EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Search for a new Z' gauge boson in 4μ events with the ATLAS experiment

The ATLAS Collaboration

This paper presents a search for a new Z' vector gauge boson with the ATLAS experiment at the Large Hadron Collider using pp collision data collected at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 139 fb^{-1}. The new gauge boson Z' is predicted by $L_\mu - L_\tau$ models to address observed phenomena that can not be explained by the Standard Model. The search examines the four-muon (4μ) final state, using a deep learning neural network classifier to separate the Z' signal from the Standard Model background events. The di-muon invariant masses in the 4μ events are used to extract the Z' resonance signature. No significant excess of events is observed over the predicted background. Upper limits at a 95% confidence level on the Z' production cross-section times the decay branching fraction of $pp \to Z'\mu\mu \to 4\mu$ are set from 0.31 to 4.3 fb for the Z' mass ranging from 5 to 81 GeV. The corresponding common coupling strengths, $g_{Z'}$, of the Z' boson to the second and third generation leptons above 0.003 – 0.2 have been excluded.

1 Introduction

This paper presents a search for a new vector boson Z' in the four-muon (4μ) final state with data recorded in proton-proton (pp) collisions at $\sqrt{s}$ = 13 TeV by the ATLAS detector [1] at the Large Hadron Collider (LHC), corresponding to an integrated luminosity of 139 fb^{-1}. The new gauge boson Z', which only interacts with the second and third generation leptons, is predicted by $L_\mu - L_\tau$ models [2] which extend the Standard Model (SM) with an additional $U(1)_{L_\mu - L_\tau}$ symmetry. In recent years searching for such a new gauge boson Z' has attracted attention in both theoretical and experimental communities [3–5] because it could address the observed muon magnetic dipole moment (g-2) anomaly [6–9] through the Z' loop corrections without contradicting other existing data. The new Z' is also ideally suited to address the lepton flavour anomalies measured with the ratio of B-meson decays to muons and electrons, $R_K = \mathcal{B}(B^+ \to K^+\mu^+\mu^-)/\mathcal{B}(B^+ \to K^+e^+e^-)$ [10–14]. Such anomalies suggest possible lepton flavor universality violation. The $L_\mu - L_\tau$ models could provide an explanation for the observed results, since by construction, the Z' arising from the models only couples to muons and taus, and does not couple to electrons. These models also aim to probe outstanding questions related to dark matter and neutrino mass [15–17].

The Z' kinematics, mass, and interactions (with the same coupling strength to the second and third lepton families), are described by the Lagrangian below:

$$L_{Z'} = -\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta} + \frac{1}{2}M^2_{Z'}Z'^\alpha Z'_\alpha - g_{Z'}Z'_\alpha(\bar{\ell}_2\gamma^\alpha\ell_2 + \bar{\mu}\gamma^\alpha\mu - \bar{\ell}_3\gamma^\alpha\ell_3 - \bar{\tau}\gamma^\alpha\tau),$$

where $F_{\alpha\beta} = \partial_\alpha Z'_\beta - \partial_\beta Z'_\alpha$ is the Z' field strength tensor; $\ell_i = (\nu_i, e_i)^T$ (i = 2, 3, denoting the second and the third generation left-handed lepton doublets); μ and τ represent the right-handed muon and tau singlets; and $g_{Z'}$ (which will be referred to as g in the rest of this paper) is the interaction coupling constant. The parameter space of ($m_{Z'}, g$) has not been strongly constrained in experiments since the Z' does not directly couple to the electron nor to any quarks, hence it could not be directly produced from e^+e^- and pp colliding beams.

In pp collisions at the LHC, the Z' could be produced from final state radiation of μ or τ pairs of the Drell-Yan (DY) process as shown in Figure 1(a) with a 4μ final state, which provides the cleanest signature to search for the Z'. For relatively low Z' mass, the most promising experimental signature would be an excess of 4μ events with the invariant mass of one $\mu^+\mu^-$ pair peaking around the Z' mass. The major background comes from the SM 4μ production processes shown in Figure 1(b) to 1(d). The Z' could also be produced in W production through the DY process, $pp \to W \to Z'\mu\nu \to 3\mu + \nu$. The experimental signature would have a final state of 3μ plus large missing transverse energy. This final state is not included in this analysis.

Both the ATLAS and CMS Collaborations have measured the cross-sections of the SM $Z \to 4\mu$ process [18–20]. The measurement by ATLAS was used by theorists to set limits in the parameter space of the $L_\mu - L_\tau$ model [15]. The CMS Collaboration has directly searched for the Z' boson in the mass region between 5 to 70 GeV with the 4μ final state using 77.3 fb^{-1} of data [21], and set upper limits on the Z' to muon coupling strength, g, of $0.004 - 0.3$ at 95% confidence level, depending on the Z' mass.
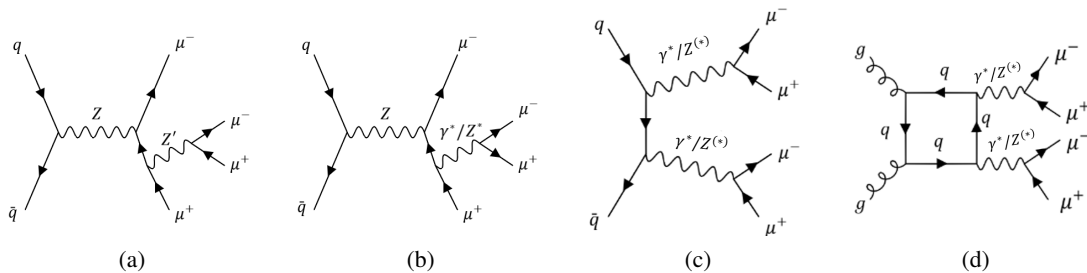
Figure 1: Feynman diagrams of Z' production through radiation in a Drell-Yan process (a), and of the corresponding SM background processes (b - d) with a 4μ final state.

The organization of this paper is as follows. The ATLAS detector is described in the next section. The dataset and Monte Carlo samples used in this analysis are detailed in Section 3. Event reconstruction and selection, followed by background estimation from data, are described in Section 4 and Section 5. Event classification using a deep learning approach is presented in Section 6. Systematic uncertainties and the statistical approach to interpret data to obtain the results are reported in Section 7 and Section 8. The conclusion is given in the final section.

2 ATLAS Detector

The ATLAS detector [1] at the LHC covers nearly the entire solid angle around the collision point.[1] It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadron calorimeters, and a muon spectrometer incorporating three large superconducting air-core toroidal magnets.

The inner-detector system (ID) is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first hit normally being in the insertable B-layer (IBL) installed before Run 2 [22, 23]. It is followed by the silicon microstrip tracker (SCT), which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$. The TRT also provides electron identification information based on the fraction of hits (typically 30 in total) above a higher energy-deposit threshold corresponding to transition radiation.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) calorimeters, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for

---

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

energy loss in material upstream of the calorimeters. Hadron calorimetry is provided by the steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadron endcap calorimeters. The solid angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic energy measurements respectively.

The muon spectrometer (MS) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by the superconducting air-core toroidal magnets. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. Three layers of precision chambers, each consisting of layers of monitored drift tubes, covers the region $|\eta| < 2.7$, complemented by cathode-strip chambers in the forward region, where the background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate chambers in the barrel, and thin-gap chambers in the endcap regions.

Interesting events are selected by the first-level trigger system implemented in custom hardware, followed by selections made by algorithms implemented in software in the high-level trigger [24]. The first-level trigger accepts events from the 40 MHz bunch crossings at a rate below 100 kHz, which the high-level trigger further reduces in order to record events to disk at about 1 kHz.

An extensive software suite [25] is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

3 Dataset and Monte Carlo Simulations

The data used for this analysis were recorded using single-muon and multi-muon triggers, corresponding to an integrated luminosity of 139 fb^{-1} after the application of data quality requirements [26]. The transverse momentum (p_T) thresholds for the single-muon trigger vary from 20 to 26 GeV, for the di-muon trigger from 10 to 14 GeV, and for the tri-muon trigger from 4 to 6 GeV, depending on the data-taking periods [27]. The overall trigger efficiency for the 4μ events selected in this analysis is higher than 98%.

