Data Analysis of the Drell–Yan Process in the CMS Experiment at the LHC

A Fourth-Semester Dissertation Submitted to

Department of Physics and Astrophysics
University of Delhi, Delhi – 110007

In partial fulfillment of the requirements for the degree of

Master of Science in Physics

Anuj Raghav
M.Sc. Physics, Semester–IV, Group – F4
University Roll No: 23222762018

Under the supervision of

Dr. Arun Kumar

May, 2025

Undertaking

This dissertation is submitted in partial fulfillment of the requirements for the Master of Science degree in Physics and was conducted under the supervision of Dr. Arun Kumar.

I declare that this dissertation has been composed solely by myself and is free from any form of plagiarism. Except where otherwise indicated by reference or acknowledgment, all work presented here is entirely my own.

Anuj Raghav
(Candidate)

Dr. Arun Kumar
(Supervisor)

Abstract

The Drell–Yan process, in which a quark and an antiquark annihilate to produce a lepton–antilepton pair, is one of the cleanest and most well-understood processes studied at the Large Hadron Collider (LHC). This dissertation focuses on analyzing the Drell–Yan process in the di-electron (electron-positron) final state using simulated data from the CMS experiment at a center-of-mass energy of 13 TeV.

The goal of this study is to identify the Drell–Yan signal and separate it from other Standard Model background processes, such as top quark production and di-boson events, which can mimic the same final state. To achieve this, various kinematic variables are examined, and a series of selection cuts are applied to enhance the sensitivity of the signal.

The analysis is based on Monte Carlo simulation samples processed under CMS standard conditions for the 2016 data-taking period available under CMS opendata. The overall approach is simple and relies on well-motivated, step-by-step techniques that are commonly used in real experimental analyses. This work serves as a hands-on exploration of how high-energy physics data is processed and how meaningful results can be extracted from complex collision events.

Acknowledgment

I extend my deepest gratitude to my supervisor, Dr. Arun Kumar, Assistant Professor, Department of Physics and Astrophysics, University of Delhi, for his positive attitude, constructive guidance, and unwavering support throughout the course of this dissertation. His dedication and passion for High Energy Physics have been a continuous source of motivation throughout this journey.

I would also like to express my heartfelt thanks to Prof. Kirti Ranjan, Professor, Department of Physics and Astrophysics, University of Delhi, for his invaluable experience and insights, which have greatly inspired me to pursue this field.

Additionally, I am deeply grateful to Prof. Debajyoti Choudhury, Head of the Department, Department of Physics and Astrophysics, for fostering a supportive academic environment and for his encouragement during this academic pursuit.

Lastly, I would like to thank the department and library staff for providing a conducive atmosphere and essential resources that have been instrumental in shaping my understanding and research in High Energy Physics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview of the Study

The study of the Drell-Yan (DY) process at the Compact Muon Solenoid (CMS) experiment is essential for understanding fundamental particle interactions at the Large Hadron Collider (LHC). In this study, we focus specifically on the DY process in the electron-positron (e^+e^-) also referred as di-electron final state, where a quark and an antiquark annihilate to produce a Z boson, which subsequently decays into an electron-positron pair. This process constitutes a fundamental aspect of the Standard Model (SM), and a thorough understanding of it is essential in high-energy physics as well as in the pursuit of new physics. It serves as a critical benchmark for investigating high-energy collisions and for testing the validity of theoretical models.

Although the DY process has high cross-section, it is accompanied by various background processes that can mimic the electron-positron final state. These background events must be carefully identified and mitigated to ensure a clean signal extraction. The presence of top-quarks pair production, single top quark production, and di-boson events complicates the extraction of the pure signal. Therefore, improving the signal purity is necessary to obtain more accurate measurements.

1.2 Summary of Previous Work

In the previous semester, the focus was on laying the theoretical foundation for the study. A detailed examination of the DY process was conducted. Along with the interaction of different radiations with the matter. Additionally, an overview of the CMS experiment and its key detector components was presented, highlighting the role of tracking systems, calorimeters in electron identification and reconstruction.

Preliminary data analysis was performed using Monte Carlo (MC) simulations

from the CMS Open Data repository [1]. This included generating initial distributions of key kinematic variables of leptons such as transverse momentum (p_T), pseudorapidity (η), and invariant mass ($m_{\ell\ell}$). Background contributions from top-quark pairs, single top quark, and di-boson processes were also identified. These studies provided the necessary groundwork for implementing optimization strategies in the current semester.

1.3 Focus of the Present Study

Building upon the previous work, the present study shifts towards refining data-driven techniques to enhance the significance of the Drell-Yan signal. The primary objectives for this semester include:

- **Signal and Background Analysis:** Characterizing the signal and background contributions using CMS Open Data and simulated MC samples.
- **Event Selection and Optimization:** Defining and implementing selection criteria to improve signal purity.
- **Scaling and Normalization:** Applying cross-section weighting and luminosity scaling to ensure accurate event yield comparisons.

Chapter 2

Signal and Backgrounds Characterization

During proton-proton (pp) collisions at the LHC, various processes can lead to the production of electron-positron pairs. While the DY process is the primary mechanism responsible for such final states, several background processes can produce similar signatures and may be misidentified as signal events. These background processes must be thoroughly understood to ensure accurate signal extraction. Although carefully optimized selection criteria can significantly reduce the contribution of background events, a residual background remains that must be accurately modeled. To enable a reliable comparison with signal MC predictions, the remaining background is subtracted from the observed data. Consequently, precise estimation of these background processes—typically based on MC simulations—is essential to avoid bias in the signal measurement and ensure the integrity of the analysis.

2.1 Drell-Yan Process

The DY production is a fundamental process in high-energy particle physics, describes the lepton pair production in hadron-hadron collisions, and was first suggested by Sidney D. Drell and Tung-Mow Yan in 1970. [2]

The quark from one proton and an antiquark from another proton annihilate to form a Z boson, which subsequently decays into a lepton-antilepton pair. This process is one of the most significant channels for studying electroweak interactions and serves as a crucial benchmark for precision measurements at the Large Hadron Collider (LHC).

In this study, we focus on the DY process in the di-electron (e^+e^-) final state, see Fig. 2.1

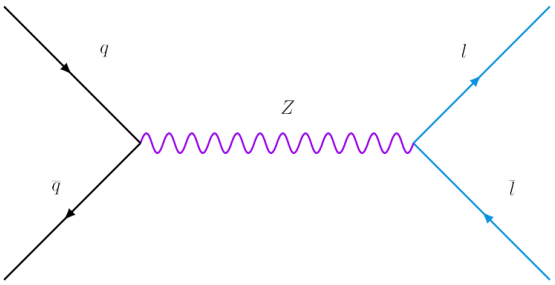


Figure 2.1: The Feynman diagram of the Drell-Yan process

The reaction follows the schematic form:

$$q\bar{q} \longrightarrow Z \longrightarrow e^{+}e^{-}$$

where q and $\bar{q}$ represent the initial-state quark and antiquark, respectively. The Z boson decays into an electron-positron pair, which is then detected by the CMS detector through inner trackers and Electromagnetic calorimeter (ECAL).

Despite being a clean signal process, the identification of DY events is complicated by various background contributions that can mimic the electron-positron final state. These background events must be carefully studied and suppressed to ensure a high-purity signal sample.

2.2 Background Processes

In this analysis, several dominant background processes that impact the purity of the signal were taken into consideration. Specifically, processes such as top quark pair production, single top quark production, and di-boson production were included. Each of these background sources will be examined individually to understand their characteristics and evaluate their potential to interfere with the signal.

