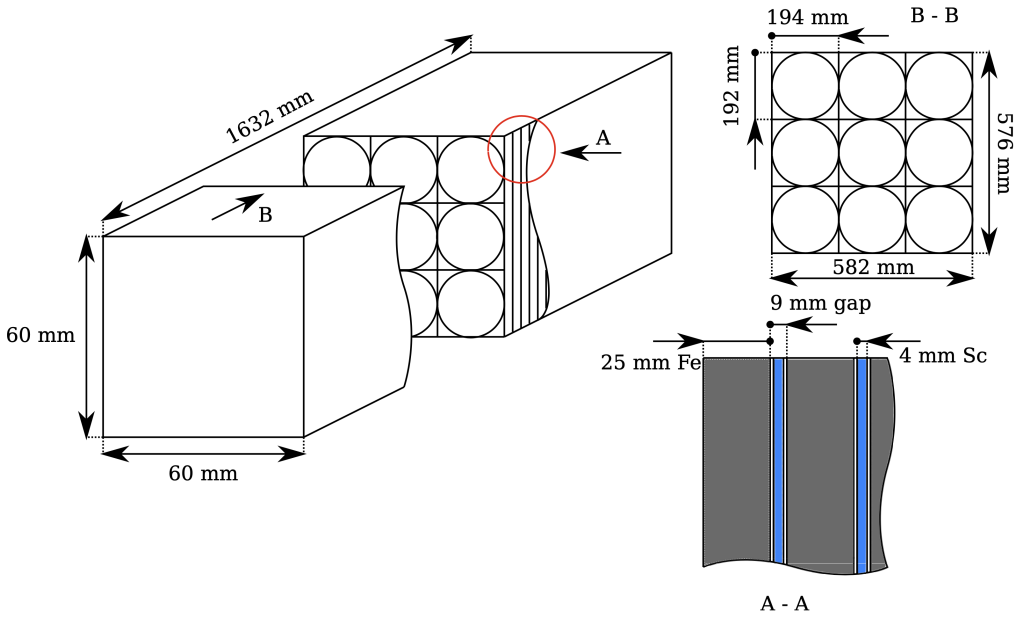


Figure 5.3: [21] Schematic of the 3×3 HCAL module. Each cell is composed of layers alternating between absorber (Fe) and scintillator material [21].

The last segment of the setup includes a series of tracking detectors, specifically straw detectors and Micromegas, positioned both before (MM_5, STD_5, STD_6) and after (MM_6, MM_7, STD_{11}, STD_{12}) the final magnet ($MBPL_3$). The magnet is of great importance, making it possible to perform a comprehensive momentum reconstruction and particle seperation of the muons. The rest of the detectors are primarily tasked with identifying muons, utilizing the magnet to effectively differentiate between positive and negative muons. These tracking detectors are crucial as they play a key role in the analysis of the targeted model and dimuons. The downstream spectrometers are particularly significant for the DLS model.

Straw detectors function using thin tubes filled with a gas mixture, typically 80% argon and 20% carbon dioxide. Inside each tube, there's a central anode wire. When a charged particle, such as a muon or hadron, passes through the tube, it ionizes the gas. The ions and electrons produced in this process are attracted to the tube walls and the central anode wire, respectively. This movement generates an electric signal along the wire, which is then detected and processed to determine the presence and properties of the passing particle. This mechanism allows straw detectors to track the paths of particles with precision. To enhance precision and reduce noise interference, the setup incorporates dual layers of straw tubes, as depicted in Figure 5.4. This configuration requires a coincident detection—a simultaneous hit in both the preceding and succeeding tubes—to confirm the presence of a particle.

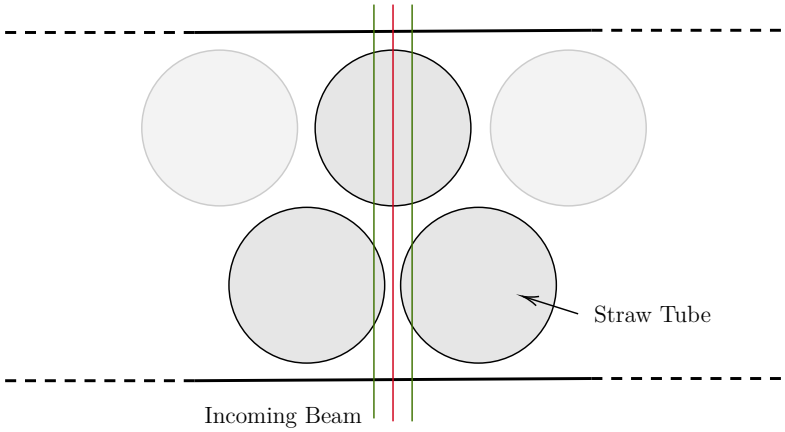


Figure 5.4: Schematic of straw detector, consisting of two layers of straw tubes. Single hit per event in straw detector are not saved (red beam), two hits are required in order to save event (green beam) [42].

5.2 Detector techniques for $\mu\mu$ and LFC searches

The techniques for finding dimuons and LFC particles mainly depend on the downstream detectors consisting of $HCAL_0$-$HCAL_2$, VETO in front of HCAL, MM_5-MM_7 and STD_{11}-STD_{12}. The trackers are used for a proper momentum reconstruction of the muons necessary in order to maintain a fully hermetic detection. By using a combination of these detectors, stringent selection criteria are applied to isolate events that indicate the presence of DLS particles. The selection cuts focus on energy deposition patterns and the presence of coinciding hits in the tracking detectors, maximizing the signal yield while minimizing background. In other words, if φ does exist, the analysis should aim to maximize the probability of its discovery while minimizing the risk of producing a false positive.

The comperhensive study on dimuons is twofold:

1. Validating simulation results with experimental data.
2. Experimental data on dimuons provide a reliable means to benchmark the implementation of the NN.

Additionally, simulations of other SM events that could be mistaken for $\mu\mu$ signals are evaluated to ensure robust selection criteria. More advanced techniques, such as NN trained models on filtered data from MC simulations, are also employed to enhance the comparison between traditional selection criteria and the model. This approach aims to improve the detection sensitivity for LFC searches. This method will be benchmarked using dimuons and more information about this procedure is presented in section 5.4.

In this section, we delve into the expected signal signatures in the experiment and the corresponding background events.

5.2.1 Phenomenology of $e^-Z \to \mu^- Z\varphi$ events

The theoretical framework of the DLS model is presented in chapter 3, providing the corresponding Lagrangian, decay widths and expected final state particles which are sought after at NA64. The weak coupling interaction of the dark scalar mediator is expected to take place in the ECAL where most of the energy deposition of the electron takes places. This could then result in a flavour change with $e \to \mu$ (see Feynman diagram 3.2). Muons are minimum ionizing particles (MIP) in the ∼GeV energy range. This unique property is key in NA64 because muons pass through detectors without leaving much energy or scattering behind, allowing more muons to survive. The expected energy deposition of a muon passing through a HCAL module is ~ 2.5 GeV, however varies slightly due to resolution and stochastic reasons [2]. This means that muons produced at energies below ~ 7.5 GeV won't surpass the stopping energy of the HCALs. The expected signal signature is illustrated in the below figure 5.5.

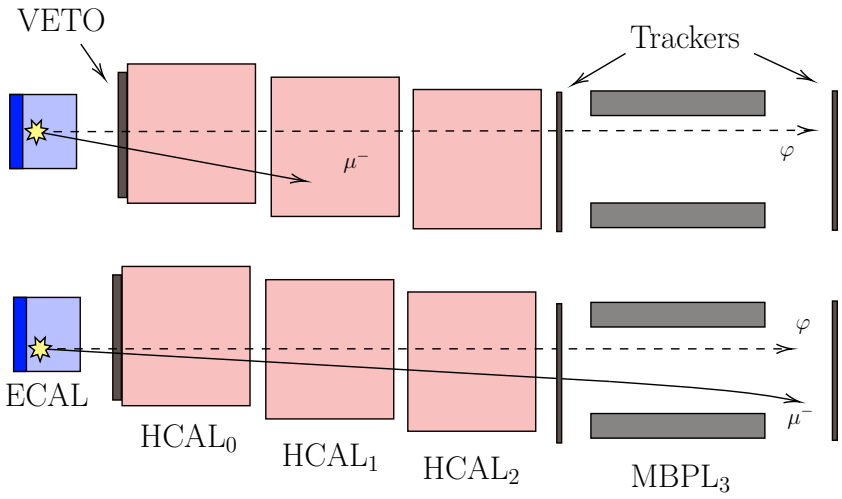


Figure 5.5: Signal signature for a dark leptonic scalar, denoted as φ. The LFC process generates a DLS φ alongside a μ^- in the ECAL. The figure below shows how the μ^- proceeds through the HCALs and is subsequently deflected, being detected in the tracker detectors. The figure above illustrates how a low energetic muon could be absorbed in the HCALs and corresponds to events that are not sufficient to be deemed as signal and are subsequently rejected.

Since the produced dark scalar φ will be undetected, the main signal that is sought after is the μ^-. The dark leptonic scalar φ mostly decays invisible to dark particles, $\Gamma(\varphi \to \text{invisible})/\Gamma_{tot} \sim 1$, like e.g. dark fermions $\varphi \to \Psi\overline{\Psi}$. Searching for visible or semi-visible signatures where the DLS φ interacts with a SM particle and goes back to SM is therefore very unlikely and not studied in this thesis. The signature is instead based on the first tree level interaction, with a μ^- passing certain requirements on the detectors, optimized by the simulation implementation which will be presented in more depth in section 5.3.1. The above figure in 5.5 also illustrates how a low-energetic μ^- might be absorbed in one of the HCAL modules, subsequently

[2]The energy resolution of each HCAL module is $\Delta_E/E \approx 0.62/\sqrt{E}$ [21].

not hitting the tracking detectors. These events are rejected since background estimation for these events are too high. Another case is if the μ^- is produced with a large production angle, making it likely to miss the tracking detectors entirely or produce a non-ideal hit pattern.

In order to be able to make a clear distinction between signal and background events, stringent selection criteria are applied. These criteria focus on minimizing the background while maximizing the signal yield, ensuring the selection of events most likely to indicate the presence of the dark leptonic scalar φ. In other words, one wants to maximize the significance, defined as $S = s/\sqrt{s+b}$, where s is the number of signal events and b is the number of background events. The specific selection cuts, which have been optimized are defined as follows (more on the methods used in section 5.3):

1. Characteristic energy deposition of a single μ minimum ionizing particle in **each** HCAL module, i.e. $E_{HCAL_i} < 3.5$ GeV, $i = 0, 1, 2$.

2. Characteristic energy deposition of a single μ minimum ionizing particle in **each** VETO cell in front of the HCALs, i.e. $E_{VETO} < 0.016$ GeV.

3. Coincidence hit of μ^- in MMs and STDs with only one hit in the expected region for a negative muon.

4. Missing energy in ECAL larger than 50 GeV, i.e. $E_{missing} = E_0 - E_{ECAL} < 50$ GeV.

5. Total event missing energy larger than 20 GeV, i.e. $E_{missing} = E_0 - E_{ECAL} - E_{HCAL} - E_{mom} > 50$ GeV.

By applying these cuts, the analysis aims to isolate events where the presence of the dark leptonic scalar φ can be inferred from the detected μ^- and the associated missing energy, while rejecting background events that do not match the expected signal signature.

5.2.2 Phenomenology of $e^-Z \to \mu^-\mu^+$ events

The process of pair production, which has a cross-section on the order of $\sigma_{\text{pair}} \sim 4Z^2\alpha r_c^2$ (where $r_c = \frac{e^2}{4\pi\epsilon_0 mc^2}$), is applicable to all charged leptons. However, due to the larger muon mass, the production of $\mu^+\mu^-$ pairs within an electromagnetic shower—following a photon interaction with a nucleus in the target—occurs less frequently compared to e^+e^- pair production. Specifically, this suppression factor is $(m_e/m_\mu)^2 = 2.34 \times 10^{-5}$. The interaction $e^-Z \to e^-Z\gamma; \gamma \to \mu^+\mu^-$ is particularly rare. However, since the total number of EOTs at NA64 is in the order of $\sim 10^{11}$, dimuons appear very distinctly in the results. This process is crucial for both validating the MC simulation (section 5.3) and for adjusting the expected signal yield, as it shares significant similarities with events that involve the DLS model.

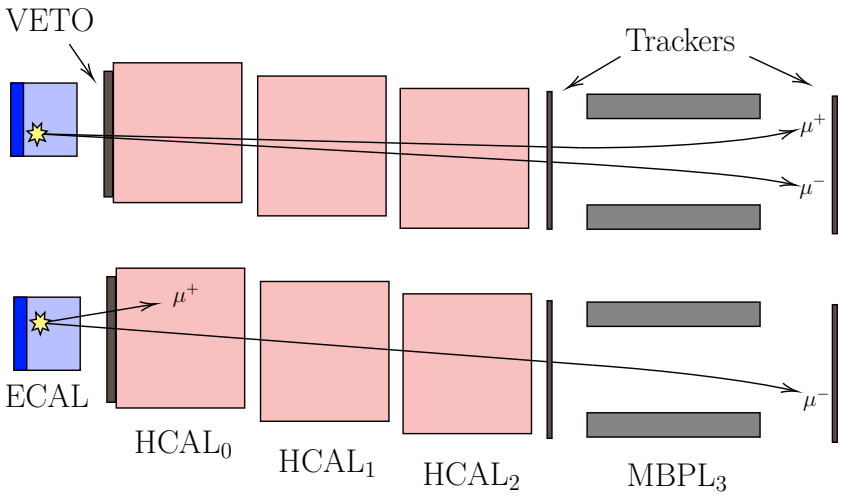


Figure 5.6: The above figure depicts the expected dimuon signature in experiment. The figure below is a dangerous event, illustrating a dimuon process that might produce the same expected signal signature as a DLS φ particle.

