The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/198118

Please be advised that this information was generated on 2019-04-20 and may be subject to change.
Search for lepton-flavor-violating decays of the Z boson into a $\tau$ lepton and a light lepton with the ATLAS detector

M. Aaboud et al. (ATLAS Collaboration)

Direct searches for lepton flavor violation in decays of the Z boson with the ATLAS detector at the LHC are presented. Decays of the Z boson into an electron or muon and a hadronically decaying $\tau$ lepton are considered. The searches are based on a data sample of proton-proton collisions collected by the ATLAS detector in 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb$^{-1}$ at a center-of-mass energy of $\sqrt{s}$ = 13 TeV. No statistically significant excess of events above the expected background is observed, and upper limits on the branching ratios of lepton-flavor-violating decays are set at the 95% confidence level: $B(Z \rightarrow \ell\ell) < 5.8 \times 10^{-5}$ and $B(Z \rightarrow \mu\tau) < 2.4 \times 10^{-5}$. This is the first limit on $B(Z \rightarrow \ell\ell)$ with ATLAS data. The upper limit on $B(Z \rightarrow \mu\tau)$ is combined with a previous ATLAS result based on 20.3 fb$^{-1}$ of proton-proton collision data at a center-of-mass energy of $\sqrt{s}$ = 8 TeV and the combined upper limit at 95% confidence level is $B(Z \rightarrow \mu\tau) < 1.3 \times 10^{-5}$.

DOI: 10.1103/PhysRevD.98.092010

I. INTRODUCTION

One of the main goals of the physics program of the Large Hadron Collider (LHC) at CERN is to discover physics beyond the Standard Model (SM). The observation of lepton flavor violation in decays of the Z boson into a pair of leptons of different flavors would provide a clear indication for new physics. These decays can occur within the SM only via neutrino oscillations and would have a rate too small to be detected [1]. This paper presents searches by the ATLAS Collaboration for the decays of the Z boson into a $\tau$ lepton and an electron or a muon, hereafter referred to as a light lepton or $\ell$. Only final states with a hadronically decaying $\tau$ lepton are considered.

Lepton-flavor-violating (LFV) Z boson decays are predicted by models with heavy neutrinos [2], extended gauge models [3] and supersymmetry [4]. The most stringent bounds on such decays with a $\tau$ lepton in the final state are set by the LEP experiments: $B(Z \rightarrow \ell\ell) < 9.8 \times 10^{-6}$ [5] and $B(Z \rightarrow \mu\tau) < 1.2 \times 10^{-5}$ [6] at 95% confidence level (C.L.). The ATLAS experiment has set the upper limit $B(Z \rightarrow \mu\tau) < 1.7 \times 10^{-5}$ at 95% C.L. [7] by analyzing 20.3 fb$^{-1}$ of proton-proton collisions at a center-of-mass energy of 8 TeV. There are no previously published limits on $B(Z \rightarrow \ell\ell)$ with ATLAS data. Regarding the LFV $Z \rightarrow e\mu$ decays, at the LHC the CMS experiment set the upper bound at $B(Z \rightarrow e\mu) < 7.3 \times 10^{-7}$ at 95% C.L. [8]. The ATLAS experiment obtained a similar result with the upper bound set at $B(Z \rightarrow e\mu) < 7.5 \times 10^{-7}$ at 95% C.L. [9].

The searches for LFV $Z$ decays presented in this paper use a data sample of proton-proton collisions collected at a center-of-mass energy of $\sqrt{s}$ = 13 TeV with the ATLAS detector at the LHC. These data correspond to an integrated luminosity of 36.1 fb$^{-1}$. The signal model used assumes unpolarized $\tau$ leptons. Events are classified using neural networks, and the output distribution is used in a template fit to data to extract the Z boson lepton-flavor-violating branching ratios, or otherwise set upper limits on these values. The major backgrounds to the search are reducible backgrounds such as $W$+jets, top-quark pair production and $Z \rightarrow \ell\ell$, and the irreducible background $Z \rightarrow \ell\ell \rightarrow \ell\ell +$ hadrons + $3\nu$. Reducible backgrounds from events with a quark- or gluon-initiated jet misidentified as a hadronically decaying $\tau$ lepton, so-called “fakes,” are estimated via a data-driven method. The reducible backgrounds from events with a quark- or gluon-initiated jet misidentified as a hadronically decaying $\tau$ lepton, so-called “fakes,” are estimated via a data-driven method. The irreducible background from $Z \rightarrow \ell\ell$, where one light lepton fakes a hadronic $\tau$-lepton decay signature, are estimated using simulation. An event selection specifically designed to reduce the contribution from this background is applied. The shape of the template for the irreducible background from $Z \rightarrow \ell\ell$ is estimated via simulations and its magnitude is determined in the fit to data.

The results of the search for the LFV $Z \rightarrow \mu\tau$ decays presented in this paper are combined with the previous ATLAS results based on 8 TeV data.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP³.
This paper is structured as follows. Section II briefly describes the ATLAS detector and the reconstruction of the detected particles. Section III details the data sample and the simulations used in the analysis. Section IV describes the event selection and classification criteria. Section V discusses the methodology used to estimate the yield of events from background sources, and Sec. VI lists the experimental and theoretical systematic uncertainties affecting the analysis. The statistical interpretation of the observed data and the results are presented in Sec. VII. The combination of the result in the observed data and the results are presented in Sec. VII.

II. THE ATLAS DETECTOR AND OBJECT RECONSTRUCTION

The ATLAS detector\(^1\) [10] at the LHC is a multipurpose particle detector with a forward-backward symmetric cylindrical geometry and a nearly \(4\pi\) coverage in solid angle. It consists of an inner tracking detector, electromagnetic and hadronic calorimeters, and a muon spectrometer. The inner detector (ID), immersed in a 2 T axial magnetic field provided by a thin superconducting solenoid, includes silicon pixel and microstrip detectors, which provide precision tracking in the pseudorapidity range \(|\eta| < 3.2\), and a hadronic calorimeter with a steel/scintillator-tile barrel section for \(|\eta| < 1.7\) and two copper/LAr end caps for \(1.5 < |\eta| < 3.2\). The forward region is covered by additional LAr calorimeters with coarser granularity up to \(|\eta| = 4.9\). The calorimeter is surrounded by the muon spectrometer, which is based on three large superconducting toroid magnets each containing eight coils. Precise momentum measurements for muons with pseudorapidity up to \(|\eta| = 2.7\) are provided by three layers of tracking chambers. The muon spectrometer also includes separate trigger chambers covering \(|\eta| < 2.4\). A two-level trigger system [12] was used during the \(\sqrt{s} = 13\) data-taking period. The first-level trigger (L1) is implemented in hardware and uses a subset of the detector information. This is followed by a software-based level which runs algorithms similar to the offline reconstruction software, reducing the event rate to approximately 1 kHz on average from the maximum L1 rate of 100 kHz.

Electron candidates are reconstructed from energy deposits in the electromagnetic calorimeter which are matched to a charged-particle track measured in the inner detector. These candidates are required to satisfy "medium" likelihood-based identification criteria [13], to have a transverse momentum \(p_T > 30\) GeV and to be in the acceptance region \(|\eta| < 2.47\) of the inner detector. Candidates in the transition region \(1.37 < |\eta| < 1.52\) between the barrel and end cap calorimeters are excluded.