2.2.1 Top Quark production

Top quark pair production constitutes a major background in DY analyses due to its ability to produce high-energy electrons and positrons. The cross-section of top quark process for Next-to-Leading Order (NLO) is 87.31 pb. This process initiates when a quark-antiquark pair or two gluons interact, resulting in the creation of a top quark and an antitop quark:

$$q\bar{q} \text{ or } gg \longrightarrow t\bar{t}$$

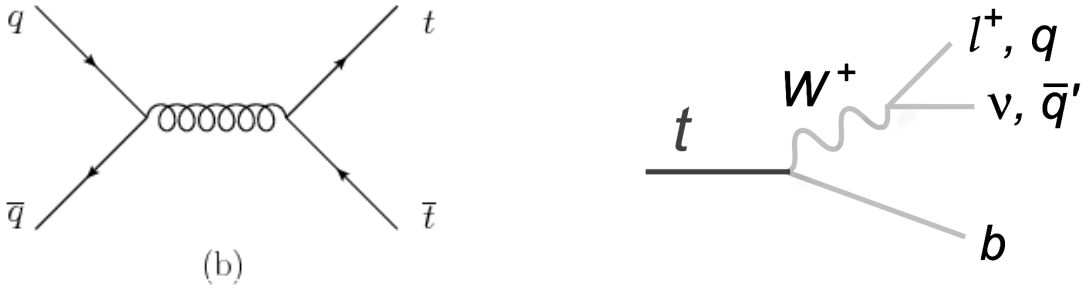
Each top quark decays almost exclusively into a W boson and a bottom quark:

$$t \longrightarrow W^+ b, \qquad \bar{t} \longrightarrow W^- \bar{b}$$

If both W bosons decay leptonically, the decay products include an electron, a positron, and two neutrinos:

$$W^+ \longrightarrow e^+ \nu_e, \qquad W^- \longrightarrow e^- \bar{\nu_e}$$

The final state thus contains an e^+e^- pair, neutrinos, and b-quark jets. The electron-positron pair can mimic the DY signal, particularly in events where the accompanying b-jets and missing transverse energy from neutrinos are poorly reconstructed or misidentified.



(a) Top quarks production

(b) Decay of top quark

Figure 2.2: Feynman diagrams for top quark production and decay.

2.2.2 Single Top Quark Production

Single top production is a less significant background and occurs via three principal mechanisms: the t-channel, s-channel, and tW-channel with a combined cross-section of 145.27 pb. Each of these topologies introduces distinct final-state characteristics.

In the **t-channel**, a light quark exchanges a W boson with a bottom quark, resulting in a top quark and another light quark:

$$qb \longrightarrow q't$$

The top quark subsequently decays as follows:

$$t \longrightarrow W^+ b \longrightarrow e^+ \nu_e b$$

This leads to a final state composed of an electron, a neutrino, a b-jet, and a light-quark jet.

In the **s-channel**, a quark-antiquark pair annihilates to form a virtual W boson, which decays into a top quark and a bottom antiquark:

$$q\bar{q}' \longrightarrow W^* \longrightarrow tb$$

The decay products, similar to the t-channel, include an electron, a neutrino, and b-jets.

In the **tW-channel**, a gluon and a bottom quark produce a top quark and an onshell W boson:

$$gb \longrightarrow tW$$

If both the top quark and the W boson decay leptonically:

$$t \longrightarrow W^+b \longrightarrow e^+\nu_e b, \qquad W^- \longrightarrow e^-\bar{\nu_e}$$

This results in an e^+e^- pair, neutrinos, and b-jet in the final state. In all single top production channels, the leptonically decaying W bosons introduce electrons or positions that can closely imitate DY signatures, particularly when additional jets and missing energy are not sufficiently distniguished.

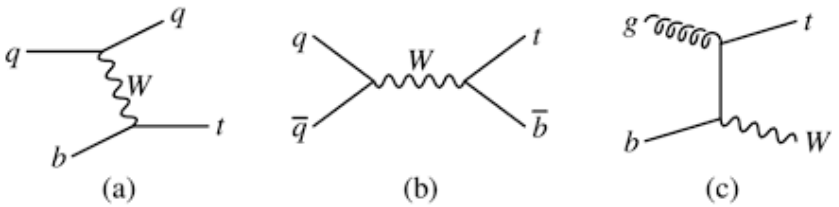


Figure 2.3: Feynman diagram of single top quark production. (a) Through t-channel, (b) Through s-channel, and (c) Through tW-channel.

2.2.3 Di-boson Production

Di-boson production processes, involving combinations of W and Z bosons, also contribute to DY-like final states due to their frequent decays into leptons.

In **WZ production** with a cross-section of 47.13 pb, a quark-antiquark pair annihilates to yield a W and a Z boson:

$$q\bar{q}' \longrightarrow WZ$$

The Z boson can decay into an e^+e^- pair, and the W boson decays leptonically into a charged lepton and a neutrino:

$$Z \longrightarrow e^+e^-, \qquad W \longrightarrow \ell\nu$$

The final state consists of an e^+e^- pair, a third lepton, and a neutrino. such events can resemble DY processes, especially when only the electron-positron pair is fully reconstruced.

In **ZZ production** with a cross-section of 16.52 pb, the interaction produces two Z bosons:

$$q\bar{q} \longrightarrow ZZ$$

One Z boson decays into an electron-positron pair, while the second typically decays into neutrinos:

$$Z \longrightarrow e^+e^-, \qquad Z \longrightarrow \nu\bar{\nu}$$

This leads to an e^+e^- pair and significant missing transverse energy from the neutrinos, a configuration that can overlap kinematically with DY signatures.

In **WW production** with a cross-section of 12.18 pb, two W bosons are produced from quark-antiquark annihilation:

$$q\bar{q'} \longrightarrow WW$$

Both W bosons may decay leptonically:

$$W^+ \longrightarrow e^+\nu_e, \qquad W^- \longrightarrow e^-\bar{\nu}_e$$

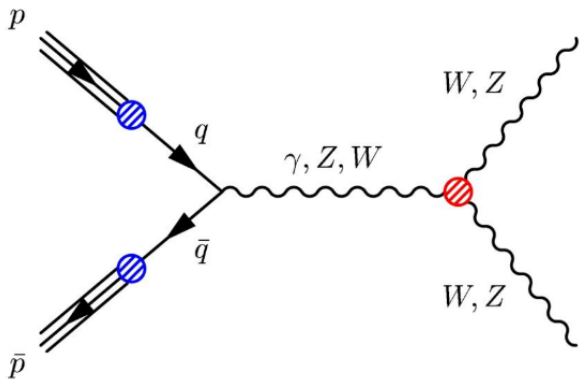


Figure 2.4: Feynman diagram for di-boson production.

2.3 Description of MC Samples

In this analysis, MC simulations are extensively employed to model both the signal and the various background processes relevant to pp collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV, as recorded during the 2016 data-taking period by the CMS detector at the LHC. All simulated samples are produced using

the NANOAODSIM data format [3], optimized for analysis-level access to high-level physics objects, and processed under the UltraLegacy (UL) 2016 campaign conditions.

The signal process under investigation is the DY production of lepton pairs via an intermediate virtual photon or Z boson. It is generated with the MadGraph5_aMC generator. The hadronization and parton showering are modeled using Pythia8 [4], employing the CP5 tune to better reflect the underlying event activity. This dataset focuses on inclusive production of lepton pairs with invariant mass above 50 GeV, thereby targeting the $Z/\gamma^* \to \ell^+\ell^-$ (with $\ell = e, \mu$) decay channels, which form a clean and well-understood benchmark for precision studies and new physics searches.

To comprehensively account for Standard Model background contributions, a diverse set of simulated datasets are utilized. The dominant irreducible background arises from top quark pair production ($t\bar{t}$), generated using the POWHEG v2 generator, interfaced with Pythia8 for parton shower simulation.

Single top quark processes constitute additional important backgrounds, particularly due to their similar final-state kinematics. These are modeled using several exclusive production modes: the t-channel top and antitop processes are generated using the POWHEG + Pythia8 framework. Additionally, the tW associated production is included also simulated at NLO with POWHEG. For completeness, the s-channel single top production which uses the MG5_aMC@NLO generator and includes leptonic decay modes of the top quark.