In situations where the energy of the μ^+ is very low and most of the energy has been transferred to the μ^-, the signature depicted in Figure 5.6 may be observed. This scenario is considered dangerous if there is also missing energy from the ECAL, as this would precisely match the signal signature of a DLS particle. Consequently, some $\mu\mu$ events are classified as background and must be carefully evaluated to ensure a comprehensive sensitivity study. The selection cuts applied on the downstream spectrometers are defined as follows:

1. Characteristic energy deposition of a dimuon $\mu\mu$ in **each** HCAL module, i.e. $8 > E_{HCAL_i} > 4$ GeV, $i = 0, 1, 2$.

2. Characteristic energy deposition of a single μ minimum ionizing particle in **each** VETO cell in front of the HCALs, i.e. $0.032 > E_{VETO} > 0.016$ GeV.

3. Coincidence hit of μ^- in MMs and STDs with only two hits in the expected region for the $\mu\mu$ pair.

These cuts are specifically used to select dimuons. This is not something we want in the end, however, simply applying the opposite of this selection will reject these events.

5.2.3 Phenomenology of $\pi^- \to \mu^-\nu_\mu$ and $K^- \to \mu^-\nu_\mu$ events

The last source of background for the LFC model are hadronic decays coming from the initial beam. Pion π^- and kaon K^- decaying into a μ^-; $\pi^- \to \mu^-\nu_\mu$ and $K^- \to \mu^-\nu_\mu$ serves as the most dangerous events. Pions and kaons come naturally from the beam source and if they leave undetected from the SRD after the vacuum tube, they could in principle then decay between the end of the vacuum tube to the beginning of the ECAL (see figure 5.1). The methods used for particle identification effectively

suppresses muonic in-flight decays that occur before the SRD. Consequently, the events of concern are those that take place between the SRD and the ECAL. If the muon carries most of the energy, the missing energy cut will exclude those events. However, events where the neutrino carries significant missing energy, exceeding 30 GeV according to simulations, will pass the selection criteria and could be potential background [36].

5.3 Methodological Framework

In order to be able to make predictions of new unseen particles interacting with our detector setup, the use of precise and fast simulations is a necessity. The conventional method of simulating these interactions for highly complicated detector setups is using methods of sampling. `GEANT4` is a particle simulator software developed at CERN [4] that utilizes MC simulations, and is used in `C++`. The main objective is to be able to reproduce the signal and background events expected for our detector setup. This means studying the following equation:

$$N_i^{Tot} = \sum_{j=runs} n_i(\epsilon, m_i, E_0) \cdot \epsilon_j(\epsilon, m_i, E_0) \cdot N_j^{EOT}, \quad i = \mu\mu, \varphi \tag{5.1}$$

where

$$n_i(\epsilon, m_i, E_0) = \frac{\rho N_A}{A_{Pb}} \sum_j \sigma_{WW}^i(E_e) \cdot n(E_0, E_e, s) \Delta s_j, \tag{5.2}$$

Here, N_i^{Tot} represents the total signal events for either $i = \mu\mu$ or $i = \varphi$. The variable n_i denotes the production rate of i given the selection criteria $\epsilon_j(m_i, \epsilon)$, and N_j^{EOT} is the number of electrons on target (EOT) for run j. Additionally, ρ is the density of the Pb target, N_A is Avogadro's number, A_{Pb} is the atomic mass of Pb, $n(E_0; E_e; s)$ is the number of electrons with energy E_e in the electromagnetic shower at depth s (in radiation lengths) within the target, and $\sigma^i(E_e)$ is the cross section for the production of i, as implemented in the `GEANT4` environment, where the kinematically allowed region extends up to the electron energy E_e.

In order for all of these factors to be properly implemented, careful analysis and comparison of the simulated- and experimental data is performed, where updates and further corrections are implemented to the MC simulation. The cross sections are calculated using the Weizsäcker Williams approximation (WW); the $\mu\mu$ production, being a QED interaction, is reproduced reliably by the simulation, and the DLS φ is implemented using the `GEANT4`-based `DMG4` package, aimed at simulating dark matter particles at fixed target experiments [16]. The primary methods used in this project are standardized and have been performed in previous studies at NA64; this is illustrated as a schematic in figure 5.7. The procedure is implemented in steps and is summarized as an overview here:

1. A MC simulated implementation of the 2023 setup is used to simulate the

interactions. Information and calibration runs are used directly from detector measurements to ensure improved performances in simulation.

2. The raw simulation data is further corrected using reconstruction algorithms to further mimic the detector behaviours and defects.

3. After the reconstruction of data, two main approaches are studied in this thesis (a) Traditional $\mu\mu$ selection and (b) $\mu\mu$ selection using a NN:

   (a) Selection based on optimized selection cuts for $\mu\mu$ selection, see Appendix B.
   (b) Selection based on a trained NN model implemented in `PyTorch` using simulated data, see section 5.4.

4. The selection is compared with selection done on blinded experimental data and is used to further correct the MC simulation.

5. Finally, traditional selection and NN selection techniques are compared with $\sim 10\%$ of experimental data where one looks at detector behaviours and $\mu\mu$ efficiencies.

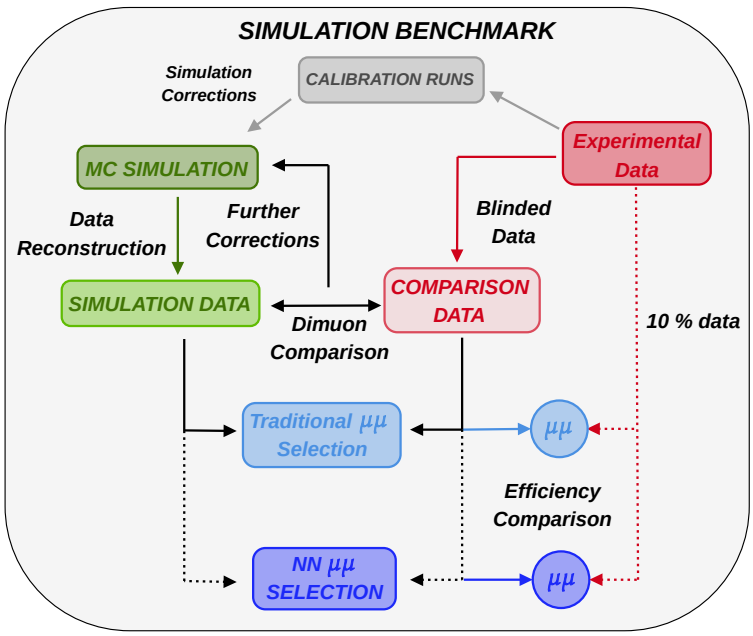


Figure 5.7: A schematic illustrating the benchmarking procedure of the `GEANT4` simulation using $\mu\mu$. Traditional selections on dimuons are performed and used for improving the implementation of simulation. A NN is also trained and applied to data, then compared with the traditional method of selection using $\sim 10\%$ of unblinded data as a means of benchmarking the NN approach.

For a more detailed and comprehensive study similar to the procedures used in this

project, refer to [21]. The code and implementation is also saved in the repository: NN-dimuon-DLS-NA64.

5.3.1 GEANT4 simulation at NA64

The setup of the simulation is described using various C++ classes within the GEANT4 framework, a comprehensive toolkit developed for the simulation of particle interactions with matter [3]. Each detector in the simulation is represented by a logical volume that details its geometry and material properties, including atomic number (Z), mass number (A), and material density. GEANT4 provides extensive capabilities for modeling the geometry and physical processes of detectors. Each logical volume is created with specific geometric shapes, such as cylinders, spheres, or rectangular boxes, and is positioned accurately within the simulation environment. The position of the detectors are based on the measurement survey performed by the survey engineers at CERN with a precision of 0.5 mm at 1σ level.

The logical volumes not only define the geometry but also encapsulate the material characteristics of the detectors. This detailed specification allows GEANT4 to simulate a wide range of physical interactions, including electromagnetic, hadronic, and optical processes. GEANT4's physics models cover an extensive energy range and include various interaction processes, providing a realistic and detailed simulation of particle behavior.

The output of GEANT4 simulations includes energy deposition in scintillators or calorimeters, hit information in tracking chambers, and MC truth information [4]. Energy deposition is recorded and corrected using calibration constants. For tracking detectors, detailed particle information is saved at each step, forming an MChit structure used in reconstruction algorithms. This involves removing low-energy hits, merging close hits, rotating coordinates, mapping detector strips, and adding noise. The reconstructed hits are used for tracking and vertexing, with results saved in a ROOT tree structure for further analysis. This design allows consistent analysis of both simulated and real data.

5.3.1.1 Method of biasing

The cross section and probability of producing a dimuon pair $\mu\mu$ at NA64 is on the order of $\sim 10^{-4} - 10^{-5}$. And given that the experimentally produced number of EOT is on the order of $\sim 10^{11}$, simulating these events without any adjustments would be computationally expensive. To address this issue, a method of biasing is implemented in the simulation code. This technique artificially increases the cross section by a factor of β, thereby enhancing the likelihood of generating these rare events in the simulation. For the dimuon production simulations in this project, a biasing factor of $\beta = 15$ is used. This factor artificially amplifies the production

[3] The simulation tasks were primarily executed using the computing resources of CERN's lxplus infrastructure.

[4] In the simulation setup, true detectors are also placed to accurately record particle interactions without relying on the physical detectors that are implemented in G4.

rate, making it feasible to simulate enough events for a meaningful analysis.

During the analysis phase, this biasing factor is taken into account to correct for the increase in the production rate. Specifically, the results obtained from the biased simulations are scaled down by the same factor ($\beta = 15$) to reflect the true production rate. This correction ensures that the simulated data accurately represents the expected number of events per EOT. Biasing might also introduce unnatural interactions happening outside of the ECAL or more than one dimuon production. By using true information from the produced dimuons, one can remove these events leading to a more realistic data sample.

5.3.1.2 Method of π^- and K^- production

Since the production of the DLS φ background from π^- and K^- is highly unlikely, directly simulating these events would be computationally expensive and inefficient. Instead, a method called *Force Decay* is employed. This technique artificially extends the lifetime of π^- and K^- until they reach the decay region. The decays are then uniformly sampled, with their decay probabilities saved to ensure proper sampling. This method can be divided into two main parts:

1. Calculate the decay probability and path length, and, if necessary, set the particle to be long-lived.

2. Force the decay at the appropriate location.

For π^- and K^- to produce a μ^- such that the muon is unidentified, they need to decay between the end of the SRD (Z_{start}) and the start of the ECAL (Z_{End}) (see section 5.2.3 for more details). The position for the particle decay is denoted Z_{Prod} and the probability for particle decay is given by:

$$P(t) = e^{-t/\gamma\tau} \tag{5.3}$$

where τ is the mean lifetime of the particle at rest and $\gamma = 1/\sqrt{1-\beta^2}$ is the Lorentz factor, $\beta = v/c = E/m$ where E is the total energy of the particle, m is the rest mass of the particle and c is the speed of light. The path length X can be calculated as $X = \tau \cdot \gamma \cdot c \cdot \beta$ and the probabilities as:

$$\begin{aligned} P_{Start} &= e^{-(Z_{Start}-Z_{Prod})/X} \\ 1 - P_{Decay} &= e^{-(Z_{End}-Z_{Start})/X} \\ P_{End} &= e^{-(Z_{End}-Z_{Prod})/X}. \end{aligned}$$

The path to decay X_{decay} is then given as:

$$X_{decay} = -X \cdot \log\left[P_{End} + \mathcal{U}\left(P_{Start} - P_{End}\right)\right] \tag{5.4}$$

where $\mathcal{U}$ denotes a uniform sampling in the decay volume. Since it is more likely for the decay to happen in the beginning of the decay volume, one needs to account

for this when uniformly sampling the decay volume. For this a weight $P_{Weight} = P_{Start} \cdot P_{Decay}$ is saved for each event. The force decay implementation involves first checking conditions to determine if the decay can occur within the designated region. If the particle is set to long-lived, the path length is calculated to determine the probability of decay within the volume. The decay is then forced at the calculated position by adjusting the particle's lifetime, ensuring it decays at the desired point. This process involves several steps:

1. Checking initial conditions to determine if the particle can decay within the specified region.

2. Using the particle's energy and momentum to calculate the path length and decay probability.

3. Sampling the probability distribution to find the decay position within the detector volume.

4. Forcing the decay by adjusting the particle's lifetime to ensure it decays at the chosen position.

This method ensures accurate simulation of rare decay events without the need for excessive computational resources, allowing for efficient and effective π^- and K^- decays within the experimental setup.