Muon candidates are reconstructed from track segments in the muon spectrometer which are matched to tracks reconstructed in the inner detector which satisfy \(|\eta| < 2.5\). The matched tracks are re-fitted using the complete track information from both detector subsystems. Muon candidates are required to satisfy "medium" identification criteria [14] and to have a \(p_T > 30\) GeV.

Isolation criteria are applied to both the electrons and muons using calorimeter- and track-based information. The calorimeter-based variables are corrected for the average energy contributions from additional proton-proton interactions in the same and neighboring bunch crossings ("pileup") [13]. A standard working point is used, where a 90% efficiency is obtained for leptons with \(p_T = 25\) GeV, rising to 99% efficiency at \(p_T = 60\) GeV in \(Z \rightarrow \ell\ell\) events. Events with electrons or muons that fail the isolation criteria are used to estimate background contributions, as described in Sec. V.

Topological clusters of energy deposits in the calorimeter are reconstructed into jets with the anti-\(k_t\) algorithm [15] and radius parameter \(R = 0.4\) using the FASTJET software package [16]. Energy deposits from pileup are subtracted using an average pileup energy density and the jet area. Jets are then calibrated as described in Ref. [17]. Jet candidates are required to have \(p_T > 20\) GeV and \(|\eta| < 2.5\). To further reduce the effect of pileup, a jet vertex tagger (JVT) algorithm is used for jets with \(p_T < 60\) GeV and \(|\eta| < 2.4\). The JVT algorithm employs a multivariate technique based on jet energy, vertexing, and tracking variables in order to determine the likelihood that jets originate from or are heavily contaminated by pileup [18].

In order to identify jets containing \(b\)-hadrons (\(b\)-jets), a multivariate algorithm is used that depends on the presence of tracks with a large impact parameter with respect to the primary vertex [19], on the presence of displaced secondary vertices, and on the reconstructed flight paths of \(b\)- and \(c\)-hadrons associated with the jet [20]. Using this algorithm, jets are \(b\)-tagged if they satisfy criteria of a standard working point, which provides a \(b\)-jet efficiency of 77%
and a light-jet rejection rate of about 134 in simulated $t\bar{t}$ events.

Hadronic $\tau$-lepton decays result in a neutrino and a set of visible decay products ($\tau_{\text{had-vis}}$), typically one or three charged pions and up to two neutral pions [21]. The reconstruction of the visible decay products [22] is seeded by jets. Selected $\tau_{\text{had-vis}}$ candidates are required to have $p_T > 20$ GeV, $|\eta| < 2.5$ excluding 1.37 < $|\eta| < 1.52$, one (1-prong) or three (3-prong) associated tracks with $p_T > 1$ GeV, and an electric charge of ±1. A boosted decision tree (BDT) identification procedure that is based on calorimetric shower shapes and tracking information is used to discriminate $\tau$-lepton decays from jet backgrounds [23,24]. All events used in this analysis must have a $\tau_{\text{had-vis}}$ candidate that passes the “loose” identification working point. For events in the signal region, the $\tau_{\text{had-vis}}$ candidate must satisfy the “tight” identification criterion. Selected events that are not in the signal region are used to estimate backgrounds (Sec. V). The combined reconstruction and identification efficiencies for loose and tight criteria are 60% (50%) and 45% (30%) for 1-prong (3-prong) hadronic $\tau$-lepton decays, and are independent of the $\tau_{\text{had-vis}}$ $p_T$ and the number of pileup interactions. To reduce the number of muons misidentified as $\tau_{\text{had-vis}}$, a $\tau_{\text{had-vis}}$ candidate is excluded if it is within $\Delta R = 0.2$ of a reconstructed muon with $p_T > 2$ GeV. An additional BDT, denoted hereafter by eBDT, is used to reduce the number of electrons misidentified as $\tau_{\text{had-vis}}$ providing 85% (95%) efficiency for 1-prong (3-prong) hadronic $\tau$-lepton decays. The leading-$p_T$ candidate is selected as the $\tau_{\text{had-vis}}$ candidate, while any other candidates are considered to be jets.

To avoid potential ambiguities among objects, light lepton and $\tau_{\text{had-vis}}$ candidates are required to be separated from each other and from jets in the following order: (a) jets within $\Delta R = 0.2$ of selected $\tau_{\text{had-vis}}$ candidates are excluded, (b) jets within $\Delta R = 0.4$ of an electron or a muon are excluded, (c) any $\tau_{\text{had-vis}}$ within $\Delta R = 0.2$ of an electron or a muon is excluded, and (d) electrons within $\Delta R = 0.2$ of a muon are excluded.

The missing transverse momentum, with magnitude $E^\text{miss}_T$, is calculated as the negative vectorial sum of the transverse momenta of all fully reconstructed and calibrated (“hard”) physics objects and inner-detector tracks that originate from the hard-scattering vertex but are not matched to a reconstructed object (“soft term”) [25]. The soft term is an important contribution for improving both the $E^\text{miss}_T$ scale and its resolution.

III. DATA AND SIMULATED EVENT SAMPLES

This search analyzes proton-proton collisions recorded by the ATLAS detector at the LHC during 2015 and 2016 at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The data correspond to a total integrated luminosity of 36.1 fb$^{-1}$ after requiring that all relevant components of the ATLAS detector were in good working condition during data collection. The uncertainty in the combined 2015 and 2016 integrated luminosity is 2.1%. It was estimated following a methodology similar to the one described in Ref. [26]. The events considered for the $e\tau$ ($\mu\tau$) channel were selected by single-lepton triggers which require the presence of at least one electron (muon) candidate with transverse momentum above 24 GeV (20 GeV) in 2015 data and 26 GeV (26 GeV) in 2016 data. These triggers apply isolation criteria for electron (muon) candidates with $p_T$ below 60 GeV (40 GeV in 2015 and 50 GeV in 2016). These isolation requirements are looser than the ones applied offline in the light-lepton selections used in this analysis.

Simulated Monte Carlo (MC) samples are used to predict the $Z/\gamma^* \rightarrow e\tau$ signal and the background contributions from $Z/\gamma^* +$ jets, $W +$ jets, $t\bar{t}$, single top-quark, Higgs boson and diboson (WW, WZ and ZZ) production.

Signal samples were simulated using Pythia 8.186 [27] with the NNPDF2.3 parton distribution function (PDF) set [28] and a set of tuned parameters called the A14 tune [29]. The lepton-flavor-violating $Z/\gamma^*$ decay was modeled assuming unpolarized $\tau$ leptons in the final state. To ensure that both the hypothetical signal $Z \rightarrow e\tau$ and the main background $Z/\gamma^* \rightarrow \tau\tau$ are normalized to the same production cross section, event weights computed as a function of the true boson transverse momentum are applied to the signal events to match the more accurate modeling of the $Z/\gamma^*$ production in the $Z/\gamma^* \rightarrow \tau\tau$ simulation described in the following. After this reweighting procedure, the signal events, together with the $Z/\gamma^* \rightarrow \tau\tau$ events, are normalized to the $Z/\gamma^*$ production cross section determined from data in the template fit described in Sec. VII. Therefore, the analysis is independent of the theoretical uncertainty in the $Z/\gamma^*$ production cross section. The SM value of this cross section is 2.1 nb, calculated at NNLO accuracy [30].

The production of $Z/\gamma^* \rightarrow \tau\tau$ events was simulated with Sherpa 2.2.1 [31]. The NNPDF 3.0 NNLO PDF set [32] was used for both the matrix element calculation and the dedicated parton-shower tuning developed by the authors of Sherpa. The event generation utilized Comix [33] and OpenLoops [34] for the matrix element calculation, which was then matched to the Sherpa parton shower using the ME+PS@NLO prescription [35]. The matrix elements were calculated for up to two additional partons at NLO and for three and four partons at LO in QCD. As stated above, the normalization of this background process, together with the signal events, is determined in a fit to data.

The $Z/\gamma^* \rightarrow \mu\mu$, $ee$ events were simulated with Powheg-Box [36–38] using the CT10 PDF set [39] and the AZNLO tune [40], and interfaced to Pythia 8.186. The normalization of the $Z/\gamma^* \rightarrow \mu\mu$, $ee$ events is determined from data in a dedicated region enhanced in $Z \rightarrow \mu\mu$ events (Sec. V) as a function of the reconstructed transverse momentum of the $Z/\gamma^*$ boson.
The other simulated processes account for only a small fraction (less than 0.3%) of the background events. Samples of $W(\rightarrow \ell \nu) + \text{jets}$ events were simulated with SHERPA 2.2.1. Events with a top-quark pair or a single top quark produced via electroweak $t$-channel, $s$-channel and $Wt$-channel processes were simulated with POWHEG-BOX using the CT10 PDF set. The parton shower, fragmentation and underlying event were simulated using PYTHIA 6.428 [41] with the Perugia 2012 tune [42]. EVTGEN [43] was used to decay bottom and charm hadrons. Diboson processes were simulated with SHERPA 2.1.0 with the CT10 PDF set. Higgs boson events, $H \rightarrow WW, \tau\tau, \ell\ell$, produced via gluon–gluon fusion and vector-boson fusion were simulated with POWHEG-BOX.

Simulated minimum-bias events were overlaid on all simulated samples to include the effect of pileup. These minimum-bias events were generated with PYTHIA 8.186, using the A2 tune [44] and the MSTW2008LO PDF set [45]. Each simulated event was processed using the GEANT-based ATLAS detector simulation [46,47] and the same event reconstruction algorithms used for the data. Reconstruction and identification efficiencies, as well as energy calibrations for all selected objects in simulated events, are corrected to match those measured in data.

IV. EVENT SELECTION AND CLASSIFICATION

Of the events satisfying the trigger and the quality criteria described in Sec. III, the events selected in this analysis are required to contain exactly one isolated electron or muon that is geometrically matched to the object that fired the trigger, and no additional light leptons. These events must also contain at least one $\tau_{\text{had-vis}}$ candidate that passes the tight identification. The isolated light lepton and the $\tau_{\text{had-vis}}$ candidate are required to have opposite charge, $q_\ell q_{\text{had-vis}} = -1$. Events with one or more $b$-tagged jets are removed to reject background events with a top-quark pair or a singly produced top quark. To reduce the $Z \rightarrow \ell\ell$ background, events with 1-prong $\tau_{\text{had-vis}}$ candidates that satisfy $|\eta(\tau_{\text{had-vis}})| > 2.2$ for the $\ell\tau$ channel or $|\eta(\tau_{\text{had-vis}})| < 0.1$ for the $\mu\tau$ channel are rejected. These regions of the detector are excluded because they are insufficiently instrumented and therefore affected by higher $\ell' \rightarrow \tau$ misreconstruction and misidentification rates. The selection described here, denoted hereafter as preselection, defines the sample of events used for the training of the neural network.

Further kinematic selections are applied to define the sample of events in the “signal region” (SR) which are used in the final template fit. Orthogonal sets of events in the so-called “calibration regions” (CR) are defined by inverting some of the preselection or SR selection requirements and used to estimate background contributions in the SR, as described in Sec. V.

Events accepted in the SR must satisfy the preselection and the following selections. The transverse mass,

$$m_T(\tau_{\text{had-vis}}, E^{\text{miss}}_T) \equiv \sqrt{2p_T(\tau_{\text{had-vis}})E^{\text{miss}}_T[1 - \cos(\Delta\phi(\tau_{\text{had-vis}}, E^{\text{miss}}_T))]} \quad (1)$$

defined as the transverse momentum from the neutrino in a direction close to the $\tau_{\text{had-vis}}$ candidate, resulting in small $m_T(\tau_{\text{had-vis}}, E^{\text{miss}}_T)$ values. The $W(\rightarrow \ell\nu/\tau\nu) + \text{jets}$ events and some of the $Z/\gamma \rightarrow \ell\ell$ events have instead higher $m_T(\tau_{\text{had-vis}}, E^{\text{miss}}_T)$ values. This selection allows the definition of a CR that is dominated by $W + \text{jets}$ events, which are the major contribution to fakes. The selection is illustrated in Fig. 1. In events with a 1-prong $\tau_{\text{had-vis}}$ candidate, an additional selection is applied to further reduce the $Z \rightarrow \ell\ell$ background. In most of these events, the momentum of the track matched to the 1-prong $\tau_{\text{had-vis}}$ candidate corresponds to the original momentum of the light lepton misidentified as $\tau_{\text{had-vis}}$, while the energy deposited in the calorimeter used to estimate the energy of the $\tau_{\text{had-vis}}$ originates from

![ATLAS Simulation](attachment:ATLAS_Simulation.png)

**FIG. 1.** Expected distributions of $m_T(\tau_{\text{had-vis}}, E^{\text{miss}}_T)$ in $Z/\gamma \rightarrow \ell\ell, W(\rightarrow \ell\nu/\tau\nu) + \text{jets}$ and signal events in the $\ell\tau$ (left) and $\mu\tau$ (right) channels after preselection requirements. The $Z/\gamma \rightarrow \ell\ell$ and $W(\rightarrow \ell\nu/\tau\nu) + \text{jets}$ distributions also include the contributions to fakes from the corresponding processes as predicted by MC simulations. All distributions are normalized to unity.
radiation (light-lepton bremsstrahlung) or other sources. Therefore, events in which the invariant mass of the $\tau_{\text{had-vis}}$ track and the light lepton ($m(\text{track}, \ell')$) is compatible with the $Z$ boson mass are rejected. In particular, events with a 1-prong $\tau_{\text{had-vis}}$ candidate are accepted when $m(\text{track}, \ell') < 84$ GeV or $m(\text{track}, \ell') > 105$ GeV if $|\eta(\tau_{\text{had-vis}})| < 2.0$, and when $m(\text{track}, \ell') < 80$ GeV or $m(\text{track}, \ell') > 105$ GeV if $|\eta(\tau_{\text{had-vis}})| > 2.0$. A wider range in $m(\text{track}, \ell')$ is rejected at high $|\eta(\tau_{\text{had-vis}})|$ because of the smaller signal contribution and the higher $Z \rightarrow \ell \ell'$ background rate.

Moreover, events in which the invariant mass of the 1-prong $\tau_{\text{had-vis}}$ candidate and the light lepton satisfies $80$ GeV < $m(\tau_{\text{had-vis}}, \ell')$ < $100$ GeV are required to have $m(\text{track}, \ell') > 40$ GeV. These selections have been optimized in the $m(\tau_{\text{had-vis}}, \ell') - m(\text{track}, \ell')$ plane to specifically reject the $Z \rightarrow \tau \tau$, $Z \rightarrow \ell \ell'$ and $W \rightarrow \ell \nu + \text{jets}$ background selection efficiencies are, including their contributions to fakes as predicted by MC simulations, $7.1 \times 10^{-3}$ ($6.5 \times 10^{-3}$), $9.4 \times 10^{-5}$ ($5.9 \times 10^{-5}$) and $2.6 \times 10^{-5}$ ($2.9 \times 10^{-5}$) respectively. A summary of the event selection criteria is given in Table I.

Events accepted in the SR are classified using neural networks (NNs) trained to discriminate $Z \rightarrow \ell \tau$ signal from $Z \rightarrow \tau \tau$, $Z \rightarrow \ell \ell'$ and $W \rightarrow \ell \nu + \text{jets}$ background events. The classification is based on event kinematic properties that are extracted by the NN from the reconstructed momenta of the selected particles, as well as from other event variables. The NN achieves good performance using low-level variables, such as the particle momentum components, due to the network’s capability to build non-linear relations between input variables.

Three types of NN classifiers, “$Z$”, “$Zll$” and “$W$”, are trained to distinguish signal from $Z \rightarrow \tau \tau$, $Z \rightarrow \ell \ell'$ and $W \rightarrow \ell \nu$ backgrounds, respectively. These classifiers are trained separately in the $e\tau$ and $\mu\tau$ channels because of the different detector acceptances, but combine 1-prong and 3-prong $\tau_{\text{had-vis}}$ candidates. Simulated events passing the preselection (Table I) are used to train, optimize and

**FIG. 2.** Expected distributions of $m(\text{track}, \ell')$ versus $m(\tau_{\text{had-vis}}, \ell')$ in signal (left) and $Z \rightarrow \ell \ell'$ (right) events with 1-prong $\tau_{\text{had-vis}}$ candidates in the $e\tau$ (top) and $\mu\tau$ (bottom) channels after the SR selection except for the cuts on these two variables (Table I).
validate the classifiers. In order to increase the size of the
available training samples for \(Z \to \ell \tau\) and \(Z \to \tau \tau\) pro-
cesses with a true hadronic \(\ell\)-lepton decay, all events with a
\(\tau\) candidate that passes the loose identification are used. Moreover, in the events used for the Zll classifiers, the
misreconstructed \(\tau\) is required to be either a true muon or
electron. With these requirements, about 40 000 signal
events, 200 000 \(Z \to \tau \tau\) events and 80 000 \(W \to \ell \nu\) events are used for training in each channel. For \(Z \to \ell \ell\), about
30 000 events are used in the \(\ell \tau\) channel and only 5000 events in the \(\mu \tau\) channel. The limited number of \(Z \to \mu \mu\) events
is due to the low \(\mu \to \tau\) misreconstruction rate, and leads to poor classification power for the Zll NN in the \(\mu \tau\)
channel. However, the \(Z \to \mu \mu\) background is effectively reduced by the selection on \(m(\text{track}, \ell)\) and \(m(\tau\text{had-vis}, \ell)\)
described earlier.

The input variables common to all the classifiers are the
light lepton, \(\tau\text{had-vis}\) and \(E_T^{\text{miss}}\) momentum components, assuming vanishing masses; the collinear mass \(m_{\text{coll}}\),
defined as the invariant mass of the \(\ell - \tau\text{had-vis} - \nu\) system,
where \(\nu\) is the neutrino selected at trigger level, which is assumed
to have a momentum that is equal in the transverse plane to
the measured \(E_T^{\text{miss}}\) and collinear in \(\eta\) with the \(\tau\text{had-vis}\)
candidate; and \(\Delta \alpha\) [48]:

\[
\Delta \alpha = \frac{1}{2} \frac{m_Z^2 - m_\tau^2}{p(\tau\text{had-vis}) \cdot p(\ell)} - \frac{p(\ell)}{p_T(\tau\text{had-vis})},
\]

where \(p(\tau\text{had-vis})\) and \(p(\ell)\) are the four-momenta of the
\(\tau\) and the light-lepton candidates respectively, and the rest masses \(m_Z\) and \(m_\tau\) take on values reported by the
Particle Data Group [21]. The variable \(\Delta \alpha\) helps to
discriminate signal events, expected to be around \(\Delta \alpha = 0\), from \(Z \to \tau \tau\) events, where \(\Delta \alpha\) is negative due
to the presence of additional neutrinos. Even though not
specifically targeted by this variable, \(Z \to \ell \ell\) and \(W \to \ell \nu\)
events tend to be at vanishing and positive values of \(\Delta \alpha\),
respectively, as shown later in Figs. 5–8. The invariant mass
\(m(\ell, \tau\text{had-vis})\) is also used in the Zll classifier. In the limit of
very large training statistics, the light lepton, \(\tau\text{had-vis}\) and
\(E_T^{\text{miss}}\) momentum components would be sufficient for the
NN to learn the full event kinematics. However, with the
available training samples, the high-level variables \(m_{\text{coll}}\),
\(\Delta \alpha\) and \(m(\ell, \tau\text{had-vis})\) were found to be able to improve the
NN classification power and were therefore included
among the NN inputs.

The NN inputs are preprocessed to harmonize their
magnitudes and to remove known symmetries as is required
for optimal training. The preprocessing consists of the
following steps:

1. Boost: after computing \(m_{\text{coll}}\), \(\Delta \alpha\) and \(p_{\text{tot}} = p(\ell) + p(\tau\text{had-vis}) + E_T^{\text{miss}}\) in the lab frame, the light lepton,
\(\tau\text{had-vis}\) and \(E_T^{\text{miss}}\) momenta are boosted to the frame in which their total momentum vanishes. The
longitudinal component of the three-momentum of \(E_T^{\text{miss}}\) is zero in the lab frame.

2. Rotation: the light lepton, \(\tau\text{had-vis}\) and \(E_T^{\text{miss}}\) momenta are first rotated so that the three-momentum of the
light lepton is along the positive \(z\) axis. A second rotation about the \(z\) axis is applied so that the \(\tau\text{had-vis}\) momentum has a vanishing component on the \(y\) axis.

3. “Feature scaling”: each input variable is scaled by subtracting its mean and by dividing by its standard
deviation, where the mean and the standard deviation are
computed on the set of signal and background
events used in the training of each classifier.

The boost and the rotation are used to remove the
degeneracy among apparently different events which are
instead equivalent under Lorentz transformation. “Feature scaling” is needed because the network works best with
input variables of the same magnitude. The same prepro-
cessing procedure, with the same mean and standard
deviation values, is applied to all the events on which
the classifiers are evaluated. After preprocessing, six of the
twelve components of the light lepton, \(\tau\text{had-vis}\) and \(E_T^{\text{miss}}\)
momenta are either vanishing or redundant, and therefore
not included in the network inputs. The resulting lists of
input variables are given in Table II. The transverse
component, \(p_T^{\text{tot}}\), of the total momentum \(p_{\text{tot}}\) in the lab
frame is also included as otherwise this information would
be lost after the preprocessing. The distributions of some of
the NN input variables are shown in Sec. VII.

The NN classifiers are sequential models optimized
for binary classification. They are based on the
TABLE II. Input variables for the NN classifiers. The first six quantities are in the boosted and rotated frame described in the text; the last four are in the laboratory frame.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Z NN</th>
<th>Zll NN</th>
<th>W NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{E}_{\text{lep}} )</td>
<td>light-lepton energy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \hat{p}_z^{\text{had-vis}} )</td>
<td>( \tau_{\text{had-vis}} ) p_\perp</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \hat{p}_z^{\text{QCD-vis}} )</td>
<td>( \tau_{\text{had-vis}} ) p_\perp</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \hat{E}_{\text{Had-vis}} )</td>
<td>( \tau_{\text{had-vis}} ) energy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \hat{p}_z^{\text{miss}} )</td>
<td>( E_{\text{miss}} ) component along z axis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \hat{E}_{\text{miss}} )</td>
<td>magnitude of ( E_{\text{miss}} )</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( p_{\text{T}} )</td>
<td>transverse component of total momentum</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( m_{\text{coll}} )</td>
<td>collinear mass</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \Delta \alpha )</td>
<td>see Eq. (1) [48]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( m(\ell, \tau_{\text{had-vis}}) )</td>
<td>invariant mass of light lepton and ( \tau_{\text{had-vis}} )</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

combined output (3P) = \( 1 - \sqrt{(1 - \text{output}_{W})^2 + (1 - \text{output}_{Z})^2 + (1 - \text{output}_{Zll})^2} \).  

combined output (1P) = \( 1 - \sqrt{(1 - \text{output}_{W})^2 + (1 - \text{output}_{Z})^2 + (1 - \text{output}_{Zll})^2} \).  

The chosen procedure to combine the individual NN outputs reduces the dimensionality of the classifiers while maintaining the correlations among these classifiers for each event. The binned distributions of these combined classifiers for the events selected in the SR are used in the final template fit, as discussed in Sec. VII.

V. BACKGROUND ESTIMATION

Background processes are categorized according to the origin of the \( \tau_{\text{had-vis}} \) candidate, which can be a true \( \tau \) lepton, or a misidentified light lepton, or a misidentified quark- or gluon-initiated jet. Different techniques are used to estimate these background contributions in the SR, as well as to model their expected combined NN output distributions, which are used in the template fit to data (Sec. VII). As described in the following, the shapes of all components are determined prior to the fit, as are the normalizations for all but the Z → ττ and fake components, which are determined in the fit.

Backgrounds from processes with a true hadronically decaying \( \tau \) lepton are estimated from simulation. The Z → ττ decays are the dominant source of these events. As detailed in Sec. III, they are modeled via simulation but their total yield in the SR is left unconstrained in the template fit to data in order to remove the theoretical systematic uncertainties in the Z production cross section.

Processes where the \( \tau_{\text{had-vis}} \) candidate is a misidentified light lepton are also estimated from simulation. These are mostly Z → \( \ell \ell \) events. The simulated rate for misidentifying electrons as 1-prong \( \tau_{\text{had-vis}} \) candidates is corrected using data [24]. Due to the lack of dedicated measurements of the rates of misidentifying electrons as 3-prong \( \tau_{\text{had-vis}} \) candidates and muons as 1-prong \( \tau_{\text{had-vis}} \) candidates, conservative uncertainties are assigned which have negligible impact on the precision of the measured \( B(Z \to \ell \tau) \).

The normalization of the Z → \( \ell \ell \) events is determined from data with a sample of events with an opposite-charge muon pair with 81 GeV < \( m_{\mu\mu} \) < 101 GeV. The preselection requirements on the leading muon, the absence of b-tagged jets and the veto on additional light leptons are imposed. A correction factor derived as the relative difference between the predicted and observed numbers of Z → \( \mu\mu \) events is applied to both the Z → ee and Z → \( \mu\mu \) yields in the SR. This correction is applied as a function of the reconstructed transverse momentum of the Z/\( \gamma^* \) boson to correct the overall normalization as well as the \( p_{\text{T}}(Z/\gamma^*) \) distribution of the simulated Z → \( \ell \ell \) events. In the Z → \( \mu\mu \)-enhanced region, the Z/\( \gamma^* \) boson momentum is computed as the vector sum of the muon pair, while in the SR it is the vector sum of the misidentified \( \tau_{\text{had-vis}} \) candidate and the remaining light lepton. The uncertainty in this correction is statistical only. Differences between the electron and muon acceptances are covered by the systematic uncertainties in the electron and muon selections, which are accounted for in the Z → \( \ell \ell \) predictions in the SR.

Events where the \( \tau_{\text{had-vis}} \) candidate originates from a quark- or gluon-initiated jet are estimated from data. This contribution is referred to as “fakes” and is dominated by W + jets and multijet processes. A data-driven fake-factor technique is used to estimate this contribution. It uses events in the so-called “fail sideband,” which is the set of events passing all but one of the SR selection requirements:
the \( \tau_{\text{had-vis}} \) candidate is required to fail the tight identification requirement. This is a set of events orthogonal to the ones selected in the SR and enhanced with fakes. The yield of these events is corrected by the fake factor, which is the transfer factor needed to scale the fail sideband sample to the yield of these events. The fake factors are measured separately for \( \tau_{\text{had-vis}} \) candidates with one and with three associated tracks. For 1-prong candidates, they are estimated in two-dimensional bins of \( \tau_{\text{had-vis}} \) \( p_T \) and \( \tau_{\text{had-vis}} \) track \( p_T \), since the associated track momentum is used in the selection of these candidates, while for 3-prong candidates they are estimated only in bins of \( \tau_{\text{had-vis}} \) \( p_T \). The choice of bin boundaries is optimized to capture the statistically significant variations of the fake factors as a function of the \( \tau_{\text{had-vis}} \) properties, while retaining enough events per bin. An additional binning as a function of \( \tau_{\text{had-vis}} \) \( |\eta| \) was found to be unnecessary. The measured fake factors are shown in Table IV. For events with low \( \tau_{\text{had-vis}} \) \( p_T \) and high \( \tau_{\text{had-vis}} \) track \( p_T \), the fake factors are large and have large statistical uncertainties because there are few events in the calibration regions. However, these fake factors are applied only to a small fraction of events in the sidebands.

The number of events from fakes in the SR is

\[
N_{\text{fake}} = \sum_k (N_{\text{fail}}^{\text{SR, data}} - N_{\text{fail}}^{\text{SR, MC, not jet-\tau}}) \times F_k,
\]

where \( F_k \) is the fake factor corresponding to the \( p_T \) (and track \( p_T \) for 1-prong \( \tau_{\text{had-vis}} \)) bin \( k \), \( N_{\text{fail}}^{\text{SR, data}} \) is the number of

<table>
<thead>
<tr>
<th>Region</th>
<th>Change relative to SR selection</th>
<th>Purity [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRZll</td>
<td>Two same-flavor opposite-charge light leptons with ( 81 &lt; m_{\ell\ell} &lt; 101 \text{ GeV} )</td>
<td>98 98</td>
</tr>
<tr>
<td>CRW</td>
<td>( m_T(\ell, E_T^{\text{miss}}) &gt; 40 \text{ GeV} ) and ( m_T(\tau_{\text{had-vis}}, E_T^{\text{miss}}) &gt; 35(30) \text{ GeV} ) in ( \eta ) (( \mu \rho )) events</td>
<td>84 85</td>
</tr>
<tr>
<td>CRT</td>
<td>( N_{p\text{-jets}} \geq 2 )</td>
<td>98 98</td>
</tr>
<tr>
<td>CRQ</td>
<td>Inverted light-lepton isolation</td>
<td>75 37</td>
</tr>
</tbody>
</table>

The factors measured in the SR for events with low \( \tau_{\text{had-vis}} \) \( p_T \) and high \( \tau_{\text{had-vis}} \) track \( p_T \), the fake factors are large and have large statistical uncertainties because there are few events in the calibration regions. However, these fake factors are applied only to a small fraction of events in the sidebands.

The number of events from fakes in the SR is

\[
N_{\text{fake}} = \sum_k (N_{\text{fail}}^{\text{SR, data}} - N_{\text{fail}}^{\text{SR, MC, not jet-\tau}}) \times F_k,
\]

where \( F_k \) is the fake factor corresponding to the \( p_T \) (and track \( p_T \) for 1-prong \( \tau_{\text{had-vis}} \)) bin \( k \), \( N_{\text{fail}}^{\text{SR, data}} \) is the number of events from fakes in the SR.

<table>
<thead>
<tr>
<th>( \tau_{\text{had-vis}} ) ( p_T )</th>
<th>( \eta ) events</th>
<th>( \mu \rho ) events</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_{\text{had-vis}} ) track ( p_T )</td>
<td>20–30 GeV</td>
<td>30–40 GeV</td>
</tr>
<tr>
<td>1–15 GeV</td>
<td>0.29 ± 0.02</td>
<td>0.32 ± 0.04</td>
</tr>
<tr>
<td>15–20 GeV</td>
<td>0.54 ± 0.06</td>
<td>0.46 ± 0.07</td>
</tr>
<tr>
<td>20–60 GeV</td>
<td>1.34 ± 0.18</td>
<td>0.80 ± 0.15</td>
</tr>
<tr>
<td>&gt;60 GeV</td>
<td>1.0 ± 1.0</td>
<td>2.6^{+5.3}_{-2.6}</td>
</tr>
<tr>
<td>( \tau_{\text{had-vis}} ) ( p_T )</td>
<td>20–30 GeV</td>
<td>30–40 GeV</td>
</tr>
<tr>
<td>20–30 GeV</td>
<td>0.21 ± 0.01</td>
<td>0.22 ± 0.02</td>
</tr>
<tr>
<td>30–40 GeV</td>
<td>0.20 ± 0.03</td>
<td>0.24 ± 0.06</td>
</tr>
<tr>
<td>&gt;40 GeV</td>
<td>0.20 ± 0.03</td>
<td>0.24 ± 0.06</td>
</tr>
</tbody>
</table>
data events in the fail sideband in bin $k$, and $N_{\text{fail}, \text{MC, not jet} \rightarrow \tau}$ is the number of events in the fail sideband in bin $k$ for which the $r_{\text{had-vis}}$ candidate did not originate from a jet as predicted by simulation.

The sources of uncertainty in the estimate of the fake background are the statistical uncertainties in the $F$ measurements in each bin, the statistical uncertainties of the data in the fail sideband and the uncertainty in $R_i$. All statistical uncertainties are treated as independent. The uncertainty in $R_i$ is estimated by varying the estimated $R_W$ by 50%, although this has a negligible impact on the sensitivity.

The simulation and the data-driven techniques used to model the signal and background processes were validated although this has a negligible impact on the sensitivity. The systematic uncertainties in

$$
\text{FIG. 3. Expected uncertainties in the total background predictions in the SR as a function of the combined NN output for the dominant systematic uncertainties in } e\tau \text{ (top) and } \mu\tau \text{ (bottom) channels with 1-prong (left) and 3-prong (right) } r_{\text{had-vis}} \text{ candidates. The uncertainties in the normalizations of the } Z \text{ and fake components are based on the expected statistical power of the fit described in Sec. VII. "Muon efficiency statistics" refers to the statistical uncertainty of the corrections applied to the simulated muon reconstruction efficiency [14]. "Tau energy scale in situ" refers to the uncertainty of the corrections applied to the energy of the } r_{\text{had-vis}} \text{ candidate based on measurements with } Z \rightarrow \tau\tau \text{ data [24].}
$$
efficiencies and the energy calibration. These are applied only to $\tau_{\text{had-vis}}$ candidates from hadronically decaying $\tau$ leptons. For misidentified $\tau_{\text{had-vis}}$ candidates originating from an electron or a muon, systematic uncertainties in the misidentification rates are assigned using a data-driven method, as detailed in Sec. V. For the simulation of electron and muon candidates, uncertainties in the trigger, reconstruction, identification and isolation efficiencies are accounted for. The effect of uncertainties in the light-lepton momentum scale and resolution is also evaluated. For jets, uncertainties in the jet momentum scale and resolution, as well as in the $b$-tagging (in)efficiencies are accounted for. All experimental uncertainties are propagated to the $E_T^{\text{miss}}$ calculation. In addition, uncertainties in the energy scale and resolution of the $E_T^{\text{miss}}$ soft term are considered.

The 2.1% uncertainty in the measured luminosity (Sec. III) is only considered for the simulated $W$ + jets, top, diboson and Higgs boson contributions, whose normalizations are based purely on simulation, without any data-driven estimate.

Data-driven techniques are used to estimate the background contributions from events with a $\tau_{\text{had-vis}}$ candidate originating from either a light lepton or a quark- or gluon-initiated jet. The systematic uncertainties in these methods are described in Sec. V.

To illustrate the sizes of the systematic uncertainties, Fig. 3 shows the relative uncertainties of the total background predictions as a function of the combined NN output for the dominant systematic uncertainties. The uncertainties in the normalizations of the $Z$ and fake components, estimated from the expected statistical power of the fit described in Sec. VII, and the statistical uncertainty in the fake factor are the largest sources of systematic uncertainty, contributing on average between 3% and 6%. The systematic uncertainty in $R_W$ is also relevant and ranges between 1% and 6% over the different final states. All other systematic uncertainties affect the total background prediction by less than one percent.

**VII. RESULTS AND STATISTICAL INTERPRETATION**

A binned maximum-likelihood fit to data, performed with the statistical analysis packages RooFit [51], RooStats [52] and HistFitter [53], is used to compare the observed binned distributions of the combined NN classifiers in the SR with the model, and to extract evidence of signal events. The parameter of interest in such fit is the signal strength modifier $\mu$ which quantifies the size of the LFV decay branching fraction $B(\tau \rightarrow \ell \nu)$.

Fits are performed independently for the $\tau \tau$ and $\ell \ell$ channels, and in each fit events with a 1-prong $\tau_{\text{had-vis}}$ candidate and those with a 3-prong candidate are considered separately. In the fits of events with 1-prong $\tau_{\text{had-vis}}$ candidates, because of the way the NN classifiers are combined, only a few background-like events have an NN output value below 0.15; these are excluded. Independent templates, estimated as described in previous sections, are used for signal, $Z \rightarrow \tau \tau$, fakes, $Z \rightarrow \ell \ell$, top, and $W(\rightarrow \nu \nu) + \text{jets}$ events. The small contributions from Higgs boson and diboson events are summed into a single template, referred to as “Other.”

The likelihood is the product of Poisson probability density functions describing the observed number of events in each bin. It also includes Gaussian, Poisson and log-normal distributions to constrain the nuisance parameters associated with the systematic, statistical and theoretical uncertainties in the predicted number of events, respectively. In addition to the parameter of interest and the nuisance parameters, three normalization parameters are included: $\mu(Z)$ determines the normalizations of the $Z \rightarrow \tau \tau$ and signal events while $\mu(\text{fakes1P})$ and $\mu(\text{fakes3P})$ control the normalization of the fake component in events with a 1-prong or a 3-prong $\tau_{\text{had-vis}}$ candidate, respectively. These parameters are fit independently in the $\tau \tau$ and $\ell \ell$ channels. Within the same channel, the same $\mu(Z)$ is used to fit events with 1-prong and 3-prong $\tau_{\text{had-vis}}$ candidates, while $\mu(\text{fakes1P})$ and $\mu(\text{fakes3P})$ are used to fit independently the corresponding contributions from fakes. By fitting the overall normalizations of the $Z \rightarrow \tau \tau$ and signal event yields, the $\mu(Z)$ parameter accounts for uncertainties in these contributions due to theoretical uncertainties on the $Z$ production cross section.

**TABLE V.** The total observed number of events and postfit event yields in the SR for the $\tau \tau$ (top) and $\ell \ell$ (bottom) channels after a fit to data. The uncertainties include both the statistical and systematic contributions. The correlations between the uncertainties in individual contributions are accounted for in the quoted uncertainties in the total postfit event yields.

<table>
<thead>
<tr>
<th></th>
<th>1-prong</th>
<th>3-prong</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total observed $\tau \tau$ events</strong></td>
<td>89,294</td>
<td>35,148</td>
</tr>
<tr>
<td><strong>Total postfit $\tau \tau$ event yield</strong></td>
<td>$89,300 \pm 300$</td>
<td>$35,200 \pm 200$</td>
</tr>
<tr>
<td>Fakes</td>
<td>$57,000 \pm 1000$</td>
<td>$21,500 \pm 500$</td>
</tr>
<tr>
<td>$Z \rightarrow \tau \tau$</td>
<td>$26,000 \pm 1000$</td>
<td>$11,500 \pm 500$</td>
</tr>
<tr>
<td>$Z \rightarrow \ell \ell$</td>
<td>$3200 \pm 100$</td>
<td>$250 \pm 150$</td>
</tr>
<tr>
<td>Top</td>
<td>$770 \pm 120$</td>
<td>$440 \pm 70$</td>
</tr>
<tr>
<td>$W + \text{jets}$</td>
<td>$540 \pm 100$</td>
<td>$950 \pm 180$</td>
</tr>
<tr>
<td>Other</td>
<td>$340 \pm 70$</td>
<td>$150 \pm 30$</td>
</tr>
<tr>
<td><strong>Total observed $\ell \ell$ signal</strong></td>
<td>$900 \pm 400$</td>
<td>$390 \pm 160$</td>
</tr>
<tr>
<td><strong>Total observed $\ell \ell$ events</strong></td>
<td>79,744</td>
<td>25,050</td>
</tr>
<tr>
<td><strong>Total postfit $\ell \ell$ event yield</strong></td>
<td>$79,700 \pm 500$</td>
<td>$25,100 \pm 700$</td>
</tr>
<tr>
<td>Fakes</td>
<td>$52,000 \pm 1000$</td>
<td>$13,600 \pm 800$</td>
</tr>
<tr>
<td>$Z \rightarrow \tau \tau$</td>
<td>$26,000 \pm 1000$</td>
<td>$10,300 \pm 300$</td>
</tr>
<tr>
<td>$Z \rightarrow \ell \ell$</td>
<td>$240 \pm 110$</td>
<td>$80 \pm 40$</td>
</tr>
<tr>
<td>Top</td>
<td>$890 \pm 140$</td>
<td>$360 \pm 60$</td>
</tr>
<tr>
<td>$W + \text{jets}$</td>
<td>$610 \pm 120$</td>
<td>$680 \pm 130$</td>
</tr>
<tr>
<td>Other</td>
<td>$290 \pm 70$</td>
<td>$110 \pm 20$</td>
</tr>
<tr>
<td><strong>$Z \rightarrow \mu \tau$ signal</strong></td>
<td>$-20 \pm 360$</td>
<td>$-10 \pm 140$</td>
</tr>
</tbody>
</table>
as well as experimental uncertainties in the measurement of the integrated luminosity and in the acceptance times efficiency of the $\ell\tau$ final state (uncertainties due to trigger, reconstruction, isolation and identification efficiencies). Therefore, $\mu(Z)$ ensures that the same $Z$ production cross section and the same $\ell\tau$ acceptance efficiencies are used in the predictions of the signal and the $Z \to \tau\tau$ background contribution. The normalization with $\mu(Z) = 1$ corresponds to the $Z$ production cross section of 2.1 nb, the SM value calculated at NNLO accuracy, multiplied by the nominal detector acceptances and the measured integrated luminosity. The normalization parameters $\mu(\text{fakes}1P)$ and $\mu(\text{fakes}3P)$ account for the systematic uncertainties in the overall normalizations of the fake contributions, so that

FIG. 4. Observed and expected postfit distributions of the combined NN output in SR for the $e\tau$ (top) and $\mu\tau$ (bottom) channels, for 1-prong (left) and 3-prong (right) $\tau_{\text{had-vis}}$ candidates. The filled histogram stacked on top of the backgrounds represents the signal normalized to the best-fit $B(Z \to \ell\tau)$. The overlaid dashed line represents the expected distribution for the signal normalized to $B(Z \to \ell\tau) = 10^{-3}$. In the panels below each plot, the ratios of the observed data (dots) and the postfit background plus signal (solid line) to the postfit background are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The first and last bins include underflow and overflow events, respectively.

092010-11
The uncertainties include both the statistical and systematic contributions. The fit is based on profiling on the combined NN classifier, but not directly on these variables. The filled histogram stacked on top of the observed (expected) upper limit at 95% C.L.

<table>
<thead>
<tr>
<th>$B(Z \to \ell\tau)$</th>
<th>$\epsilon\tau$</th>
<th>$\mu\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(3.3^{+1.5}_{-0.9}) \times 10^{-5}$</td>
<td>$(-0.1^{+1.2}_{-0.9}) \times 10^{-5}$</td>
</tr>
<tr>
<td>$\mu(Z)$</td>
<td>$0.83^{+0.07}_{-0.09}$</td>
<td>$0.87^{+0.09}_{-0.08}$</td>
</tr>
<tr>
<td>$\mu(\text{fakes1P})$</td>
<td>$1.18^{+0.06}_{-0.05}$</td>
<td>$1.12^{+0.09}_{-0.13}$</td>
</tr>
<tr>
<td>$\mu(\text{fakes3P})$</td>
<td>$1.01^{+0.06}_{-0.05}$</td>
<td>$1.09^{+0.13}_{-0.14}$</td>
</tr>
</tbody>
</table>

Table V reports the total observed number of events and postfit event yields in the SR after a fit to data. The observed and postfit expected distributions of the combined NN output are shown in Fig. 4. As reported in Table VI, the best-fit values for $\mu(Z)$, $\mu(\text{fakes1P})$ and $\mu(\text{fakes3P})$ are consistent between the $\epsilon\tau$ and $\mu\tau$ channels, while the best-fit value for $B(Z \to \ell\tau)$ is consistent with zero in the $\mu\tau$ channel with 1-prong $\tau_{\text{had-veto}}$ candidates. The fit is based on profiling the combined NN classifier, but not directly on these variables. The filled histogram stacked on top of the backgrounds represents the signal normalized to the best-fit $B(Z \to \ell\tau)$. The overlaid dashed line represents the expected distribution for the signal normalized to $B(Z \to \ell\tau) = 10^{-3}$. In the panels below each plot, the ratios of the observed data (dots) and the postfit background plus signal (solid line) to the postfit background are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The first and last bins include underflow and overflow events, respectively.
channel, \( B(Z \rightarrow \mu\tau) = (0.1^{+1.3}_{-1.5}) \times 10^{-5} \), and slightly deviates from zero in the \( e\tau \) channel, \( B(Z \rightarrow e\tau) = (3.3^{+1.9}_{-1.5}) \times 10^{-5} \).

Observed and expected postfit distributions of the unscaled NN inputs of the events in the SR are shown in Figs. 5–8. The postfit distributions are compatible with data. An alternative fit combining the \( e\tau \) and \( \mu\tau \) channels with two independent parameters of interest and the same shared free parameter \( \mu(Z) \) yielded the same results for the signal branching fractions and for the background normalization parameters as in the nominal fit. The compatibility of the normalizations of the \( Z \rightarrow \tau\tau \) background in events with 1-prong or 3-prong \( \tau_{\text{had-vis}} \) candidates was also tested by fitting these two contributions independently. The best-fit values for all the normalization parameters were compatible in less than 1σ with the values obtained in the nominal fit and no significant differences were observed in the upper limits on the signal branching ratios.

After the fit, the probabilities of compatibility between the data and the background-plus-signal and background-only hypotheses are assessed using the profile log-likelihood ratio method [54], where the nuisance parameters are profiled as a function of the parameter of interest. The normalization parameters are not profiled. As no significant deviation from the background-only hypothesis is observed, exclusion upper limits on \( B(Z \rightarrow e\tau) \) are set using the CLs method [55]. The resulting observed (expected) upper limits at 95% C.L. are \( B(Z \rightarrow e\tau) < 5.8 \times 10^{-5} \) (2.8 \times 10^{-5} \) and \( B(Z \rightarrow \mu\tau) < 2.4 \times 10^{-5} \) (2.4 \times 10^{-5} \). The significance of the excess in the \( e\tau \) channel is 2.3σ.

![Graphs](image.png)

**FIG. 6.** Observed and expected postfit distributions of unscaled NN inputs in SR for the \( e\tau \) channel with 3-prong \( \tau_{\text{had-vis}} \) candidates. The fit is based on profiling on the combined NN classifier, but not directly on these variables. The filled histogram stacked on top of the backgrounds represents the signal normalized to the best-fit \( B(Z \rightarrow e\tau) \). The overlaid dashed line represents the expected distribution for the signal normalized to \( B(Z \rightarrow e\tau) = 10^{-3} \). In the panels below each plot, the ratios of the observed data (dots) and the postfit background plus signal (solid line) to the postfit background are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The first and last bins include underflow and overflow events, respectively.
and was normalized to data. The reducible background was estimated using so-called embedded events \( \rightarrow Z \) This is a likelihood-based mass estimator optimized for as reconstructed by using the missing mass calculator [56].

The result of the search for \( Z \rightarrow \mu\tau \) decays presented here is combined with the result published by ATLAS with 20.3 fb\(^{-1}\) of data at a center-of-mass energy of \( \sqrt{s} = 8 \text{ TeV} \) [7]. In this previous analysis, a 95% C.L. upper limit was set at \( B(Z \rightarrow \mu\tau) < 1.7 \times 10^{-5} \). The expected upper limit was \( 2.6 \times 10^{-5} \).

The analysis of the 8 TeV data was based on a template fit to the observed distributions in data of the \( m_{\text{MMC}} \) mass, as reconstructed by using the missing mass calculator [56]. This is a likelihood-based mass estimator optimized for \( Z \rightarrow \tau\tau \) events. The dominant irreducible \( Z \rightarrow \tau\tau \) background was estimated using so-called embedded events [57] and was normalized to data. The reducible background of events with \( \tau_{\text{had-vis}} \) candidates originating from misidentified jets was also estimated from data using events with \( \mu\tau \) pairs with the same electric charges. The other smaller background contributions were estimated from simulation. The \( Z \rightarrow \mu\tau \) signal was simulated and was normalized using the predicted \( Z \) production cross section at 8 TeV.

The 8 TeV and 13 TeV analyses are combined using the same parameter of interest, but assuming no other correlation. Indeed, the estimates of the two dominant sources of background, \( Z \rightarrow \tau\tau \) and fakes, are based on different data and different methods. The signal predictions are also uncorrelated since the same parameter of interest, but assuming no other correlation.
on different data. If these modeling uncertainties are set to zero, the combined upper limit changes by only 3%. This 3% represents an upper bound on how much the combined limit can change if different assumptions are made about correlations in systematic uncertainties related to detector modeling.

The combined best-fit value of $\mathcal{B}(Z \to \mu\tau)$ is $(-0.8^{+0.9}_{-0.8}) \times 10^{-5}$ and the combined observed (expected) 95% C.L. upper limit is $\mathcal{B}(Z \to \mu\tau) < 1.3 \ (1.8) \times 10^{-5}$.

VIII. CONCLUSIONS

Direct searches for lepton flavor violation in decays of the $Z$ boson are performed using a data sample of proton-proton collisions recorded by the ATLAS detector at the LHC corresponding to an integrated luminosity of 36.1 fb$^{-1}$ at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The analysis selects events consistent with the decay of a $Z$ boson into an electron or muon and a hadronically decaying $\tau$ lepton. In these decays the $\tau$ lepton is assumed to be unpolarized. Neural network classifiers are used to discriminate signal from backgrounds, and the NN output distributions are analyzed in a template fit to data.

No significant excess of events above the expected background is observed and upper limits on the lepton-flavor-violating branching ratios are set at the 95% confidence level using the CL$_s$ method: $\mathcal{B}(Z \to \mu\tau) < 2.4 \times 10^{-5}$ and $\mathcal{B}(Z \to e\tau) < 5.8 \times 10^{-5}$. The corresponding expected upper limits are $2.4 \times 10^{-5}$ and $2.8 \times 10^{-5}$, respectively. An excess of data over the expected backgrounds is
observed in the $e\tau$ final state with a significance of $2.3\sigma$.

No upper limits on $\mathcal{B}(Z \to e\tau)$ from ATLAS data have been published previously. The current best upper limit is from LEP at $\mathcal{B}(Z \to e\tau) < 0.98 \times 10^{-5}$.

The result on $\mathcal{B}(Z \to \mu\tau)$ presented here is combined with the previous ATLAS result based on 20.3 fb$^{-1}$ of data at a center-of-mass energy of $\sqrt{s} = 8$ TeV. The combined 95% C.L. upper limit is $\mathcal{B}(Z \to \mu\tau) < 1.3 \times 10^{-5}$, to be compared with LEP upper limit of $\mathcal{B}(Z \to \mu\tau) < 1.2 \times 10^{-5}$.

ACKNOWLEDGMENTS

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently. We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; STFC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DRF/IRFU, France; SRNSF, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, the Canada Council, CANARIE, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, ERDF, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BMBF, Gif and Minerva, Israel; BRF, Norway; CERCA Programme Generalitat de Catalunya, Generalitat Valenciana, Spain; the Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/ GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [58].