Electroweak di-boson processes are included as subdominant but non-negligible backgrounds, particularly in di-lepton final states. Pythia8 generator is used for WZ and ZZ samples, while WW production is generated with NLO accuracy using POWHEG. These processes can mimic the signal through the production of multiple isolated leptons and missing transverse momentum.

All samples are normalized to the most accurate theoretical cross sections available, incorporating higher-order QCD and electroweak corrections as applicable. They are processed with the UL2016 detector simulation and reconstruction conditions to ensure consistency with real data conditions. The datasets are publicly available via the CERN Open Data Portal and have been produced by the CMS Collaboration in accordance with standard simulation and validation workflows.

Process	**Dataset Name**	**Cross Section [pb]**	$N_{\textbf{events}}$
Drell–Yan	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 [5]	6189.39	81.6M
$t\bar{t}$	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8 [6]	87.31	42.2M
ST (t-channel, top)	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8 [7]	44.33	63.1M
ST (t-channel, antitop)	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8 [8]	26.38	29.7M
ST (tW, top)	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 [9]	35.60	2.1M
ST (tW, antitop)	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 [10]	35.60	2.5M
ST (s-channel)	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8 [11]	3.36	5.5M
WZ	WZ_TuneCP5_13TeV-pythia8 [12]	47.13	7.6M
ZZ	ZZ_TuneCP5_13TeV-pythia8 [13]	16.52	1.1M
WW	WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8 [14]	12.18	2.9M

Table 2.1: Summary of Monte Carlo simulated datasets used for the signal and background processes along with their cross sections and number of events (M stands for million). All samples correspond to the 2016 CMS data-taking period and are processed using the UL16 NANOAODSIM workflow. Cross-sections are taken from [15]

2.4 Scaling and Normalization of Samples

The signal and background samples used for MC simulation are listed in Table 2.1.

To enable a meaningful comparison between observed data and simulated events, it is essential to normalize the simulated MC samples. This normalization ensures that the simulated distributions accurately reflect the same average number of pp collisions per bunch crossing as observed in real experimental conditions. Discrepancies in these conditions between data and simulation can lead to systematic biases in physics analyses, especially for observables that are sensitive to event multiplicity and detector occupancy.

The normalization factor, denoted as ω_{norm}, is defined by the following relation:

$$\omega_{\text{norm}} = \frac{\sigma \cdot \mathcal{L}}{N_{\text{events}}} \tag{2.1}$$

- σ represents the theoretical cross section of the process under consideration,
- $\mathcal{L}$ is the integrated luminosity of the dataset (taken to be 1 fb^{-1} in this study),
- N_{events} is the total number of events in the corresponding MC sample before any selection is applied.

The cross-section values used for each process are listed in Table 2.1, and are taken from [15].

This normalization procedure allows simulated samples to be scaled appropriately, providing a realistic expectation of the number of events one would observe for a given process within the same luminosity conditions as the actual data. Failure to apply such normalization can result in misleading comparisons and incorrect interpretation of the underlying physics processes.

Chapter 3

Event Selection and Kinematic variables

3.1 Kinematic Variables and their Definitions

In the analysis of pp collisions, a range of kinematic variables is used to describe the motion and spatial orientation of particles produced in the final state. These variables are essential tools for understanding the event dynamics and form the basis for distinguishing between different physical processes. In the context of this analysis, which focuses on the DY process in the dielectron final state, the most relevant kinematic observables are defined below:

- **Transverse Momentum (p_T):** The transverse momentum of a particle is defined as the component of its momentum perpendicular to the beam axis (z-axis) or simply in the transverse plane ($x-y$ plane). It is calculated as:

$$p_T = \sqrt{p_x^2 \ + \ p_y^2} \tag{3.1}$$

  where p_x and p_y are the momenta in x and y axis respectively.

- **Pseurdorapidty (η):** Pseudorapidity is a spatial coordinate that describes the angle of a particle relative to the beam axis. It is defined as:

$$\eta \ = \ -\ln\left(tan\frac{\theta}{2}\right) \tag{3.2}$$

  where θ is the polar angle with respect to the beam axis. Pseudorapidity is commonly used due to its invariance under Lorentz boosts along the beam direction.

- **Azimuthal Angle (ϕ):** It is the angle of the particle in the transverse plane, measured from the x-axis. It ranges from to $-\pi$ to π.

- **Invariant mass of the di-electron ($m_{\ell\ell}$):** The invariant mass of a di-electron system is calculated using the four momenta of two electrons.

$$m_{\ell\ell} = \sqrt{(E_1 + E_2)^2 - (\vec{p_1} + \vec{p_2})^2} \tag{3.3}$$

  where:

  - m_{ll} is the invariant mass of the di-electron system.
  - E_1 and E_2 are the energies of the two electrons.
  - $\vec{p_1}$ and $\vec{p_2}$ are the three-momenta of the two electrons.

- **Transverse momentum of di-electron ($p_T\ell\ell$):** The transverse momentum of the di-electron system, denoted as $p_T\ell\ell$, is given by:

$$p_T\ell\ell = \sqrt{(p_x^1 + p_x^2)^2 + (p_y^1 + p_y^2)^2} \tag{3.4}$$

  where, p_x^i and p_y^i are the momentum of leptons in x and y directions.

- **Azimuthal angle difference ($\Delta\phi_{\ell\ell}$):** The azimuthal angle difference of the di-electron system, denoted as $\Delta\phi_{\ell\ell}$, is given by:

$$\Delta\phi_{\ell\ell} = |\phi_1 - \phi_2| \tag{3.5}$$

  where:

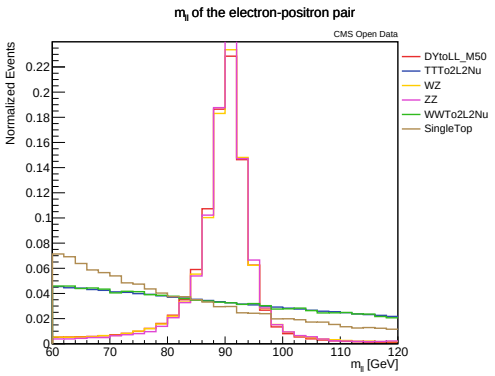
  - ϕ_1 and ϕ_2 are the azimuthal angles of the two electrons in the transverse plane, defined as:

$$\phi_1 = \tan^{-1}\left(\frac{p_{y1}}{p_{x1}}\right), \quad \phi_2 = \tan^{-1}\left(\frac{p_{y2}}{p_{x2}}\right) \tag{3.6}$$

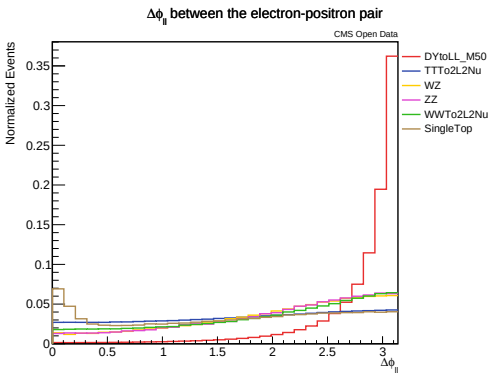
- **Missing Transverse Energy (MET):** MET represents the magnitude of the vector sum of transverse momenta of all detected particles, with its negative value indicating momentum imbalance due to undetected particles like neutrinos:

$$MET \ = \ |-\Sigma p_T^{visible}| \tag{3.7}$$

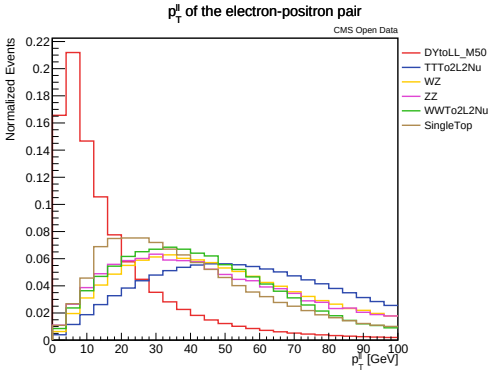
  A better variable than traditional MET is *Proj-MET* which gives the projection of MET along $\phi_{electron}$.