5.4 Neural Network-based selection of dimuons at NA64

The method of using a NN selection still follows the steps described in section 5.3. However, not all detectors are relevant for selecting muons. The chosen detectors, referred to as the features of our model, are numbered as $i = 1, \ldots, N$. Training an NN model directly on the raw data, h_i, is not ideal. Therefore, an initial filtering process, similar to data selection, is performed to create a dataset more suited for optimizing the selection of muons. This filtered dataset is represented as $\{x_i\}$. The selection process ensures that most of the events are electrons by utilizing rejection methods to purify the beam. These selected events are then embedded into a format, X_{event}, that is well-suited for the NN. The NN is then trained on this prepared dataset, resulting in a final model, $f^{\text{event}}(\{x_i\})$. A schematic displaying this process is shown in figure 5.8.

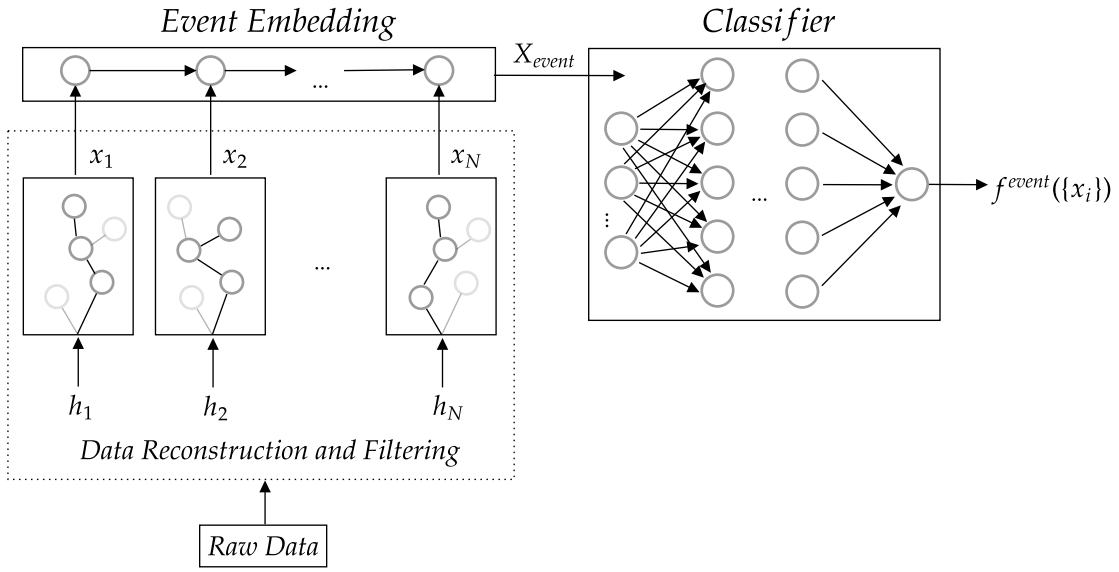


Figure 5.8: Schematic showing the implementation of the classifier using a NN. Initial step show how the raw data h_i undergoes a primary filtering and selection resulting in trainable data x_i. This data is then embedded and combined before it is fed into training the classifier, resulting in the model f^{event}.

5.4.1 Data reconstruction and filtering

The raw data originates from separate MC simulations producing dimuons, pions, kaons and *standard events*[5] as described in section 5.3.1. These simulations are then combined in balanced datasets, meaning that distribution of data across the different classes is approximately equal, aimed at improving the overall performance and reliability of the model [29]. This is refered to as the *raw data* for the training of the NN.

The raw data then undergoes a primary data reconstruction and filtering, simply to ensure proper identification of electrons and to reduce pile-up events. The relevant features for selecting muons are derived mostly from the downstream detectors, these are: the ECAL (active target), VETO_{01}, VETO_{23}, VETO_{45} (before the HCALs), the central cells of the HCAL_{0-2}, STD_{11}, and STD_{12}. Among these, specifically the VETO, STD and HCAL modules, play a crucial role in the selection process of $\mu\mu$ and signal events. The HCAL modules are essential for detecting the double-MIP signature characteristic of muon pairs, and the STDs are critical for tracking and identifying the events. The precision and efficiency of these downstream detectors ensure the effective selection and analysis of dimuons, significantly enhancing the accuracy of signal event detection. More on the phenomenology and expected signal signatures in section 5.2.

[5]Standard events are primarily electrons that strike the ECAL and come to a stop, without contributing to signal or background events.

5.4.2 Architecture and implementation

The NN consists of five fully connected layers $\{10, 128, 64, 32, 16, 1\}$ with 10, 128, 64, 32, 16, and 1 units respectively, the input layer takes in 10 variables and is connected to 4 hidden layers each followed by ReLU activation and dropout, with a Sigmoid activation in the final layer. The Sigmoid activation in the final layer outputs probabilities that facilitate binary classification by predicting whether each input belongs to one of two classes, signal or not signal. This is implemented using the `PyTorch` library. The network is trained using the Adam optimization algorithm and the BCE loss function (see section 4 for more information). The training performance is compared across different models looking at different architectures, optimizers and loss functions. This is benchmarked by evaluating the Receiver Operating Characteristic (ROC) curve and its Area Under Curve (AUC) score, as well as looking at the accuracy score for both the training and test datasets. The ROC curve displays the True Positive Rate (TPR) over the False Positive Rate (FPR) for all the different classification thresholds between $[0, 1]$, and the AUC score is the area of this curve. This approach is a good way of measuring the performance of the binary classification and helps to identify potential overfitting effects.

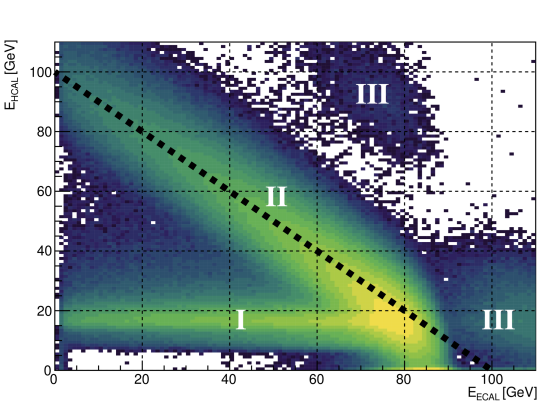
Table 5.1: Information about the NN and the different parameters that are used to train the model.

NN information	
Optimizer	**ADAM**
Node activation	**ReLu**
Output node activation	**Sigmoid**
Number of input variables	**10**
Number of hidden layers	**4**
Total number of epochs	**1300**
Size of test sample [%]	**0.2**
Stopping threshold	$\mathbf{10^{-7}}$
Learning rate	**0.001**
Dropout percentage	**0.5**
Number of trainable parameters	**12289**

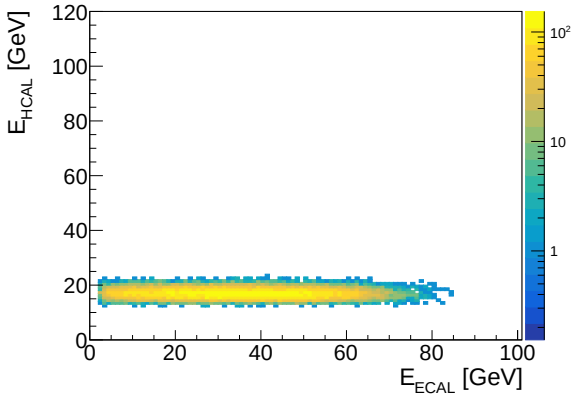
6 | Results

The following section presents the results from the MC simulations with comparisons to the data taken in the 2023 electron run. The comparison on dimuons serves as a validation and benchmark for the MC simulation implemented in `GEANT4` which is important to complete before studying the selection criteria, the signal efficiency, the background level and before training the NN. The results mainly show the energy distribution that is deposited on our downstream detectors as well as the total efficiencies of dimuon production. The configuration of these detectors are also analyzed and optimized for a higher dimuon yield. Lastly is the results on the trained NN model for $\mu\mu$ selection.

To provide an overview of the experimental results, hermeticity plots are used to illustrate different events by plotting the $[E_{ECAL}, E_{HCAL}]$ plane for a sample of selected data, as shown in Figure 6.1. An easy way to understand this plot is to look at region (II), which represents energy-conserving events where $E_{ECAL} + E_{HCAL} \simeq 100$ GeV. These events refer to hadronic production and other interactions that leave some energy deposition in the ECAL and have a produced hadron impinging on the HCAL, resulting in a total energy deposition of ~ 100 GeV. The most common events are those where the electron simply stops and is fully absorbed in the ECAL. Region (III) refers to *pile-up* events, where more than one electron passes through the selection, leaving more energy than expected. The last events, that are most relevant for this project, are the dimuons labeled as (I). These events show a clear total energy deposition in the HCAL of ~ 15 GeV, where the muons carry away some of the energy, leading to a stripe of events with energy deposition in the ECAL between $3 - 80$ GeV. The *blank* region, seen in figure 6.1a, is the blinded region. This is where the anticipated signal events should appear after all data selection. Following a thorough analysis of the selection criteria, the selection should be applied to see if any events appear in the signal box, as defined in Section 5.2.1. The goal now is to optimize the selection of dimuons, ensuring that events like (II) and (III) are rejected. The primary objective is to investigate whether a NN trained on simulated data can more accurately select dimuons. If successful, this NN could be used to train a NN for selecting the DLS φ, potentially leading to a higher signal yield.



(a) Hermeticity plot produced by NA64 [33] where (I) refers to dimuons, (II) refers to hadronic production, mostly pions and kaons in ECAL and (III) are pileup events.



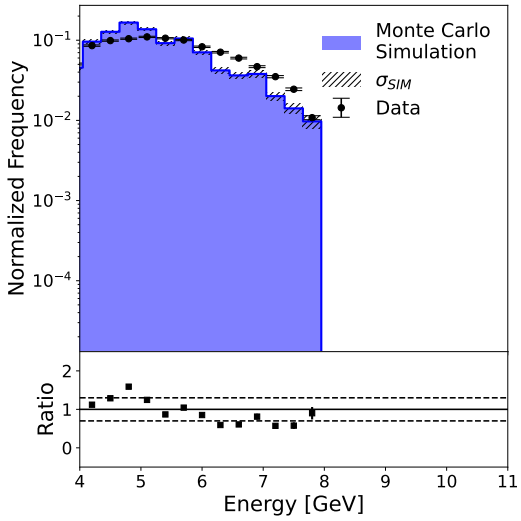
(b) Small subset of blinded experimental data taken 2023. Focus is on dimuon selection, with the selection criteria described in appendix B.

Figure 6.1: Two hermiticity plots depicting different event selection, (a) is a general event selection (b) depicts a more specific selection of dimouns. Each dot in the plot denotes an event with total HCAL energies E_{HCAL_i}, $i = 0, 1, 2$ on the y-axis and total ECAL energy on the x-axis.

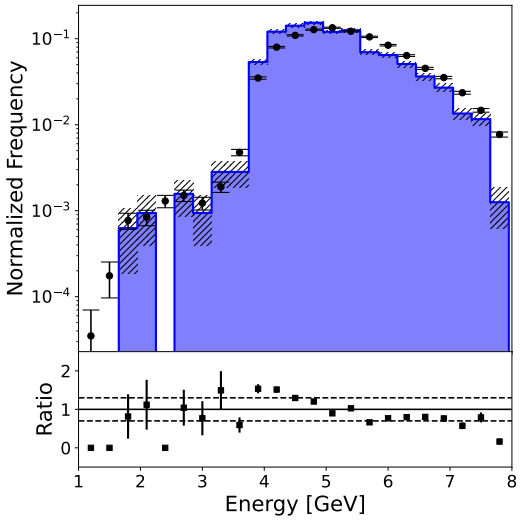
6.1 Comparison and validation of Monte Carlo simulation

The comparison between the simulated and real data is done with the most common and crucial detectors being the HCALs and STDs. The simulation is performed producing a total of $\sim 1.5 \cdot 10^{10}$ EOT. The real data is taken from the 2023 data from period $2-5$ with a total of $4.55 \cdot 10^{11}$ EOT, where a blinding is performed on each HCAL cell that requires a minimum of 4 GeV energy deposition. This is to blind the signal from models like the dark photon A', Axion Like Particles (ALP) and the DLS φ. The complete selection criteria is described in appendix B. The error of each bin is calculated as $\sqrt{N_i}$, where N_i is the total number of elements in bin i, and is denoted σ_{SIM} for the simulation data. The ratio plot describes the ratio N_{SIM}/N_{EXP} for each bin. The results for the dimuon selection show two distinct peaks, referred to as the single and double MIP peaks, appearing around ~ 2.5 GeV and ~ 5 GeV respectively. These peaks correspond to dimuon events, with the requirement that their combined total energy matches the production energy. However, the individual energies of the muons passing through are not always identical. The single MIP peak appearing in HCAL_1 and HCAL_2 indicates that one of the muons was absorbed in the HCAL before it, resulting in a single MIP. Note that the uncertainty in the single MIP measurement is larger, due to the lower probability of a muon being absorbed. An important detail is that the plots only show the energy deposition in the central

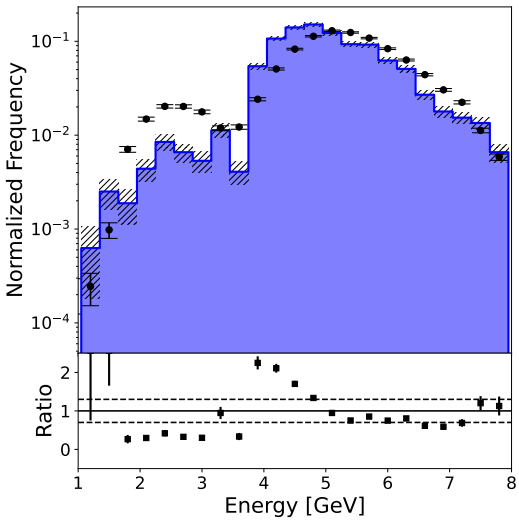
cell of the HCALs, focusing on this area due to larger discrepancies observed in the lateral cells, as shown in figure 6.3.