3.1 Simulation of Z' Production

The $L_\mu - L_\tau$ model is used for Monte Carlo (MC) Z' signal sample production, where the Z' couples to the left-handed (LH) muon or tau leptons and their corresponding neutrinos, and to the right-handed (RH) muon or tau leptons. In the model, the Z' couplings to the first lepton families (electron and its neutrino) and all quarks are set to zero. The couplings to second and third generation leptons are assumed to be identical, therefore the branching fractions of the Z' decay to a pair of muons and a pair of muon neutrinos are set to $\frac{1}{3}$ and $\frac{1}{6}$, respectively. The signal from tau decays in the 4μ final state is found to be negligible in this analysis and is not included as signal.

4

The signal events are generated with MadGraph5_aMC@NLO 2.7.3 [28] at leading-order (LO) accuracy in QCD by using the *Universal FeynRules Output* (UFO) format [29, 30]. The interactions [31] mediated by a resonance Z' which couples to the second and third generation leptons are used for four-muon signal event generation. Based on theoretical calculations [32, 33] for related processes, the appropriate NNLO/LO correction factor K of 1.3 is used to correct the MC LO signal cross-sections. In the signal production simulations, the NNPDF2.3NLO set [34] is used as parton distribution function (PDF) for the pp collisions. Photos++ 3.61 [35] is used as the photon emissions from electroweak vertices and charged leptons via a QED process.

The MC simulated events are generated for a range of masses and coupling parameters of the $L_\mu - L_\tau$ model. The Z' mass ranges from 5 to 81 GeV, and the value of coupling constant g ranges from 0.008 to 0.316, as summarized in Table 1. The value of the g at each Z' mass was chosen to allow a sensitive search for a very small Z' signal in the full Run 2 dataset.

With the chosen g, the natural width Γ of the Z' (smaller than the experimental resolution) and the cross-section, $\sigma \, (pp \to Z'\mu^+\mu^- \to \mu^+\mu^-\mu^+\mu^-)$, are calculated as listed in Table 1.

Table 1: Summary of the chosen Z' hypotheses and corresponding coupling, width, and cross-section (calculated at LO accuracy in QCD) at each mass point.

$m_{Z'}$ [GeV]	g	Γ [GeV]	σ [fb]	$m_{Z'}$ [GeV]	g	Γ [GeV]	σ [fb]
5	0.0080	2.45×10^{-5}	9.96	42	0.0900	2.71×10^{-2}	13.38
7	0.0085	3.99×10^{-5}	7.06	45	0.1000	3.58×10^{-2}	11.72
9	0.0090	5.78×10^{-5}	5.60	48	0.1100	4.62×10^{-2}	9.96
11	0.0095	7.89×10^{-5}	4.65	51	0.1200	5.84×10^{-2}	8.24
13	0.0100	1.03×10^{-4}	3.95	54	0.1600	1.10×10^{-1}	10.07
15	0.0120	1.72×10^{-4}	4.45	57	0.2000	1.81×10^{-1}	10.73
17	0.0140	2.65×10^{-4}	4.80	60	0.2665	3.39×10^{-1}	12.92
19	0.0160	3.87×10^{-4}	5.00	63	0.2680	3.60×10^{-1}	8.84
23	0.0240	1.05×10^{-4}	7.30	66	0.2780	4.06×10^{-1}	6.50
27	0.0320	2.20×10^{-3}	8.50	69	0.2890	4.59×10^{-1}	4.89
31	0.0400	3.95×10^{-3}	8.72	72	0.3000	5.15×10^{-1}	3.80
35	0.0600	1.00×10^{-2}	12.82	75	0.3000	5.37×10^{-1}	2.88
39	0.0800	1.99×10^{-2}	14.77	78	0.3080	5.89×10^{-1}	2.40
				81	0.3160	6.44×10^{-1}	2.08

The Z' signal samples were simulated with the ATLAS fast simulation framework (Atlfast-II) [36] to produce predictions that can be directly compared with the data.

3.2 Simulation of Background Events

Dominant SM backgrounds in this analysis come from the SM $Z \to 4\mu$ processes where the four leptons have an invariant mass close to that of the Z boson. In the higher mass region the ZZ^* production contributes a sizable number of prompt 4μ events. In addition, there are very small contributions from the Higgs boson, $t\bar{t}V$ ($V = W, Z$), and tri-boson (VVV) production processes. These events are estimated with MC simulations. The non-prompt muon background events,

mostly coming from Z + jets, $t\bar{t}$ and single-top-quark production processes, are estimated from data in this analysis, as described in Section 5. The Z + jets and $t\bar{t}$ MC samples are also produced for background studies.

Samples of diboson final states ($q\bar{q} \rightarrow VV^{(*)}$), including the processes shown in Figures 1(b) and 1(c), were simulated with the SHERPA 2.2.2 [37] generator, including off-shell effects and Higgs boson contributions, where appropriate. Fully leptonic final states and semileptonic final states, where one boson decays leptonically and the other hadronically, were generated using matrix elements at next-to-leading-order (NLO) accuracy in QCD for up to one additional parton and at LO accuracy for up to three additional parton emissions. Samples for the loop-induced processes $gg \rightarrow VV^{(*)}$, shown in Figure 1(d), were generated using LO-accurate matrix elements for up to one additional parton emission for both the cases of fully leptonic and semileptonic final states. The matrix element calculations were matched and merged with the SHERPA parton shower based on Catani–Seymour dipole factorisation [38, 39] using the MEPS@NLO prescription [40–43]. The virtual QCD corrections were provided by the OPENLOOPS library [44–46]. The NNPDF3.0NNLO set of PDFs was used [47], along with the dedicated set of parton-shower parameters (tune) developed by the SHERPA authors.

The production of $t\bar{t}$ events was modelled using the POWHEG BOX v2 [48–51] generator at NLO in QCD with the NNPDF3.0NLO PDF set and the h_{damp} parameter[2] set to 1.5 m_{top} [52]. The events were interfaced to PYTHIA 8.230 [53] to model the parton shower, hadronisation, and underlying event, with parameters set according to the A14 tune [54] and using the NNPDF2.3LO set of PDFs. The associated production of top quarks with W bosons (tW) was modelled by the POWHEG BOX v2 [49–51, 55] generator at NLO in QCD using the five-flavour scheme and the NNPDF3.0NLO set of PDFs. The diagram removal scheme [56] was used to remove interference and overlap with $t\bar{t}$ production. The events were interfaced to PYTHIA 8.230 using the A14 tune and the NNPDF2.3LO set of PDFs.

The production of V+jets was simulated with the SHERPA 2.2.1 generator using NLO matrix elements for up to two partons, and LO matrix elements for up to four partons, calculated with the Comix and OPENLOOPS libraries. They were matched with the SHERPA parton shower using the MEPS@NLO prescription with the set of tuned parameters developed by the SHERPA authors. The NNPDF3.0NNLO set of PDFs was used and the samples were normalised to a next-to-next-to-leading-order (NNLO) prediction [57].

The production of $t\bar{t}V$ events was modelled using the MADGRAPH5_aMC@NLO 2.3.3 generator at NLO with the NNPDF3.0NLO PDF. The events were interfaced to PYTHIA 8.210 using the A14 tune and the NNPDF2.3LO PDF set. The production of tri-boson (VVV) events was simulated with the SHERPA 2.2.2 generator. Matrix elements accurate to LO in QCD for up to one additional parton emission were matched and merged with the SHERPA parton shower based on Catani–Seymour dipole factorisation using the MEPS@NLO prescription. Samples were generated using

---

[2] The h_{damp} parameter is a resummation damping factor and one of the parameters that controls the matching of POWHEG matrix elements to the parton shower and thus effectively regulates the high-p_{T} radiation against which the $t\bar{t}$ system recoils.

the NNPDF3.0nnlo PDF set, along with the dedicated set of tuned parton-shower parameters developed by the Sherpa authors.