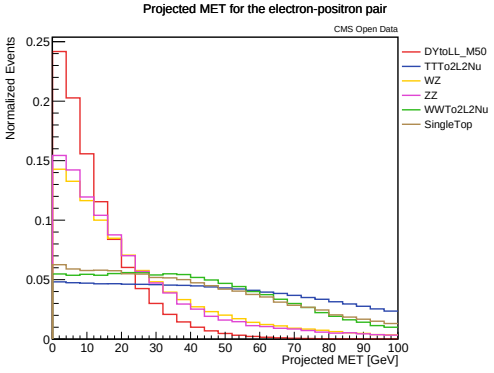
(a) Invariant mass ($m_{\ell\ell}$) of electron pair



(b) Azimuthal separation ($\Delta\phi_{\ell\ell}$)



(c) Transverse momentum ($p_{T\ell\ell}$) of electron pair



(d) Projected missing transverse energy (MET)

Figure 3.1: Distributions of key di-electron variables for the signal and background processes superimposed.

3.2 Events Selections

The distributions of the aforementioned kinematic variables for both signal and background processes are presented in the above plots. These distributions provide critical insights that can be leveraged to define effective event selection strategies. By analyzing the distinguishing features between signal and background, it becomes possible to implement optimization techniques aimed at enhancing signal purity and improving the overall performance of the analysis.

3.2.1 Signal Processes

The DY signal, characterized by the process $q\bar{q} \to Z \to e^+e^-$, is clearly observed in the distributions presented in Figure 3.1. These plots highlight several distinguishing features of the DY process. The electron-positron pair typically exhibits low p_T, and the two leptons are emitted nearly back-to-back in the detector, resulting $\Delta\phi$ close to π.

Additionally, the final state predominantly contains two well-identified electrons with negligible MET. Any observed MET arises primarily from detector-related mismeasurements rather than the physics process itself. The invariant mass distribution of the dielectron system peaks near 90 GeV, reflecting the decay of the intermediate Z boson. Another notable feature is the η distribution, where the signal is largely contained within $|\eta| < 2.5$, corresponding to the acceptance range of the electromagnetic calorimeter (ECAL).

3.2.2 Background Processes

In the context of background processes, several key features can be utilized to discriminate them from the signal during the optimization of selection criteria. One of the prominent background processes is top quark pair production. This process is characterized by a significantly high MET, primarily due to the presence of neutrinos in the final state. Additionally, events from this process tend to exhibit high p_T of the final-state particles and relatively $m_{\ell\ell}$. These distinct features, as illustrated in Figure 3.1, offer valuable handles for suppressing this background in the signal region.

Another relevant background arises from single top quark production. Similar to the top quark pair production, the single top process also involves a neutrino in the final state, resulting in elevated MET. Furthermore, the $m_{\ell\ell}$ distribution in this process remains predominantly in the lower mass region. A notable characteristic that sets this background apart from the DY process is the $\Delta\phi_{\ell\ell}$, which exhibits a broader distribution and deviates from the more back-to-back configuration typically observed in DY events.

In the case of WW, the final state contains two neutrinos, which contribute to a considerably high MET. Additionally, similar to the top-related backgrounds, the $m_{\ell\ell}$ in WW events tends to lie in the lower energy range, distinctly different from the DY process where $m_{\ell\ell}$ is centered around the Z boson mass peak at approximately 90 GeV due to the intermediate Z boson.

For the WZ process, the presence of an on-shell Z boson results in a $m_{\ell\ell}$ distribution that peaks around 90 GeV, similar to DY. However, this process can still be distinguished based on its comparatively high MET resulting from the undetected neutrino, as well as a tendency for the $\Delta\phi_{\ell\ell}$ to peak closer to zero, unlike the more back-to-back lepton configuration in signal events. These properties can be exploited to effectively suppress this background.

Lastly, the ZZ, although it involves two Z bosons in the final state, can still mimic the signal under certain conditions. However, it typically features intermediate levels of MET and $p_T\ell\ell$. These distinguishing kinematic signatures allow for the suppression of the ZZ contribution in the selected signal region.

3.3 Selection Cuts

To purify the signal and suppress background contributions, a set of selection criteria is applied to the dataset. These selection cuts are divided into two categories based on their purpose and the nature of the variables involved: **pre-selection cuts** and **Final cuts**.

The pre-selection cuts ensure that basic event requirements are satisfied before applying more refined selections. Specifically, events are required to contain exactly two electrons ($N_{\text{electron}} = 2$), both of which must pass the Medium identification criteria based on a multivariate analysis. In addition, the two electrons must have opposite electric charges, consistent with the expected signature of processes such as DY process.

After the pre-selection step, final cuts are applied to further refine the phase space of the selected leptons. These cuts include constraints on the p_T and η of each electron, ensuring that the leptons are within optimal regions of the detector for efficient reconstruction and acceptance.

The kinematic variables considered in this step include the $m_{\ell\ell}$, $p_T\ell\ell$, $\Delta\phi_{\ell\ell}$, and the Proj-MET. Each of these observables adds additional discriminative power, enhancing the ability to isolate signal events and improve the overall selection efficiency.

In the following subsections, each category of cuts is discussed in detail, along with the motivation for the choice of variables and the impact of the applied

thresholds.

3.3.1 Pre-Selection Cuts

Since the final state of the process $q\bar{q} \to Z \to e^+e^-$ consists of exactly two electrons (or one electron and one positron), it is essential to select events containing precisely two reconstructed electrons. This requirement is implemented by applying the condition `nElectron == 2`.

MC simulated samples typically include all possible lepton flavor combinations. To isolate events corresponding specifically to an electron–positron final state, a selection based on the Particle Data Group (PDG) identification codes is applied [16].

Furthermore, to ensure that only true, well-identified electrons are selected, each electron must satisfy the `Electron_mvaFall17V2Iso_WP90` working point. This multivariate identification criterion combines several observables to optimize electron identification efficiency while suppressing backgrounds [17, 18].

3.3.2 Final Cuts

Electrons with higher transverse momentum are generally more useful for analysis, as they are more likely to originate from the signal process and are reconstructed with greater efficiency. Therefore, the two selected electrons—referred to as the leading and subleading electrons—are required to have transverse momenta greater than 25 GeV and 20 GeV, respectively.

Although electrons can be produced in all directions, the ECAL provides reliable reconstruction only within its geometric acceptance. To ensure the selected electrons are within the ECAL's coverage, a requirement on the η is imposed: $|\eta| < 2.5$.

Once the selected electrons have passed both the previous cuts, additional cuts are applied to exploit the global characteristics of the event. These cuts are particularly effective in enhancing the purity of signal by leveraging features that are specific to the DY process.

The $m_{\ell\ell}$ in DY events is centered around the mass of the Z-boson, approximately 90 GeV. Events in this mass region are more reliable and less contaminated by background processes, particularly since the dominant backgrounds considered tend to populate lower invariant mass regions. Therefore, a window of 60 GeV $< m_{\ell\ell} <$ 120 GeV is applied to select signal-like events.

In addition, the DY process typically produces electron pairs with low p_T. Consequently, background processes—which often exhibit higher $p_T^{\ell\ell}$—can be effectively suppressed by requiring the transverse momentum of the di-electron system to satisfy $p_T\ell\ell < 40$ GeV.

Another important discriminating variable is the Proj-MET. In signal events, genuine missing energy is negligible, and any observed MET arises primarily from detector mismeasurement. In contrast, background processes involving neutrinos can result in significant missing energy. To exploit this distinction, events with Proj-MET < 25 GeV are selected.

Lastly, electrons produced in the DY process tend to be emitted in opposite directions due to momentum conservation, resulting in a large $\Delta\phi$ between them. Thus, requiring the $\Delta\phi_{\ell\ell}$ to be close to π helps to further reduce background contamination, particularly from processes that produce leptons more isotropically.