(a) Central cell of $HCAL_0$.



(b) Central cell of $HCAL_1$.

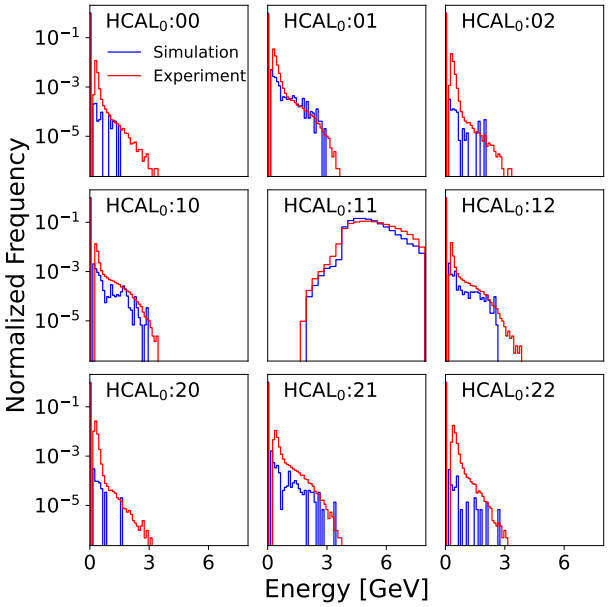


(c) Central cell of $HCAL_2$.

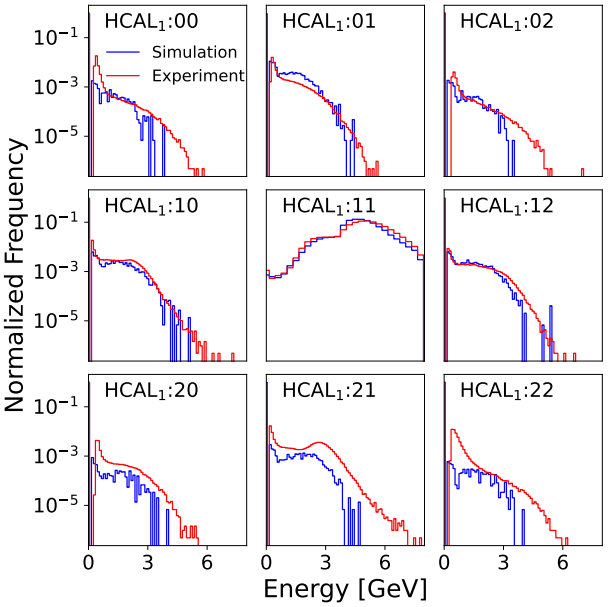
Figure 6.2: Comparison between MC simulated data and real data of $\mu\mu$ selection of the central cells of HCALs.

The results show good agreement for the central cells, as seen in figure 6.2. However, when examining the total energy for the HCAL, a discrepancy is visible. In the peripheral cells, the simulated data for each HCAL generally shows slightly lower energy deposition compared to the real data, resulting in a shift of around $\sim$ 2 GeV in total energy when summing all cells, as seen in figure 6.3d. Additionally, one can see a slight discrepancy also in the central cells, suggesting that the energy deposition does not fully match experimental results. This discrepancy could be

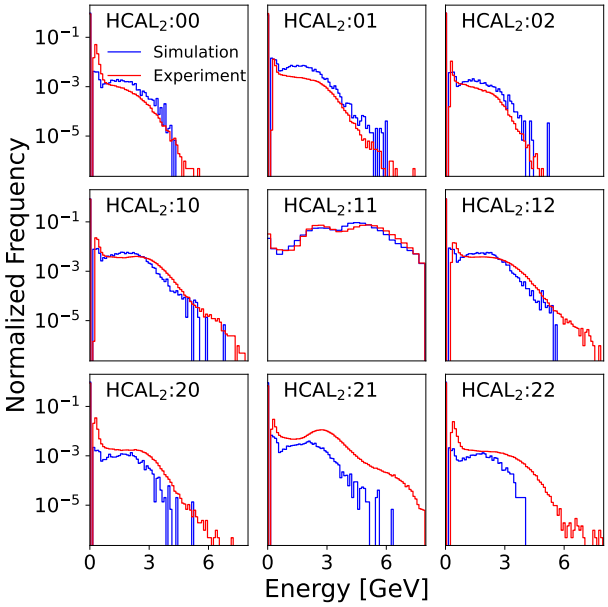
due to slight misalignments of the HCALs, even though they were implemented according to the measurement survey.



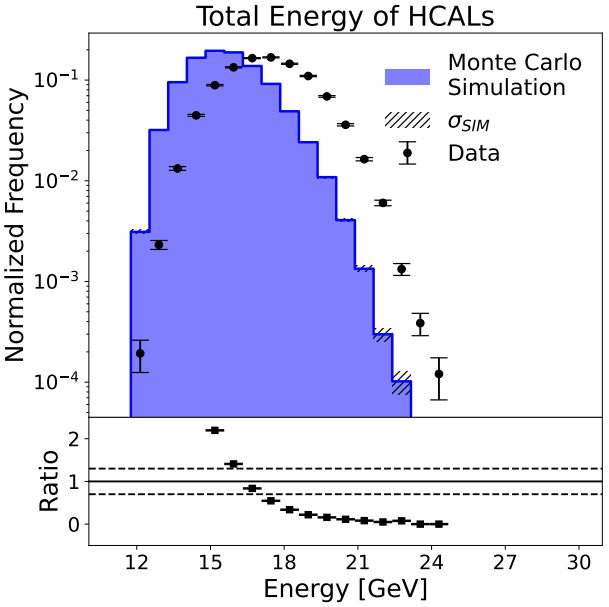
(a) Energy distribution of all the cells in HCAL$_0$.



(b) Energy distribution of all the cells in HCAL$_1$.



(c) Energy distribution of all the cells in HCAL$_2$.



(d) Total Energy of HCAL$_{0-2}$

Figure 6.3: Comparison between $\mu\mu$ selection on the modules from all the HCALs using Monte-Carlo simulated data (blue) and experimental data from 2023, using period 2-5.

Another possible reason is improper calibration of the HCAL cells during data collection. Proper calibration runs at these energy scales could correct the errors and reduce the overall discrepancies.

Accurate simulation behavior and results are crucial for using the straw detectors in training the NN. The STD_{11} and STD_{12} were positioned to align their peaks. Additionally, the magnetic fields bending the beam were adjusted to match the hits in the straw detectors. The first magnet modifies the central position of the spots, while the second magnet alters the separation between the two peaks. Notably, the simulated magnets do not correspond one-to-one with the actual magnets; no field map has been implemented, and a simple uniform magnetic field is currently applied. The best results are obtained with magnetic field strengths of $B_1 = -1.85$ T/m for the first magnet and $B_2 = -2.15$ T/m for the second magnet implemented in simulation. These adjustments lead to good agreement and behavior, as shown in Figure 6.4. Additionally, the iron blocks surrounding the magnets, which would absorb muons in the actual setup, are not included in the simulation. To address this, a simple cut is applied in the simulation to account for the muon absorption.

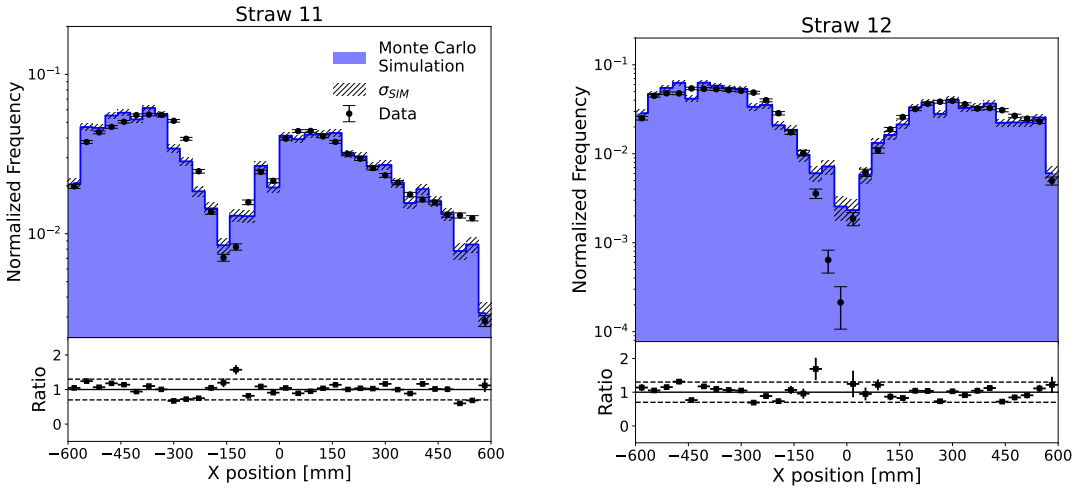


Figure 6.4: Comparison between MC simulated $\mu\mu$ hits on ST_{11} & ST_{12} with experimental data from 2023, period 2-5. $\mu\mu$ selection is applied on both dataset.

Most of the selection that are being applied before the HCALs have not changed in the process of this project. These selections simply makes sure that the events that hit the ECAL are electrons, where they reduce pile-ups and purifies the beam. The full event selection can be seen in appendix B. The main selection that have been of importance to this project are the selections done in the VETO, HCAL and STD. A simple cutflow illustrating the most important selection for this project is shown in table 6.1, in addition are extra figures for the ECAL, MM and VETO detector shown in appendix C.

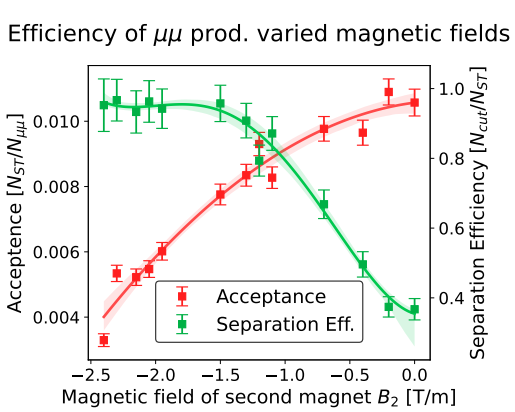
Table 6.1: Cutflow of $\mu\mu$ events using a magnetic field of $B_2 = -1.5$ T/m.

	Cut Information	Events	Rel. Eff.	Total $\mu\mu$ eff.
Total EOTs	Bias = 15	$1.35 \cdot 10^9$	–	–
Total $\mu\mu$ events	1 $\mu\mu$ production	56368	$\sim 10^{-5}$	100 %
After HCAL	$8 > \text{HCAL}_i > 4$ GeV, $i = 0, 1, 2$	9252	16.4 %	16.4 %
Magnet Acceptance	$130 > \text{Y}(\text{STD}_{11}) >$ -130 mm	2728	29 %	4.8 %
Straw Acceptance	No hit	289	11 %	–
	Background	630	23 %	–
	Single μ hit	1368	50 %	–
	$\mu\mu$ hit	441	16 %	0.78 %
$\mu\mu$ Separation	μ^- right, μ^+ left	414	92 %	0.73 %

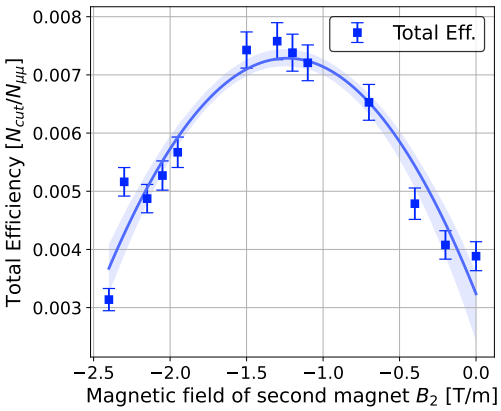
Table 6.1 show clearly that the biggest bottleneck is coming from the HCAL and Magnet acceptance. This is simply due to the fact that many dimuons are absorbed in the HCALs and that the lower energetic beam gets more and more deflected making it less probable for the dimuon pair to hit the straw detectors in the sought after positions.

