Z-Boson + Jet Analysis

Emre Yağız Öztürk

22 July 2025

Abstract

This study uses proton-proton collision data from CERN's ATLAS experiment to investigate the production of the Z boson via the dilepton ($\ell^+\ell^-$) decay channel.The invariant mass histogram of the Z boson was plotted and fitted with the Breit Weigner function to verify its accuracy. Furthermore, the analysis examined the kinematic distributions of leptons and jet properties.

Within the scope of the analysis, not only leptons but also the kinematic distributions of the jets generated in the events were investigated. Variables such as the jets' transverse momentum and azimuthal angle were analyzed using histograms to evaluate the traces of hadronization processes and quark-gluon interactions in the collisions.

This study both tests electroweak interactions and contributes to the evaluation of the experimental results of QCD. Concepts such as jet physics, hadronization, and quark-gluon plasma form the theoretical background of the study and contribute to the understanding of the particle distributions observed in high-energy collisions.

1 Introduction

1.1 Quantum Chromodynamics (QCD)

Quantum Chromodynamics (QCD) is the theory that describes the strong interaction between quarks, mediated by gluons. Quarks carry a special charge, called "color charge", which is carried by gluons [1]. QCD has two key properties:

- **Color Confinement:** It states that quarks cannot be directly observed in their free state but can only be found in hadrons.
- **Asymptotic freedom:** It states that quarks interact weakly when they get very close to each other, but the interaction becomes stronger as they move away from each other.

These properties play a crucial role in understanding the behavior of particles and the observed jet structure resulting from high-energy collisions. QCD forms

the theoretical basis for many phenomena observed in high-energy physics experiments, such as jets, hadronization processes, and quark-gluon plasmas (QGP).

1.2 Quark–Gluon Plasma (QGP)

It is a state of matter formed by the combination of quarks and gluons, the most fundamental constituents of matter, and emerges under extraordinary conditions of temperature and density. Under normal conditions, quarks cannot be observed directly because, according to Quantum Chromodynamics (QCD), they are always found within hadrons due to the "color confinement" resulting from the strong interaction [2]. However, QCD predicts that when the temperature or energy density exceeds a certain threshold, the confinement property disappears, and quarks can exist with gluons in a free environment. The state of matter formed in this environment is called Quark-Gluon Plasma.

QGP cannot be observed directly because this state is very short-lived and rapidly decays into hadrons. However, some experimental observations provide indirect evidence for the existence of QGP. Heavy-ion collisions, particularly those performed at accelerators such as the LHC at CERN and RHIC at Brookhaven, are among the most effective ways to produce QGP in the laboratory. These experiments test both the extreme limits of QCD and contribute to our understanding of the fundamental properties of matter.

1.3 Hadronization

Hadronization is a fundamental process that allows free quarks and gluons, produced in high-energy collisions, to transform into observable particles. This phenomenon is directly related to the principle of color confinement, a key consequence of Quantum Chromodynamics (QCD). According to QCD, quarks and gluons carrying color charge cannot be directly observed; these particles become stable in nature only by forming color-neutral systems. Therefore, a free quark or gluon emerging in a high-energy collision can form other quark-antiquark pairs a short distance away, creating new hadrons, before reaching a detector.

The jets that emerge after the collision are actually structures formed by the propagation of a large number of hadrons in a specific direction, resulting from hadronization. Hadronization plays a critical role not only in high-energy physics but also in understanding the transition regions of QCD and the transformation of particles from quark-gluon plasmas into observable forms.

1.4 Jets

Jets are particle clusters formed by the aggregation of many hadrons moving in a narrow direction resulting from high-energy particle collisions. These structures are considered indirect traces left in detectors by high-energy quarks and gluons that cannot be directly observed. Because the quarks and gluons resulting from the collisions carry a color charge, they cannot exist freely in a vacuum.

These particles undergo hadronization for a short distance before reaching the detector, producing a large number of hadrons. Because the resulting hadrons generally disperse in a direction close to the original direction of the quark or gluon, detectors observe jet structures that appear to be a single particle but actually contain dozens of hadrons [3].