Category	**Variable**	**Selection Criterion**
Pre-selection	Number of electrons	`nElectron == 2`
	Real electrons	`Electron_mvaFall17V2Iso_WP90`
	Opposite-sign requirement	$\ell_1 \cdot \ell_2 = -1$
Final	p_T	Leading > 25 GeV, Subleading > 20 GeV
	η	$\lvert\eta\rvert < 2.5$
	$m_{\ell\ell}$	60 GeV $< m_{\ell\ell} <$ 120 GeV
	$p_T\ell\ell$	$p_T\ell\ell < 40$ GeV
	Proj-MET	Proj-MET < 25 GeV
	$\Delta\phi_{\ell\ell}$	$\Delta\phi_{\ell\ell} > 2.5$

Table 3.1: Summary of all selection cuts applied in the analysis, grouped by category.

Chapter 4

Analysis Strategy and Result

In high-energy pp collision experiments, it is essential to carefully select events that are most likely to originate from the signal process under study. While initial selection criteria already reduce a significant portion of background contributions, additional refinement of the event selection is necessary to further improve the quality of the signal sample.

In this analysis, which focuses on the DY production of electron-positron pairs, several background processes—such as top-quark pair production, single-top production, and di-boson events—can result in final states similar to the signal. To distinguish genuine signal events from such backgrounds, a series of additional cuts based on key kinematic variables is applied.

The optimization strategy aims to improve the overall purity of the selected event sample while maintaining a high level of efficiency for true signal events. By carefully designing and applying these cuts, a more reliable and interpretable data sample is obtained for the subsequent stages of the analysis.

4.1 Sequential Cut flow and Event Yield

In order to refine the selection of DY events and suppress contributions from background processes, a sequence of additional cuts is applied. Each cut is designed to exploit specific kinematic features of the signal while reducing the presence of backgrounds that exhibit different characteristics.

The performance of these cuts is evaluated by monitoring the number of surviving events after each selection stage for both signal and background processes.

In addition to monitoring the evolution of event yields throughout the sequence of selection criteria, it is essential to quantitatively evaluate the impact of each individual cut using the concepts of **retention efficiency** and **rejection efficiency**.

The *retention efficiency* (ϵ_{ret}) represents the fraction of events that pass a

specific selection, relative to the number of events that passed the pre-selection cuts. It is defined as:

$$\epsilon_{\text{ret}} = \frac{N_{\text{after cut}}}{N_{\text{pre-selection}}}, \tag{4.1}$$

where $N_{\text{after cut}}$ is the number of events that pass the current cut as well as all previous cuts, and $N_{\text{pre-selection}}$ is the number of events that passed the pre-selection cuts. A high retention efficiency for the signal indicates that the selection criteria effectively preserve genuine Drell–Yan events.

In contrast, the *rejection efficiency* (ϵ_{rej}) measures the fraction of events rejected by the cut and is given by:

$$\epsilon_{\text{rej}} = 1 - \epsilon_{\text{ret}}. \tag{4.2}$$

Maximizing the rejection efficiency for background processes is critical, as it signifies strong suppression of background events while minimizing the impact on signal retention.

A systematic analysis of the retention, rejection, and baseline efficiencies at each stage of the selection provides critical insight into the performance and optimization of the overall event selection strategy.

4.1.1 Pre-Selection Cuts

- **Requirement of Two Electrons with Medium Identification and Opposite Charge**

  The first step in the event selection is to require exactly two reconstructed electrons in each event. This selection targets the final-state topology of the DY process, characterized by the production of an electron-positron pair from the decay of an intermediate Z boson. Events with fewer or more than two electrons are rejected to suppress contributions from processes involving additional leptons, jets misidentified as electrons, or incomplete event reconstruction.

  To ensure the high quality of the selected electron candidates, both electrons are required to pass the multivariate (MVA)-based electron identification algorithm, developed for the CMS experiment, combines several discriminating variables—including track quality, calorimeter shower shapes, and isolation information—into a single score. This identification strategy effectively suppresses backgrounds from hadronic fakes and ensures a high-purity electron selection.

  Additionally, an opposite-charge requirement is imposed on the electron pair to match the expected e^+e^- final state signature from neutral boson decays.

This condition provides further rejection of backgrounds arising the mis-assignment of charge.

These baseline criteria form the foundation for isolating Drell–Yan signal events with high purity and efficiency.

Process	**Before cuts**	**After Cuts**
Drell–Yan	6,189,390	702,364
$t\bar{t}$	87,310	6,651
Single top quark	145,270	225
WZ	47,130	702
ZZ	16,520	511
WW	12,180	707

Table 4.1: Number of events before and after pre-selection cuts, scaled by the cross-section of each process.

4.1.2 Final Cuts

- **p_T and η:** Following the identification of two tightly selected, oppositely charged electrons, additional kinematic requirements are imposed. These cuts are applied to the events that have already passed the previous selection criteria. Results of this cut is shown in Table 4.2

Process	**Before Cuts**	**After Cuts**	$\epsilon_{\mathbf{ret}}$ [%]	$\epsilon_{\mathbf{rej}}$ [%]
DY	702,364	591,809	84.25	15.75
$\boldsymbol{t\bar{t}}$	6,651	4,289	64.48	35.52
Single top quark	225	52	23.11	76.89
WZ	702	546	77.78	22.22
ZZ	511	382	74.75	25.25
WW	707	465	65.77	34.23

Table 4.2: Event yields before and after applying the p_T and η requirements, along with the corresponding rejection and retention efficiencies for each process. The "after cuts" column includes the cumulative effect of the prior cuts as well as the p_T and η cuts.

The leading electron is required to satisfy $p_T > 25\,\mathrm{GeV}$, while the sub-leading electron must have $p_T > 20\,\mathrm{GeV}$.

In addition to the p_T thresholds, both electrons are required to lie within the detector's central acceptance region, defined by the condition $|\eta| < 2.5$.

- $m_{\ell\ell}$: In addition to the P_T and η requirements, we further impose a $m_{\ell\ell}$ cut to ensure the selection of events consistent with the DY process. We apply a requirement of $60\,\text{GeV} < m_{ll} < 120\,\text{GeV}$.

  Table 4.3 below shows the event yields before and after applying the m_{ll} cut, along with the corresponding rejection and retention efficiencies for each process.

Process	Before Cuts	After Cuts	ϵ_{ret} [%]	ϵ_{rej} [%]
Drell–Yan	702,364	576,077	82.02	17.98
Top quarks	6,651	1,780	26.76	73.24
Single top quark	225	24	10.67	89.33
WZ production	702	518	73.79	26.21
ZZ production	511	361	70.64	29.35
WW production	707	208	32.25	67.75

Table 4.3: Event yields before and after applying the pre-selection, p_T, η, and m_{ll} requirements, along with the corresponding rejection and retention efficiencies for each process. The "after cuts" column reflects the cumulative effect of all cuts.

- **Proj-MET:** To enhance signal purity and further suppress background contamination, a selection is applied on the Proj-MET. Various Proj-MET thresholds were systematically tested, and the cut value that maximized signal retention while ensuring strong background rejection was chosen. The details of this optimization procedure are summarized in Table 4.4.

Threshold	**30 GeV**	**25 GeV**	**20 GeV**
Signal Yield	530,329	**503,826**	462,562
Background Yield	1,275	**1,123**	938
Signal ϵ_{ret} [%]	75.50	**71.73**	65.85
Backgrounds ϵ_{ret} [%]	14.49	**12.76**	10.66

Table 4.4: Summary of the optimization for Proj-MET.

The Table 4.5 shows the event yields before and after applying the Proj-MET cut, along with the corresponding rejection and retention efficiencies for each process.