6.1.1 Analysis on straw detectors and magnet for optimization of $\mu\mu$ production

The comparison between the straw peaks in experiment and simulation seemed to agree very well when the magnetic field of the first magnet is -1.85 T/m and second magnet is -2.15 T/m. However, this is not necessarily the most optimized configuration for a high dimuon yield. By sweeping over different magnetic field strengths B_2, one can find an optimized field strength. The acceptance of the selection will obviously increase for lower B_2, this is due to lower energetic particles not being bent away. However, having a too low magnetic field will also make it harder for the separation of the dimuons. The two effects are therefore investigated and the results are shown in figure 6.5. Each simulation uses a total of $1.5 \cdot 10^6$ giving a total of $\sim 9.75 \cdot 10^6$ EOT when accounting for the bias $\beta = 15$.



(a) Acceptence and seperation efficiency for different magnetic field strengths of the second magnet B_2.



(b) Total efficiency for different magnetic field strengths of the second magnet B_2. The peak seems to lie between $1.0 - 1.5$ T/m.

Figure 6.5: Total $\mu\mu$ efficiency is presented in figure (b) which is produced from the (separation $\times$ acceptance) efficiency, shown in (a).

The separation efficiency is seen to reach almost perfect separation of $\sim$ 94% efficiency with $B_2 < -1.5$ T/m, beyond which it starts to saturate. Increasing the magnetic field strength does not improve the separation. Ideally one would use a value of ~ -1.5 T/m to ensure a high separation efficiency, one could use ~ -1.0 T/m, which would give a similar total efficiency. However, the separation efficiency is lower which would increase the background of other events.

Table 6.2 displays the true cutflow of events hitting the STD_{11} and STD_{12}. This analysis uses true detectors and shows a decrease in *No hit* events, while the *background* and *Single μ hit* events remain mostly unchanged. This suggests an overall increase in acceptance, with $\mu\mu$ efficiency improving from 0.48% to 0.73%. It is important to note, however, that this selection pertains to $\mu\mu$ events, not DLS φ events. The efficiency may change when considering single muon events. Furthermore, the energy distribution of the muons is crucial for determining the appropriate magnetic field strength. According to [37], the expected muons from DLS φ production typically have energies in the range of 20-80 GeV, whereas dimuons produced in the ECAL peak at around 10-20 GeV.

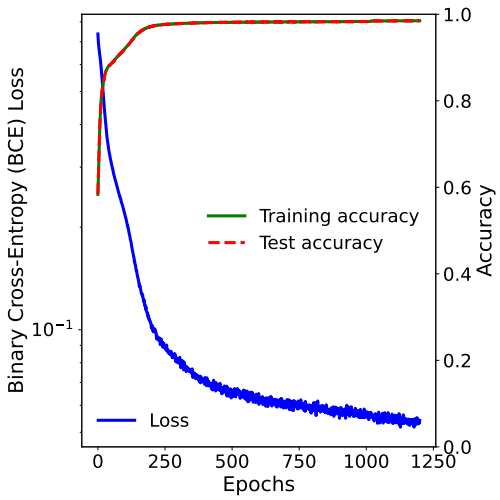
Table 6.2: Cutflow of $\mu\mu$ events using a magnetic field of $B_2 = -1.5$ T/m and $B_2 = -2.15$ T/m. Red and green markings show the improvement between the 2023 and the optimized detector setup.

$B_2 = -1.5$ T/m	Cut Information	Events	Rel. eff.	Total $\mu\mu$ eff.
Straw Acceptance	No hit	289	11 %	–
	Background	630	23 %	–
	Single μ hit	1368	50 %	–
	$\mu\mu$ hit	441	16 %	0.78 %
$\mu\mu$ Separation	μ^- right, μ^+ left	414	92 %	**0.73 %**
$B_2 = -2.15$ T/m		Events	Rel. eff.	Total $\mu\mu$ eff.
Straw Acceptance	No hit	639	22 %	–
	Background	558	19 %	–
	Single μ hit	1453	49 %	–
	$\mu\mu$ hit	295	10 %	0.51 %
$\mu\mu$ Separation	μ^- right, μ^+ left	280	94 %	**0.48 %**

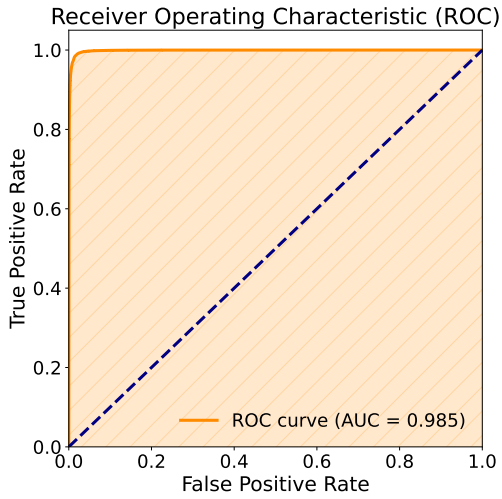
By considering $\mu\mu$ events as signal and kaon (κ) and pion (π) events as background, the signal-to-background ratio can be calculated and optimized. By weighting the events as described in 5.3.1 and selecting those that maximize this ratio, the final result shows a true positive rate of approximately 97% and a false positive rate of 0.51%

6.2 Neural Network-based selection of $\mu\mu$ events

The training of the NN is based on the properties described in section 5.4 and 4. The results of the NN training is displayed in figure 6.6 in terms of the loss function and accuracy scores for each epoch. Additionally, a ROC curve is also shown for evaluating the binary classification. In this analysis, both the training and test accuracies are displayed over the course of approximately 1300 epochs. The test accuracy reaches a high score of 0.98, while the loss decreases steadily until the stopping threshold is met at 10^{-7}. The result of this is shown in figure 6.6a.



(a) Accuracy score and BCE loss of the NN model.



(b) ROC curve of the trained NN model.

Figure 6.6: Benchmarking results of a NN trained to select $\mu\mu$ events.

The ROC curve illustrates the performance of the trained model by plotting the TPR against the FPR at various classification thresholds. In a binary classification model like our NN, the classification threshold is the predicted value that the model evaluates given the event, ranging between $[0, 1]$. Adjusting this threshold can control the trade-off between detecting true positives and minimizing false positives, with a higher threshold typically reducing the background noise. If the model's predictions were entirely random, the ROC curve would align with the blue dashed line, indicating random selection of events. However, a well-performing model will have a ROC curve that rises towards the top-left corner of the plot, indicating a high TPR for a wide range of classification thresholds. The effectiveness of the model is quantified by the AUC score; a higher AUC signifies that the model maintains a high TPR while keeping the FPR low across different thresholds. The NN trained model for this project reaches an AUC score of 0.985 and the ROC curve is displayed in figure 6.6b showing an overall good result.

To further analyze the model, we plot the ratio between dimuon and non-dimuon selections for various classification thresholds (Predicted Probability), as shown in Figure 6.7. In binary classification, achieving good separation between signal and background is a primary objective. This plot illustrates how the background presence varies with the chosen classification threshold. It clearly demonstrates that a higher threshold increases the signal purity and reduces the background noise. However, setting the threshold too high can result in too few selected events. This flexibility is a significant advantage of using a trained binary classification model over traditional event selection methods. With traditional methods, systematically increasing the data while monitoring the background is not straightforward. In contrast, a NN model allows for easy adjustment of the threshold to control the amount of background, with the background level being quantifiable at all times. This abil-

ity to dynamically adjust the threshold provides a powerful tool for optimizing the balance between signal and background.

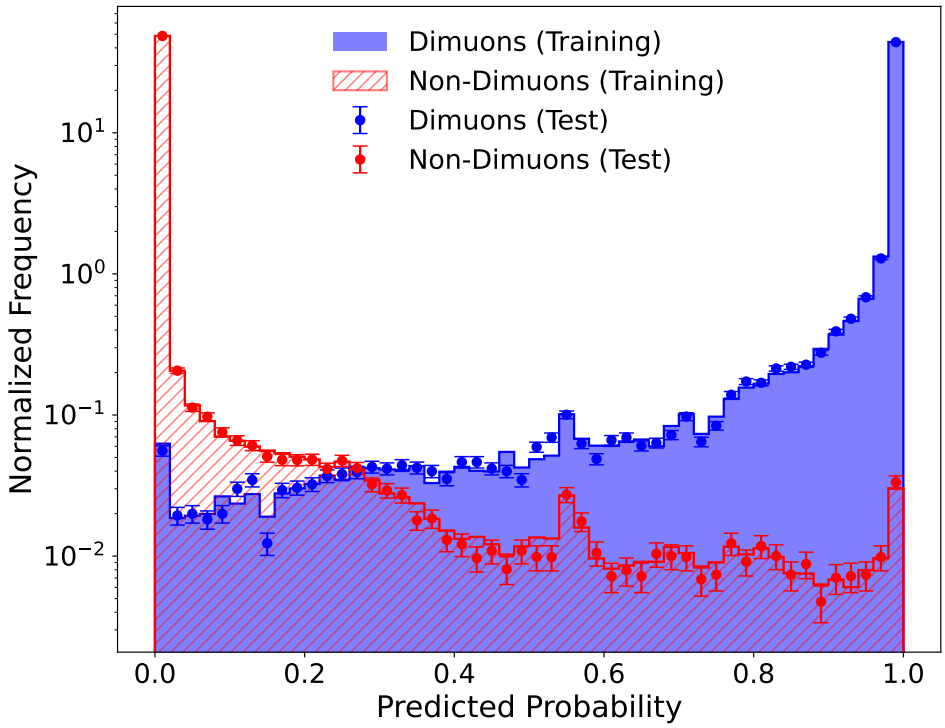
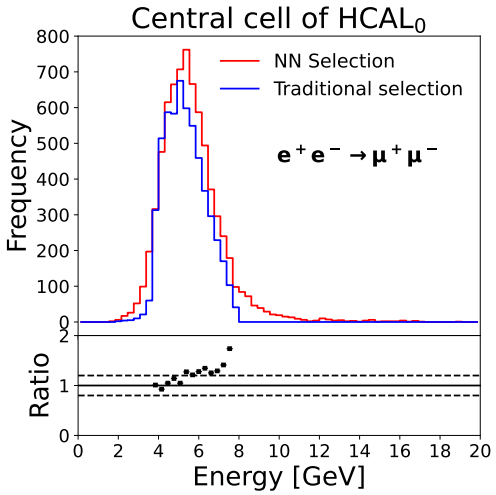
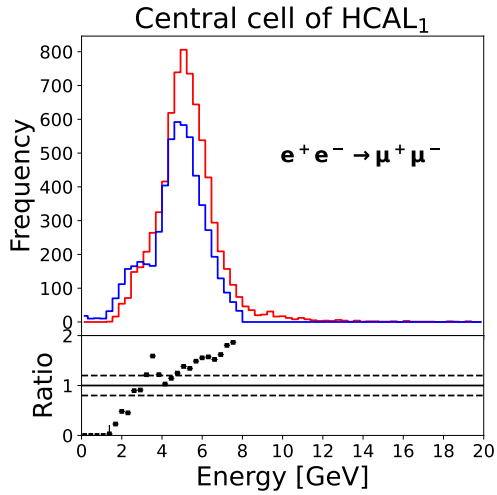


Figure 6.7: Background versus signal selection of the NN model's $\mu\mu$ selection for different classification thresholds. Note that the y-axis is displayed in log-scale.

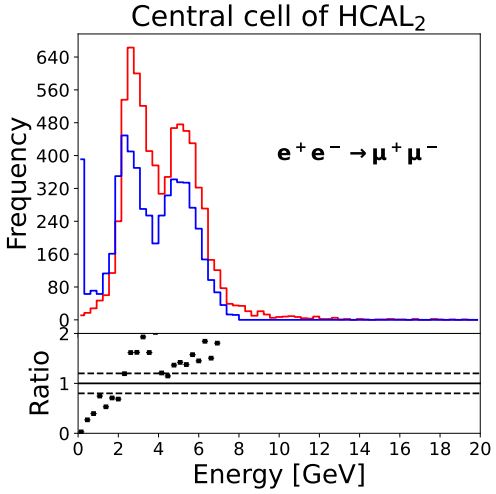
The most definitive way to evaluate the model is to apply it to real data. Using a small sample of unblinded data from run 8747, both the optimized traditional selection and the trained NN can be compared to determine if there are any improvements and a higher $\mu\mu$ yield. To ensure that the model performs as intended, the sample was first pre-selected using the same filtering and selection that was used for training the simulation data, see section 5.4. This is to ensure that there are no pile-up and other unwanted events as described previously. The result of this is shown in figure 6.8, which shows the comparison between the the HCALs and VETO_{23} detectors. An overall increase in the number of events is seen.



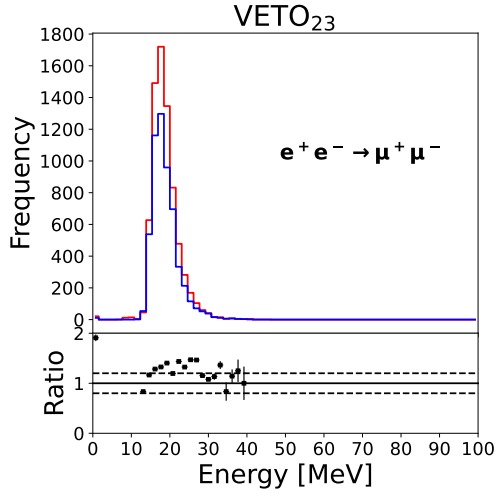
(a) Central cell of $HCAL_0$.