The generated background MC samples were produced through the full ATLAS detector simulation based on Geant4 [58]. The effect of multiple interactions in the same and neighbouring bunch crossings (pile-up) was modelled by overlaying the simulated hard-scattering event with inelastic pp events generated with Pythia 8.186 [59] using the NNPDF2.3lo set of PDF and the A3 set of tuned parameters [60]. Simulated events were reweighted to match the pile-up conditions in the data. All simulated events were processed using the same reconstruction algorithms and triggering requirements as used in data.

4 Event Reconstruction and Pre-selection

Proton-proton interaction vertices are reconstructed in events with at least two tracks, each with $p_T > 0.5$ GeV. The primary hard-scatter vertex is defined as the one with the largest value of the sum of squared track transverse momenta.

Muons are identified by matching tracks reconstructed in the MS to tracks reconstructed in the ID (referred to as combined muons). To increase the muon reconstruction efficiency non-combined muon identification algorithms are also used in the analysis, including using the MS stand-alone tracks in the region $2.5 < |\eta| < 2.7$, and matching the ID tracks with calorimeter hit information within $|\eta| < 0.1$, as well as using the ID tracks associated with at least one local track segment in the MS. In the 4μ event selection at most one of the selected muons can be a non-combined muon. Each muon is then required to satisfy the 'loose' identification criteria [61]. Muons are required to be isolated using a particle-flow algorithm [62] and associated with the primary hard-scatter vertex by satisfying $|\frac{d_0}{\sigma_{d_0}}| < 3$ and $|z_0 \times \sin\theta| < 0.5$ mm, where d_0 is the transverse impact parameter calculated with respect to the measured beam-line position, σ_{d_0} its uncertainty, and z_0 is the longitudinal distance between the point at which d_0 is measured and the primary vertex. The minimum muon p_T threshold is 3 GeV.

In addition to muons, electrons, jets and missing transverse momentum (E_T^{miss}) are also used to select control samples for background estimation in this analysis. The reconstructions of these objects are described below.

Electrons are identified with a likelihood discriminator built from the shower shapes of electron energy loss in the calorimeter, track-cluster matching, and some of the track quality distributions. Each electron is required to satisfy the 'medium' likelihood identification criteria [63], as well as similar vertex and isolation requirements as muons. The reconstructed electrons are required to have $p_T > 7$ GeV and $|\eta| < 2.5$, excluding the transition region between the barrel and endcap calorimeters, $1.37 < |\eta| < 1.52$.

Jets are reconstructed with the anti-k_t algorithm [64, 65] with a radius parameter of $R = 0.4$. The jet clustering input objects are based on particle-flow [66] in the ID and the calorimeter. Jets are required to have $p_T > 20$ GeV and $|\eta| < 2.5$. Jets containing B hadrons, referred to as

b-jets, are identified with a multivariate discriminant [67]. To reduce the effect of pile-up an additional quality requirement based on the 'Jet-Vertex-Tagger' algorithm (JVT) [68] is applied in jet identification.

E_T^{miss} is determined as the magnitude of the negative vectorial sum of the transverse momenta of the selected and calibrated physics objects (including muons, electrons, photons, and jets including hadronically decaying tau-leptons) and the ID tracks must come from the main vertex and not associated with any physics object (soft term) [69].

Events containing at least four muons with kinematics consistent with $Z(Z^*/\gamma^*) \to 4\mu$ production are then selected as follows. The four leading p_T-ordered muons are required to pass the p_T thresholds of 20, 15, 8, and 3 GeV, respectively. If a muon is selected as a non-combined muon, its p_T must be greater than 15 GeV. Any di-muon pair in the event must have an invariant mass $m_{\mu\mu}$ greater than 4 GeV and an angular separation ΔR larger than 0.2. To search for the $Z' \to \mu^+\mu^-$ signature, two muon pairs are selected based on their invariant mass values. The first pair (referred to as Z_1) is selected from all the possible $\mu^+\mu^-$ pairs to have the smallest mass difference between the Z_1 mass and the Z mass, $|m_Z - m_{Z_1}|$. The second $\mu^+\mu^-$ pair is selected from the remaining muons that has the highest invariant mass (referred to as Z_2). The correct signal di-muon pairing fraction varies with the Z' mass, where the selected di-muon pair that forms m_{Z1} or m_{Z2} originates from the Z'. For example, for $m_{Z'}$ = 5, 42, 63, 72, and 81 GeV, the correct di-muon pairing fractions are about 78%, 50%, 88%, 82%, and 90%, respectively. Finally, the selected four muons must have an invariant mass in a range of 80 to 180 GeV, but excluding the Higgs boson resonance mass region of 110 to 130 GeV.

The Z' signal efficiency at various stages of the event selection is shown in Table 2 for five representative mass points. The event selection efficiencies vary significantly depending on the Z' mass. At generator level, an MC filter is applied, which requires at least four muons with $p_T > 2$ GeV and $|\eta| < 3.0$. The MC filter efficiencies of these representative Z' signal samples are listed in the table as well.

Table 2: The Z' signal event selection efficiencies compared to the events passing the previous cut level for several representative mass points. The overall signal efficiencies are the products of the 4μ MC filter and the combined event selection efficiencies.

$m_{Z'}$ [GeV]	5	42	63	72	81
MC filter efficiency	32.8%	57.7%	61.0%	65.3%	70.0%
Number of identified muons ≥ 4	47.3%	74.1%	70.8%	72.4%	75.4%
$p_T^i (i = 1, 4) > 20, 15, 8, 3$ GeV	60.0%	82.6%,	90.3%	93.6%	98.2%
$\Delta R(\mu_i, \mu_j) > 0.2$ & vertex requirement	87.2%	95.4%	96.2%	96.6%	97.2%
Isolation	54.2%	76.9%	79.2%	84.1%	87.5%
$m_{4\mu}$ within [80, 110] or [130, 180] GeV	91.9%	88.8%	58.9%	33.5%	16.8%
Combined event selection efficiency	12.3%	39.9%	28.7%	18.4%	10.6%
Overall 4μ signal efficiency	4.1%	23.0%	17.5%	11.9%	7.4%

The Z' production signature is searched for in the Z_1 or Z_2 mass spectrum depending on the Z' mass. The relatively high-mass Z' signals mostly appear as a peak in the Z_1 spectrum while the relatively low-mass signals mostly appear as a peak in the Z_2 spectrum. Representative examples of the predicted signal over background, after further selection with a deep learning approach which will be described in Section 6, are shown in Figure 4. In the analysis the Z_1 and Z_2 mass spectra are scanned to search for a Z' with mass greater or smaller than 42 GeV, respectively. The boundary value of 42 GeV is chosen based on the studies to optimize the search sensitivity.

The numbers of 4μ events in data and the estimated background yields are given in Table 3. More details about the estimation of the reducible backgrounds containing non-prompt muons can be found in Section 5. The total uncertainties of simulated backgrounds are also listed in the table. The evaluations of systematic uncertainties will be described in Section 7.

Table 3: The selected 4μ events in data and the estimated backgrounds and their combined statistical and systematic uncertainties.

Data	Total background	$qq \to ZZ^*$ from MC	$gg \to ZZ^*$ from MC	$ttV + VVV + H$ from MC	Reducible background from data
1131	1148 ± 70	1065^{+70}_{-69}	15.6 ± 2.5	6.2 ± 2.9	$61.1^{+8.3}_{-9.1}$

5 Estimation of Reducible Background from Data

A data-driven technique is used to estimate the reducible background from Z + jets and $t\bar{t}$ production. Events from these processes may contain two prompt leptons from W or Z boson decays, together with additional activity such as heavy-flavor jets or misidentified components of jets yielding reconstructed muons, collectively referred to as non-prompt muons. Such backgrounds are estimated from data using a control sample of $\mu^+\mu^-\mu_j\mu_j$ events, selected with the standard signal requirements except that non-prompt muons, μ_j, are selected in place of two of the signal muons. Non-prompt muons are defined as muon candidates using the standard selection but fail the requirements on isolation and impact parameter. However, they are required to pass a much looser impact parameter significance requirement, $|\frac{d_0}{\sigma_{d_0}}| < 10$. The background in the signal sample is estimated by scaling each event in the selected $\mu^+\mu^-\mu_j\mu_j$ control sample by $f_1 \times f_2$, where the $f_i (i = 1, 2)$ is referred to as a "fake-factor" for each of the two non-prompt muons and is computed as a function of their p_T and $|\eta|$.