The existence of jets is directly related to the principles of asymptotic freedom and confinement, fundamental predictions of Quantum Chromodynamics (QCD). In dynamics where these two principles work together, jets emerge as a natural consequence of quarks first trying to move away from each other and then transforming into hadrons. Jets are not only theoretically important; they also play a central role in many experimental applications in high-energy physics, such as event selection, signal/background separation, kinematic analysis, and detector calibration. Jets, particularly observed in proton-proton collisions, are indispensable tools for understanding particle distributions and unraveling fundamental interaction processes.

In this study, the production of the Z boson in the dilepton decay channel was analyzed, while variables such as the number of jets per event, the jet transverse momentum (p_T), the azimuthal angle (ϕ), and the energy (E) were also examined. The resulting histograms reveal the kinematic properties of the jets and provide information about the internal structure of the event. This analysis is important for evaluating the observational traces of QCD processes and provides a powerful example of how jets are used in experimental analyses in particle physics. Furthermore, by examining the distribution of the jets, inferences can be made regarding whether the event is centralized or more diffuse, and the extent of background effects. In this respect, jet analysis contributes not only to the understanding of Z boson production but also to the broader interpretation of collision events.

2 Methodology

Lepton Selection Criteria and Event Filtering

To determine the events to be included in the analysis, various selection criteria were applied to the leptons. First, the number of leptons present in each event, their transverse momenta (p_T), pseudorapidity (η), electric charge, type, whether they met the strict identification criteria.

Based on this information, the following selections were applied in this order:

- Exactly two leptons must be present in each event,
- Both leptons must satisfy the $p_T > 20\,\text{GeV}$ condition,
- The pseudorapidity of both leptons must be in the range $|\eta| < 2.5$,
- The total electric charge of the two leptons must be zero (i.e., they are oppositely charged pairs),

- Both leptons must be muons,
- The tight ID criteria must be met,
- The simulation must match the correct particle (truth matching).

Events that simultaneously met all of these criteria were selected for analysis and filtered with a mask called `combined_mask`. Thus, only events that were physically meaningful, reliably identified in the detector, and corresponding to the Z boson decay were included in the analysis.

Invariant Mass Calculation

Among the events that met the selection criteria, those containing two leptons each were filtered, and invariant mass calculations were performed on these events. To this end, the transverse momentum (p_T), azimuthal angle (ϕ), pseudorapidity (η), and energy (E) of the selected leptons were obtained. Using these parameters, the Cartesian momentum components of each lepton were calculated as follows:

$$p_x = p_T \cos\phi, \quad p_y = p_T \sin\phi, \quad p_z = p_T \sinh\eta$$

Using the four-vector components (E, p_x, p_y, p_z) thus obtained, the invariant mass of the system of two leptons is calculated by means of the formula known from special relativity:

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

The calculated invariant mass values for all suitable events are visualized in a histogram. In this distribution, the peak corresponding to the decay of the Z boson is clearly visible. The mass is given in on the horizontal axis of the histogram, and the number of events is shown on the vertical axis.

Breit-Wigner Fit to the Z Boson Peak

To more accurately determine the characteristic mass peak of the Z boson, a curve fit was performed to the invariant mass distribution using the Breit-Wigner function. This function is defined as follows [4]:

$$f(m) = \frac{A\,\Gamma^2}{(m^2 - M_0^2)^2 + M_0^2\Gamma^2}$$

Here, m represents the invariant mass, A the scaling constant, M_0 the resonance peak (the expected Z boson mass) and Γ the decay width of the Z boson.

For the fitting process, an invariant mass histogram was first plotted and the centroid and event count for each bin in the histogram were calculated. These data were used as input to the curve fitting function. The initial parameters were:

- The maximum number of events in the data distribution for A
- 91.2 GeV, which is approximately the Z boson mass for M_0,
- 2.5 GeV for Γ.

The fitting process was performed using the `scipy.optimize.curve_fit` function, and the parameters were optimized. The resulting fitting curve was visualized with the invariant mass histogram.

Analysis of Signal and Background Events

To numerically separate the events corresponding to the Z boson, the signal and background regions were defined on the invariant mass distribution. The signal region was defined as the interval $[86, 96]$ GeV, which includes the mass peak of the Z boson. This interval corresponds to the peak in the data distribution and provides sufficient coverage considering the width of the Z boson (Γ).