(b) Central cell of $HCAL_1$.



(c) Central cell of $HCAL_2$.



(d) $VETO_{23}$.

Figure 6.8: Comparison between traditional- and NN selection of dimuons with $HCAL_i$, $i = 0, 1, 2$ and the $VETO_{23}$ on experimental data from 2023 on run 8747. The NN model utilizes a classification threshold of 0.5 giving 98.2% TPR and 0.55% FPR.

With a classification threshold of 0.5 giving 98.2% TPR and 0.55% FPR, the NN selects a total of 1955 events, whereas the traditional selection selects 1442 events. Furthermore, by requiring dual hits in both STD_{11} and STD_{12}, the NN produces 31 events compared to 17 events from the traditional selection. This demonstrates an overall increase in the number of selected dimuons using the NN approach.

7 | Conclusion

The search for a particle candidate that could explain the origin of dark matter is a central goal in modern astro-particle physics. Numerous experiments employing various measurement strategies are being developed in an attempt to understand better understand this phenomenon. The primary purpose of this project was studying the background for a DLS model using a highly accurate MC simulation for the NA64 experiment. More precisely, the `GEANT4` software was utilized to simulate the experimental setup used in 2023. The results of these simulations were then compared with real data taken in 2023. The results show good agreement when looking at the central cells of the HCAL modules shown in figure 6.2 along with the VETO_{01}, VETO_{23}, VETO_{45}, ECAL which are shown in appendix C and STD_{11} and STD_{12} which are shown in figure 6.4. However, to ensure better results one needs to make sure that the full HCAL modules show better agreement with real data, this was illustrated in figure 6.3. Improvements on this can be made by implementing calibration runs for the HCAL modules.

The simulation results were important in refining the methods of event selection. The primary sources of background for DLS particle φ include $\mu\mu$ production, kaon κ, and pion π decay. These processes were simulated and subsequently used to train a NN for selecting $\mu\mu$ events, serving as a benchmark for the method. The proper use of the NN would be to select a DLS φ. This is illustrated in a schematic in figure 7.1. The selection of $\mu\mu$ events using the trained NN was then compared to traditional selection methods which resulted in an overall increase of 36% of the total number of data. When also requiring subsequent hits in both ST_{11} and ST_{12} we see an increase from 17 final events when using the traditional selection to 31 final events when using the NN selection. It is important to note that the signal-to-background ratio for both methods were comparable, but the NN gives more data which means that the NN did a better job in selecting signal. The detector results are shown in figure 6.7. This demonstrates the potential of NN's to enhance event selection processes. A future study could focus on simulating the DLS φ particles and training a NN to optimize event selection.

The issue of working with a trained deep learning model becomes apparent when trying to assess the results and analyze the selection process, especially when selecting DM candidates that have never been observed. This challenge is known as the 'black box' problem [22], where the internal decision-making process of the NN is not fully understood. What is clear is that the model's free parameters (weights and bias) are adjusted to minimize a loss function based on simulation data. Some might argue

that optimized data selection should be possible using logic and reasoning, which has significant merit, however, if a well-crafted simulation is conducted—where all the necessary logic and reasoning is embedded—is it essential to understand how the data is selected during the analysis stage? It is much more time efficient to simply train a model to select on the well made simulation. The counter argument on this however, is that no simulation can be perfect. Therefore a lot of emphasis should be placed in understanding how discrepancies between simulation data and real data affect the event selection using the NN model.

One way to keep a level of impartialness would be to implementing pre-defined 'safe' or 'loose' cuts. These should correspond to the acceptable thresholds that each detector can accommodate for a given signal signature, ensuring that any data falling outside these boundaries is excluded from the NN's output. This approach helps safeguard the model against biased selection, providing a more reliable and interpretable selection process.

In conclusion, a NN model trained on simulated data has demonstrated the potential to enhance event selection yield compared to traditional methods. However, thorough and careful analysis of the model is essential before it can be applied to unblinded real data. Future research should focus on training an NN specifically on the DLS φ signal in conjunction with the already simulated backgrounds from $\mu\mu$ events, kaons κ, and pions π. This approach could further refine the selection process and improve the overall detection efficiency for DM candidates.

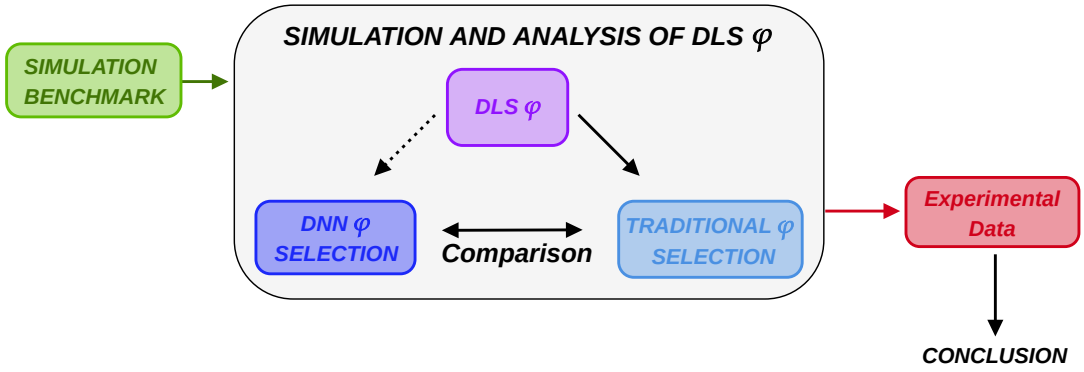


Figure 7.1: Schematic showing that the benchmarked simulation setup can be used to simulate the DLS φ. Traditional event selection is analyzed and optimized, followed by training and comparison of a NN model with simulated data. The final analysis applies either the model or traditional selection to data, and a sensitivity study is conducted to draw conclusions of a discovery.

Bibliography

[1] S. Abachi, B. Abbott, M. Abolins, et al. "Observation of the Top Quark". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632.

[2] F. Abe, H. Akimoto, A. Akopian, et al. "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626.

[3] P. A. R. Ade et al. "Planck2015 results: XIII. Cosmological parameters". In: *Astronomy; Astrophysics* 594 (Sept. 2016), A13. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201525830.

[4] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 2003), pp. 250–303. ISSN: 0168-9002. DOI: 10.1016/s0168-9002(03)01368-8.

[5] D. P. Aguillard et al. "Detailed report on the measurement of the positive muon anomalous magnetic moment to 0.20 ppm". In: *Physical Review D* 110.3 (Aug. 2024). ISSN: 2470-0029. DOI: 10.1103/physrevd.110.032009. URL: http://dx.doi.org/10.1103/PhysRevD.110.032009.

[6] Jim Alexander et al. *Dark Sectors 2016 Workshop: Community Report*. 2016. DOI: 10.48550/ARXIV.1608.08632.

[7] Yu. M. Andreev et al. "Search for Light Dark Matter with NA64 at CERN". In: *Phys. Rev. Lett.* 131 (16 Oct. 2023), p. 161801. DOI: 10.1103/PhysRevLett.131.161801.

[8] G. Arnison, A. Astbury, B. Aubert, et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at $s = 540$ GeV". In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91177-2.

[9] G. Arnison, A. Astbury, B. Aubert, et al. "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider". In: *Physics Letters B* 126.5 (July 1983), pp. 398–410. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)90188-0.

[10] P. Bagnaia, M. Banner, R. Battiston, et al. "Evidence for $Z_0 \to e^+e^-$ at the CERN p collider". In: *Physics Letters B* 129.1–2 (Sept. 1983), pp. 130–140. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)90744-x.

[11] C. A. Baker et al. "Improved Experimental Limit on the Electric Dipole Moment of the Neutron". In: *Physical Review Letters* 97.13 (2006), p. 131801.

[12] D. Banerjee et al. “Search for vector mediator of dark matter production in invisible decay mode”. In: *Physical Review D* 97.7 (Apr. 2018). ISSN: 2470-0029. DOI: 10.1103/physrevd.97.072002.

[13] M. Banner, R. Battiston, Bloch, et al. “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN p collider”. In: *Physics Letters B* 122.5–6 (Mar. 1983), pp. 476–485. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)91605-2.

[14] James B. R. Battat. “Resource Letter DM1: Dark matter: An overview of theory and experiment”. In: *American Journal of Physics* 92.4 (Apr. 2024), pp. 247–257. ISSN: 1943-2909. DOI: 10.1119/5.0187346.

[15] Gianfranco Bertone and Dan Hooper. “A History of Dark Matter”. In: (2016). DOI: 10.48550/ARXIV.1605.04909.

[16] M. Bondi et al. “Fully Geant4 compatible package for the simulation of Dark Matter in fixed target experiments”. In: *Computer Physics Communications* 269 (Dec. 2021), p. 108129. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2021.108129.

[17] S. Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021.

[18] CMS Collaboration. *Dark sector searches with the CMS experiment*. 2024. DOI: 10.48550/ARXIV.2405.13778.

[19] Richard Connor et al. “Correlations of Cross-Entropy Loss in Machine Learning”. In: *Entropy* 26.6 (June 2024), p. 491. ISSN: 1099-4300. DOI: 10.3390/e26060491.

[20] PyTorch Contributors. *torch.optim.Adam*. https://pytorch.org/docs/stable/generated/torch.optim.Adam.html. Accessed: 2024-07-10. 2023.

[21] Depero, Emilio. “Searching for Light Dark Matter and a new X17 boson with the NA64 experiment at the CERN SPS”. en. PhD thesis. 2020. DOI: 10.3929/ETHZ-B-000458684.

[22] James E. Dobson. “On reading and interpreting black box deep neural networks”. In: *International Journal of Digital Humanities* 5.2–3 (Nov. 2023), pp. 431–449. ISSN: 2524-7840. DOI: 10.1007/s42803-023-00075-w. URL: http://dx.doi.org/10.1007/s42803-023-00075-w.

[23] T. Faber and M. Visser. “Combining rotation curves and gravitational lensing: how to measure the equation of state of dark matter in the galactic halo”. In: *Monthly Notices of the Royal Astronomical Society* 372.1 (Oct. 2006), pp. 136–142. ISSN: 1365-2966. DOI: 10.1111/j.1365-2966.2006.10845.x.

[24] Katherine Freese. “Status of Dark Matter in the Universe”. In: (2017). DOI: 10.48550/ARXIV.1701.01840.

[25] Y. Fukuda, T. Hayakawa, E. Ichihara, et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[26] S.N. Gninenko and N.V. Krasnikov. “Leptonic scalar portal: Origin of muon g-2 anomaly and dark matter?” In: *Physical Review D* 106.1 (July 2022). ISSN: 2470-0029. DOI: 10.1103/physrevd.106.015003.

[27] Dan Guest, Kyle Cranmer, and Daniel Whiteson. "Deep Learning and its Application to LHC Physics". In: (2018). DOI: 10.48550/ARXIV.1806.11484.

[28] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13.16 (1964), p. 508.

[29] Justin M. Johnson and Taghi M. Khoshgoftaar. "Survey on deep learning with class imbalance". In: *Journal of Big Data* 6.1 (Mar. 2019). ISSN: 2196-1115. DOI: 10.1186/s40537-019-0192-5.

[30] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. DOI: 10.48550/ARXIV.1412.6980.

[31] Yu-Sheng Liu, David McKeen, and Gerald A. Miller. "Validity of the Weizsäcker-Williams approximation and the analysis of beam dump experiments: Production of a new scalar boson". In: *Phys. Rev. D* 95 (3 Feb. 2017), p. 036010. DOI: 10.1103/PhysRevD.95.036010.

[32] R. N. Mohapatra and Palash B. Pal. *Massive Neutrinos in Physics and Astrophysics*. Third. Singapore: World Scientific, 2006.

[33] Laura Molina Bueno, Paolo Crivelli, and The NA64 collaboration. *NA64 status report 2023*. Tech. rep. Geneva: CERN, 2023. URL: https://cds.cern.ch/record/2868332.

[34] Roberto D. Peccei and Helen R. Quinn. "CP Conservation in the Presence of Pseudoparticles". In: *Physical Review Letters* 38 (1977), pp. 1440–1443.

[35] Planck Collaboration et al. "Planck 2018 results. VI. Cosmological parameters". In: (2018). DOI: 10.48550/ARXIV.1807.06209.

[36] A. Ponten et al. *Probing Hidden Leptonic Scalar Portals using the NA64 Experiment at CERN*. 2024. DOI: 10.48550/ARXIV.2404.15931.

[37] Pontén, Axel. "Sensitivity study of a dark leptonic scalar portal at NA64". en. In: (2022). DOI: 10.3929/ETHZ-B-000584300.

[38] Stefano Profumo. *TASI 2012 Lectures on Astrophysical Probes of Dark Matter*. 2013. DOI: 10.48550/ARXIV.1301.0952.

[39] R. et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020.