The fake-factor f is derived from two independent non-prompt-muon enriched data control samples dominated by Z + jets or $t\bar{t}$ events. The Z + jets ($t\bar{t}$) sample contains non-prompt muons dominantly from light jets (heavy-flavor jets). The Z + jets sample is tagged by two isolated prompt leptons from Z decays ($Z \to \mu^+\mu^-$, e^+e^-). The p_T-threshold for the two prompt leptons are 27 and 25 GeV and their invariant mass must be within 10 GeV of the Z mass.

The E_T^{miss} of the event must be less than 25 GeV. The $t\bar{t}$ sample is tagged by two prompt isolated high-p_T $e^\pm\mu^\mp$ leptons with p_T thresholds of 28 and 25 GeV and associated with at least one b-jet.

The event is required to have $E_T^{miss} > 50$ GeV. If there is an additional isolated high-p_T lepton in the event, the transverse mass calculated with the E_T^{miss} and the third lepton is required to be less than 60 GeV to reduce the contribution from the $t\bar{t}W$ events. Events with four muons that satisfy the signal selection criteria are removed from both Z + jets and $t\bar{t}$ control samples. Additional reconstructed muon objects, passing either the signal-muon or the non-prompt-muon selection criteria are counted in the control samples. The fake-factor f is calculated as the ratio of the probability for a non-prompt muon to satisfy the signal muon selection criteria to the probability for a non-prompt muon to fail the signal muon selection. The factor f is derived as a function of $|\eta|$ and p_T of the non-prompt muons, which varies from 0.11 to 0.76 (0.01 to 0.27) in the Z + jets ($t\bar{t}$) control sample. Then, the fake-factors derived from two data control samples are combined using the m_{Z1} spectra obtained from simulated 4μ background events from Z + jets and $t\bar{t}$ processes. A simultaneous fit of the m_{Z1} spectra of the MC events to data selected in the $\mu^+\mu^-\mu_j\mu_j$ control sample is performed to determine the fractions of the Z + jets and $t\bar{t}$ events in each bin of the mass spectrum.

The overall systematic uncertainty of f is about 14%. It is determined with alternative non-prompt muon selections, such as changing the muon isolation criteria in data control samples (7.6%), and the uncertainties of the Z + jets and $t\bar{t}$ event fractions when combining the fake-factors derived from two control samples (8.6%). The statistical uncertainties of the control samples are also accounted as part of the systematic uncertainties (6.1%).

Additional reducible background contributions come from the WZ process. These events contain three prompt muons from W and Z decays and one non-prompt muon. This background is estimated by scaling a $3\mu + \mu_j$ control sample by the f derived from the Z + jets sample. The total estimated number of reducible background events, $61.1^{+8.3}_{-9.1}$, is listed in Table 3.

6 Event Classification with Deep Learning Approach

The selected 4μ events are classified with a "Deep Learning" approach to further separate the Z' signal from the SM background. A parametrized deep neural network (pDNN) architecture [70] is used in the analysis. This algorithm allows the training of a single classifier for multiple signal mass hypotheses in the search range. In practice, the kinematic inputs together with the Z' mass parameters of signal and background are used for training. The MC generated Z' masses (listed in Table 1) are used as the multiple signal mass parameters, while the distribution of the mass parameter for background events is randomly drawn from the same distribution as for the signal class. The algorithm was implemented in the Keras [71] framework with the TensorFlow [72] backend. Two classifiers are trained for low (high) Z' mass searches using MC mass parameters lower (higher) than 40 GeV. During the training of the classifier, the training samples were composed of simulated signal and background 4μ events using the pre-selection described in Section 4. A set of kinematic distributions were used for pDNN training input features. They are the p_T and η of each muon, the p_T of the Z_1 and Z_2, the mass difference of the Z_1 and Z_2, ΔR and $\Delta \eta$ of each muon pair that forms the Z_1 and Z_2, and the p_T and mass of the 4μ system. Examples of the input variable distributions, both for representative signals ($m_{Z'} = 15$, and 51

GeV) and background, are shown in Figure 2 to compare the predicted and observed transverse momenta ($p_T^{Z_1}$ and $p_T^{Z_2}$) distributions and the mass difference of the Z_1 and Z_2. In addition to the major background from the SM $Z(Z^*) \rightarrow 4\mu$ production, other backgrounds, including 4μ events containing non-prompt muons estimated from data, and from ttV, VVV, and Higgs boson production processes, are included in the plots.
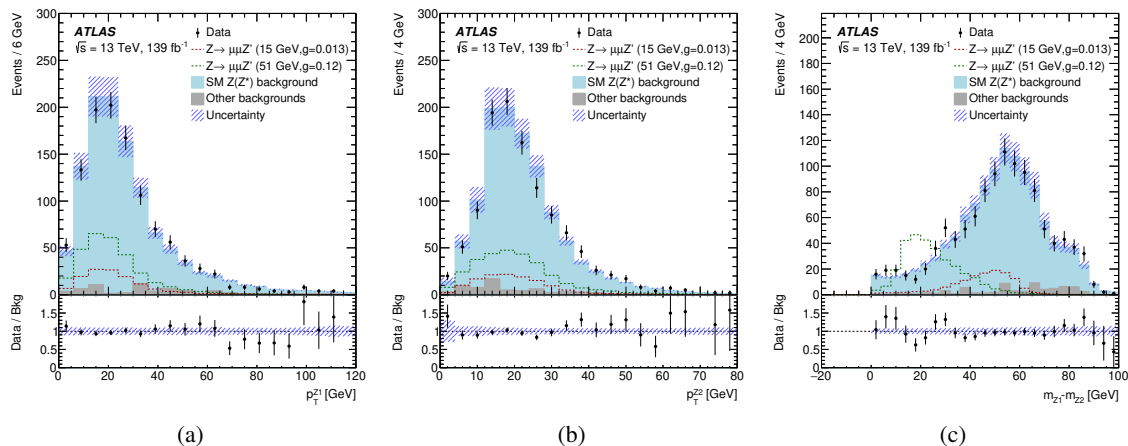

Figure 2: Distributions of $p_T^{Z_1}$ (a), and $p_T^{Z_2}$ (b), and the mass difference of the Z_1 and Z_2 candidates (c). Small background contributions are denoted as "other backgrounds", including 4μ events containing non-prompt muons estimated from data and from ttV, VVV, and Higgs boson production processes.

The training results are a set of weights of the pDNN model which are applied to real data or MC samples to obtain the pDNN output discriminating variable. The m_{Z1} or m_{Z2} are used as the mass parameter when applying the model to data.

To improve the power of the pDNN classifier the model is optimized through an automatic process developed with the package TUNE [73] and HEBO SEARCH algorithm [74]. The neural-network structure was chosen to have two fully connected hidden layers, each with 256 (32) nodes for the high (low) mass search in the analysis. Other training hyper-parameters, such as the *learning rate* with *decay* and *class weight*, for both high and low mass searches were also selected. Two pDNN output discriminant classifiers for high and low mass searches were obtained and are shown in Figure 3 for data and simulated signal and background events.

To maximize the search sensitivity, a scan of the pDNN output scores was performed to find the optimal cut values for each Z' mass hypothesis. These cut values vary from 0.42 to 0.74 (0.12 to 0.16) for the low (high) mass region. These cuts keep high signal efficiencies between 98% and 95% (90% and 50%), while the corresponding background reductions range from 10% to 50% (50% to 96%) for the low (high) mass region.