Process	Before Cuts	After Cuts	ϵ_{ret} [%]	ϵ_{rej} [%]
DY	702,364	503,826	71.73	28.27
$t\bar{t}$	6,651	458	6.89	93.11
Single top quark	225	7	3.11	96.89
WZ	702	344	49.00	50.99
ZZ	511	241	47.16	52.84
WW	707	73	10.32	89.68

Table 4.5: Event yields before and after applying the pre-selection, p_T, η, m_{ll}, and proj-MET requirements, along with the corresponding rejection and retention efficiencies for each process. The "after cuts" column reflects the cumulative effect of all cuts.

- $p_T\ell\ell$: A summary of optimization procedure chosen for testing various thresholds of $p_T\ell\ell$ is shown in Table 4.6.

Threshold	**50 GeV**	**40 GeV**	**30 GeV**
Signal Yield	470,226	**453,184**	437,349
Background Yield	522	**439**	297
Signal ϵ_{ret} [%]	66.94	**64.52**	62.27
Backgrounds ϵ_{ret} [%]	5.93	**4.98**	3.37

Table 4.6: Summary of the optimization for $p_T\ell\ell$.

Process	Before Cuts	After Cuts	ϵ_{ret} [%]	ϵ_{rej} [%]
DY	702,364	453,184	64.52	35.48
$t\bar{t}$	6,651	139	2.09	97.91
Single top quark	225	3	1.33	98.67
WZ	702	139	19.80	80.20
ZZ	511	104	20.35	79.65
WW	707	54	7.64	92.36

Table 4.7: Event yields before and after applying the pre-selection, p_T, η, m_{ll}, Proj-MET, and p_{Tll} requirements, along with the corresponding rejection and retention efficiencies for each process. The "after cuts" column reflects the cumulative effect of all cuts.

The table 4.7 shows the event yields before and after applying the $p_T\ell\ell$ cut.

- $\boldsymbol{\Delta\phi_{ll}}$: In addition to the transverse momentum and mass requirements, we further impose a cut on the $\boldsymbol{\Delta\phi_{ll}}$.

  This cut of $\boldsymbol{\Delta\phi_{ll} > 2.5}$ is chosen to reject events where the leptons are more collinear, which are characteristic of backgrounds with additional hard-scattering particles.

  The table 4.8 shows the event yields before and after applying the $\boldsymbol{\Delta\phi_{ll}}$ cut, along with the corresponding rejection and retention efficiencies for each process.

Process	**Before Cuts**	**After Cuts**	$\boldsymbol{\epsilon_{\text{ret}}}$ [%]	$\boldsymbol{\epsilon_{\text{rej}}}$ [%]
DY	702,364	431,586	61.45	38.55
$\boldsymbol{t\bar{t}}$	6,651	97	1.46	98.54
Single top quark	225	2	0.89	99.11
WZ	702	109	15.52	84.47
ZZ	511	67	13.11	86.89
WW	707	45	6.36	93.64

Table 4.8: Event yields before and after applying the pre-selection, $\boldsymbol{p_T}$, $\boldsymbol{\eta}$, $\boldsymbol{m_{ll}}$, Proj-MET, $\boldsymbol{p_{Tll}}$, and $\boldsymbol{\Delta\phi_{ll}}$ cuts, along with the corresponding rejection and retention efficiencies for each process. The "after cuts" column reflects the cumulative effect of all cuts.

4.2 Result

Table 4.9 presents the event yields for the Drell–Yan signal and dominant Standard Model background processes after each selection step. The cuts are applied sequentially, and the event counts are scaled to the respective theoretical cross-sections. Each entry includes the number of surviving events, with the percentage retention relative to the pre-selection level shown in parentheses.

The Drell–Yan signal maintains a high overall retention, with approximately 61% of events preserved after the full set of selection cuts. This indicates the strong compatibility of the selection strategy with the kinematic features of signal events. In contrast, the background contributions, such as top quark pair production and diboson processes, are significantly reduced at each stage. For example, top quark events decrease to a rejection efficiency exceeding 98%. Similar strong suppression is observed for single top and WW backgrounds.

Key variables contributing to this performance include Proj-MET, the $p_T\ell\ell$, and $\Delta\phi_{\ell\ell}$. These exploit the absence of genuine missing energy and the back-

Process	Pre-Selections	p_T, η cuts	$m_{\ell\ell}$ cuts	Proj-MET cuts	$p_T\ell\ell$ cuts	$\Delta\phi_{\ell\ell}$ cuts
Drell–Yan	702,364	591,809 (84.25%)	576,077 (82.02%)	503,826 (71.73%)	453,184 (64.52%)	431,586 (61.45%)
Top quarks	6,651	4,289 (64.48%)	1,780 (26.76%)	458 (6.89%)	139 (2.09%)	97 (1.46%)
Single top	225	52 (23.11%)	24 (10.67%)	7 (3.11%)	3 (1.33%)	2 (0.89%)
WZ	702	546 (77.78%)	518 (73.79%)	344 (49.00%)	139 (19.80%)	109 (15.52%)
ZZ	511	382 (74.75%)	361 (70.64%)	241 (47.16%)	104 (20.35%)	67 (13.11%)
WW	707	465 (65.77%)	208 (32.25%)	73 (10.32%)	54 (7.64%)	45 (6.36%)

Table 4.9: Number of events remaining after each sequential cut, scaled to the respective cross-sections. Percentages indicate the retention at each stage with respect to pre-selection yield.

Process	Pre-selection Yield	After Analysis Yield	$\epsilon_{\mathbf{ret}}$ [%]	$\epsilon_{\mathbf{rej}}$ [%]
Signal	702,364	431,586	61.45	38.55
Backgrounds	8,796	321	3.64	96.35

Table 4.10: Summary of signal (DY) and backgrounds ($t\bar{t}$, single top quark, WZ, ZZ, and WW) yields, retention efficiencies (ϵ_{ret}), and rejection efficiencies (ϵ_{rej}) after the full set of selection cuts.

to-back topology of DY events, as opposed to the broader distributions seen in background processes.

Overall, the cut-flow clearly demonstrates the effectiveness of the event selection strategy in retaining the DY signal while minimizing background contamination, thereby enhancing the purity of the selected sample for subsequent analysis.

In summary, the sequential application of cuts results in a highly purified event sample. The DY signal retains approximately 61.45% of its initial yield, while the major background processes are suppressed by more than 85–99%. These results demonstrate the effectiveness of the cut-based optimization strategy in enhancing the sensitivity of the analysis and isolating the DY signal with high purity for further study.

Chapter 5

Summary

This dissertation presents a comprehensive analysis of the DY process in the electron–positron final state using MC simulated data from the CMS experiment, corresponding to the 2016 LHC run at a center-of-mass energy of 13 TeV. The DY process, due to its clean experimental signature and robust theoretical modeling, serves as an important benchmark for validating the performance of the CMS detector.

The study begins with the characterization of both signal and background processes. While the DY process constitutes the primary signal, significant backgrounds arise from top quark pair production, single top processes, and di-boson production (WW, WZ, ZZ). Each background was studied using dedicated MC samples, with a detailed comparison of their kinematic features against those of the DY signal.

A rigorous selection strategy was implemented in two stages. The pre-selection step required events to contain exactly two tightly identified, oppositely charged electrons. This retained 61.45% DY events while significantly reducing backgrounds. Subsequently, a sequence of optimized kinematic cuts was applied to further enhance signal purity. These cuts included requirements p_T, η, $m_{\ell\ell}$, Proj-MET, $p_T\ell\ell$, and $\Delta\phi_{\ell\ell}$.

Each cut was evaluated for its impact on signal efficiency and background rejection. For example, applying Proj-MET < 25 GeV retained 71.73% of the DY signal and removed over 94% of the top background. The final cut, $\Delta\phi_{\ell\ell} > 2.5$, leveraged the back-to-back topology of DY decays, retaining 61.45% of the signal while reducing remaining top background by over 98%.

These results highlight the effectiveness of a structured, cut-based analysis approach for isolating the DY signal in the dielectron channel and demonstrate the discriminative power of well-chosen kinematic observables in high-energy physics analyses.

Bibliography

[1] CERN Open Data Portal. CMS Experiment Data. https://opendata.cern.ch/search?f=experiment%3ACMS, 2025. Accessed: 2025-05-03.

[2] Sidney D. Drell and Tung-Mow Yan. Massive lepton-pair production in hadron-hadron collisions at high energies. *Phys. Rev. Lett.*, 25:316–320, Aug 1970.

[3] CMS Collaboration. CMS NanoAOD Documentation. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAOD, 2025. Accessed: 2025-05-03.

[4] Torbjörn Sjöstrand et al. *PYTHIA 8.3 Online Manual*. The PYTHIA Collaboration, 2025. Accessed: 2025-05-03.

[5] CMS Collaboration. Simulated dataset DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.CRNB.POY1, 2024. CERN Open Data Portal.

[6] CMS Collaboration. Simulated dataset TTTo2L2Nu_TuneCP5_13TeV in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.4RTG.JPI2, 2024. CERN Open Data Portal.

[7] CMS Collaboration. Simulated dataset ST_t-channel_top _4f _InclusiveDecays _TuneCP5_13TeV in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.448S.OXEJ, 2024. CERN Open Data Portal.

[8] CMS Collaboration. Simulated dataset ST_t-channel _antitop_4f _Inclusive Decays _TuneCP5 _13TeV-powheg-madspin-pythia8 in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.D885.EHVZ, 2024. CERN Open Data Portal.

[9] CMS Collaboration. Simulated dataset ST_tW_top_5f_inclusive Decays _TuneCP5_13TeV-powheg-pythia8 in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.B9V5.WXFJ, 2024. CERN Open Data Portal.

[10] CMS Collaboration. Simulated dataset ST_tW_antitop _5f _inclusive Decays_TuneCP5 _13TeV-powheg-pythia8 in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.GOSF.G1IA, 2024. CERN Open Data Portal.

[11] CMS Collaboration. Simulated dataset ST_s-channel_4f_ leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8 in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.Y5N3.PYKM, 2024. CERN Open Data Portal.

[12] CMS Collaboration. Simulated dataset WZ_TuneCP5_13TeV-pythia8 in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.1H64.40FS, 2024. CERN Open Data Portal.

[13] CMS Collaboration. Simulated dataset ZZ_TuneCP5_13TeV-pythia8 in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.WMBQ.G35Q, 2024. CERN Open Data Portal.

[14] CMS Collaboration. Simulated dataset WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8 in NANOAODSIM format for 2016 collision data. https://doi.org/10.7483/OPENDATA.CMS.TZ1Q.TQDV, 2024. CERN Open Data Portal.

[15] Latino CMS Group. Latinoanalysis: Cross sections for 2018 mc samples in nanogardener. https://github.com/latinos/LatinoAnalysis/blob/master/NanoGardener/python/framework/samples/samplesCrossSections2018.py, 2018. Accessed: 2025-04-13.

[16] L. Garren, I.G. Knowles, T. Sjöstrand, et al. Monte carlo particle numbering scheme. *Eur. Phys. J. C*, 15:205–207, 2000.

[17] CMS Open Data Workshop. Electrons & photons - cms physics objects. https://cms-opendata-workshop.github.io/workshop2024-lesson-physics-objects/02-electrons.html, 2024. Accessed: 2025-04-20.

[18] CMS Collaboration. Electron and photon reconstruction and identification with the cms experiment at the cern lhc. *Journal of Instrumentation*, 16(05):P05014, May 2021.

Appendix A

Setting Up ROOT for Data Analysis

A.1 Introduction to ROOT

ROOT is a powerful data analysis framework developed by CERN. It provides a set of tools for processing large datasets, performing statistical analysis, generating visualizations, and working with complex data structures. ROOT is widely used in high-energy physics experiments, such as those conducted at the LHC, for analyzing data from particle collisions.

In this thesis, ROOT is used for the analysis of the Drell-Yan (DY) process in the CMS experiment at the LHC. The software is employed to handle large volumes of data, conduct statistical analysis, and generate plots that assist in the interpretation of the results.

A.2 Installing and Setting Up ROOT

A.2.1 Install ROOT Dependencies

Before installing ROOT, make sure the necessary dependencies are installed on your system. To do so, run the following commands on an Ubuntu-based Linux machine:

```
sudo apt update

sudo apt install -y build-essential gcc g++ cmake python3-all \
  libffi-dev libssl-dev libxpm-dev libxft-dev libxext-dev \
  libpng-dev libjpeg-dev libgif-dev libtiff-dev libcurl4-openssl-dev \
  libboost-all-dev
```

This installs the essential development tools and libraries that ROOT requires for compiling and running on your system.

A.2.2 Set Up ROOT

After installing the required dependencies, you can proceed to install ROOT itself. Here's how to configure the environment for ROOT:

First, download and install the ROOT source code. You can get the latest version from the official ROOT website, or use 'wget' to download the ROOT tarball:

```
wget https://root.cern.ch/download/root_v6.34.02.source.tar.gz
```

Then, extract the tarball and follow the build instructions in the directory:

```
tar -xvzf root_v6.34.02.source.tar.gz
cd root-6.34.02
mkdir build
cd build
cmake ..
make -j4
sudo make install
```

Next, configure the ROOT environment by adding the following lines to your `.bashrc` file (replace `/path/to/root` with the actual path to where ROOT is installed):

```
export ROOTSYS=/path/to/root  # Replace with actual ROOT install path
export PATH=$ROOTSYS/bin:$PATH
export LD_LIBRARY_PATH=$ROOTSYS/lib:$LD_LIBRARY_PATH
export PYTHONPATH=$ROOTSYS/lib:$PYTHONPATH
```

To apply the changes, run the following command to reload the environment:

```
source ~/.bashrc
```

A.2.3 Verify ROOT Installation

To verify that ROOT has been successfully installed, run the following command in your terminal:

If the installation was successful, you should see the following output:

```
anujraghav252@anuj-Inspiron-3501:~/Desktop/Dissertation_workshop$ root
   ------------------------------------------------------------------
  | Welcome to ROOT 6.34.02                        https://root.cern |
  | (c) 1995-2024, The ROOT Team; conception: R. Brun, F. Rademakers |
  | Built for linuxx8664gcc on Dec 28 2024, 08:39:22                 |
  | From tags/v6-34-02@v6-34-02                                      |
  | With c++ (Ubuntu 13.3.0-6ubuntu2~24.04) 13.3.0                   |
  | Try '.help'/'.?', '.demo', '.license', '.credits', '.quit'/'.q'  |
   ------------------------------------------------------------------

root [0]
```

This message confirms that ROOT is successfully installed and ready for use in analysis.

Appendix B

Kinematic Variable Calculations

This appendix outlines the computation of key kinematic variables used in the analysis of the Drell–Yan process, as well as the logic applied for event selection based on those variables.

B.1 Dilepton System Variables

For each selected event with two tight electrons of opposite charge, the following Lorentz vectors are constructed using the ROOT `TLorentzVector` class:

```
TLorentzVector e1, e2;
e1.SetPtEtaPhiM(Electron_pt[i1], Electron_eta[i1], Electron_phi[i1], 0.000511);
e2.SetPtEtaPhiM(Electron_pt[i2], Electron_eta[i2], Electron_phi[i2], 0.000511);
TLorentzVector dilepton = e1 + e2;
```

From these vectors, the following variables are computed:

- **Invariant mass** ($m_{\ell\ell}$):

$$m_{\ell\ell} = \sqrt{(E_1 + E_2)^2 - |\vec{p}_1 + \vec{p}_2|^2}$$

- **Transverse momentum of the dilepton system** ($p_{T,\ell\ell}$): calculated using `dilepton.Pt()`.
- **Azimuthal angular separation between leptons** ($\Delta\phi_{\ell\ell}$): calculated using

```
double dphill = fabs(e1.DeltaPhi(e2));
```

These variables are used to apply the following selection criteria:

- $m_{\ell\ell} \in (60, 120)$ GeV
- $p_T \ell\ell < 40$ GeV
- $\Delta\phi_{\ell\ell} > 2.5$

B.2 Projected MET Calculation

To further suppress background contributions, particularly from `WW` and $t\bar{t}$ processes, a variable called **projected MET** is used. It is defined based on the angle between the MET vector and the directions of the two leading electrons:

$$\text{projected\_MET} = \begin{cases} p_T^{\text{miss}} \cdot \sin(\Delta\phi_{\min}) & \text{if } \Delta\phi_{\min} < \pi/2 \\ p_T^{\text{miss}} & \text{otherwise} \end{cases}$$

where $\Delta\phi_{\min} = \min(\Delta\phi(\ell_1, \vec{p}_T^{\,\text{miss}}), \Delta\phi(\ell_2, \vec{p}_T^{\,\text{miss}}))$.

The $\Delta\phi$ calculation uses a utility function that ensures the result lies within $[-\pi, \pi]$:

```
float deltaPhi(float phi1, float phi2) {
    float dphi = phi1 - phi2;
    while (dphi > M_PI) dphi -= 2 * M_PI;
    while (dphi <= -M_PI) dphi += 2 * M_PI;
    return dphi;
}
```

The resulting projected MET value is stored as a new branch in the NANOAOD tree and later used in the selection cut:

```
projected_MET < 25 GeV
```

This cut significantly reduces contributions from events with genuine MET not aligned with lepton directions.

Appendix C

Code for Event Selection and Applying Cuts

In this section, the code used to select events and apply sequential cuts in the analysis of the DY process is provided. The code performs the following:

- Selects tight electrons using a set of predefined cuts.
- Applies sequential cuts based on the number of electrons, their kinematics, the invariant mass of the electron pair, and other relevant variables.
- Tracks the number of events passing each cut to evaluate the efficiency.

C.1 Code for Electron Event Selection and Cut Flow Analysis

```
#include <TFile.h>
#include <TTree.h>
#include <TCanvas.h>
#include <TH1D.h>
#include <iostream>
#include <cmath>
#include <TLorentzVector.h>

using namespace std;

void Electron_Cut_Flow(){
    TFile *file = TFile::Open("DYtoLL_M50.root"); //Root file in
       use
    if (!file || file->IsZombie()) {
        cerr << "Error opening the ROOT file" << endl;
        return;
    }
```

```
TTree *tree = (TTree*)file->Get("Events");
if (!tree) {
    cerr << "Error getting the TTree from the ROOT file" <<
       endl;
    return;
}

//Declareing variables
UInt_t nElectron;
Float_t Electron_pt[10];
Float_t Electron_eta[10];
Float_t Electron_phi[10];
Int_t Electron_pdgId[10];
Bool_t Electron_mvaFall17V2Iso_WP90[10];
Float_t projected_MET;

//Setting branch addresses
tree->SetBranchAddress("nElectron", &nElectron);
tree->SetBranchAddress("Electron_pt", Electron_pt);
tree->SetBranchAddress("Electron_eta", Electron_eta);
tree->SetBranchAddress("Electron_phi", Electron_phi);
tree->SetBranchAddress("Electron_pdgId", Electron_pdgId);
tree->SetBranchAddress("Electron_mvaFall17V2Iso_WP90",
   Electron_mvaFall17V2Iso_WP90);
tree->SetBranchAddress("projected_MET", &projected_MET);

//setting counters
int eventAfterstage1=0; //After nElectron ==2 and opposite
   charge
int eventAfterstage2=0; //After electrons pass eta and pT cuts
    plus previous
int eventAfterstage3=0; //After electrons pass mll cut and
   previous
int eventAfterstage4=0; //After electrons pass pmet cut and
   previous
int eventAfterstage5=0; //After electrons pass ptll cut and
   previous
int eventAfterstage6=0; //After electrons pass dphill cut and
   previous

Long64_t nEntries = tree->GetEntries();

for (Long64_t i = 0; i < nEntries; i++) {
    tree->GetEntry(i);

    // Select tight electrons
    vector<int> tightElectrons;
```

```
for (UInt_t j = 0; j < nElectron; j++) {
    if (Electron_mvaFall17V2Iso_WP90[j]) {
        tightElectrons.push_back(j);
    }
}

if (tightElectrons.size() != 2) continue;

int i1 = tightElectrons[0];
int i2 = tightElectrons[1];

if (Electron_pdgId[i1] * Electron_pdgId[i2] < 0) {
    eventAfterstage1++;

    int lead = (Electron_pt[i1] > Electron_pt[i2]) ? i1 :
       i2;
    int sublead = (lead == i1) ? i2 : i1;

    if ((Electron_pt[lead] > 25) && (Electron_pt[sublead]
       > 20) &&
        (fabs(Electron_eta[lead]) < 2.5) && (fabs(
           Electron_eta[sublead]) < 2.5)) {

        eventAfterstage2++;

        TLorentzVector el1, el2;
        el1.SetPtEtaPhiM(Electron_pt[lead], Electron_eta[
           lead], Electron_phi[lead], 0.000511);
        el2.SetPtEtaPhiM(Electron_pt[sublead],
           Electron_eta[sublead], Electron_phi[sublead],
           0.000511);
        TLorentzVector dilepton = el1 + el2;

        double mll = dilepton.M();
        double ptll = dilepton.Pt();
        double dphill = fabs(el1.DeltaPhi(el2));

        if (mll > 60 && mll < 120) {
            eventAfterstage3++;

            if (projected_MET < 25) {
                eventAfterstage4++;

                if (ptll < 40) {
                    eventAfterstage5++;

                    if (dphill > 2.5) {
```

```
                                    eventAfterstage6++;
                                }
                            }
                        }
                    }
                }
            }
        }

        //Print output
        cout << "\n\nCut Flow results for DYtoLL_M50.root:\n" << endl;
        cout << "Total events in file "<<nEntries<<endl;
        cout << "Events after nElectron == 2 and opposite charge: " <<
             eventAfterstage1 << endl;
        cout << "Events after |eta| < 2.5 and leading pT > 25 GeV,
           subleading pT > 20 GeV: " << eventAfterstage2 << endl;
        cout << "Events after dilepton mass 60 < mll < 120 GeV: " <<
           eventAfterstage3 << endl;
        cout << "Events after projected MET < 25 GeV: " <<
           eventAfterstage4 << endl;
        cout << "Events after pTll < 40 GeV: " << eventAfterstage5 <<
           endl;
        cout << "Events after |dphill| > 2.5: " << eventAfterstage6 <<
             endl;
        cout << "\nFinal events passing all cuts: " <<
           eventAfterstage6 << endl;
        cout << "\n\n";

        // Close the file
        delete file;
}
```

C.2 Explanation of the Cut Flow

The code performs the following steps:

1. **Electron Selection**: Only tight electrons (defined by the 'mvaFall17V2Iso_WP90' criterion) are selected. If there are exactly two tight electrons with opposite charges, the event proceeds to the next cut.
2. **Kinematic Cuts**: The leading electron must have $p_T > 25\,\text{GeV}$ and the subleading electron must have $p_T > 20\,\text{GeV}$. Both electrons should lie within the pseudo-rapidity range $|\eta| < 2.5$.
3. **Dilepton Mass Cut**: The invariant mass of the electron pair must satisfy $60 < m_{ll} < 120\,\text{GeV}$.
4. **Projected MET Cut**: The projected missing transverse energy must be less than 25 GeV.

5. **Dilepton Transverse Momentum Cut**: The transverse momentum of the dilepton system must be less than 40 GeV.
6. **Delta Phi Cut**: The azimuthal angle difference between the two electrons must be greater than 2.5.

The final output provides the number of events passing each stage of the cut flow, along with the efficiency of the selection process.

C.3 GitHub Repository

The complete code for event selection, cut flow analysis, and additional codes and useful details can be found in the following GitHub repository:*anujraghav252/msc-diss*