[40] Robert H Sanders. *The dark matter problem*. en. Cambridge, England: Cambridge University Press, Apr. 2010.

[41] *Secondary Beams & Areas*. 2024. URL: http://sba.web.cern.ch/sba/.

[42] V. Yu. Volkov et al. "Straw Chambers for the NA64 Experiment". In: *Physics of Particles and Nuclei Letters* 16.6 (Nov. 2019), pp. 847–858. ISSN: 1531-8567. DOI: 10.1134/s1547477119060554.

[43] Steven Weinberg. "A New Light Boson?" In: *Physical Review Letters* 40.4 (1978), pp. 223–226.

[44] Frank Wilczek. "Problem of Strong P and T Invariance in the Presence of Instantons". In: *Physical Review Letters* 40.5 (1978), pp. 279–282.

[45] Fritz Zwicky. "On the Masses of Nebulae and of Clusters of Nebulae". In: *Astrophysical Journal* 86 (1937), p. 217.

[46] P A Zyla et al. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020). ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104.

A | Exact tree-level calculation of DLS production $eZ \to \varphi Z \mu$

Refer to the Feynman diagram in figure 3.2. We assume that the nucleus has zero spin [31], as a result the photon-nucles vertex is given by

$$ieP^{\mu}F(t) = ie\left(P_i^{\mu} + P_f^{\mu}\right)F(t) \tag{A.1}$$

where $F^2(t) \equiv G_2^{elastic}(t)$ is the squared elastic form factor

$$F^2(t) = \approx \left[Z\left(\frac{a^2 t}{1+a^2 t}\right)\left(\frac{1}{1+t/d}\right)\right]. \tag{A.2}$$

The target at NA64 uses lead, so its safe to assume the target nucleus to have $Z = 82$, $A = 207$ and $M = 207.2$ GeV. Here we define

$$a = 111\frac{Z^{-1/3}}{m_e} \tag{A.3}$$

$$d = 0.164A^{-2/3}/\text{GeV}^{-2} \tag{A.4}$$

The amplitude at the lowest order approximation, represented by tree-level is

$$i\mathcal{M} = ihe^2\frac{F(q^2)}{q^2}P_{\mu}\bar{u}_e(p')\left[P_R\frac{(\not{p}' + \not{k}) + m_{\mu}}{\tilde{s}}\gamma^{\mu} + \gamma^{\mu}\frac{(\not{p} - \not{k}) + m_e}{\tilde{u}}P_R\right]u_{\mu}(p) \tag{A.5}$$

$h = g'/\sqrt{4\pi\alpha}$ is the coupling constant of the dark leptonic scalar φ to muons, here $\alpha = 1/137$ is the fine-structure constant. The mandelstam variables are written as

$$\begin{aligned}
\tilde{s} &= (p' + k)^2 - m_i^2 \\
\tilde{u} &= (p - k)^2 - m_f^2 \\
t_2 &= (p' - p)^2 \\
t &= -q^2 \\
m_{\varphi}^2 &= \tilde{s} + \tilde{u} + t_2 + t.
\end{aligned}$$

Calculated using `Mathematica` with `FeynCalc`, see Appendix A in [37],

$$\langle|\mathcal{M}|^2\rangle = h^2e^4\frac{F^2(t)}{t^2}|A_{2\to3}|^2 \tag{A.6}$$

where

$$|A_{2\to3}|^2 = \frac{(\tilde{s}+\tilde{u})^2}{2\tilde{s}\tilde{u}}P^2 - \frac{2t}{\tilde{s}\tilde{u}}(P\cdot k) + \frac{(\tilde{s}+\tilde{u})^2}{2\tilde{s}^2\tilde{u}^2}\Delta m^2\left[P^2 t - 4\left(\frac{\tilde{u}P\cdot p}{+}\tilde{s}P\cdot p'\tilde{s}+\tilde{u}\right)\right] \tag{A.7}$$

with Δm^2 defined as $\Delta m^2 \equiv m_\varphi^2 - m_i^2 - m_f^2$. Integrating the phase space of the final state lepton from a general $2 \to 3$ scattering process yields the resulting double differential cross section

$$\left(\frac{d^2\sigma}{dx\,d\cos\theta_k}\right) = \frac{h^2\alpha^2}{4\pi}\frac{E_i|\mathbf{k}|}{|\mathbf{V}||\mathbf{p}|}\int_{t_{min}}^{t_{max}} dt\frac{F^2(t)}{t^2}\int_0^{2\pi}\frac{d\phi_q}{2\pi}\frac{|A_{2\to3}|^2}{8M^2} \tag{A.8}$$

t_{min} and t_{max} are the values of the minimum and maximum momentum transfer, respectively, these are derived explicitly in Ref. [31] (define the angle between $\mathbf{p}$ and $\mathbf{k}$ as θ_k).

$$\cos\theta_q^0 = \frac{(E_i - E_f + q_0)^2 - |\mathbf{V}|^2 - |\mathbf{q}|^2 - m_f^2}{2|\mathbf{V}||\mathbf{q}|}$$
$$\mathbf{V} = |\mathbf{p}-\mathbf{k}| = \sqrt{\mathbf{p}^2 + \mathbf{k}^2 - 2\mathbf{p}\mathbf{k}\cos\theta_k}$$

from which we get the final cross section

$$\left(\frac{d\sigma}{dx}\right) = \frac{h^2\alpha^2}{4\pi}\frac{E_i|\mathbf{k}|}{|\mathbf{p}|}\int_0^{\theta_{max}} d\cos\theta_k\frac{1}{V}\int_{t_{min}}^{t_{max}} dt\frac{F^2(t)}{t^2}\int_0^{2\pi}\frac{d\phi_q}{2\pi}\frac{|A_{2\to3}|^2}{8M^2} \tag{A.9}$$

B | The complete selection criteria at NA64

The experimental and simulated selection cuts exhibit slight differences, although considerable effort has been made to align them as closely as possible. For comparative purposes, table B.1 and B.2 present the efficiencies relative to the signal blinding. For a more accurate comparison, the results are also evaluated based on the efficiency per EOT. After applying all the selection cuts for dimuons, the real data shows a total efficiency of $6.27 \cdot 10^{-8}$ per EOT, whereas the simulation results indicate a total efficiency of $9.25 \cdot 10^{-8}$ per EOT.

It is important to note that the selection criteria prior to the HCAL dimuon cuts have remained unchanged throughout this thesis. These pre-established cuts have already been optimized to ensure the selection of electrons impinging on the ECAL with reduced pile-up events and with proper energy reconstruction. The focus of the optimization is on the HCAL dimuon cuts and all subsequent cuts. This is where the NN can significantly enhance the selection process for selecting dimuons. Therefore, the NN training is based on the data from this point onwards.

Table B.1: Cutflow applied to the experimental data. The number of events, efficiency [%] and efficiency per EOT [%] are displayed for each cut. The total number of EOT is $4.55 \cdot 10^{11}$.

Cut Description	# Events	Eff. [%]	Eff. / EOT
Physics Trigger, filtering of bad spills and HCAL pedestal [a]	$4.03 \cdot 10^8$	–	$8.85 \cdot 10^{-4}$
Signal blinding: $E_{HCAL_i} > 4$ GeV, $i = 0, 1, 2$	$5.77 \cdot 10^6$	100	$1.27 \cdot 10^{-5}$
Time difference trigger: $113.5 \geq$ Master Time ≥ 86.5 ns	$5.75 \cdot 10^6$	99.7	$1.27 \cdot 10^{-5}$
SRD: $0.08 > E_{SRD_i} > 0.001$ GeV, $i = 0, 1, 2$	$4.99 \cdot 10^6$	86.5	$1.10 \cdot 10^{-5}$
Track Quality: $9999 > P_{\text{Beam}} > 0$ & P-value > 0.01	$4.36 \cdot 10^6$	75.5	$9.56 \cdot 10^{-6}$
Momentum: $90 \leq P_{\text{Beam}} \leq 110$ GeV	$4.16 \cdot 10^6$	72.1	$9.13 \cdot 10^{-6}$
Angle 12: In Angle < 3 [Rad]	$4.16 \cdot 10^6$	72.1	$9.13 \cdot 10^{-6}$
Straw 3 multiplicity: $5 \geq N_{Straw3X} \geq 0$ & $5 \geq N_{Straw3Y} \geq 0$	$4.03 \cdot 10^6$	69.8	$8.84 \cdot 10^{-6}$
VHCAL energy: $E_{VHCAL} \leq 1.5$ GeV	$3.89 \cdot 10^6$	67.4	$8.53 \cdot 10^{-6}$
Straw 4 multiplicity: $5 \geq N_{Straw4X} \geq 0$ & $5 \geq N_{Straw4Y} \geq 0$	$3.72 \cdot 10^6$	64.4	$8.16 \cdot 10^{-6}$
ECAL center: Max E_{ECAL} deposition in center cell $(2, 2)$	$3.66 \cdot 10^6$	63.5	$8.03 \cdot 10^{-6}$
Total ECAL: $E_{ECAL,\,\text{IN TIME}} < 30$ GeV & $E_{ECAL} < 120$ GeV & $E_{PS} > 0.7$ GeV	$3.61 \cdot 10^6$	62.6	$7.92 \cdot 10^{-6}$
ECAL periphery ratio: $(E_{ECAL} - E_{ECAL-(3\times3)})/E_{ECAL} < 0.06$	$3.61 \cdot 10^6$	62.5	$7.91 \cdot 10^{-6}$
χ^2 cut	$3.60 \cdot 10^6$	62.4	$7.90 \cdot 10^{-6}$
HCAL out-of-time: $(E_{HCAL} - E_{HCAL,\text{IN TIME}}) < 50$ GeV	$3.60 \cdot 10^6$	62.4	$7.90 \cdot 10^{-6}$
Last HCAL cut: $E_{HCAL_3} < 2$ GeV	$2.59 \cdot 10^6$	44.9	$5.69 \cdot 10^{-6}$
HCAL topo [b]	$2.59 \cdot 10^6$	44.9	$5.68 \cdot 10^{-6}$
HCAL$_0$ dimuon cut: $8 > E_{HCAL_0} > 4$ GeV	$2.35 \cdot 10^6$	40.7	$5.15 \cdot 10^{-6}$
HCAL$_1$ dimuon cut $8 > E_{HCAL_1} > 4$ GeV	$2.20 \cdot 10^6$	38.2	$4.83 \cdot 10^{-6}$
HCAL$_2$ dimuon cut $8 > E_{HCAL_2} > 4$ GeV	$2.06 \cdot 10^6$	35.7	$4.52 \cdot 10^{-6}$
STD$_{11}$ multiplicity cut: $N_{Straw11X} = 2$ & $N_{Straw11Y} = 2$	$4.14 \cdot 10^4$	0.72	$9.08 \cdot 10^{-8}$
STD$_{11}$ separation cut: (1 hit/side of $X_{Straw11} = -140$ mm) $\wedge$ $(-150 < Y_{Straw11} < 150$ mm)	$2.85 \cdot 10^4$	0.49	$6.26 \cdot 10^{-8}$

[a] HCAL pedestal cut: $E_{HCAL_i,\text{CELL}} > 0.2$ GeV, for all cells, $i = 0, 1, 2$

[b] If $\sum N(E_{HCAL_0} > 0.5 \text{ GeV}) \geq 1 \wedge E_{HCAL_0} < 2 \cdot E_{HCAL}(0, 1, 1)$ GeV || $N(E_{HCAL_0} > 0.5 \text{ GeV}) = 0$

Table B.2: Cutflow applied to the simulated data. The number of events and efficiency [%] are displayed for each cut. The total number of EOT is $4.55 \cdot 10^{11}$.