After the 4μ event selection with the pDNN classifier, the final discriminant to search for the Z' resonance signature is the Z_1 (for $m_{Z'} \geq 42$ GeV) or Z_2 (for $m_{Z'} \leq 42$ GeV) mass spectrum as shown in Figure 4. Data are compared to the estimated background together with two

representative signals with masses of 15 and 51 GeV, shown in Figure 4(a) and 4(b), respectively. The values of the gauge coupling strengths (g) for the two mass points are chosen for the purpose of illustration.

One should note that the dimuon masses (m_{Z1}, m_{Z2}) are not used as the pDNN training variables, therefore cutting on the pDNN output scores does not sculpt the dimuon invariant mass distributions of the backgrounds (see Figure 4). To interpret the observed mass spectra systematic uncertainties in the predictions are estimated in Section 7.

(a) low mass: $m_{Z'} \le 42$ GeV

(b) high mass: $m_{Z'} \ge 42$ GeV
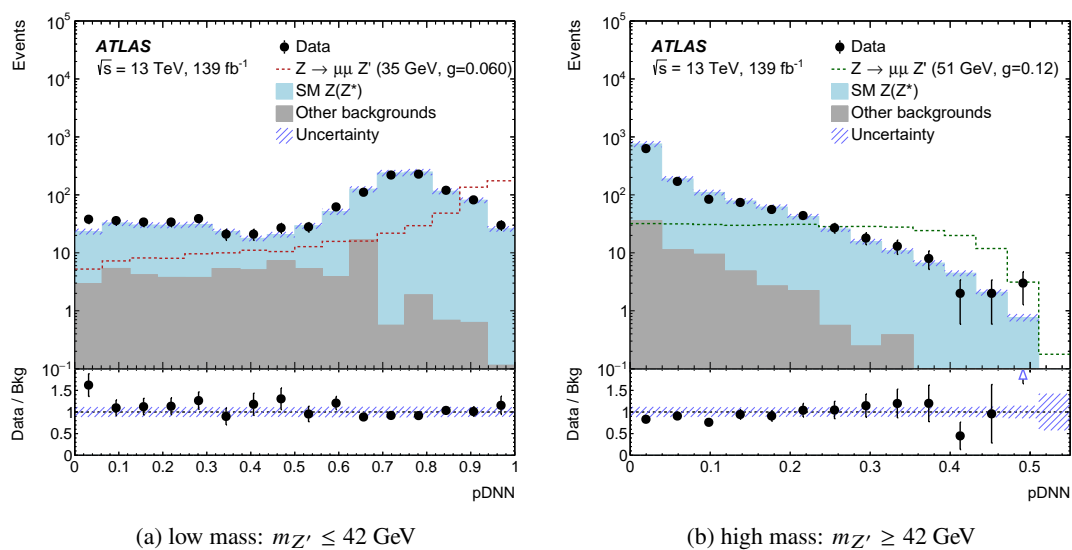
Figure 3: The pDNN output discriminant variable distributions for low mass (a) and high mass (b) with a signal sample at 35 GeV and 51 GeV, respectively.

(a)

(b)
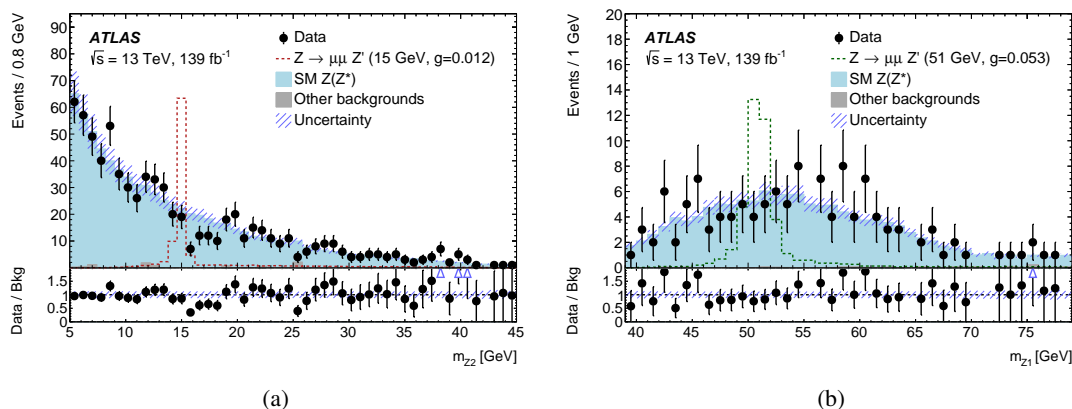
Figure 4: Mass spectra of m_{Z2} (left) and m_{Z1} (right) for the pDNN-selected events with a signal sample at 15 GeV and 51 GeV, respectively.

7 Systematic Uncertainties

Systematic uncertainties in the simulated event yields and shapes, for both signal and background processes, may arise from the calibration of the physics objects and from the theoretical modelling used in the predictions.

The major experimental uncertainties come from the muon reconstruction, identification, and isolation requirement efficiencies. These efficiencies are corrected based on studies performed in data control regions. The energy and momentum scales and resolutions of the simulated objects are calibrated to reproduce data from $Z \rightarrow \mu^+\mu^-$ and $J/\psi \rightarrow \mu^+\mu^-$ decays [61]. The uncertainties on the 4μ detection efficiency are determined by varying the nominal calibrations in the MC samples by one standard deviation, including muon momentum resolutions and scales, and the trigger, reconstruction, identification, and isolation requirement efficiencies. The overall relative experimental uncertainties in the 4μ event selection efficiency is about 3.9%, dominated by the uncertainty of the isolation efficiency (2.9%) and the low p_T calibration uncertainty (2.0%). The signal event selection efficiency uncertainties vary from 8.3% to 3.9%, depending on the Z' mass. In addition, the uncertainty in the combined 2015-2018 integrated luminosity is 1.7% [75], obtained using the LUCID-2 detector [76] for the primary luminosity measurements.

Sources of theoretical uncertainties come from the choice of QCD scales (renormalization μ_R and factorization μ_F), strong coupling constant α_S, and PDFs, as well as the parton shower models. These uncertainties affect the signal and background event selection efficiencies, normalization, and the shape of their kinematic distributions. The scales are varied independently from 0.5 to 2.0 times the nominal values and the largest deviation is chosen as the systematic uncertainty. The PDF uncertainty is estimated by comparing events generated with different PDF sets, as well as the uncertainties from the nominal PDF set itself. The maximal variation (envelope) is accounted for as the systematic uncertainty, following the PDF4LHC [77] recommendations. The α_S uncertainty is estimated by comparing events generated with different α_S values using the nominal PDF set. The parton shower uncertainty is estimated by comparing events with different parton shower parameters in the Sherpa MC samples. For the $Z(Z^*) \rightarrow 4\mu$ process, the relative uncertainties of event yields for scale, PDF, α_S, and parton shower, are 4.6%, 1.8%, 1.0%, and 1.9%, respectively. The $t\bar{t}V$, VVV, and Higgs boson, processes contribute 5.2, 0.50, and 0.53 background events, respectively. The total relative theoretical uncertainty for these background is estimated about 16%. The interference effect between the signal and background is estimated (using MC samples generated by MADGRAPH5_aMC@NLO) by evaluating the cross-section ratio, $\Delta\sigma/\sigma_{Z'}$, where $\Delta\sigma$ is the difference of the cross-section from the inclusive MC sample and the sum of the cross-sections of the signal and background samples in the Z' detection phase space. The effect varies from 1.8 to 7.5% for the Z' mass from 5 - 81 GeV, which is accounted for as additional signal yield systematic uncertainties. In addition, the MC Z' signal filter acceptance uncertainty is estimated to be about 2%, which is calculated by varying the QCD scales, the PDF-sets, and the strong coupling constant using MC events at the generator level for different Z' mass points.

All the uncertainties (from both experiment and theory) on the final discriminant di-muon mass

spectra are included as nuisance parameters in the signal extraction fitting process described in Section 8.1.

8 Data Interpretation and Results

The statistical analysis is performed by comparing the data to the sum of the background prediction and the signal to search for the Z' signature and information about the $pp \rightarrow \mu^+\mu^- Z' \rightarrow 4\mu$ signal production cross-section and the associated coupling strength. In case of no significant data excess over the background prediction, upper limits on the signal production cross-section times the decay branching fraction for different Z' masses are set at 95% confidence level (CL).

8.1 Statistical Procedure and Results

The statistical procedure used to interpret the data is described in Ref. [78]. The likelihood function is constructed from the product of Poisson probabilities:

$$L(\text{data}|\,\mu, \theta) = \prod_{i=1}^{N} \text{Poisson}(\text{data}_i|\,\mu \cdot s_i(\theta) + b_i(\theta)) \times \text{G}(\tilde{\theta}|\theta)$$

where i is the bin-index of the fitted variable distribution; μ is the signal strength, and θ denotes the nuisance parameters, which represent the uncertainties of the measurements; $\text{G}(\tilde{\theta}|\theta)$ is the Gaussian scaling function of the nuisance parameters constructed as deviations from the nominal model of the systematic uncertainties, where $\tilde{\theta}$ provides a maximum likelihood estimate for θ. The parameter of interest in the statistical analysis is the global signal strength factor μ, which acts as a scale factor on the total number of events predicted by the signal model. This factor is defined such that $\mu = 0$ corresponds to the background-only hypothesis and $\mu > 0$ corresponds to a Z' signal in addition to the background. Hypothesised values of μ are tested based on the profile likelihood ratio [79], which compares data with background-only (b) and signal+background ($s + b$) models using the following test statistic:

$$q_\mu = \begin{cases} -2\ln\frac{L(\text{data}|\,\mu, \hat{\hat{\theta}}_\mu)}{L(\text{data}|\,\hat{\mu}, \hat{\theta})} & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

where $\hat{\mu}$ and $\hat{\theta}$ are the values for μ and θ when maximizing the L with all parameters floating, which are named as the unconditional maximum-likelihood (ML) estimators; $\hat{\hat{\theta}}_\mu$ is the conditional ML estimator of θ for a fixed value of μ. This test statistic extracts the information on the signal strength from a full likelihood fit to the data. The likelihood function includes all the parameters that describe the systematic uncertainties and their correlations.

In case of no significant data excess over the background, exclusion limits are set, based on the CLs prescription [80], which calculates the ratio of the probabilities based on the $s + b$ background-only models:

$$CL_s(\mu) = \frac{CL_{s+b}}{CL_b} = \frac{P(q_\mu > q_\mu^{obs})}{P(q_0 > q_0^{obs})}$$

A value of μ is regarded as excluded at 95% confident level when CLs is less than 5%.

The significance of an excess in the data is first quantified with the local p_0, the probability that the background can produce a fluctuation greater than or equal to the excess observed in data. The equivalent formulation in terms of number of standard deviations is referred to as the local significance, which is computed from test statistic q_0:

$$p_0 = P(q_0 < q_0^{obs}).$$

The global probability for the most significant excess to be observed anywhere in a given search region is estimated with the method described in Ref. [81].

The statistical tests are performed by first searching for data excesses by means of a p_0 scan, and in case no such significant data excess is found, by setting limits on the Z' production cross-section times the branching fraction. In both steps the values and reconstructed mass distributions of hypothesized Z' mass $m_{Z'}$ are used. The steps between the mass points used in these test procedures are set to match the mass resolutions determined by reconstruction of the di-muon invariant mass. A linear interpolation of the expected event yields and shapes between the MC generated signal samples is used in the fitting process. The asymptotic approximation [79] upon which the results are based has been validated against the method (detailed in Ref. [78]) of using pseudo-experiments for several mass points.

In both steps of the statistical tests, data are fit to the m_{Z1} and m_{Z2} spectrum with background (b) only and signal+background ($s + b$) hypotheses. In the fitting process each mass spectrum is divided in the signal region (SR) and the background control region (CR). For each Z' mass point the SR is defined in a mass window of $m_{Z'} \pm 3\sigma_{m_{\mu\mu}}$ of the di-muon mass spectrum. The sidebands outside of the SR are defined as the CR. The di-muon mass resolution $\sigma_{m_{\mu\mu}}$ is determined by the fully simulated Z' mass distribution, which combines the Z' natural width and the detector resolution, ranging from 0.10 to 1.75 GeV. The mass resolution is mostly dominated by the detector resolution. Finer binning is used in the SR to enhance the sensitivity. The background CR is used to constrain the overall normalization for the background in the signal region. The shape of the major background from $Z(Z^*) \rightarrow 4\mu$ is fixed with prior uncertainties included in the fitting process, but the normalization (or strength) floats in the fit. Other background normalizations and shapes are fixed with prior uncertainties included in the fitting.

8.2 The p_0 Scan Results

The p_0-values corresponding to the background-only hypothesis are scanned in the mass range of this analysis. A binned profile-likelihood fit [79] is performed simultaneously across the Z' signal-region and the background control region using the predicted and observed mass spectrum as inputs. Data are fit to the m_{Z1} and m_{Z2} distributions for $m_{Z'} \geq 42$ GeV and $m_{Z'} \leq 42$ GeV,

respectively, with a "sliding" mass window as the defined SR changes for different Z' mass points. The chosen bin-size inside the SR is around 0.3 $\sigma_{m_{\mu\mu}}$, for each mass point. The total number of bins in the CR is 20. The fit mass range of m_{Z1} (m_{Z2}) is [30, 85] GeV ([0, 45] GeV).

The p_0-values at different Z' mass hypothesis points are computed and transformed into Gaussian standard deviations to indicate the significance as shown in Figure 5. The smallest p_0-value is at 39.6 GeV, corresponding to a local 2.65σ deviation from the background-only hypothesis, while the global deviation [81] is found to be 0.52σ, indicating that no significant data excess over the expected background is observed.
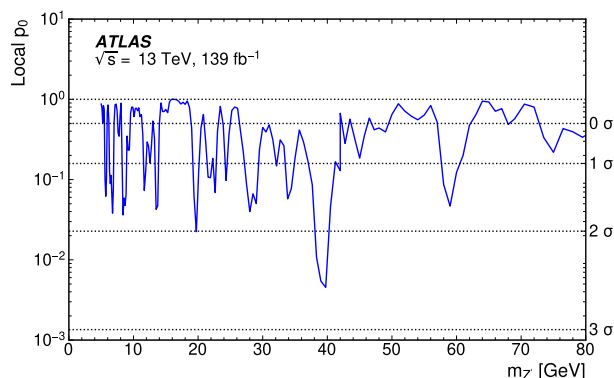
Figure 5: The p_0-value scan across the Z' mass signal regions.

8.3 Upper Limits

The upper limits on the production cross-section times branching fraction of the $pp \rightarrow \mu^+\mu^- Z' \rightarrow 4\mu$ process are calculated using a similar fitting procedure described in Section 8.2. Confidence intervals are computed based on the profile-likelihood-ratio test statistics [79]. The observed and expected upper limits at 95% CL on the cross-section times branching fraction, $\sigma(pp \rightarrow Z'\mu\mu \rightarrow 4\mu)$, are shown in Figure 6(a). Assuming the same coupling strength g of the Z' boson to the second and third lepton families and to the left- and right-handed fermions, the upper limits on the coupling parameter g are extracted from the limits of the Z' production cross-section times branching fraction using the $L_\mu - L_\tau$ model, where the branching fraction of $\mathcal{B}(Z' \rightarrow \mu^+\mu^-) = \frac{1}{3}$, which is determined by counting all the possible Z' decay modes in this model. At each generated Z' mass point, a limit on the coupling strength g has been obtained from the cross-section limit. The observed and expected upper limits on the coupling parameter g are shown in Figure 6(b). The limits on the coupling g are in the range of 0.003 (for $m_{Z'}$ = 5 GeV) to 0.2 (for $m_{Z'}$ = 81 GeV) depending on the Z' mass ranging from 5 to 81 GeV. This ensures that the ratio of the Z' natural width and mass, $\Gamma(Z')/m_{Z'}$, is well below 1% in this mass range.

Motivated by theoretical interpretations in Ref. [15], a 2-dimensional exclusion contour at 95% CL in the parameter-space of $(m_{Z'}, g)$ of the $L_\mu - L_\tau$ model from this analysis is produced and
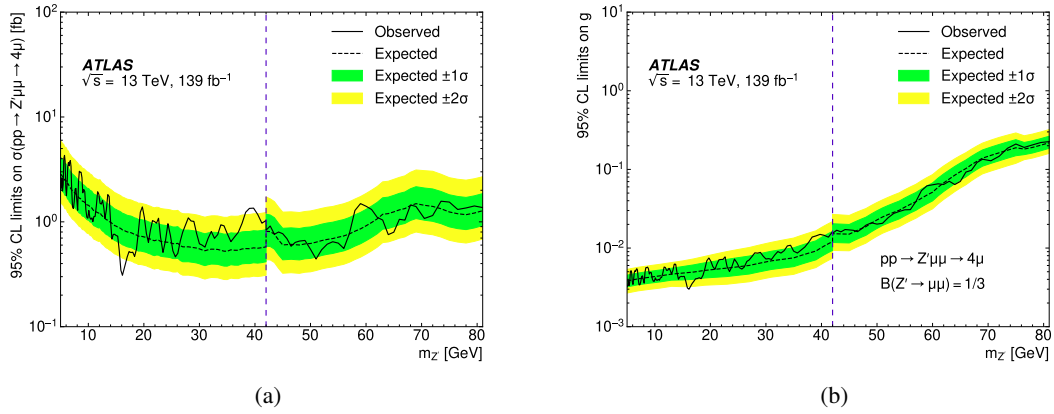
Figure 6: 95% CL upper limits (expected and observed) on the cross-sections times branching fraction (a) and coupling parameter (b). The discontinuity at 42 GeV represents the border of the low/high mass classifiers.

shown in Figure 7. The parameter space exclusion regions calculated by theorists using data from the Neutrino Trident experiment [82] and the B_s mixing measurements by a global analysis performed in Ref. [15] are also shown in Figure 7. This had left a large gap in the parameter space not yet excluded, that was allowed to explain the LHCb $b \rightarrow s\mu^+\mu^-$ anomalies [10, 11]. This gap is now largely excluded by this analysis.
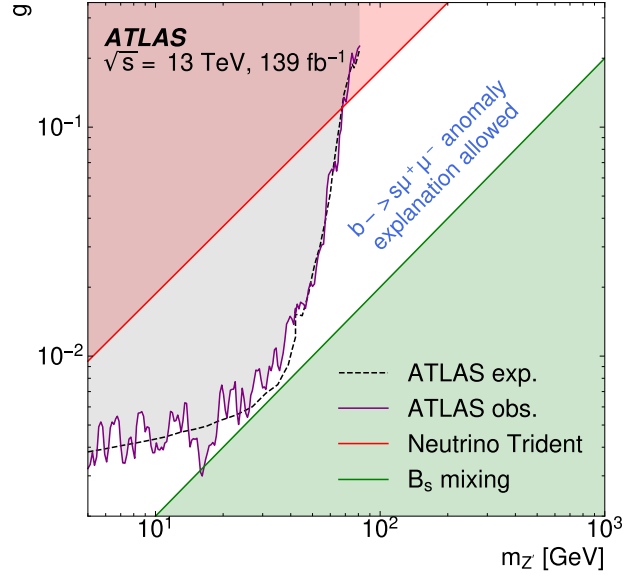
Figure 7: The coupling parameter g limits from this search as a function of the Z' mass compared to the limits [15] from the Neutrino Trident (red) and the B_s mixing (green) experimental results.

9 Conclusion

A search for a new vector gauge boson Z' predicted by the $L_\mu - L_\tau$ models has been performed with a 4μ final state in the invariant mass range of [80, 180] GeV, excluding the Higgs boson mass window [110, 130] GeV, using 139 fb^{-1} of $\sqrt{s}$ = 13 TeV proton-proton collision data collected with the ATLAS detector. No significant excess of events over the expected SM background is observed. Therefore, upper limits are set on the Z' production cross-section times the decay branching fraction of the $pp \rightarrow Z'\mu^+\mu^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ process, varying from 0.31 to 4.3 fb at 95% CL, in a Z' mass range of [5, 81] GeV, from which the coupling strength g of the Z' to muons above 0.003 to 0.2 (depending on the Z' mass) are excluded in the same mass range. This search shows significant sensitivity improvements over previous indirect and direct searches of the Z' with 4μ final state. An interesting parameter space of the $L_\mu - L_\tau$ model prediction that was not excluded before and could explain the $b \rightarrow s\mu^+\mu^-$ anomalies is now largely excluded by this search.

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[2] X.-G. He, G. C. Joshi, H. Lew and R. R. Volkas, *Simplest $Z^{'}$ model*, Phys. Rev. D **44** (1991) 2118.

[3] F. Elahi and A. Martin, *Constraints on $L_\mu - L_\tau$ interactions at the LHC and beyond*, Phys. Rev. D **93** (2016).

[4] F. Elahi and A. Martin, *Using the modified matrix element method to constrain $L_\mu - L_\tau$ interactions*, Phys. Rev. D **96** (2017).

[5] A. Crivellin, G. D'Ambrosio and J. Heeck, *Explaining $h \to \mu^\pm \tau^\mp$, $B \to K^* \mu^+ \mu^-$, and $B \to K \mu^+ \mu^- / B \to K e^+ e^-$ in a Two-Higgs-Doublet Model with Gauged $L_\mu - L_\tau$*, Phys. Rev. Lett. **114** (2015) 151801.

[6] Muon g-2 Collaboration, *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, Phys. Rev. Lett. **126** (2021) 141801, arXiv: `2104.03281 [hep-ex]`.

[7] A. Keshavarzi, *The Muon g-2 experiment at Fermilab*, EPJ Web of Conferences **212** (2019) 05003, ed. by M. Achasov, F. Ignatov and P. Krokovny.

[8] G. Venanzoni, *The Fermilab Muon g-2 Experiment*, PoS **EPS-HEP2015** (2016) 568.

[9] Muon g-2 Collaboration, *Final report of the E821 muon anomalous magnetic moment measurement at BNL*, Phys. Rev. D **73** (2006) 072003.

[10] LHCb Collaboration, *Test of lepton universality in beauty-quark decays*, Nat. Phys. **18** (2022) 277.

[11] LHCb Collaboration, *Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0} \mu^+ \mu^-$*, Phys. Rev. Lett. **111** (2013) 191801.

[12] LHCb Collaboration, *Test of lepton universality using $B^+ \to K^+ \ell^+ \ell^-$ decays*, Phys. Rev. Lett. **113** (2014) 151601.

[13] LHCb Collaboration, *Measurement of CP-Averaged Observables in the $B^0 \to K^{*0} \mu^+ \mu^-$ Decay*, Phys. Rev. Lett. **125** (2020) 011802.

[14] T. Blake, G. Lanfranchi and D. M. Straub, *Rare B Decays as Tests of the Standard Model*, Prog. Part. Nucl. Phys. **92** (2017) 50.

[15] W. Altmannshofer, S. Gori, S. Profumo and F. S. Queiroz, *Explaining dark matter and B decay anomalies with an $L_\mu - L_\tau$ model*, JHEP **12** (2016) 106, arXiv: `1609.04026 [hep-ph]`.

[16] E. Ma, D. Roy and S. Roy, *Gauged $L_\mu - L_\tau$ with large muon anomalous magnetic moment and the bimaximal mixing of neutrinos*, Phys. Lett. B **525** (2002) 101.

[17] P. Langacker, *The physics of heavy Z' gauge bosons*, Rev. Mod. Phys. **81** (2009) 1199.

[18] CMS Collaboration,
Observation of Z Decays to Four Leptons with the CMS Detector at the LHC,
JHEP **12** (2012) 034, arXiv: 1210.3844 [hep-ex].

[19] ATLAS Collaboration, *Measurements of Four-Lepton Production at the Z Resonance in pp Collisions at $\sqrt{s}$ =7 and 8 TeV with ATLAS*, Phys. Rev. Lett. **112** (2014) 231806,
arXiv: 1403.5657 [hep-ex].

[20] CMS Collaboration,
Measurements of the $pp \rightarrow ZZ$ production cross section and the $Z \rightarrow 4\ell$ branching fraction, and constraints on anomalous triple gauge couplings at $\sqrt{s} = 13$ TeV,
Eur. Phys. J. C **78** (2018) 165, arXiv: 1709.08601 [hep-ex],
Erratum: Eur. Phys. J. C **78**.6 (2018) 515.

[21] CMS Collaboration, *Search for an L_μ–L_τ gauge boson using $Z \rightarrow 4\mu$ events in proton–proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **792** (2019) 345,
arXiv: 1808.03684 [hep-ex].

[22] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*,
ATLAS-TDR-19; CERN-LHCC-2010-013, 2010,
URL: https://cds.cern.ch/record/1291633, Addendum: ATLAS-TDR-19-ADD-1;
CERN-LHCC-2012-009, 2012, URL: https://cds.cern.ch/record/1451888.

[23] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*,
JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det].

[24] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*,
Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex].

[25] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*,
ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187.

[26] ATLAS Collaboration,
ATLAS data quality operations and performance for 2015-2018 data-taking,
JINST **15** (2020) P04003.

[27] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*,
JINST **15** (2020) P09015, arXiv: 2004.13447 [hep-ex].

[28] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*,
JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph].

[29] C. Degrande et al., *UFO - The Universal FeynRules Output*,
Comput. Phys. Commun. **183** (2012) 1201, arXiv: 1108.2040 [hep-ph].

[30] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr and B. Fuks,
FeynRules 2.0 - A complete toolbox for tree-level phenomenology,
Comput. Phys. Commun. **185** (2014) 2250, arXiv: 1310.1921 [hep-ph].

[31] F. del Aguila, M. Chala, J. Santiago and Y. Yamamoto,
Collider limits on leptophilic interactions, JHEP **03** (2015) 059, arXiv: `1411.7394`.

[32] M. Grazzini, S. Kallweit and D. Rathlev,
ZZ production at the LHC: Fiducial cross sections and distributions in NNLO QCD,
Phys. Lett. B **750** (2015) 407.

[33] A. H. Ajjath et al., *Resummed Drell-Yan cross-section at N^3LL*, JHEP **10** (2020) 153.

[34] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244,
arXiv: `1207.1303 [hep-ph]`.

[35] N. Davidson, T. Przedzinski and Z. Was,
PHOTOS Interface in C++: Technical and physics documentation,
Comput. Phys. Commun. **199** (2016) 86, arXiv: `1011.0937 [hep-ph]`.

[36] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*,
Eur. Phys. J. C **70** (2010) 823, arXiv: `1005.4568 [physics.ins-det]`.

[37] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034,
arXiv: `1905.09127 [hep-ph]`.

[38] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039,
arXiv: `0808.3674 [hep-ph]`.

[39] S. Schumann and F. Krauss,
A parton shower algorithm based on Catani–Seymour dipole factorisation,
JHEP **03** (2008) 038, arXiv: `0709.1027 [hep-ph]`.

[40] S. Höche, F. Krauss, M. Schönherr and F. Siegert,
A critical appraisal of NLO+PS matching methods, JHEP **09** (2012) 049,
arXiv: `1111.1220 [hep-ph]`.

[41] S. Höche, F. Krauss, M. Schönherr and F. Siegert,
QCD matrix elements + parton showers. The NLO case, JHEP **04** (2013) 027,
arXiv: `1207.5030 [hep-ph]`.

[42] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, *QCD Matrix Elements + Parton Showers*,
JHEP **11** (2001) 063, arXiv: `hep-ph/0109231`.

[43] S. Höche, F. Krauss, S. Schumann and F. Siegert,
QCD matrix elements and truncated showers, JHEP **05** (2009) 053,
arXiv: `0903.1219 [hep-ph]`.

[44] F. Buccioni et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866,
arXiv: `1907.13071 [hep-ph]`.

[45] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering Amplitudes with Open Loops*,
Phys. Rev. Lett. **108** (2012) 111601, arXiv: `1111.5206 [hep-ph]`.

[46] A. Denner, S. Dittmaier and L. Hofer,
Collier: A fortran-based complex one-loop library in extended regularizations,
Comput. Phys. Commun. **212** (2017) 220, arXiv: `1604.06792 [hep-ph]`.

[47] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph].

[48] S. Frixione, P. Nason and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph].

[49] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146.

[50] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph].

[51] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph].

[52] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, URL: https://cds.cern.ch/record/2216168.

[53] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph].

[54] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419.

[55] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: 1009.2450 [hep-ph].

[56] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029, arXiv: 0805.3067 [hep-ph].

[57] C. Anastasiou, L. J. Dixon, K. Melnikov and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, Phys. Rev. D **69** (2004) 094008, arXiv: hep-ph/0312266.

[58] S. Agostinelli et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[59] T. Sjöstrand, S. Mrenna and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

[60] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: https://cds.cern.ch/record/2206965.

[61] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2021) 578, arXiv: 2012.00578 [hep-ex].

[62] ATLAS Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
Eur. Phys. J. C **77** (2017).

[63] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*,
JINST **14** (2019) P12006.

[64] ATLAS Collaboration, *Track assisted techniques for jet substructure*,
ATL-PHYS-PUB-2018-012, 2018, URL: https://cds.cern.ch/record/2630864.

[65] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*,
Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph].

[66] ATLAS Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex].

[67] ATLAS Collaboration,
Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run,
ATL-PHYS-PUB-2016-012, 2016, URL: https://cds.cern.ch/record/2160731.

[68] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*,
ATLAS-CONF-2014-018, 2014, URL: https://cds.cern.ch/record/1700870.

[69] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*,
Eur. Phys. J. C **78** (2018) 903, arXiv: 1802.08168 [hep-ex].

[70] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson,
Parameterized neural networks for high-energy physics, Eur. Phys. J. C **76** (2016).

[71] F. Chollet et al., *Keras*, https://keras.io, 2015.

[72] Mart'ın Abadi et al.,
TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,
Software available from tensorflow.org, 2015, URL: https://www.tensorflow.org/.

[73] R. Liaw et al., *Tune: A Research Platform for Distributed Model Selection and Training*,
(2018), arXiv: 1807.05118 [cs.LG].

[74] A. I. Cowen-Rivers et al.,
HEBO Pushing The Limits of Sample-Efficient Hyperparameter Optimisation, 2020,
arXiv: 2012.03826 [cs.LG].

[75] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019,
URL: https://cds.cern.ch/record/2677054.

[76] G. Avoni et al.,
The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS,
JINST **13** (2018) P07017.

[77] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*,
J. Phys. G **43** (2016) 023001, arXiv: `1510.03865 [hep-ph]`.

[78] ATLAS Collaboration, *Combined search for the Standard Model Higgs boson in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, Phys. Rev. D **86** (2012) 032003,
arXiv: `1207.0319 [hep-ex]`.

[79] G. Cowan, K. Cranmer, E. Gross and O. Vitells,
Asymptotic formulae for likelihood-based tests of new physics,
Eur. Phys. J. C **71** (2011) 1554, arXiv: `1007.1727 [physics.data-an]`,
Erratum: Eur. Phys. J. C **73** (2013) 2501.

[80] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693.

[81] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*,
Eur. Phys. J. C **70** (2010) 525.

[82] S. R. Mishra et al., *Neutrino tridents and W-Z interference*,
Phys. Rev. Lett. **66** (24 1991) 3117.