The region representing the background events was defined as two regions outside the Z resonance: $[60, 80]$ GeV and $[100, 120]$ GeV. These regions were chosen as the intervals where the signal's influence is minimal and represent the contribution of the continuous background.

As a result of these definitions, the number of events in the signal and background regions was counted separately.

Jet Number and Transverse Momentum Distribution

To analyze the jets produced with the Z boson, the number of jets present in each event (*jet multiplicity*) and the transverse momentum (p_T) of these jets were examined. Since there is more than one jet per event in the ATLAS dataset, the total number of jets per event was first obtained. Using this information, the distribution of jet numbers per event was visualized as a histogram.

Furthermore, the transverse momentum value of each jet was considered separately, and the jets from all events were combined to form their p_T distribution. Since the jets' transverse momentum is in MeV, it was analyzed by converting it to GeV. This distribution was used to evaluate the frequency of high-energy jets in these events.

The two separate distributions obtained—jet multiplicity and p_T histogram—were presented side by side on a single graph.

Jet-Lepton Angular Separation (ΔR) Analysis

To examine the angular separation of jets and leptons in space, the distance ΔR between the leading jet (with the highest p_T) and the leading lepton of each event was calculated. This quantity is used to measure the spatial separation of particles at the detector and is defined as follows:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

Here η is the pseudorapidity of the particle and ϕ is the azimuthal angle. Only events containing at least one jet and one lepton were considered in the analysis. Information on jet and lepton was prepared in four-vector form and analyzed using the `vector` library.

The resulting distribution ΔR is presented as a histogram to visualize the angular separation between the leader jet and the leader lepton.

Jet Abundance Analysis in the Z Boson Signal Region

To analyze the distribution of jet numbers per event for events thought to have produced the Z boson, only events with invariant mass values in the range $86 < M < 96$ GeV were selected. This filtering allowed the isolation of events corresponding to the Z boson and made the jet analysis more meaningful.

For these selected signal events, the number of jets observed in each event was counted and visualized as a histogram. This allowed us to examine the number of jets typically associated with Z boson production.

Furthermore, the event rates corresponding to specific jet numbers were calculated as percentages and made available for interpretation.

Calculating the Transverse Momentum of the Z Boson

To analyze the distribution of the transverse momentum (p_T^Z) of the Z boson, only events containing two leptons were selected. The four-vector components of each lepton within these events, transverse momentum (p_T), pseudorapidity (η), azimuthal angle (ϕ) and energy (E) were used to generate four-vectors using the `vector` library.

For each event, the four-vectors of the two leptons were added to obtain the combined four-vector corresponding to the Z boson. The transverse momentum component of this vector (p_T^Z) is an important quantity to understand the kinematic properties of the Z boson production event.

Only events with invariant masses in the range $86 < M < 96$ GeV (the signal region) were included in the analysis, and the transverse momenta of the Z bosons in these events were visualized using a histogram.

3 Results

All Events	668152
2 lepton	664613
... + $p_T > 20$ GeV	555056
... + $\|\eta\| < 2.5$	549246
... + opposite charge	541459
... + flavor == muon	270621
... + tight ID	240075
... + truth matched	240075

Table 1: Applied cut-offs and the number of events remaining thereafter

As seen in Table 1, while a total of 668152 events were available before analysis, only 240075 events were selected for analysis after a seven-stage selection criterion was applied for physical and experimental accuracy. This corresponds to a filtering rate of approximately 35.93%. The dataset was significantly simplified by eliminating leptons with low p_T or high pseudorapidity values.

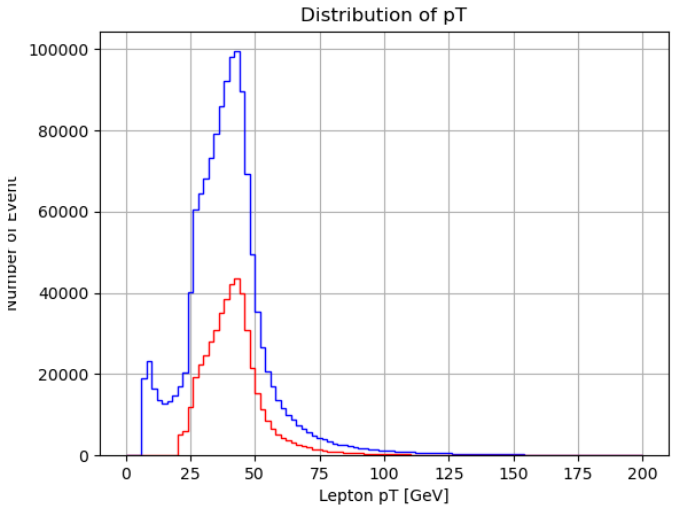

Figure 1: Pure Data (Blue) and Cut Data (Red) Histogram

Figure 1 shows a comparison of the distributions of lepton transverse momenta (p_T) for the raw data (blue) and the data past the selection criteria (red). Leptons with low p_T values are seen to be dominant in the raw data. However, the $p_T > 20$ GeV cutoff applied to the analysis excluded these low-energy leptons, and only physically reliable events were retained in the analysis.

This cutoff is important both for reducing experimental noise and for isolating decays from high-energy processes such as the Z boson. The graph clearly demonstrates the effectiveness of this filter.

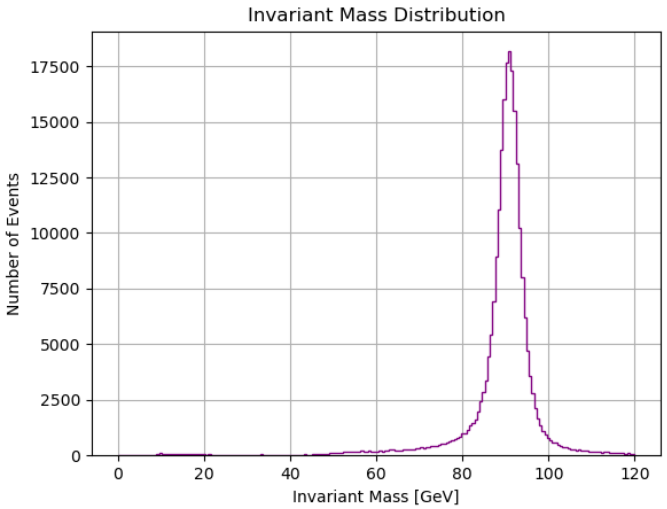

Figure 2: Z-Boson Invariant Mass Histogram

Figure 2 shows the invariant mass ($M_{\ell\ell}$) distribution obtained from the total four-momentums of two leptons. A distinct peak is observed in the histogram around 91 GeV. This peak corresponds to the mass of the Z boson and can be interpreted as a clear sign of Z boson production in the analyzed data set.

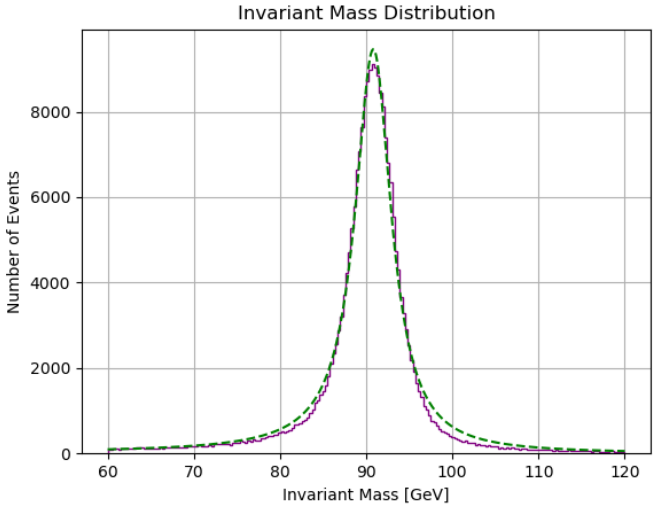

Figure 3: Breit-Wigner Fit (Green Line)

Figure 3 shows the fit curve for the invariant mass distribution obtained from lepton pairs, along with the Breit–Wigner function applied to this distribution. The peak in the histogram represents events where the Z boson was produced, and the fit process made it possible to more precisely quantify the center and width of this peak.

The resonance center obtained from the fit is approximately $M_0 \approx 91.2$ GeV, and the width parameter is around $\Gamma \approx 2.5$ GeV. These values agree well with the known Z boson properties in the literature [5]. Thus, the obtained distribution is confirmed to correspond to the expected Z resonance.

Within the scope of the analysis, a signal region ($86 < M < 96$ GeV) and two background regions ($60 < M < 80$ GeV and $100 < M < 120$ GeV) were defined on the invariant mass distribution to determine the number of events corresponding to the Z boson. These regions were selected based on the shape of the distribution and the known mass of the Z boson.

In the next step, the signal region corresponding to the Z boson on the invariant mass distribution and the background regions outside this region were defined. The signal region was chosen as the range $86 < M < 96$ GeV, which was determined to include the known resonance peak of the Z boson. The background regions were taken as the ranges 60–80 GeV and 100–120 GeV, where the Z boson contribution is minimal. Due to this regional distinction, the density of events related to Z boson production was compared with the number of events originating from background processes, and the signal clarity was numerically evaluated. Total of **179655** events were observed in the signal region, while a total of **21048** events were observed in the background regions.

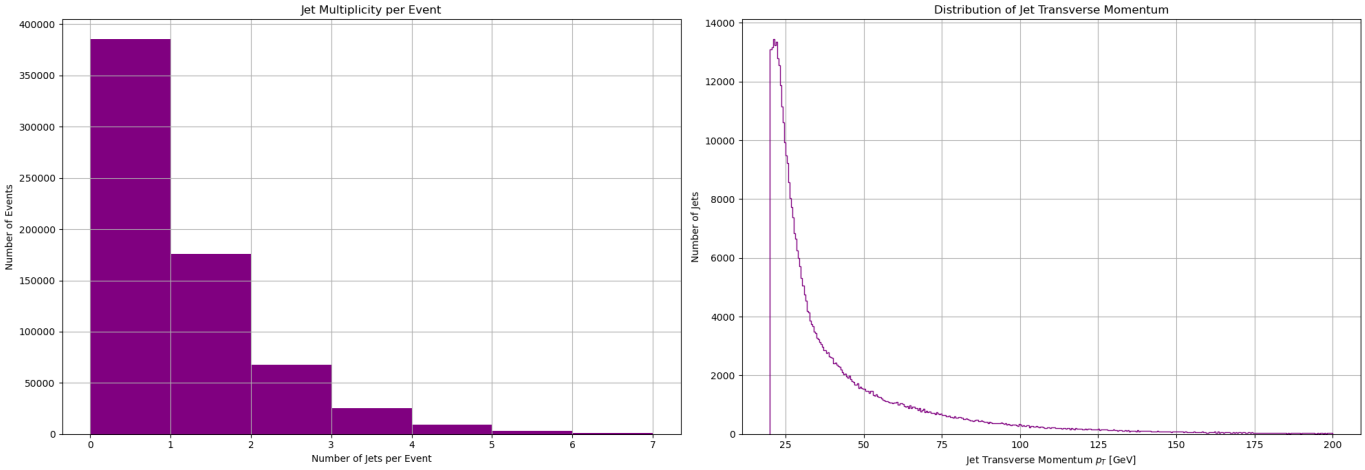

Figure 4: Jet Number and Transverse Momentum Distribution

Figure 4 presents the jet multiplicity and jet transverse momentum (p_T) distributions for all events used in the analysis. The histogram on the left shows the distribution of the number of jets observed in each event, with the majority of events containing 0 to 2 jets. This suggests that low-multiplicity hadronic activity dominates the analyzed dataset.

The histogram on the right shows the transverse momentum distribution of all jets. The distribution is concentrated at low p_T values, suggesting that high-momentum jets are less frequently observed. This trend suggests that hadronic jets are mostly produced by low-energy processes, with high p_T jets corresponding to more selective events.

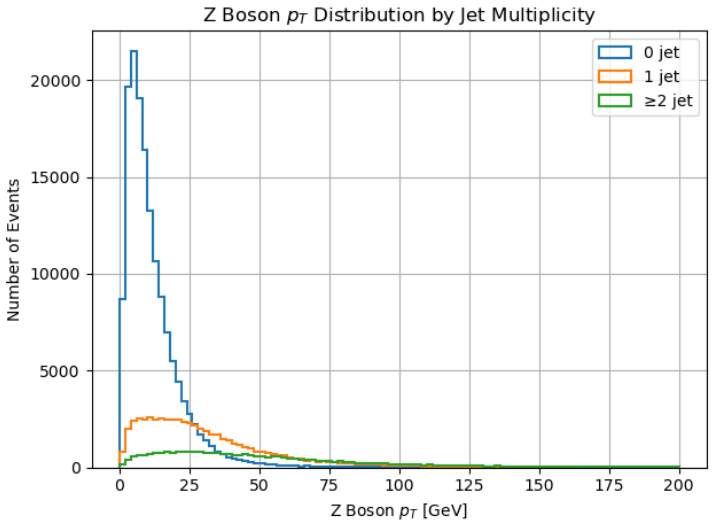

Figure 5: Jet multiplicity and jet transverse momentum (p_T) distributions

In Figure 5, the number of jets per event (left) and the transverse momentum (p_T) distribution of all jets (right) are visualized for all events in the analysis.

The jet multiplicity distribution on the left shows that 0 to 2 jets were observed in the majority of events. This suggests that the vast majority of the dataset corresponds to processes with low hadronic activity, and that events with isolated production of the Z boson are particularly dominant.

The p_T distribution on the right indicates that the jets mostly have low transverse momentum values, suggesting that hadronic activity corresponds to soft scattering rather than high-energy processes. The small number of jets with higher p_T values observed suggests that such events are rare.

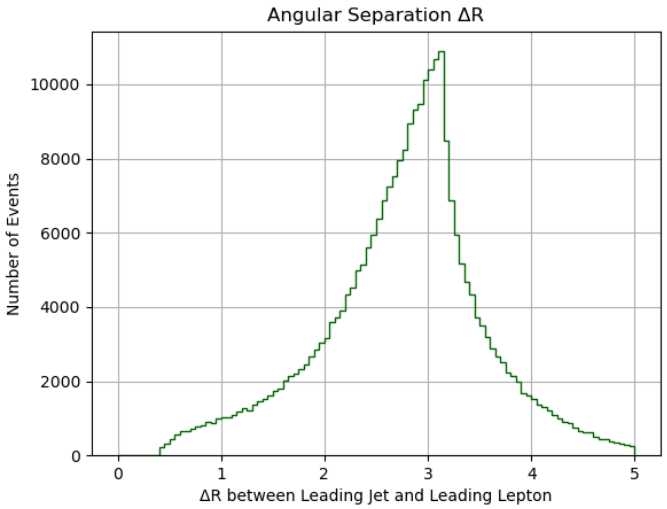

Figure 6: Jet-Lepton Angular Separation (ΔR) Analysis

Figure 6 presents the distribution of the angular distance ΔR between the leading jet with the highest transverse momentum (p_T) and the leading lepton with the highest p_T per event. The angular distance is defined using the pseudorapidity (η) and azimuthal angle (ϕ) components as follows:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

The distribution shows that the majority of leptons are found far from the jets (isolated). Approximately 99.996 % of the leptons satisfy the $\Delta R > 0.4$ condition, indicating that they are observed unaffected by hadronic activity.

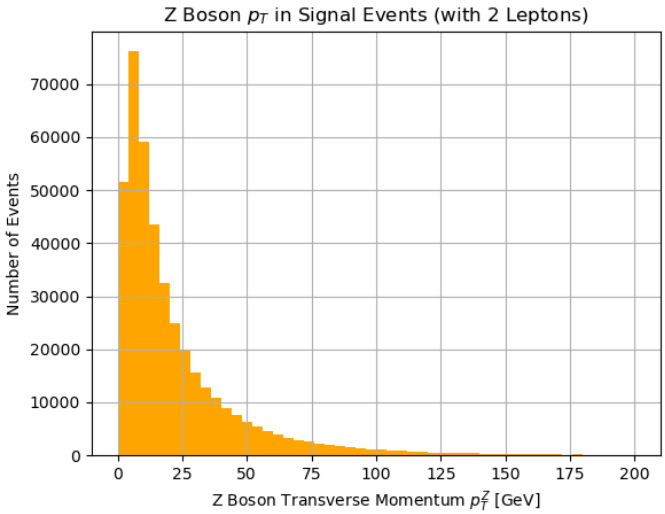

Figure 7: Transverse Momentum of the Z Boson

Figure 7 shows the distribution of the Z boson's transverse momentum (p_T^Z) for events involving exactly two leptons, corresponding to the signal region ($86 < M_{\ell\ell} < 96$ GeV). In this analysis, the four-momentum vector of the Z boson is calculated as the sum of the four-vectors of the two leptons, and the transverse component of this vector is removed.

The resulting distribution is concentrated at low p_T values, indicating that Z bosons are mostly produced with velocities near the center-of-mass frame of the system. High p_T^Z values correspond to events where the Z boson is produced with hadronic jets or partons.

4 Discussion & Conclusion

In this study, an analysis of the Z boson using the ATLAS open dataset was performed via the $\ell^+\ell^-$ decay channel. The selection criteria used isolated only physically meaningful and experimentally reliable events from the initial large and heterogeneous set of events. After the cutoff, the remaining number of events was reduced to 240075 by a filter of approximately 36%, demonstrating the highly selective and effective masks used in the analysis.

The transverse momentum distribution of the leptons revealed that the applied selections successfully eliminated low-energy and potentially noisy events. The peak around 91 GeV observed in the invariant mass distribution of the Z boson directly indicates the presence of the Z resonance, and the parameters obtained from the fitting process with the Breit–Wigner function were in good agreement with the theoretically predicted value of $M_Z = 91.1876 \pm 0.0021$ GeV [5].

When the signal and background regions were separated, a significant density of events was observed in the signal region compared to the background, demonstrating the effectiveness of the analysis methods in distinguishing resonant processes. When the number of jets and the jet transverse momentum distribution are examined, it can be said that in most of the analyzed events, 0 to 2 jets are observed, and most of the jets have low p_T values. This may be because the Z boson is produced without hadronic activity. The (ΔR) analysis clearly showed that most of the leptons are isolated. Furthermore, analyzing the number of jets in the signal region separately provided additional information regarding the characterization of the hadronic activity accompanying the Z production. The transverse momentum distribution of the Z boson also supports these findings, indicating that the Z bosons are mostly produced with low p_T.

While the analysis was generally successful, due to the limited number of variables used, some potential systematic uncertainties were not taken into account. Specifically, measurement errors, detector effects, smearing corrections, and statistical errors were excluded from the scope of this study. Furthermore, more sophisticated methods were not used to separate the signal from the background.

Because this analysis was conducted at an undergraduate level, its scope remained basic. However, more detailed methods could be used in more advanced analyses.

In conclusion, this study provides a successful example for understanding the fundamental structure of the Z boson analysis and offers important results regarding both the impact of experimental choices and the interpretation of physical distributions.

References

[1] Wit Busza, Krishna Rajagopal, and Wilke van der Schee. "Heavy Ion Collisions: The Big Picture, and the Big Questions". In: *Annual Review of Nuclear and Particle Science* 68 (2018), pp. 339–376. DOI: 10.1146/annurev-nucl-101917-020852.

[2] Fidele J. Twagirayezu. "Confinement in QCD: A Hybrid String Model with Vortex Corrections and Entanglement Entropy". In: *arXiv preprint arXiv:2308.00000* (2023).

[3] Umar Sohail Qureshi and Raghav Kunnawalkam Elayavalli. "Deep Image Reconstruction for Background Subtraction in Heavy-Ion Collisions". Preprint. July 21, 2025.

[4] Gregory Breit and Eugene P. Wigner. "Capture of Slow Neutrons". In: *Physical Review* 49 (1936), pp. 519–531. DOI: 10.1103/PhysRev.49.519.

[5] Particle Data Group. *Review of Particle Physics*. Z boson properties reviewed in Section 55. Progress of Theoretical and Experimental Physics, 2024.