Cut Description	# Events	Eff. [%]	Eff. / EOT
Filter magnet iron block and HCAL pedestal [a]	338200	–	$2.57 \cdot 10^{-5}$
Signal blinding: $E_{HCAL_i} > 4$ GeV, $i = 0, 1, 2$	102705	100	$7.79 \cdot 10^{-6}$
ECAL: $E_{ECAL} < 76$ GeV & $E_{PS} > 0.4$ GeV	79861	77.8	$6.06 \cdot 10^{-6}$
SRD: $0.08 > E_{SRD_i} > 0.001$ GeV, $i = 0, 1, 2$	72407	70.5	$5.50 \cdot 10^{-6}$
Track Quality: $9999 > P_{\text{Beam}} > 0$ & P-value > 0.01	63530	61.9	$4.82 \cdot 10^{-6}$
Momentum: $90 \leq P_{\text{Beam}} \leq 110$ GeV	61297	59.7	$4.65 \cdot 10^{-6}$
Angle 12: In Angle < 3 [Rad]	61297	59.7	$4.65 \cdot 10^{-6}$
Straw 3 multiplicity: $5 \geq N_{Straw3X} \geq 0$ & $5 \geq N_{Straw3Y} \geq 0$	60740	59.1	$4.61 \cdot 10^{-6}$
VHCAL energy: $E_{VHCAL} \leq 1.5$ GeV	60740	59.1	$4.61 \cdot 10^{-6}$
Straw 4 multiplicity: $5 \geq N_{Straw4X} \geq 0$ & $5 \geq N_{Straw4Y} \geq 0$	59822	58.2	$4.54 \cdot 10^{-6}$
ECAL center: Max E_{ECAL} deposition in center cell $(2, 2)$	59818	57.9	$4.54 \cdot 10^{-6}$
Total ECAL: $E_{ECAL} + E_{PS} < 150$ GeV & $E_{PS} > 0.7$ GeV	59448	57.9	$4.51 \cdot 10^{-6}$
ECAL periphery ratio: $(E_{ECAL} - E_{ECAL-(3\times3)})/E_{ECAL} < 0.06$	59446	54.0	$4.51 \cdot 10^{-6}$
Last HCAL cut: $E_{HCAL_3} < 2$ GeV	55501	54.0	$4.21 \cdot 10^{-6}$
HCAL topo [b]	55489	51.9	$4.21 \cdot 10^{-6}$
HCAL$_0$ dimuon cut: $8 > E_{HCAL_0} > 4$ GeV	53308	49.8	$4.05 \cdot 10^{-6}$
HCAL$_1$ dimuon cut $8 > E_{HCAL_1} > 4$ GeV	51119	47.4	$3.88 \cdot 10^{-6}$
HCAL$_2$ dimuon cut $8 > E_{HCAL_2} > 4$ GeV	48716	47.4	$3.70 \cdot 10^{-6}$
VETO cut: $0.05 > E_{VETO_{23}} > 0.01$ GeV	46074	44.9	$3.50 \cdot 10^{-6}$
STD$_{11}$ multiplicity cut: $N_{Straw11X} = 2$ & $N_{Straw11Y} = 2$	1477	1.43	$1.12 \cdot 10^{-7}$
STD$_{11}$ separation cut: (1 hit/side of $X_{Straw11} = -140$ mm) $\wedge$ $(-150 < Y_{Straw11} < 150$ mm)	1219	1.19	$9.25 \cdot 10^{-8}$

[a] HCAL pedestal cut: $E_{HCAL_i,\text{CELL}} > 0.2$ GeV, for all cells, $i = 0, 1, 2$

[b] If $\sum N(E_{HCAL_0} > 0.5 \text{ GeV}) \geq 1 \wedge E_{HCAL_0} < 2 \cdot E_{HCAL}(0,1,1)$ GeV $||$ $N(E_{HCAL_0} > 0.5 \text{ GeV}) = 0$

C | Extra figures

Below follows some additional figures comparing the simulation and real data. The results are based on the selection after the HCAL dimuon cuts. The ECAL, MM_3, MM_4, MM_5 and VETO are shown in figure C.1, C.2, C.3, C.4 and C.5 respectively. Each of the plots uses blinded data from period $2-5$ and the simulations are based on the simulation described in appendix B.

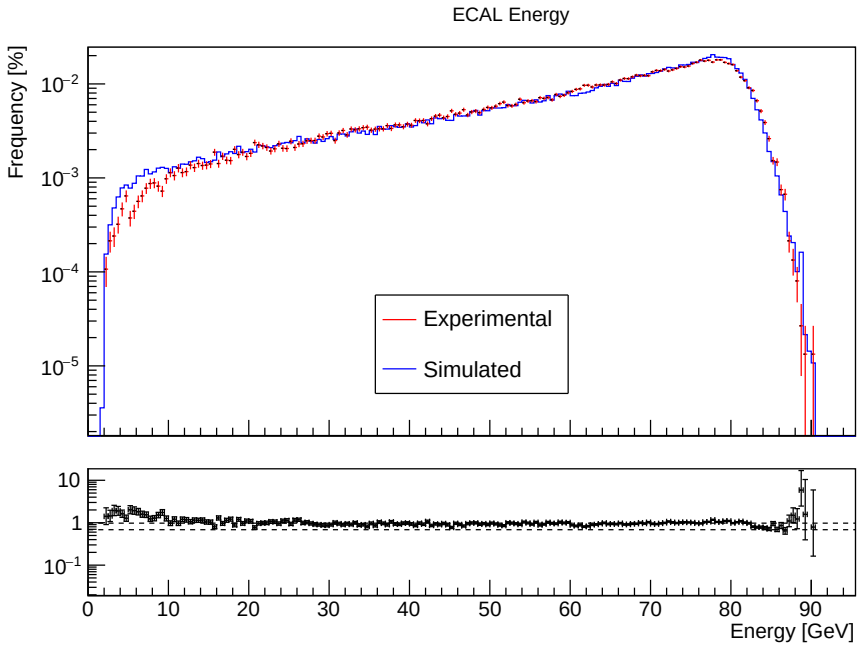


Figure C.1: Comparison between simulated and experimental results for the ECAL.

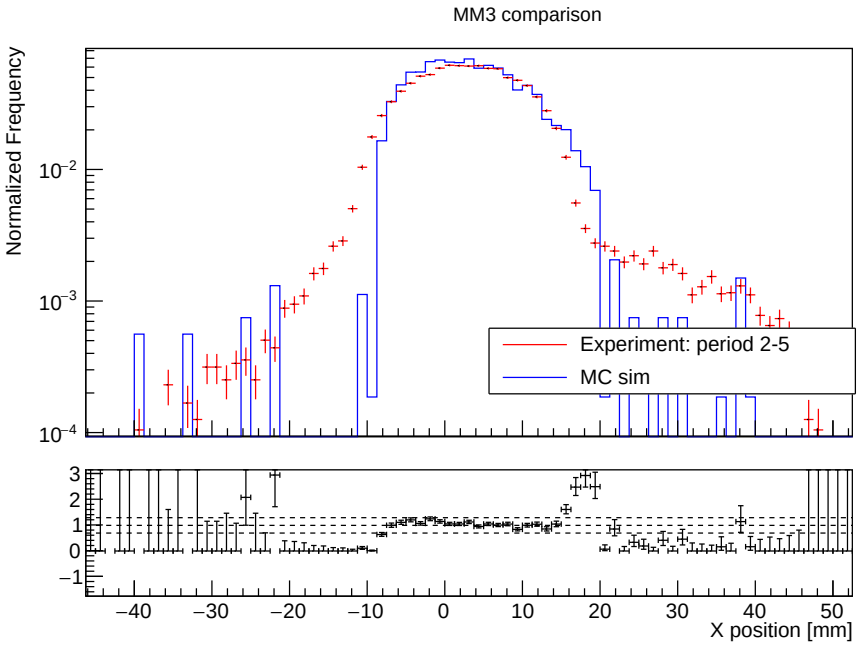


Figure C.2: Comparison between simulated and experimental results for MM_3.

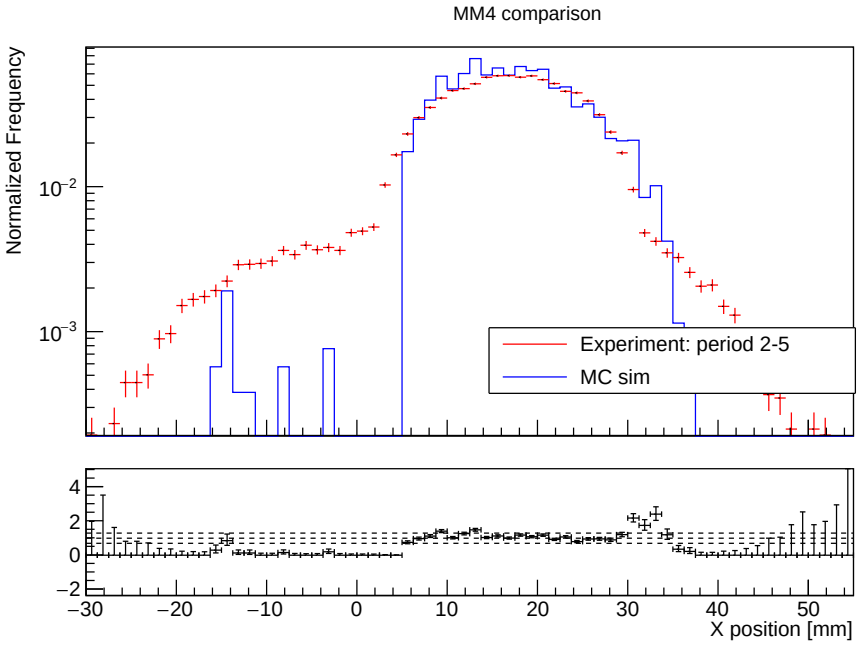


Figure C.3: Comparison between simulated and experimental results for MM_4.

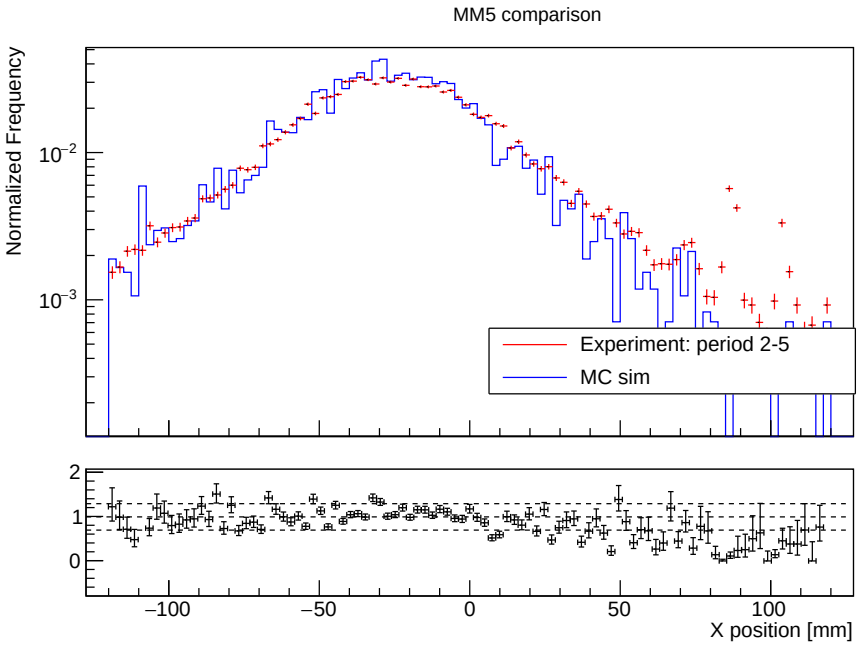
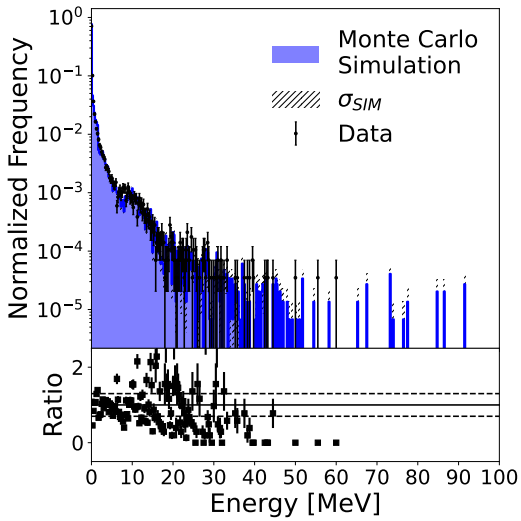
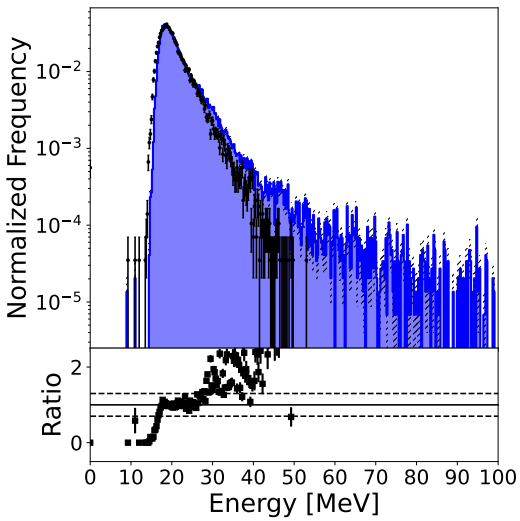


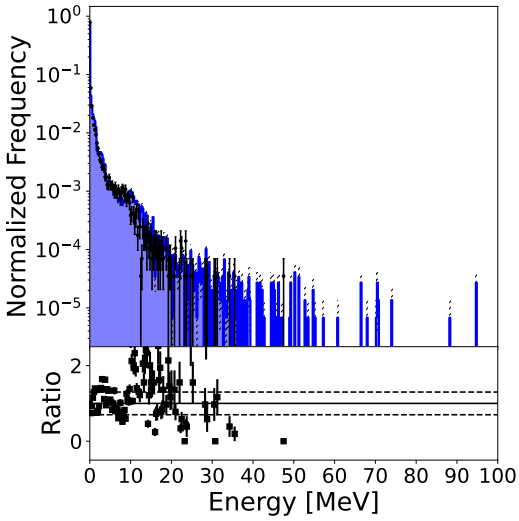
Figure C.4: Comparison between simulated and experimental results for MM_5.



(a) VETO_{01}



(b) VETO_{23}



(c) VETO_{45}

Figure C.5: Comparison of the energy deposition between simulated and experimental results for the VETO detector in front of HCAL_0.

DEPARTMENT OF PHYSICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY