Beyond Vanilla

a search for lepton-flavour-violating
$Z \to \ell\tau$ decays with the ATLAS detector

Wing Sheung Chan

Beyond Vanilla

a search for lepton-flavour-violating $Z \to \ell\tau$ decays
with the ATLAS detector

Wing Sheung Chan

ISBN: 978-94-6416-448-0

Beyond Vanilla – a search for lepton-flavour-violating $Z \to \ell\tau$ decays with the ATLAS detector

Thesis, Radboud Universiteit Nijmegen

Cover design: Mimi Szeto – *www.szetomimi.com / maniya852@gmail.com*

Printing: Ridderprint, the Netherlands

This work has been performed at the National Institute for Subatomic Physics (Nikhef) which is funded by the Dutch Research Council (NWO). The research was financially supported by the NWO Innovational Research Incentives Scheme (Vici).

Beyond Vanilla

a search for lepton-flavour-violating $Z \to \ell\tau$ decays
with the ATLAS detector

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen in het openbaar te verdedigen
op maandag 22 maart 2021
om 12:30 uur precies

door

Wing Sheung Chan

geboren op 13 augustus 1992
te Hong Kong

Promotoren:	Prof. dr. N. de Groot Prof. dr. S. Xella (Københavns Universitet, Denemarken)
Manuscriptcommissie:	Prof. dr. R.H.P. Kleiss Prof. dr. S.C.M. Bentvelsen (Universiteit van Amsterdam) Prof. dr. R.J.M. Snellings (Universiteit Utrecht) Dr. S. Caron Dr. M. Wu

To Mum and Dad,
without whom none of my success would be possible

To every Hongkonger,
who has stood and fought for what is right
while I was writing this thesis far away from my homeland

In memory of Olga Igonkina,
who inspired this thesis and whose legacy will live on
with the people and research she loved

I am not plain, or average or – God forbid – vanilla.
I am peanut butter rocky road with multicolored sprinkles,
hot fudge and a cherry on top.

— Wendy Mass, *Every Soul a Star*

Contents

Introduction

"You see things; and you say 'Why?' But I dream things that never were; and I say 'Why not?'"

— George B. Shaw, "the Serpent", *Back to Methuselah*

Physics is the study of Nature. Its main goal is to understand the universe and the world around us. As such, the ultimate destination of physics (whether or not reachable) is to obtain a self-consistent theory that accurately describes everything that exists/happens, existed/happened, or will exist/will happen in the universe. Such a hypothetical theory is commonly referred to as a theory of everything (TOE). It is without a doubt still a distant goal. Nonetheless, physicists have made remarkable progress in the past centuries. In particular, they brought us the two most important theoretical frameworks in modern physics, namely the general theory of relativity (or simply general relativity, GR) and the Standard Model of particle physics (or simply the Standard Model, SM). Together they are the closest thing we have to a TOE. Over the past century, both theories have been put under the scrutiny of countless experiments, none of which was able to refute the theories definitely. Yet, despite the enormous success, the facts that GR and the SM are incompatible with each other and that there are observations that could not be explained by the theories imply that neither of them is a complete theory in explaining Nature. Thus, understanding what is wrong or missing in these theories has become a great challenge for physicists today. Luckily, we are not searching in complete darkness. Driven and constrained by high-energy experiments, precision measurements and cosmological observations, theorists are able to conjecture beyond-the-Standard-Model (BSM) theories that better explain the universe. There are many of these theories and they are all waiting to be tested by all the ever improving, limit-pushing physics experiments.

Among others, (charged) lepton flavour violation is one of the most promising and plausible BSM phenomena. Leptons are a class of elementary particles in the SM that come in three generations, or so-called flavours. According to the SM in its current formulation, the number of charged leptons in each flavour does not change in any physical process[†]. Violation of this rule is known as lepton flavour violation. However, such a rule

[†] Unless neutrino mixing is considered. Nonetheless, violations of charged lepton flavour in point interactions due to neutrino mixing have only vanishingly small probabilities, and are negligible when actual observations are concerned. There are also ongoing debates on whether and how neutrino mixing should be considered as part of the SM. More will be discussed in the upcoming chapter.

is purely based on empirical evidence and simplicity of the model, but lacks fundamental motivations. Therefore, if observed, lepton flavour violation would be an unequivocal evidence of BSM physics, and could point us in the right direction in identifying possible "loopholes" in the SM.

In particular, the decay of a Z boson into an electron or muon and a τ lepton ($Z \to \ell\tau$) is an interesting signal of lepton flavour violation. Searches for $Z \to \ell\tau$ decays have been performed using data collected from the Large Electron-Positron Collider (LEP), and stringent limits on the probability of such decays have been set. However, new opportunities have opened up as the Large Hadron Collider (LHC) and the ATLAS detector are collecting more and more data, allowing the search for $Z \to \ell\tau$ decays to reach an unprecedented sensitivity. Currently, with the data collected by the ATLAS detector and through careful analysis, we are able to surpass the sensitivity of the LEP experiments for the first time after more than two decades since the last $Z \to \ell\tau$ search result from the LEP experiments was published. This marks the beginning of a new era for lepton flavour violation searches. This thesis is a documentation of this exciting work.

The thesis is divided into six chapters:

Chapter 1 gives a brief introduction to the Standard Model, with a focus on parts that are most relevant to this thesis. An introduction to a handful of selected BSM theories related to lepton flavour violation is also given, followed by a summary of the current experimental status of lepton flavour violation searches. At last, we discuss the motivation for the chosen search channel, $Z \to \ell\tau$.

Chapter 2 provides an overview of the LHC and the ATLAS detector.

Chapter 3 outlines the algorithms used to reconstruct and identify physics objects from the data collected by the ATLAS detector. A focus is given to the reconstruction and identification of hadronic τ decays, which are especially important to the presented analysis, and are work to which the author has made important contributions.

Chapter 4 describes how observed and simulated proton–proton collision events are selected and classified in the search for $Z \to \ell\tau$ decays. It also documents the training and usage of neural network classifiers for signal and background classification.

Chapter 5 details the methods that are used to model the signal and background events, which generate predictions that can be compared with observations from data.

Chapter 6 presents the statistical analysis method and the final results of the analysis.

These are followed by a conclusion and outlook.

The analysis described in Chapters 4–6 is the author's own, original work. The results have also been published as Reference [1].

Chapter 1.

The Standard Model and lepton flavour violation

> *"This isn't the kind of story where understanding makes you smart, or not understanding makes you dumb."*
>
> — CLAMP, "Yūko Ichihara", *xxxHOLiC*

Before we dive into the technical details of the search for lepton-flavour-violating $Z \to \ell\tau$ decays, let us first motivate ourselves by reviewing things that "we know we know" – the Standard Model, that "we know we don't know" – unsolved mysteries of the Standard Model, and that "we don't know if we know" – beyond-the-Standard-Model theories, in relation to lepton flavour violation.

In this chapter, the Standard Model will first be introduced in a conceptual and intuitive way and with a focus on the electroweak sector. We shall then discuss flavour violation in the quark and neutrino sectors of the SM, and after that, motivate ourselves to search for charged lepton flavour violation. Finally, we will discuss what makes the $Z \to \ell\tau$ decay specifically the chosen search channel for this thesis.

1.1. Units and conventions

Before the chapter actually begins, let us make some quick remarks on the units and conventions to avoid possible confusions.

Throughout this thesis, we will be using a conventional system of natural units, one that is commonly adopted in particle physics, to simplify expressions. In this system, velocities are expressed in units of the speed of light

$$c = 299\,792\,458\,\mathrm{m\,s^{-1}} \tag{1.1}$$

and actions are expressed in units of the reduced Planck constant

$$\begin{aligned}\hbar &= \frac{6.626\,070\,15 \times 10^{-34}\,\mathrm{J\,s}}{2\pi} \\ &\approx 1.055 \times 10^{-34}\,\mathrm{Js} \\ &\approx 6.582 \times 10^{-16}\,\mathrm{eV\,s}.\end{aligned} \tag{1.2}$$

Consequently, length, time, mass, energy and momentum all share a same common unit[†]. Furthermore, the unit of electric charge is chosen such that the vacuum permittivity $\varepsilon_0 = 1$. This simplifies the expression of the electromagnetic coupling to be the same as the positron charge e.

In particle physics, the common unit of mass, energy and momentum is often chosen to be electronvolt (eV) or decadic multiples of it, while for practical reasons, lengths and times in collider experiments are often still expressed in metres and seconds. Also, it might be worth noting that it is customary to express scattering cross sections (luminosities) in decadic multiples of barn (inverse barn), where barn is a unit of area defined as

$$1\,\mathrm{b} = 10^{-28}\,\mathrm{m}^2. \tag{1.3}$$

Einstein notation will be used throughout this chapter, implying summation over index variables that appear twice in a single term. Upper indices represent components of contravariant vectors, and lower indices represent components of covariant vectors. In the case of spacetime four-vectors, indices are raised or lowered by the Minkowski metric tensor with the $(+\,-\,-\,-)$ signature

$$g_{\mu\nu} = g^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{1.4}$$

Spacetime four-vectors are always represented by notations in italic face, while their spatial components are represented by the same notations in bold face.

1.2. The Standard Model

The Standard Model [2–5] was developed over decades in the last century by many physicists based on previous discoveries and understandings. Ever since it was forged into its current form in the 1970s, the model has successfully explained and predicted many phenomena and has withstood myriad experiments to high precision and accuracy. Today, even after half a century since its birth, it remains the best description of Nature at small length scales and the foundation of particle physics.

[†] Technically, the unit of length and time is the inverse of the unit of mass, energy and momentum.

In this section, the Standard Model will be introduced. For the benefits of some readers, we will begin with a very brief introduction to the basic concepts of quantum field theory, upon which the foundation of the SM is built. After that, we will go through the particle contents of the SM and describe their interactions, with a focus on the electroweak sector as it is most relevant to this thesis.

1.2.1. Quantum field theory

In quantum field theory, analogous to classical field theory, the evolution of a field can be determined by the Lagrangian density $\mathscr{L}$ (or simply Lagrangian for short). Consider a classical free scalar field $\phi(\mathbf{x}, t)$ as an example, the Lagrangian density can be written as

$$\mathscr{L} = \frac{1}{2} \left(\partial_\mu \phi\right) \left(\partial^\mu \phi\right) - \frac{1}{2} m^2 \phi^2, \tag{1.5}$$

where m is a real constant. The equation of motion for the field can then be derived using the principle of least action by requiring vanishing variation of the action S:

$$\delta S = \delta \int \mathscr{L} \, \mathrm{d}^3\mathbf{x} \, \mathrm{d}t = 0. \tag{1.6}$$

In our example, this gives rise to the famous Klein-Gordon equation:

$$\partial_t \phi - \nabla^2 \phi + m^2 \phi = 0. \tag{1.7}$$

The formulation of QFT is similar to its classical counterpart but differs in two main ways, each of which has important implication that makes quantum fields fundamentally different from classical fields. The first difference is that fields in QFT are promoted into field operators in a process known as canonical quantisation or second quantisation. For our free scalar field example, the field operator would be

$$\phi(x) = \int \frac{1}{\sqrt{2E}} \left(a_{\mathbf{p}} e^{-ip_\mu x^\mu} + a^\dagger_{\mathbf{p}} e^{ip_\mu x^\mu}\right) \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3}, \tag{1.8}$$

where x and p are position and momentum four-vectors, $E = \sqrt{|\mathbf{p}|^2 + m^2}$ is the time-component of p, and $a^\dagger_{\mathbf{p}}$ and $a_{\mathbf{p}}$ are the creation and annihilation operators for momentum $\mathbf{p}$, respectively. This implies that there exists a vacuum state $|0\rangle$ where $a_{\mathbf{p}} |0\rangle = 0$ for all $\mathbf{p}$, and that there can be an arbitrary number of particles, which can be interpreted as excitations of the field, at any given time. This contrasts classical theories, where "vacuum" is interpreted as the state in which all components of a field are zero, and the number of particles is always conserved.

The second difference is that evolutions, or paths, that do not correspond to the least action can also contribute to the transition amplitude between two states at two given times. This can be formulated using Feynman path integrals, where the overall transition

amplitude is the sum over e^{iS} for every possible path. The principle of least action only applies as a limiting case, in which the classical results can be reproduced.

1.2.2. Elementary particles and their interactions

The complete Standard Model Lagrangian (before electroweak symmetry breaking) can be expressed as

$$\begin{aligned}\mathscr{L}_{\mathrm{SM}} = &-\frac{1}{4}G^{\alpha}_{\mu\nu}G^{\mu\nu}_{\alpha} - \frac{1}{4}\mathbf{W}_{\boldsymbol{\mu\nu}}\cdot\mathbf{W}^{\boldsymbol{\mu\nu}} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ &+ i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - \frac{1}{2}\bar{\psi}_{q}\gamma^{\mu}\left(g_{\mathrm{s}}\lambda_{\alpha}G^{\alpha}_{\mu}\right)\psi_{\mathrm{q}} \\ &- \bar{\Psi}_{\mathrm{L}}\gamma^{\mu}\left(g\mathbf{T}\cdot\mathbf{W}_{\boldsymbol{\mu}} + \frac{1}{2}g'YB_{\mu}\right)\Psi_{\mathrm{L}} - \frac{1}{2}\bar{\psi}_{\mathrm{R}}\gamma^{\mu}\left(g'YB_{\mu}\right)\psi_{\mathrm{R}} \\ &+ \frac{1}{2}\left|\left(i\partial_{\mu} - \frac{1}{2}g\mathbf{T}\cdot\mathbf{W}_{\boldsymbol{\mu}} - \frac{1}{2}g'YB_{\mu}\right)\phi\right|^{2} - V\left(\phi\right) \\ &+ \frac{1}{2}\left(y\bar{\Psi}_{\mathrm{L}}\phi\psi_{\mathrm{R}} + \mathrm{h.c.}\right),\end{aligned} \tag{1.9}$$

where

G	are the gluon gauge fields,
$\mathbf{W}$	are the weak isospin gauge fields,
B	is the weak hypercharge gauge field,
$\psi_{(\mathrm{R}/q)}$	are the (right-handed/quark) fermion fields,
Ψ_{L}	are the left-handed doublet fermion fields,
ϕ	is the Higgs field,
g_{s}	is the strong coupling,
g	is the weak isospin coupling,
g'	is the weak hypercharge coupling,
y	is the Yukawa coupling,
$\mathbf{T}$	is the weak isospin operator,
Y	is the weak hypercharge operator,
γ	are the Dirac matrices,
λ	are the Gell-Mann matrices,
V	is the Higgs potential, and
h.c.	stands for hermitian conjugate of the previous terms.

The above equation elegantly encapsulates all elementary particles and their interactions known to date. However, it is admittedly obscure and far from easy to understand at first glance. In this and the upcoming sections, we shall try to break down this cryptic

equation into smaller sectors where we list out the field (particle) contents of the model and discuss their properties and interactions in a more accessible and conceptual way.

Overview

The fields that appear in Equation (1.9) can be interpreted as elementary particles in the SM. These particles interact with each other in ways that are completely determined by the Lagrangian. Figure 1.1 summarises the properties of all the elementary particles and the interactions that they participate in. For each of these particles, there also exists an antiparticle that has the same properties but carries opposite charges. Analogous to the periodic table of chemical elements, these elementary particles can be categorised into several classes according to their properties. And just like how patterns in the periodic table have lead to the discovery of the atomic structure, symmetries in the SM that are now considered to be fundamental might one day also lead to a deeper understanding of Nature.

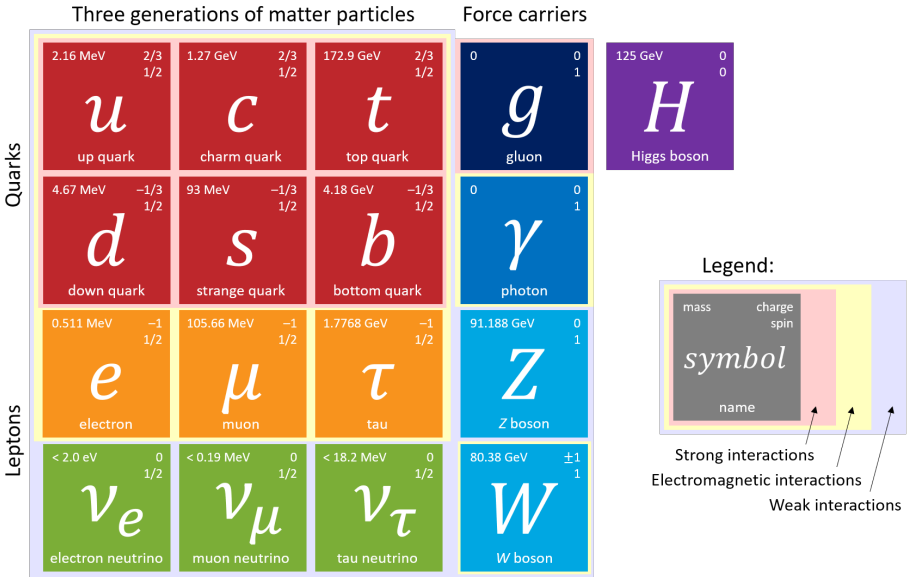


Figure 1.1.: Elementary particles in the Standard Model of particle physics.

Matter particles

In the SM, all known matters are composed of elementary fermions (f). They are the fundamental building blocks of our everyday world and the directly observable universe. Without counting antiparticles and different colour flavours, there are 12 unique elementary

Table 1.2.: Properties of leptons [6]. Limit on the electron mean life is measured at 90% confidence level. Limits on the neutrino rest masses are measured at 95% confidence level and are on the mass expectation values of the weak eigenstates.

Generation	Lepton	Rest mass [MeV]	Mean life [s]	Electric charge [e]
1	e^-	$0.511 \pm 3.1 \times 10^{-9}$	$> 2.1 \times 10^{36}$	-1
	ν_e	$< 2.0 \times 10^{-6}$		0
2	μ^-	$105 \pm 2.4 \times 10^{-6}$	$2.20 \times 10^{-6} \pm 2.2 \times 10^{-12}$	-1
	ν_μ	< 0.19		0
3	τ^-	1776.9 ± 0.12	$(290.3 \pm 0.5) \times 10^{-15}$	-1
	ν_τ	< 18.2		0

fermions in the SM. All of them are spin-$\frac{1}{2}$. They can be further classified into leptons and quarks.

Leptons are fermions that only participate in electroweak interactions but not in strong interactions. There are three generations of leptons, each consist of a charged lepton (ℓ) and a neutrino (ν). Charged leptons carry electric charges. The electron (e^-), the first discovered elementary particle, is the first-generation charged lepton. It is also the only stable charged lepton. The muon (μ^-) and tau (τ^-) are the second- and third-generation charged leptons respectively. In the SM, the charged leptons are distinguished only by their rest masses. Charged leptons of higher generations have higher rest masses. Associated with each charged lepton, there is a neutrino of the same flavour. The neutrinos are named accordingly as the electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). Neutrinos are special in the SM because they are the only fermions that do not carry charges (hence the name) and exist with only one chirality: there are only left-handed neutrinos, but no right-handed ones. Neutrinos only participate in weak (and gravitational) interactions.

A summary of the properties of leptons [6] is given in Table 1.2. The charged leptons, especially the τ lepton, are going to play an important role in this thesis.

Quarks (q) are another class of elementary fermions in the SM. Similar to leptons, there are three generations of quarks in the SM. Each generation consists of an up-type quark that carries an electric charge of $\frac{2}{3}e$ and a down-type quark that carries an electric charge of $-\frac{1}{3}e$. The three up-type quarks are named up (u), charm (c) and top (t), while the three down-type quarks are named down (d), strange (s) and bottom (b). A summary of the properties of quarks [6] is given in Table 1.3.

Quarks differ from leptons in that they carry colour charges. As a result, they do not only participate in electroweak interactions, but also in strong interactions. Quarks can be bound together by strong interactions to form compound particles known as hadrons.

Table 1.3.: Properties of quarks [6].

Generation	Quark	Rest mass	Electric charge [e]
1	u	$2.16^{+0.49}_{-0.26}$ MeV	$2/3$
	d	$4.67^{+0.48}_{-0.17}$ MeV	$-1/3$
2	c	1.27 ± 0.02 GeV	$2/3$
	s	93^{+11}_{-5} MeV	$-1/3$
3	t	172.9 ± 0.4 GeV	$2/3$
	b	$4.18^{+0.03}_{-0.02}$ GeV	$-1/3$

Table 1.4.: Properties of some commonly observed hadrons [6, 7]. The limit on the proton mean life is measured at 90% confidence level.

Hadron	Rest mass [MeV]	Mean life [s]	Major decay mode (branching fraction†)
p	$938 \pm 3 \times 10^{-7}$	$> 6.6 \times 10^{36}$	
n	$940 \pm 5 \times 10^{-7}$	879.4 ± 0.6	$pe^-\bar{\nu}_e$ (100.00%)
π^+	$139 \pm 2.4 \times 10^{-4}$	$(2.6033 \pm 0.0005) \times 10^{-8}$	$\mu^+\nu_\mu$ (99.98%)
π^0	$135 \pm 5 \times 10^{-4}$	$(8.52 \pm 0.18) \times 10^{-17}$	$\gamma\gamma$ (98.82%)

Hadrons can be further classified into baryons and mesons depending on the baryon number, which is defined as (number of quarks − number of antiquarks)/3. Baryons have a non-zero baryon number, while mesons have a zero baryon number. Each single quark can carry one of the three colour charges, arbitrarily labelled red, blue and green. A quark system is colour neutral if it has three quarks or three antiquarks that carry all the three different colour charges, or a quark and an antiquark that carry the same colour charge (but with opposite signs), or is a combination of these systems. In fact, an isolated quark has never been observed experimentally, and hadrons are only known to exist in colour-neutral states. This phenomenon is referred to as the colour confinement. Physicists could only rely on the spectroscopy of hadrons to provide evidence for the existence of quarks and indirectly measure their properties. Due to colour confinement, quarks produced in a collider experiment are never detected as isolated particles, but instead bunches of hadrons clustered together. These observed bunches of hadrons are called jets and the process of their production is called hadronisation. The proton (p), neutron (n), charged pion ($\pi^\pm$) and neutral pion (π^0) are some of the most commonly observed hadrons in everyday environment as well as in a collider experiment. A summary of the properties of these hadrons [6, 7] is given in Table 1.4.

† The branching fraction for a decay is the fraction of particles which decay by an individual decay mode with respect to the total number of particles which decay.

Table 1.5.: Properties of gauge bosons [6].

Gauge boson	Rest mass [GeV]	Electric charge [e]	Assicoated interaction
γ	0	0	Electromagnetic
g	0	0	Strong
$W^{\pm}$	80.379 ± 0.012	± 1	Weak (charged-current)
Z	91.1876 ± 0.0021	0	Weak (neutral-current)

Force carriers

Another class of elementary particles in the SM is the gauge bosons. They are the force carriers for the electroweak and strong interactions. Leptons and quarks can interact with each other by exchanging gauge bosons. There are four different gauge bosons in the SM: the photon (γ), the gluon (g), the W and the Z bosons. Some of them have finite rest masses while some of them are massless. All of them have spin one. A summary of their properties [6] are shown in Table 1.5.

Photons are neutral massless gauge bosons that mediate the electromagnetic (EM) force. They interact with any particles that carry electric charges, including charged leptons, quarks and the W bosons. The zero rest mass and the lack of charges of the photon implies that EM interaction is a long-range interaction and is observable at macroscopic levels. The theory that describes the interaction between photons and electrically charged particles is called quantum electrodynamics (QED). The first reasonably complete theory of QED was created by Paul Dirac in 1927 [8]. Since then, QED has demonstrated huge success in explaining and predicting experimental results. One particularly remarkable achievement is the excellent agreement of the experimentally measured value of the electron gyromagnetic ratio with the theoretical value calculated from QED. The relative difference between the values is in the order of merely 10^{-12} [9].

The W and Z bosons are the mediators of the weak interactions. They can interact with any fermions in the Standard Model, including neutrinos, which otherwise do not participate in any other fundamental interactions. Both the W and Z bosons are relatively heavy particles. Their high masses limit the range of the weak interaction, which is consistent with the observed short range of the weak nuclear force. The W boson the only charged gauge boson in the SM. It is responsible for the charged-current interactions, which couple left-handed up-type quarks to left-handed down-type quarks, or left-handed charged leptons to left-handed neutrinos. Right-handed fermions do not interact with the W bosons at all. The Z boson is responsible for the neutral-current interactions. In the SM, neutral-current interactions can only transfer momentum from a particle to another, without changing the flavours and charges of the interacting particles. These interactions will be discussed in more details in the next section when the electroweak theory is introduced.

The gluon is an electrically neutral massless gauge boson that mediates the strong force. Gluons interact only with particles that carry colour charges, which include all the quarks as well as gluons themselves. Gluons and quarks are constituents of hadrons. In such context, they are both referred to as partons. The theory that describes the strong interaction in the Standard Model is called quantum chromodynamics (QCD).

The Higgs boson

The last elementary particle to be introduced is the Higgs boson (H), a neutral spin-0 boson. It is neither a matter particle or a force carrier, and is the only scalar particle in the SM. The existence of the Higgs boson has been conjectured more than half a century ago [10–15]. It was discovered by the ATLAS and CMS experiments in 2012 [16], which completed the search for all elementary particles in the SM. The discovered Higgs boson has a rest mass of $(125.1 \pm 0.2)\,\text{GeV}$ and has been observed to behave as the SM predicted so far. The discovery is a direct evidence of the Higgs mechanism, the mechanism that naturally explains why elementary fermions and the W and Z bosons appear to have rest masses, despite the lack of an explicit mass term in the SM Lagrangian. The Higgs field interacts with the SM fermions via the Yukawa coupling, which is hypothesised to be proportional to the observed fermion masses. At an energy below the electroweak scale ($\approx 250\,\text{GeV}$), the Higgs field acquires a vacuum expectation value, spontaneously breaking the weak isospin and weak hypercharge SU(2)⊗SU(1) symmetry and giving the otherwise massless elementary particles their apparent masses.

In the SM, the Higgs boson mass is a measured parameter. The SM provides no explanation to the observed value. Naturally, one could expect the Higgs boson mass to be comparable to the Planck mass[†] due to quantum corrections. The fact that the observed Higgs boson mass is so much smaller than the Planck mass thus poses a fine-tuning problem, which is known as the hierarchy problem in particle physics.

1.2.3. The electroweak theory

The reader might have noticed that even though we claimed that the photons, the W and the Z bosons are the fundamental force carriers in the SM, there are no corresponding gauge fields explicitly written in Equation (1.9). This is because these physical gauge fields are only results of the spontaneous breaking of the SU(2)⊗SU(1) symmetry. Since it is not directly relevant to this thesis, we will refrain from going into the details of how the symmetry is broken. However, we can show that, given a broken symmetry, the $\mathbf{W}$ and B fields in Equation (1.9) could indeed be recast into the physical photon, $W^{\pm}$ and Z fields. This was first shown by Weinberg, Salam and Glashow in the 1960s [2–4, 17], and is now known as the GWS theory or the electroweak theory.

[†] The Planck mass is a natural unit of mass defined as $\sqrt{\hbar c/G}$, where G is the gravitational constant. It is roughly equal to $1.22 \times 10^{19}\,\text{GeV}$. Physical quantities similar in magnitude to the Planck mass are said to be at the Planck scale.

The interaction terms in the electroweak sector of the SM Lagrangian are

$$\mathscr{L}_{\text{EW,int}} = -\bar{\Psi}_{\text{L}}\gamma^{\mu}\left(g\mathbf{T}\cdot\mathbf{W}_{\boldsymbol{\mu}} + \frac{1}{2}g'YB_{\mu}\right)\Psi_{\text{L}} - \frac{1}{2}\bar{\psi}_{\text{R}}\gamma^{\mu}\left(g'YB_{\mu}\right)\psi_{\text{R}}. \tag{1.10}$$

The capitalised notation for the left-handed fermion fields is to emphasise that they are indeed doublets of up-type and down-type quarks, or neutrinos and charged leptons:

$$\Psi_{\text{L}} = \begin{pmatrix} u_{\text{L}} \\ d'_{\text{L}} \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \nu'_{\text{L}} \\ \ell_{\text{L}} \end{pmatrix}. \tag{1.11}$$

Here, we denoted any up-type and down-type quarks collectively by u and d respectively. The primed symbols indicate that they are the weak interaction eigenstates instead of the mass eigenstates, the meaning of which will be elaborated later in Section 1.2.5.

$W^{\pm}$ bosons and the charged-current interaction

We choose a basis such that the weak isospin operator can be expressed in terms of the Pauli matrices $\boldsymbol{\tau}$: $T_i = \tau_i/2$. Since the first two Pauli matrices,

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{1.12}$$

mix the components of the fermion field doublets, the corresponding W^1 and W^2 fields cannot be physical. However, if we define $\tau_{\pm} = (\tau_1 \pm i\tau_2)/2$ and $W^{\pm} = (W^1 \mp iW^2)/\sqrt{2}$, we can rewrite the relevant terms in the Lagrangian into:

$$\begin{aligned} \mathscr{L}_{\text{charged current}} &= -g\bar{\Psi}_{\text{L}}\gamma^{\mu}\left(T_1W^1_{\mu} + T_2W^2_{\mu}\right)\Psi_{\text{L}} \\ &= -\frac{g}{\sqrt{2}}\bar{\Psi}_{\text{L}}\gamma^{\mu}\left(\tau_+W^+_{\mu} + \tau_-W^-_{\mu}\right)\Psi_{\text{L}} \\ &= -\frac{g}{2\sqrt{2}}\bar{\Psi}\gamma^{\mu}\left(1-\gamma^5\right)\left(\tau_+W^+_{\mu} + \tau_-W^-_{\mu}\right)\Psi. \end{aligned} \tag{1.13}$$

In our representation,

$$\tau_+ = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \tau_- = \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{1.14}$$

In this basis, components of the fermion field doublets are not mixed anymore. It becomes immediately apparent that interactions with the W^+ field changes a down-type quark into an up-type quark, or a charged lepton into a neutrino, raising the electric charge by one in the process. Vice versa, interactions with the W^- field lower the electric charge of a an up-type quark or a neutrino, turning them into a down-type quark or a charged lepton respectively.

Z boson and the neutral-current interaction

The remaining component of the weak isospin gauge field W^3 and the weak hypercharge gauge field B both lead to neutral-current interactions. As a result, the physical neutral-current fields, A_μ and Z_μ, can be a linear combination of the W^3 and B fields:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_\mathrm{w} & \sin\theta_\mathrm{w} \\ -\sin\theta_\mathrm{w} & \cos\theta_\mathrm{w} \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}, \tag{1.15}$$

where the parameter θ_w is known as the weak mixing angle. The Higgs mechanism predicts that if A is to be the massless physical photon field, then θ_w is related to the weak isospin and weak hypercharge coupings by

$$\tan\theta_\mathrm{w} = \frac{g'}{g}, \tag{1.16}$$

and the electromagnetic charge Q is related to the third component of weak isospin and the weak hypercharge by

$$Q = T_3 + \frac{Y}{2}. \tag{1.17}$$

For the Z field, the coupling becomes

$$g_Z = \frac{g}{\cos\theta_\mathrm{w}}. \tag{1.18}$$

The interaction Lagrangian terms in the electroweak sector can now be written as

$$\begin{aligned} \mathscr{L}_\mathrm{EW,int} = &- eQ\bar{\psi}\gamma^\mu A_\mu \psi \\ &- \frac{g}{2\sqrt{2}}\bar{\Psi}\gamma^\mu \left(1-\gamma^5\right)\left(\tau_+ W^+_\mu + \tau_- W^-_\mu\right)\Psi \\ &- \frac{g}{2\cos\theta_\mathrm{w}}\bar{\psi}\gamma^\mu\left(C_\mathrm{V} - C_\mathrm{A}\gamma^5\right)\psi, \end{aligned} \tag{1.19}$$

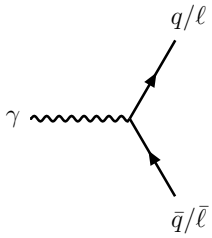
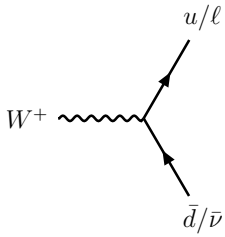
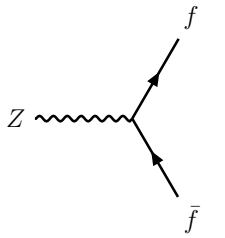
where C_V and C_A are the vector and axial vector couplings related to the weak isospin and the electric charge of the fermions by

$$C_\mathrm{V} = T_3 - 2Q\sin^2\theta_\mathrm{w}, \tag{1.20}$$

$$C_\mathrm{A} = T_3. \tag{1.21}$$

Table 1.6 shows the Feynman rules [18] that represent the possible interactions expressed in Equation (1.19).

Table 1.6.: Possible vertices for electroweak interactions in the Standard Model and their corresponding vertex factors.

Vertex	Vertex factor
	$-ieQ\gamma^\mu$
	$-i\frac{g}{2\sqrt{2}}\gamma^\mu\left(1-\gamma^5\right)$
	$-i\frac{g}{2\cos\theta_\mathrm{w}}\gamma^\mu\left(C_\mathrm{V}-C_\mathrm{A}\gamma^5\right)$

1.2.4. Symmetries and conservations

The Standard Model exhibits several symmetries. According to Noether's theorem [19], every symmetry is associated to a conservation law. These symmetries are the essence of the Standard Model. They are either postulated symmetries that were used to construct and constrain the model in the first place, or are observed symmetries that have survived many experiments so far. The conservation laws also help physicists make predictions and design experiments more easily.

First, there is the external, global symmetry for translation, rotation and Lorentz transformation. It is also known as the Poincaré symmetry. Without this symmetry, energy and momentum would no longer be conserved and the SM would violate special relativity. Then, there is also the internal, local symmetry for the gluon, weak isospin and

weak hypercharge gauge fields. It implies the conservation of colour charge, weak isospin, weak hypercharge, and consequently the electric charge.

Other than the Poincaré and gauge symmetries, there are also several accidental symmetries in the SM that are not postulated when the SM was formulated. They correspond to the global phase invariance of the quark fields as a whole, and that of the electron, muon and tau fields individually. The phase invariance of the quark fields implies conservation of baryon number. Similarly, the phase invariance of the lepton fields implies conservation of electron number, muon number and tau number independently. The electron, muon and tau numbers are defined as (number of neutrinos and charged leptons − number of antineutrinos and charged antileptons) of the respective flavours. They are collectively known as the lepton family numbers. To date, in no experiments has the conservation of baryon number been violated. Meanwhile, the observation of neutrino oscillations has shown that lepton family numbers are indeed not conserved in Nature. It is one of the pieces of evidence that exposed the incompleteness of the SM.

As a summary, Table 1.7 shows all of the symmetries and the associated conserved quantities in the SM.

Table 1.7.: Symmetries in the Standard Model and the associated conserved quantities.

Symmetry	Lie group	Type	Conserved quantities
Poincaré	$\mathbf{R}^{1,3}\otimes$O(1,3)	external, global	energy, momentum
Gauge	SU(3)⊗SU(2)⊗U(1)	internal, local	colour charge, weak isospin, weak hypercharge, electric charge
Quark phase	U(1)	accidental, global	baryon number
Lepton phase	U(1)	accidental, global	lepton family numbers

1.2.5. Flavour violation in the Standard Model

The weak charged-current interaction is the only process in the SM that does not conserve fermion flavours. Not only does it couple up- and down-type quarks, or neutrinos and charged leptons in the same generation. It also mixes fermions in different generations. This is directly due to the mismatch of the mass eigenstates and the weak interaction eigenstates of fermions.

Quark mixing

The quark doublets that participate in charged-current interaction do not just consist of up- and down-type quarks of the same generation, but superpositions of quarks in

different generations. These superposition states are the weak interaction eigenstates. By convention, a basis is chosen such that the mass eigenstates of up-type quarks are also the weak interaction eigenstates. This forces the weak interaction eigenstates of down-type quarks to be superpositions of the mass eigenstates. The quark doublets can be written as

$$\Psi_{\mathrm{L}} = \begin{pmatrix} u_{\mathrm{L}} \\ d'_{\mathrm{L}} \end{pmatrix}, \quad \begin{pmatrix} c_{\mathrm{L}} \\ s'_{\mathrm{L}} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} t_{\mathrm{L}} \\ b'_{\mathrm{L}} \end{pmatrix}, \tag{1.22}$$

where the primed fields are the interaction eigenstates and are related to the mass eigenstates by the Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{CKM}:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\mathrm{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{1.23}$$

The CKM matrix is a unitary matrix and its elements are fundamental parameters of the SM. The magnitudes of the CKM matrix elements have been measured by experiments [6] to be

$$\left|V_{\mathrm{CKM}}^{ij}\right| = \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359 \pm 0.00011 & 0.04214 \pm 0.00076 \\ 0.00896 \pm 0.00024 & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix}. \tag{1.24}$$

The off-diagonal elements of the matrix are relatively small, implying that the probability of transition of a quark from one generation to another is relatively low.

Neutrino mixing

The observation of neutrino oscillations [20, 21] has indisputably proven that neutrinos have finite rest masses, however small they are. And similar to the quarks, there is a mismatch between the neutrino mass eigenstates and the neutrino weak interaction eigenstates. In the conventional basis where the charged lepton mass eigenstates are also the interaction eigenstates, the neutrino eigenstates are related by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix U_{PMNS}, analogous to the CKM matrix for quark mixing:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U_{\mathrm{PMNS}} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \tag{1.25}$$

where ν_1, ν_2 and ν_3 are the neutrino mass eigenstates. The PMNS matrix is also unitary and can be parameterised by three mixing angles, θ_{12}, θ_{23} and θ_{13}, and one phase angle, δ:

$$U_{\text{PMNS}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta} & c_{13}c_{23} \end{pmatrix}, \tag{1.26}$$

where $c_{jk} = \cos\theta_{jk}$ and $s_{jk} = \sin\theta_{jk}$. These parameters have been measured by various neutrino oscillation experiments [22]. They take on different values depending on the mass ordering assumption: the Normal Ordering where $m_{\nu_1} < m_{\nu_2} < m_{\nu_3}$ or the Inverted Ordering where $m_{\nu_3} < m_{\nu_1} < m_{\nu_2}$. The current best-fit values are summarised in Table 1.8. Unlike the CKM matrix which is approximately diagonal, the PMNS matrix has much larger off-diagonal elements.

Table 1.8.: Current best-fit values of the mixing angles and the phase angle of the PMNS matrix [22].

Parameter	**Best-fit value $\pm 1\sigma$ [°]**	
	Normal Ordering	**Inverted Ordering**
θ_{12}	$34.5^{+1.2}_{-1.0}$	$34.5^{+1.2}_{-1.0}$
θ_{23}	$47.7^{+1.2}_{-1.7}$	$47.9^{+1.0}_{-1.7}$
θ_{13}	$8.45^{+0.16}_{-0.14}$	$8.53^{+0.14}_{-0.15}$
δ	218^{+38}_{-27}	281^{+23}_{-27}

One noteworthy clarification is that while quark flavours usually refer to the quark mass eigenstates, neutrino flavours always refer to the interaction eigenstates. This confusing terminology is related to the facts that the neutrino masses are not exactly known due to experimental limitations and that historically neutrinos were discovered in weak interaction processes and were therefore defined by the flavours of their charged lepton interaction partners.

In this thesis, from this point onward, neutrino mixing and the PMNS matrix will be considered as part of the SM, with neutrino masses treated as Dirac masses. In the literature, this is also known as the Neutrino Minimal Standard Model (νMSM) [23] with zero Majorana masses. The scenario where neutrinos have Majorana masses will be discussed later in Section 1.3.1.

1.2.6. Incompleteness of the Standard Model

Despite its success in explaining many physical phenomena, the SM is still an incomplete theory. This is known as a fact due to a number of unsolved problems in physics to which the SM fails to provide satisfactory answers.

Some of these unsolved problems are:

Quantum gravity Despite being one of the four known fundamental forces of Nature, gravity is not a part of the SM. Currently, the best theory that describes gravity is the general theory of relativity (GR). Even though GR is a classical field theory like electromagnetism in classical mechanics, attempts to construct a corresponding quantum field theory like it was done for QED has been met with two main obstacles. The first of which is the fact that spacetime is dynamic in GR, while quantum field theory depends on a fixed spacetime background (Minkowski spacetime). The second is that gravity seems to be non-renormalisable in perturbation theory, suggesting that infinitely many independent parameters are needed to meaningfully define the theory [24].

Dark matter There is an abundance of evidence that the universe consists of not only ordinary baryonic matter, but also some kind of invisible matter that does not interact, or only interact extremely weakly, with the ordinary matter [25–27]. This unknown, hypothetical type of matter is referred to as dark matter, and there is no particle candidate for such matter in the SM. The existence of dark matter can only be inferred by astronomically or cosmologically observed gravitational phenomena. This implies that dark matter could be just a manifestation of our possibly inaccurate understanding of gravity, instead of undiscovered particles. In all cases, it is clear that the dark matter phenomena cannot be explained by the SM.

The hierarchy problem As also mentioned in Section 1.2.2, the mass of the Higgs boson is expected to be comparable with the Planck mass due to quantum corrections. The fact that it is not poses a fine-tuning problem known as the hierarchy problem. Since the Planck mass is defined using the gravitational constant as a natural unit, the hierarchy problem can therefore also be understood as the question of why the gravitational force is so much weaker than the electroweak force.

The matter-antimatter asymmetry It is apparent that the observable universe consists of radically more (baryonic) matter than antimatter. Extrapolating back in time, this implies an asymmetry in the matter and antimatter generation process in the early universe. To create such an asymmetry, baryon-generating interactions must satisfy three criteria known as the Sakharov conditions [28]: they violate baryon number conservation, they violate the charge (C) and charge-parity (CP) symmetries, and they must be out of thermal equilibrium. Baryon number and CP are indeed violated in the SM. However, the violation as we understand it is too small to account for the observed asymmetry.

Neutrino mass With the observation of neutrino oscillations, it is now evident that neutrinos have finite rest masses. Nonetheless, the mechanism responsible for neutrinos to acquire mass still remains unknown. It is possible that neutrinos acquire their masses in a way similar to the other SM fermions. However, if that is indeed the case, extra explanation will be needed for the huge difference between the neutrino masses and the other SM fermion masses, which turns it into another fine-tuning problem.

Evidence against lepton flavour universality In the SM, the three generations of leptons have exactly the same behaviours and properties except for their masses (or in other words, their interactions with the Higgs field). This is known as lepton flavour universality (LFU). In recent decades, multiple experiments have shown clues that point to the possibility of the violation of LFU. The most intriguing ones are the anomalies in semi-leptonic B-meson decays: deviations of the values of $R(D^*) = \mathcal{B}(B^0 \to D^{*0}\tau^-\bar{\nu})/\mathcal{B}(B^0 \to D^{*0}\mu^-\bar{\nu})$ and $R(K^*) = \mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)/\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)$ measured by the BaBar, Belle and LHCb experiments [29–34] from the SM-predicted values. The combined measurement for $R(D^*)$ has a discrepancy of $3.08\,\sigma$ with the SM prediction [35]. These are sometimes referred to as the B-anomalies. Although these deviations are not statistically significant enough to conclusively rule out LFU, they do hint at a possible flaw in the SM and a direction for us to look for New Physics.

These problems have motivated the construction of many beyond-the-Standard-Model theories. Some of them will be discussed in the following section.

1.3. Lepton flavour violation in BSM theories

Neutrino mixing implies that the lepton family numbers could not be conserved quantities. However, when only point interactions of charged leptons are concerned, lepton family numbers are approximately conserved. Violation of this approximate conservation is termed (charged) lepton flavour violation. Examples of lepton-flavour-violating (LFV) processes include $\mu \to e\gamma$, $\tau \to \mu\mu\mu$, and the processes that are in the spotlight of this thesis: $Z \to e\tau$ and $Z \to \mu\tau$. The transition amplitudes of these processes are heavily suppressed by a factor of m_ν^2/m_W^2 when they are solely induced by neutrino mixing in the SM. As a result, LFV decays typically have an expected branching fraction $\ll 1 \times 10^{-50}$ [36, 37]. Such a vanishingly small branching fraction is beyond reach of any experiments, even in the distant future. This makes any observation of LFV processes an unambiguous signal of BSM physics.

There are many interesting BSM theories that can give rise to much more sizable amplitudes for LFV processes. These theories are often motivated by unsolved problems in physics, such as those listed in Section 1.2.6. Searches of LFV processes could help us constrain these theories and potentially point us in the right direction of solving these problems. In this section, the main ideas and motivations of a handful of popular (historically or currently) types of models will be discussed: heavy neutrinos, supersymmetry and extended Higgs sector.

1.3.1. Heavy neutrinos and the seesaw mechanism

As mentioned earlier, experiments have confirmed that neutrinos indeed have finite rest masses. Nonetheless, the origin of their masses is still unknown. In QFT, fermions can

acquire their masses from two types of mass terms in the Lagrangian, namely the Dirac mass term and the Majorana mass term.

A Dirac mass term has the basic structure of $m\bar{\psi}\psi$ in the Lagrangian. Since it is an interaction of a field with its conjugation, the mass term conserves charges, flavours and number of particles. Any fermions could have a Dirac mass term. In the SM, the Yukawa interaction of the Higgs field and charged fermion fields is a Dirac mass term.

A Majorana mass term has the basic structure of $m\psi^{\mathrm{T}}\psi$, without the conjugation as in a Dirac mass term. As apparent, any quantum numbers that ψ carries are not conserved with the Majorana mass term. If the conservation of charge is not to be violated, only particles that are antiparticles of themselves could have Majorana mass. Fermions with such a property are called Majorana fermions. In the SM, only neutrinos could possibly be Majorana fermions.

If neutrinos only have Dirac mass like the other fermions in the SM, then the Yukawa coupling of the neutrinos must be many orders of magnitude smaller than that of the other SM fermions. Hence, there is a fine-tuning problem in such a model: there lacks a natural explanation for the smallness of the neutrino-Higgs Yukawa coupling. However, the problem could be avoided if neutrinos are indeed Majorana fermions and there exist some mechanisms other than the SM Yukawa interactions that give mass to the neutrinos.

One type of commonly studied models are models with heavy right-handed neutrinos [38]. Right-handed neutrinos can couple with the left-handed neutrinos via the SM-like Yukawa coupling, giving a Dirac mass term. On top of that, if neutrinos are Majorana fermions, the right-handed neutrinos may also have a Majorana mass term. The relevant terms in the Lagrangian after the electroweak symmetry breaking can be expressed as:

$$\mathscr{L}_\nu = -\frac{1}{2} m_{\mathrm{D}} \bar{\nu}_{\mathrm{L}} N_{\mathrm{R}} - \frac{1}{2} M_{\mathrm{R}} N_{\mathrm{R}}^{\mathrm{T}} N_{\mathrm{R}} + \mathrm{h.c.}, \tag{1.27}$$

where N_{R} denotes the conjectured right-handed neutrinos. This could also be represented using a mass (block) matrix in the basis of $\{\nu_{\mathrm{L}}, N_{\mathrm{R}}\}$:

$$\mathcal{M}_\nu = \begin{pmatrix} 0 & m_{\mathrm{D}} \\ m_{\mathrm{D}}^T & M_{\mathrm{R}} \end{pmatrix}. \tag{1.28}$$

The null matrix at the top left represents that there is no Majorana mass term for the left-handed neutrinos at tree level. This is important since such a term will clearly violate the isospin SU(2) symmetry for left-handed leptons. The off-diagonal terms are the Dirac masses, which have an origin similar to the charged fermion masses in the SM. It is natural to assume that these masses are also at the electroweak scale. The bottom-right Majorana mass term for right-handed neutrinos could have an origin from BSM physics at a much higher energy scale. This could be, for example, the SU(5), SO(10) or E_6 model of a Grand Unification Theory. The values of M_{R} will then naturally be at the scale where New Physics appears.

The actual masses of the neutrinos are then result of the diagonalisation of $\mathcal{M}_\nu$. Disregarding the multiple flavours and treating the entries of the mass matrix as plain numbers, one can immediately show that the eigenvalues of the matrix are

$$m_\pm = \frac{M_\mathrm{R} \pm \sqrt{M_\mathrm{R}^2 + 4m_\mathrm{D}^2}}{2}. \tag{1.29}$$

In the regime of $M_\mathrm{R} \gg m_\mathrm{D}$,

$$m_+ \approx M_\mathrm{R} \qquad \text{and} \qquad m_- \approx -\frac{m_\mathrm{D}^2}{M_\mathrm{R}}. \tag{1.30}$$

This can be interpreted as the right-handed neutrino retaining its heavy Majorana mass, while the left-handed neutrino acquires a very small mass that is inversely proportional to the right-handed neutrino mass. The same argument still holds when multiple flavours are considered if the condition $\sqrt{\mathrm{Tr}(|m_\mathrm{D} M_\mathrm{R}^{-1}|^2)} \ll 1$ is met. By this, the model naturally explains the smallness of the left-handed neutrino masses without fine-tuning the neutrino-Higgs Yukawa coupling. This is known as the (type-I) seesaw mechanism. The name "seesaw" refers to the feature that the larger the right-handed neutrino masses are, the smaller the left-handed neutrino masses will be. There are also some more sophisticated and complete models that make use of similar mechanisms (type-II and type-III seesaw mechanisms), but the basic working principle is the same.

Models with heavy Majorana neutrinos are certainly attractive since they could "complete" the SM with the "missing" right-handed neutrinos and explain the smallness of the SM neutrino masses at the same time. But there is another reason for them to be favoured. That is the possibility for these models to fit with leptogenesis models in cosmology. Leptogenesis models are models constructed to explain the matter-antimatter asymmetry problem (see Section 1.2.6). As mentioned earlier, CP violation in the SM is not enough to explain the observed abundance of matter in the universe. An attempt to resolve this problem is to introduce conjectured physical processes that generate leptons and antileptons asymmetrically in the early universe. These processes are collectively referred to as leptogenesis. Since heavy neutrinos with Majorana mass origins clearly violate CP at a high energy scale, their interactions could also contribute to leptogenesis.

Although the interactions of heavy neutrinos are new physics at high energy scale, their existence could still induce observable effects at low energy scale. These effects include the possible enhancement of LFV processes. There have been theoretical studies that calculated the expected branching fraction of LFV Z decays assuming the existence of Majorana neutrinos [39, 40]. These studies built their models in a way that incorporates or averts existing experimental constraints, such as the stringent μ–e transition constraint from $\mu \to e\gamma$ experiments. The calculations show that, with heavy neutrinos at the TeV scale, the branching fractions $\mathcal{B}(Z \to e\tau)$ and $\mathcal{B}(Z \to \mu\tau)$ could be as large as 10^{-8}–10^{-7}.

1.3.2. Supersymmetry

Supersymmetry (SUSY) [41, 42] is a conjectured symmetry that relates fermions and bosons, the two fundamental types of particles. It is an extension to the SM Poincaré symmetry of spacetime. The fundamental idea of SUSY is that for each SM particle, there exists a superpartner, also known as sparticle. The SM particles and their corresponding superpartners carry exactly the same quantum numbers, except for their spins, which are always different by a half-integer. This means that superpartners of fermions are always bosons, and vice versa. If it exists, SUSY could provide an elegant solution to many of the puzzling problems in modern physics. One of which is the hierarchy problem. In a supersymmetric theory, contributions from SM particles and their superpartners to the quantum corrections of the Higgs boson mass can cancel each other. This allows the smallness of the observed Higgs boson mass compared to the Planck scale to be explained without a miraculous fine tuning. With an additional discrete symmetry known as the R-parity, which forbids the decay of the lightest sparticle, SUSY could also provide us with a promising dark matter candidate. Supersymmetric theories typically have many free parameters. Even in the Minimal Supersymmetric Standard Model (MSSM), which only extends the Standard Model minimally by adding the smallest possible number of supersymmetric particles and interactions, has over 100 parameters. The high dimensionality of SUSY models make them very difficult to search for or constrain. The lack of evidence so far has made SUSY much less popular among physicists than it was decades ago. Nonetheless, popularity is no indication of truth, and SUSY still remains a possible candidate of undiscovered BSM physics.

Like in the SM, superpartners of leptons, or sleptons ($\tilde{\ell}$), can also have flavour mixing. In some particular models, such as the minimal supergravity model (mSUGRA) with seesaw mechanism, charged slepton flavour mixing can be sizeable [43]. The large flavour mixing can be realised by their additional contributions to the amplitudes of LFV processes. A study has shown that, subject to constraints from existing measurement of neutrino oscillation and $\ell \to \ell'\gamma$ decays, slepton flavour mixing could enhance the branching fraction of LFV Z decays to up to 10^{-8} [44].

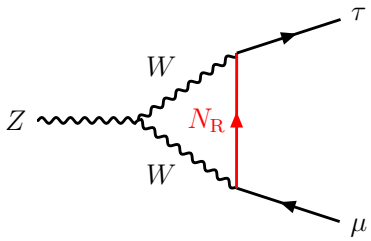
1.3.3. Extended Higgs sector

The SM has only one single Higgs doublet. However, that is merely the simplest hypothesis one can assume in order to explain the electroweak symmetry breaking and the mass generation of fermions. There is no reason why Nature cannot be more complex and have multiple Higgs doublets or even multiplets. In fact, there are a number of BSM models that have extended Higgs sectors, which include, for example, the aforementioned MSSM. The most studied model with an extended Higgs sector is probably the Two-Higgs-doublet model (2HDM) [45], the next simplest case to having only one Higgs doublet. Just by adding an extra Higgs doublet, the 2HDM could already allow many more phenomena than by the SM. This can be appreciated from the fact that there are five physical Higgs boson in the 2HDM, compared to just one in the SM. The five bosons consist of two CP-even

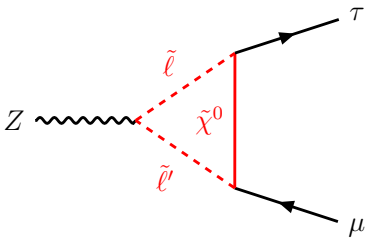
neutral Higgs bosons (h^0 and H^0, one of which is the observed 125-GeV Higgs boson), a CP-odd pseudoscalar (A) and two charged Higgs bosons ($H^\pm$). Two-Higgs-doublet models can be characterised by six independent parameters, whereas the SM Higgs sector only needs two.

In the SM, the Yukawa coupling matrix can always be diagonalised, meaning that the Higgs boson cannot induce flavour-changing neutral currents. However, with two Higgs doublets and thus two Yukawa coupling matrices in 2HDM, simultaneous diagonalisation of both matrices is no longer trivial. In fact, in the most generic scenario where the two Higgs doublets can both couple to all the up-type quarks, down-type quarks and charged leptons, simultaneous diagonalisation of the Yukawa couplings is generally impossible [46]. This is the so-called Type-III 2HDM. Flavour-changing neutral currents are possible at tree level in Type-III 2HDM. At the one-loop level, LFV Z interactions can also be induced. A study has shown that with constraints from proton–proton (pp) collision data at a centre-of-mass energy $\sqrt{s} = 8\,\text{TeV}$, the branching fraction of the LFV $Z \to \mu\tau$ decay is bound from above by 10^{-6} [47].

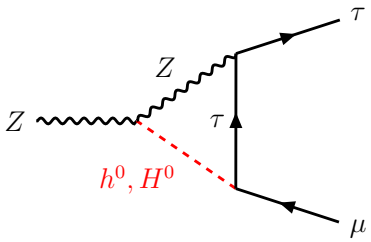
As examples, a few Feynman diagrams for the mentioned possible BSM LFV processes are shown in Figure 1.2.



(a) Heavy right-handed neutrinos



(b) SUSY (slepton flavour mixing)



(c) Type-III 2HDM

Figure 1.2.: Examples of Feynman diagrams for possible BSM LFV processes. BSM particles are highlighted in red.

1.4. Lepton-flavour-violating $Z \to \ell\tau$ decays

Hopefully by now, a compelling argument has been made that lepton flavour violation is an interesting BSM phenomenon and searches for it are important for finding or constraining New Physics. But with that comes the natural question of where and how to search for lepton flavour violation. BSM models predict various possible LFV processes and physicists have designed and run experiments to search for many of them. These searches are sensitive to different BSM models, mass scales and lepton flavours, and complement each other. Ultimately, it is the interplay of these searches that teaches us about potential BSM physics. Table 1.9 summarises some of the search channels and their current most stringent limits. While all of the searches are important in their own ways, this thesis focuses particularly on the $Z \to \ell\tau$ decays (here and from now on, ℓ will be used to denote a light lepton, i.e. e or μ, but not a τ lepton). These decays are interesting for a number of reasons.

LFV Z decays are interesting both theoretically and experimentally. Theoretically, the massiveness of the Z boson relative to leptons or hadrons would make its decays much more sensitive to BSM physics at a higher energy scale. Assuming that LFV Z decays are induced by loops involving undiscovered particles that are much heavier than the Z rest mass m_Z, it can be shown that the effective LFV coupling (branching fraction) would be proportional to m_Z^2 (m_Z^4), whereas in $\tau \to \mu\mu\mu$ it would be proportional to the much smaller m_τ^2 (m_τ^4) [61]. Experimentally, Z decays are interesting because of the abundance of Z bosons produced at the Large Hadron Collider. The Z boson production cross section in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ times the single-charged-lepton-flavour branching fraction has been measured to be 1.91 nb [62]. This means that roughly eight billion Z bosons have been produced in the second operational run of the Large Hadron Collider for the ATLAS detector. Such an abundance is essential in providing the necessary statistical power for rare decay searches. The massiveness of the Z boson is also an experimental advantage. While the Large Hadron Collider produces a lot of hadrons and τ leptons, for ATLAS and CMS, their momenta are often too low for meaningful analysis. The high mass of the Z boson, however, ensures that its decay products are boosted in the laboratory frame of reference. This allows for a much higher trigger and reconstruction efficiency compared to hadrons or τ decays searches.

The Z boson has three possible LFV two-lepton decay modes, namely $e\mu$, $e\tau$ and $\mu\tau$. There is no necessary correlation between these decay modes. The $e\mu$ decay mode generally receives much less attention than the other two. This is mainly because in most models, the precise results from low-energy $\mu \to e\gamma$ experiments have already set an indirect limit on $\mathcal{B}(Z \to e\mu)$ that is much more stringent than a direct search at the Large Hadron Collider can possibly achieve. Using Effective Field Theory, the limit on $\mathcal{B}(\mu \to e\gamma)$ can be translated into an indirect limit of $\mathcal{B}(Z \to e\mu) \lesssim 10^{-10}$ [61]. Nevertheless, direct searches of $Z \to e\mu$ should still be appreciated as concrete confirmation of the $\mu \to e\gamma$ indirect limit. In this thesis, however, focus is given to the other two decay modes which have a τ lepton in the final state. With the sensitivities of existing LFV τ decay searches, there is currently no stringent indirect limit on $\mathcal{B}(Z \to \ell\tau)$. This makes the search for $Z \to \ell\tau$

Table 1.9.: Current experimental limits on the branching fractions of various LFV processes. The limits are set at 90% confidence level for μ, τ or hadron decays and 95% for Z and H decays. The limits on the $Z \to \ell\tau$ decays were the most stringent before the work in this thesis. More limits on LFV hadron decays can be found summarised in Reference [6].

Violation	Process	Upper limit on branching fraction	Experiment
e and μ lepton flavours	$\mu \to e\gamma$	4.2×10^{-13} [48]	MEG
	$\mu \to eee$	1.0×10^{-12} [49]	SINDRUM
	$\pi^0 \to e\mu$	3.6×10^{-10} [50]	KTeV
	$J/\psi \to e\mu$	1.6×10^{-7} [51]	BESIII
	$H \to e\mu$	3.5×10^{-4} [52]	CMS
	$Z \to e\mu$	7.3×10^{-7} [53]	CMS
e and τ lepton flavours	$\tau \to e\gamma$	3.3×10^{-8} [54]	BaBar
	$\tau \to eee$	2.7×10^{-8} [55]	Belle
	$\tau \to e^{\pm}\mu^{\mp}\mu^{\pm}$	2.7×10^{-8} [55]	Belle
	$J/\psi \to e\tau$	8.3×10^{-6} [56]	BES
	$H \to e\tau$	4.7×10^{-3} [57]	ATLAS
	$Z \to e\tau$	9.8×10^{-6} [58]	OPAL
μ and τ lepton flavours	$\tau \to \mu\gamma$	4.4×10^{-8} [54]	BaBar
	$\tau \to e^{\pm}e^{\mp}\mu^{\pm}$	1.8×10^{-8} [55]	Belle
	$\tau \to \mu\mu\mu$	2.1×10^{-8} [55]	Belle
	$J/\psi \to \mu\tau$	2.0×10^{-6} [56]	BES
	$H \to \mu\tau$	2.5×10^{-3} [59]	CMS
	$Z \to \mu\tau$	1.2×10^{-5} [60]	DELPHI

decays especially exciting, since it can be expected to either find evidence for New Physics or be able to set new constraints on relevant BSM models. Moreover, the B-anomalies (see Section 1.2.6) seem to hint at something different between the muon and the τ lepton that we have yet to understand, making the $Z \to \mu\tau$ channel even more intriguing.

Experimentally, a τ lepton in the final state is a challenge for reconstruction and identification due to its short life time (mean life = 290 fs, corresponding to a proper decay length of 86.9 µm) and many decay modes. But on the other hand, the neutrino(s) from the τ lepton decay, which escapes the detector and becomes missing energy in reconstruction, gives us a unique signature to identify the signal events. In the ATLAS (or CMS) detector, τ leptons produced at the collisions almost always decay before reaching the detector,

either leptonically (τ_{lep}) or hadronically (τ_{had}). While there are efforts in both the $\ell\tau_{\text{lep}}$ and $\ell\tau_{\text{had}}$ search channels, this thesis will focus mainly on the $\ell\tau_{\text{had}}$ channel.

The existing limits: $\mathcal{B}(Z \to e\tau) < 9.8 \times 10^{-6}$ and $\mathcal{B}(Z \to \mu\tau) < 1.2 \times 10^{-5}$ were both set by experiments at the Large Electron-Positron Collider (LEP) [58, 60]. More than 20 years have passed since these limits were set and they had remained the most stringent. However, in this thesis, it is the author's main goal to push the limits and finally reach new sensitivities using data from the Large Hadron Collider and the ATLAS detector.

Chapter 2.

The Large Hadron Collider and the ATLAS detector

Experiment is the only means of knowledge at our disposal. Everything else is poetry, imagination.

— Max K. E. L. Planck

To solve the mysteries in fundamental physics, we need to carefully test our current knowledge of Nature in order to find clues of New Physics. Our current knowledge of Nature is best summarised by the two theoretical frameworks, the general theory of relativity and the Standard Model. To test the former, one must look to astronomical-scale objects and events, where the effect of gravity becomes prominent. For that, the most recent breakthrough is the detection of gravitational waves by the LIGO and Virgo experiments [63], which has found no evidence against the predictions of GR so far [64]. And to test the latter, one must instead look to events at the smallest scale. However, that by no means implies that the experiments for such tests are small in scale as well. On the contrary, due to the fundamental quantum-mechanical relation of

$$L = \frac{\hbar c}{E}, \tag{2.1}$$

where L and E are the length and energy scales respectively, these experiments tend to be physically huge to pursue high energy. In fact, the largest single machine in the world, the Large Hadron Collider, is exactly built for such a purpose. The high precision and performance required to observe such high-energy events also push the size and complexity of the detectors used in these experiments. The ATLAS detector, designed to test multiple aspects of the SM and search for evidence of various BSM phenomena, is once again the largest particle detector in the world. In this chapter, the design and workings of the Large Hadron Collider and the ATLAS detector will be described. These state-of-the-art apparatus are what provided the data that make the work in this thesis possible.

2.1. The CERN accelerator complex and the LHC

The European Organization for Nuclear Research, also known as CERN, is a world-leading institute for frontier particle physics research. Among other experiments, it maintains and operates the world's most powerful particle accelerator and collider, the Large Hadron Collider (LHC).

2.1.1. The CERN accelerator complex

Located at the French-Swiss border close to Geneva, CERN's main site is the home to many particle accelerators built and used throughout the 66-year history of CERN. Although many of these accelerators are no longer the most energetic accelerators as they once were, physicists capitalise on them by using them to pre-accelerate particles before feeding them into the LHC. These accelerators form a chain that accelerates particles in successive stages of increasingly higher energies. This chain of accelerators is known as the CERN accelerator complex. Figure 2.1 provides a schematic representation of the CERN accelerator complex during the second operational run (Run 2) of the LHC.

When the complex is operating for proton-proton (pp) collisions, protons are first extracted from hydrogen atoms by ionisation and injected into the linear accelerator LINAC 2 for initial acceleration. After accelerated to 50 MeV, the protons are then directed into the Proton Synchrotron Booster (PSB or BOOSTER) where they are further accelerated to 1.4 GeV. Following that, the proton beam enters the Proton Synchrotron (PS) and subsequently the Super Proton Synchrotron (SPS), which drive the beam energy up to 25 GeV and 450 GeV per proton respectively. Finally, the protons are split into two beams which are then fed into the LHC, where they are accelerated in opposite directions up to the maximum energy of 6.5 TeV.

2.1.2. The Large Hadron Collider

The LHC is a circular collider with a circumference of 26.7 km located 50–170 m underground. It uses radiofrequency (RF) cavities to accelerate, dipole magnets to steer, and quadrupole magnets to focus charged particle beams. All of these components are maintained in a superconducting state when operating to avoid energy dissipation due to electrical resistance. After the beams have reached the desired energy, they are made to collide at one of the four designated collision points where particle detectors are situated.

To accelerate the charged particles, oscillating electromagnetic fields are supplied to the 16 RF cavities (eight per beam) on the LHC. These RF cavities are superconducting chambers made in specific shape and size that allow the electromagnetic field inside to resonate and build up at a designated frequency. When timed correctly, the oscillating field in a RF cavity allows charged particles that pass through to experience a forward electric force only. They also work as a modulator as charged particles with energies

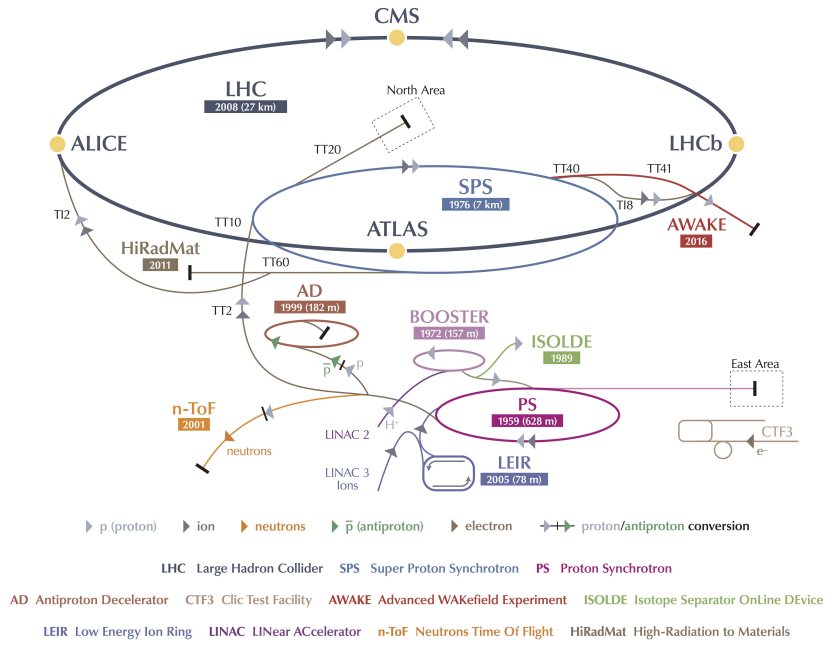


Figure 2.1.: A schematic representation of the CERN accelerator complex during the second operational run of the LHC. Circular accelerators are represented by ellipses labelled with their names and the years that they were built in. The LHC can be seen as the largest ellipse at the top. The four circles on the LHC represent the collision points where the four main detectors are located. The colour-coded arrows indicate the types and directions of particles being transmitted or accelerated in each part of the complex. This picture is adapted from Reference [65].

slightly higher (lower) than the ideal energy will travel faster (slower) through the cavities and will therefore be decelerated (accelerated) by the electromagnetic field due to the offset timing. As a result, charged particles are grouped in discrete packets known as "bunches", which increases the chance of collision. On the LHC, the RF cavities are driven by high-power klystrons that modulate at a frequency of 400 MHz. Each cavity can build an electric potential up to 2 MV.

Thousands of superconducting electromagnets are deployed along the LHC to steer, stabilise and focus the particle beams. Dipole magnets are used to provide strong magnetic fields that bends the particle beams, keeping them in a (nearly) circular orbit. The main dipole magnets on the LHC could generate magnetic fields of 8.3 T. Quadrupole magnets are used to tighten the beams by squeezing them in the vertical and horizontal directions alternately. Sextupole, octupole and decapole magnets are also installed to counteract higher-order effects such as the inhomogeneity of the main dipole magnets or electromagnetic interactions between bunches. Additionally, eight sets of so-called inner

triplets, each of which is a system of three quadrupole magnets, are placed on each sides of the four collision points. Their purpose is to tighten the beams as much as possible before crossing the beams in order to maximise the chance of collision.

There are currently eight different experiments at the LHC: ALICE, ATLAS, CMS, LHCb, LHCf, MOEDAL, TOTEM, and the recently approved FASER. ALICE, ATLAS, CMS and LHCb are the four main experiments. Each of these experiments have their detector constructed in one of the underground caverns at the four collision points. ATLAS and CMS are the two general-purpose detectors at the LHC. They are the experiments that discovered the Higgs boson in 2012 using pp collision data from the first operational run (Run 1) of the LHC. ALICE and LHCb are more specialised detectors. ALICE is designed for studying strong interactions in quark-gluon plasma, while LHCb is specialised in measuring CP violation of b-hadrons. The remaining four experiments are considerably smaller experiments, both in terms of the physical size of their detectors and the size of the collaborations. The data analysed in this thesis were collected by the ATLAS detector, which will be described in more details later in Section 2.2.

2.1.3. Luminosity and pile-up

For any given process, the expected number of events occurred at the LHC N is the product of the integrated luminosity L and the interaction cross section σ for that process:

$$N = L\sigma. \tag{2.2}$$

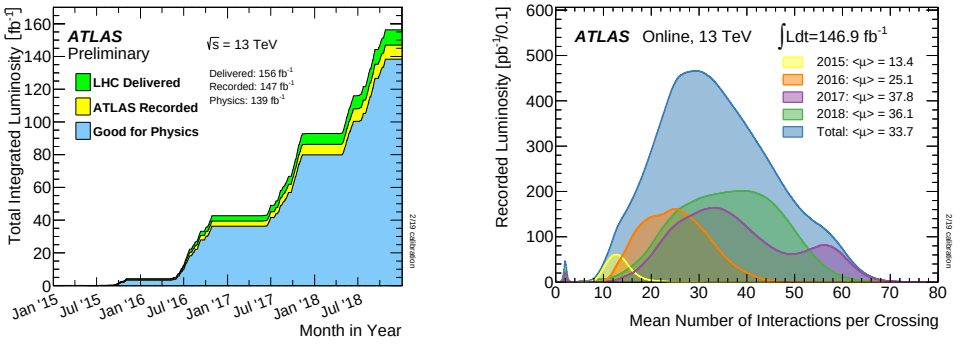
The integrated luminosity can be understood as the expected number of "encounters" between particles in the colliding beams, while the interaction cross section is a measure of the probability that a certain process will occur during an encounter. The integrated luminosity is the time-integrated value of the instantaneous luminosity $\mathcal{L}$:

$$L = \int \mathcal{L}\,\mathrm{d}t. \tag{2.3}$$

The instantaneous luminosity can be interpreted as the rate of particle encounters, and can be approximated by

$$\mathcal{L} = \frac{f_\mathrm{r} n_\mathrm{b} N_\mathrm{p}^2 \gamma}{4\pi\varepsilon_\mathrm{n}\beta^*} F(\theta_\mathrm{c}, \sigma_x, \sigma_z), \tag{2.4}$$

where f_r is the LHC revolution frequency, n_b is the number of bunches in each beam, N_p is the number of particles per bunch, γ is the relativistic Lorentz factor, ε_n is the normalised transverse emittance, β^* is the amplitude function at the interaction point, and F is the geometric reduction factor, which is a function of the crossing angle θ_c and the beam cross-sectional sizes σ_x and σ_z. During Run 2, the LHC has delivered pp collisions to ATLAS for a total integrated luminosity of 156 fb^{-1}, of which 147 fb^{-1} were recorded by ATLAS, and of which 139 fb^{-1} were good for physics analyses. The integrated luminosities over time throughout Run 2 are shown in Figure 2.2a.



(a) (b)

Figure 2.2.: The (a) integrated luminosities over time and (b) pile-up profile for *pp* collisions delivered to ATLAS during Run 2 of the LHC [66].

During bunch crossings, multiple collisions can happen at the same time. However, in physics analyses, typically only one of these collisions, usually the hardest scattering, is of interest. The other additional, usually softer, interactions are collectively referred to as (in-time) "pile-up". Other interactions from preceding or subsequent bunch crossings can also sometimes be registered as events occurring at the same time as the collision of interest. This could happen when components of a detector have a relaxation time longer than 25 ns. These additional signals are known as out-of-time pile-up. Since pile-up events also produce particles that leave tracks and deposit energies in the detector, having a large number of them is a challenge to the detector's trigger and readout system, as well as to the physics analyses. Pile-up depends mainly on the number of interactions per crossing, which follows a Poisson distribution with mean μ. The mean μ is proportional to the instantaneous luminosity and can be calculated as

$$\mu = \frac{\mathcal{L}\sigma_{\text{inel}}}{f_{\text{r}} n_{\text{b}}}, \tag{2.5}$$

where σ_{inel} is the cross section for inelastic scatterings. Therefore, the measurement of instantaneous luminosity is important to physics analyses as it determines the amount of pile-up. Pile-up changes over time as the instananeous luminosity decreases as particles in the beams are lost in collisions. The average value of μ over a period of time, denoted as $\langle\mu\rangle$, is also used to quantify pile-up. Figure 2.2b shows the pile-up profile for *pp* collisions delivered to ATLAS during Run 2. In the beginning of Run 2, the LHC operated at relatively low instantaneous luminosities with $\langle\mu\rangle = 13.4$ in 2015. The operating luminosity increased in the following years with $\langle\mu\rangle = 25.1$ in 2016, $\langle\mu\rangle = 37.8$ in 2017 and $\langle\mu\rangle = 36.1$ in 2018. The overall value of $\langle\mu\rangle$ over the entire data-taking period of Run 2 is 33.7 [66].

2.2. The ATLAS detector

The ATLAS detector [67] is one of the two general-purpose detectors at the LHC. It was designed to measure various parameters of the SM and search for BSM phenomena by probing pp or heavy-ion collisions at the LHC. The idea is to detect, identify and reconstruct particles produced in pp collisions, and use those particles to reconstruct the collision events. To achieve that, the detector must be able to (1) track particles to determine where they came from and in what directions they were travelling, (2) measure the energies or momenta of the particles and (3) produce differential responses to different types of particles in order to identify them. Equipped with a number of subsystems, each specialised in a certain type of measurement, the ATLAS detector is able to accomplish all of the above. On top of these, the detector also has to be efficient, meaning that it should collect as much useful data as possible for a given integrated luminosity delivered by the LHC. Therefore, the ATLAS detector is designed to cover a solid angle of almost 4π to capture as many particles as possible, and have a fast and efficient trigger and read-out system to record as many useful events as possible.

2.2.1. Coordinate system

A coordinate system is defined and used consistently in the ATLAS experiment. The origin of the coordinate system is defined to be the nominal collision point. The Cartesian positive x-axis points toward the centre of the LHC ring while the positive y-axis points vertically upward. The z-axis is the beam axis and its direction is defined such that the coordinate system is right-handed. The x–y plane is referred to as the transverse plane. Components of a spatial vector in the transverse plane are denoted by a subscript T as in $\mathbf{p_T}$ of vector $\mathbf{p}$.

Because of the symmetry of the detector as well as the expected symmetry of the physical events, cylindrical and spherical coordinates are also often used. As in common mathematical conventions, the azimuthal angle is denoted by ϕ and is measured from the positive x-axis on the transverse plane, while the polar angle is denoted by θ and is measured from the positive z-axis. The radius in cylindrical coordinates is denoted by R. The pseudorapidity η is conventionally used in place of θ and is a monotonic transformation of θ:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{2.6}$$

The magnitude of a spatial vector $\mathbf{p}$ and its components can be related by the pseudorapidity with the formulae

$$\tanh\eta = \frac{p_z}{|\mathbf{p}|} \tag{2.7}$$

and

$$\cosh \eta = \frac{p_{\mathrm{T}}}{|\mathbf{p}|}. \tag{2.8}$$

In the ultrarelativistic limit, the pseudorapidity of a particle converges to the rapidity y[†]:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{2.9}$$

where E is the energy of the particle. Differences in rapidity are Lorentz invariant under boosts in the z-direction, and thus transform additively. The distance in the rapidity–azimuth space ΔR[‡]:

$$\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2} \tag{2.10}$$

is often used to quantify the directional proximity of two measured particle tracks.

2.2.2. Overview of the detector

The ATLAS detector is a massive machine made up of layers of detector components that surround the collision point (Figure 2.3). The entire detector is 25 m high and 44 m long and weighs approximately 7000 t. Components of the detector are arranged in a cylindrical manner, with the axis of symmetry being the beam axis. The detector has a forward-backward reflectional symmetry and a nominal eight-fold azimuthal symmetry.

The ATLAS detector consists of three main detector subsystems: the inner detector, the calorimeters and the muon spectrometer. The inner detector is the innermost layer of the detector and its mission is to trace charged particles that traverse it. Outside of the inner detector are the calorimeters. The calorimeters measure the energies of particles when the particles interact with them. The outermost part of the detector is the muon spectrometer, which is designed to detect muons and measure their momenta. Besides the three detector subsystems, the ATLAS detector is also equipped with a magnet system that bends charged particle trajectories, which helps the identification and measurement of these particles.

The subsystems, which will be discussed in greater detail in the following subsections, are designed to work together to reconstruct and identify particles in the following way: Particles produced by the collisions are first met by the inner detector. If the particles are charged, they interact with the inner detector and leave "hits" while traversing it. These hits can then be used to reconstruct the trajectories, or "tracks", of the particles. Alone, a track is useful in determining the travelling direction of an individual charged particle. Together, tracks can be used to reconstruct the "primary vertex", the precise position where

[†] Not to be confused with the Cartesian y-coordinate, even though they conventionally, and inconveniently, share the same symbol.

[‡] Not to be confused with changes or differences in R, the radius in cylindrical coordinates.

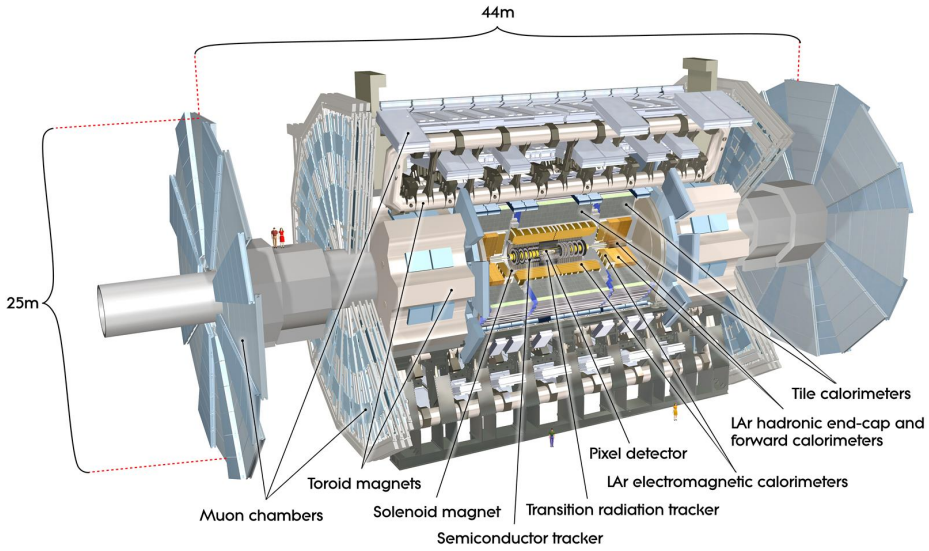


Figure 2.3.: A computer-generated cutaway view of the ATLAS detector [68].

the collision occurred, or "secondary vertices", the position where short-lived† particles decayed into other particles before reaching the detector. Tracks are bent by the magnetic field generated by the superconducting solenoid surrounding the inner detector. This allows the determination of the charges and momenta of particles. Leaving the inner detector, particles will then travel into the calorimeters where they interact with the materials and deposit energy into them. The deposited energy is then measured. There are multiple calorimeters in the ATLAS detector. They interact differently with different types of particles, producing differential responses that help to identify the particles. By associating tracks in the inner detector with energy deposits in the calorimeters, both the direction and the energy of a particle can be precisely determined. While most particles are stopped by the calorimeters where they deposit all of their energies, muons usually pass through the calorimeters with only very limited interactions. Muons leaving the calorimeters are met by the muon spectrometer. Somewhat similar to the inner detector, the muon spectrometer registers hits when muons travel through it and reconstruct tracks from it. The muon spectrometer is provided with a magnetic field generated by a superconducting toroidal magnet. Similar to the case in the inner detector, the magnetic field in the muon spectrometer bends muon trajectories so that their charges and momenta can be measured. Figure 2.4 provides a neat summary of how different particles can be detected by the ATLAS detector.

† But long-lived enough to have travelled a measurable distance before they decay

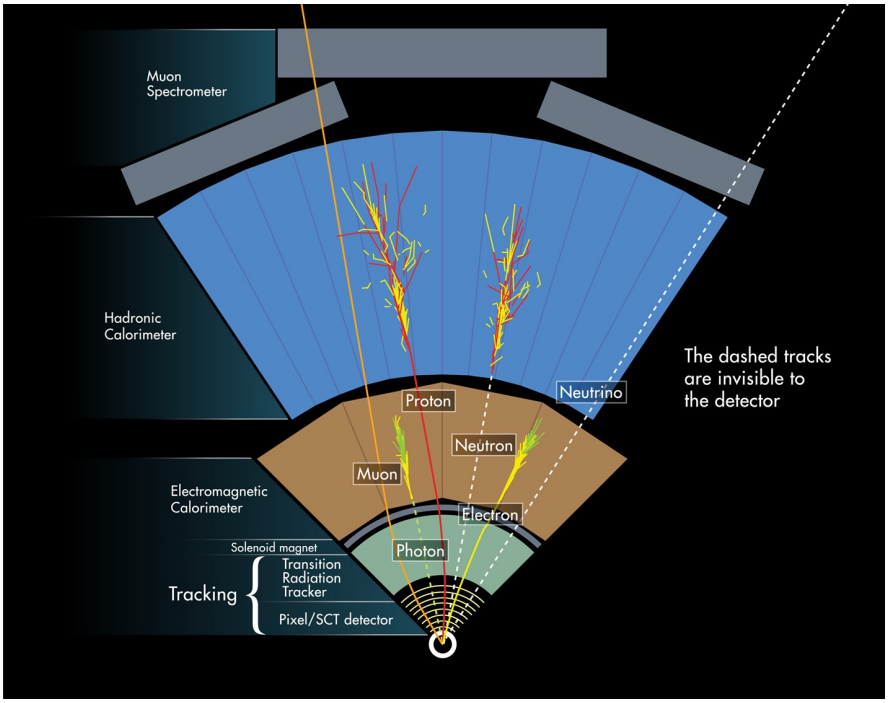
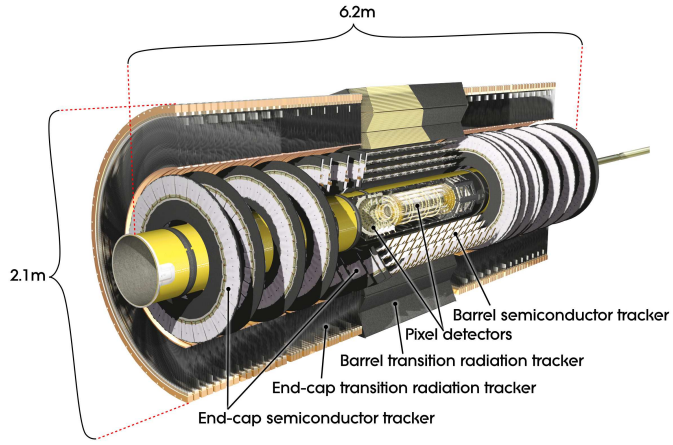


Figure 2.4.: A schematic illustration showing a sector of a cross section of the ATLAS detector and how different particles are typically detected [69].

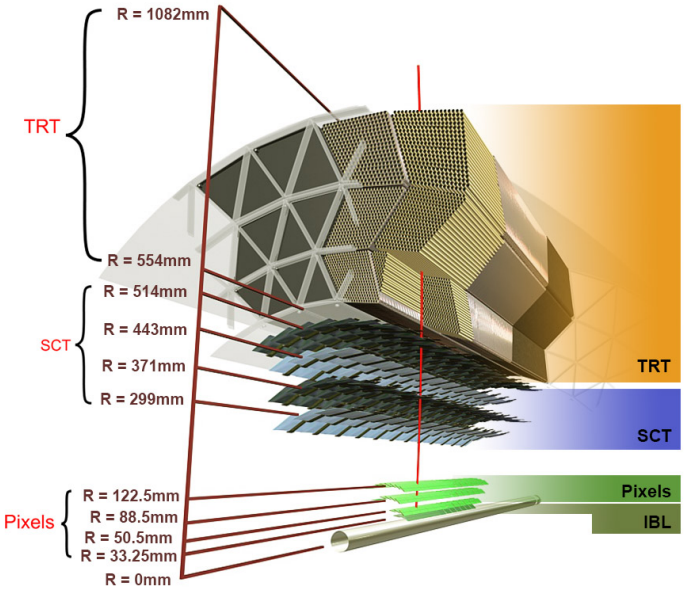
In Run 2, the LHC delivered pp collisions at a rate of 40 million bunch crossings per second. With such a rate, it is practically impossible to process and store all of the data collected. Because of this, the ATLAS detector has a trigger system that decides what data are necessary for analysis and should therefore be kept for further processing, while discarding the rest to save the data processors and storage from being flooded by the overwhelming amount of data.

2.2.3. The inner detector

The inner detector (ID) is the innermost subsystem that records trajectories of charged particles as "hits", which are then reconstructed as "tracks". The ID is placed closest to the collisions, so that it can precisely determine the interaction vertices. Since the main purpose of the ID is not to measure the energies of particles, it is designed to work with a minimal amount of material that particles can interact with, so that the energies of the particles are mostly unaltered when they leave the ID and enter the calorimeters. The ID consists of four subdetectors: the insertable b-layer (IBL), the pixel detector (Pixels), the semiconductor tracker (SCT) and the transition radiation tracker (TRT). Cutaway views of the ID are illustrated in Figure 2.5.



(a)



(b)

Figure 2.5.: Computer-generated cutaway views of the ATLAS inner detector (a) before and (b) barrel after the insertion of the IBL [67, 70].

Insertable b-layer

The IBL is the innermost layer of the ID. It was added to the ID during the long shutdown of the LHC in 2014. The main reason for the addition is to minimise the risk of tracking efficiency lost due to radiation damage to the original innermost layer of the ID. The addition also allows more precise tracking and vertexing, which is especially beneficial to physics performance that depends on the precise measurements of displaced secondary vertices or impact parameters, such as the identification of b-quark-initiated jets or hadronically decaying τ leptons. The IBL is a cylinder of silicon pixel sensors. Each pixel has a size of $(50\,\mu\text{m}, 250\,\mu\text{m})$ in the (ϕ, z) direction and the centres of the sensors are at a radius of 33.25 mm from the beam axis.

Pixel detector

The pixel detector is a precision tracking detector that covers the region $|\eta| < 2.5$. It has three concentric cylindrical layers of silicon pixel sensors in the "barrel" region and three disks of the same sensors on each side in the "end-cap" region. In the barrel region, the three cylindrical layers are 50.5 mm, 88.5 mm and 122.5 mm away from the beam axis respectively and cover the region $|z| < 400.5$ mm. In the end-cap region, the centres of the three disks are 495 mm, 580 mm and 650 mm away from the nominal collision point respectively and cover the region $88.8\,\text{mm} < R < 149.6\,\text{mm}$. Every pixel sensor is identical and the minimum pixel size is $50\,\mu\text{m} \times 400\,\mu\text{m}$. Arranged in a specific configuration, the layers of sensors together have an intrinsic accuracy of $(10\,\mu\text{m}, 115\,\mu\text{m})$ in the (ϕ, z) direction in the barrel region and in the (ϕ, R) direction in the end-cap region.

Semiconductor tracker

The SCT is another precision tracking detector but has slightly less granularity than the pixel detector. It relies on layers of silicon strip detectors placed in alternating orientations to measure every spatial coordinate. Each track is expected to cross eight layers, which marks four space points. In the barrel region, strips are placed parallel to the beam axis or along the ϕ direction and cover the region $299\,\text{mm} < R < 514\,\text{mm}$ and $|z| < 749$ mm. In the end-cap region, strips are placed running radially or along the ϕ direction and cover the region $275\,\text{mm} < R < 560\,\text{mm}$ and $839\,\text{mm} < |z| < 2735\,\text{mm}$. The intrinsic accuracy of the SCT is $(17\,\mu\text{m}, 580\,\mu\text{m})$ in the (ϕ, z) direction in the barrel region and in the (ϕ, R) direction in the end-cap region.

Transition radiation tracker

The TRT is the outermost layer of the ID. The TRT registers hits using cylindrical drift tubes (straws) that extend up to $|\eta| = 2.0$. Each straw is 4 mm in diameter and filled with 70% xenon, 27% carbon dioxide and 3% oxygen. At the centre of each straw, a 30-µm-diameter gold-plated tungsten wire serves as the anode for the drift field. The

straws in the barrel region are 144 cm long and are parallel to the beam direction, while the straws in the end-cap region are 37 cm long and are arranged radially in wheels. Although the TRT does not provide as much spatial information as the precision tracking detectors, it provides a valuable particle identification capability. Since the probability for a particle to emit transition radiation depends on the relativistic Lorentz factor, ultrarelativistic particles such as electrons leave drastically more hits in the TRT than heavier particles such as pions. The information provided by the TRT is important in, for example, distinguishing electrons from hadronically decaying τ leptons.

2.2.4. The calorimeters

The ATLAS detector is built with a set of calorimeters surrounding the ID. The main mission of the calorimeters is to trigger particles to produce showers and subsequently measure their energy and structure. The calorimeters in ATLAS can be divided into the electromagnetic calorimeter (EMCal) and the hadronic calorimeters (HCal), which are designed to measure and contain EM showers and hadronic showers respectively. Figure 2.6 shows a cutaway view of the calorimeters.

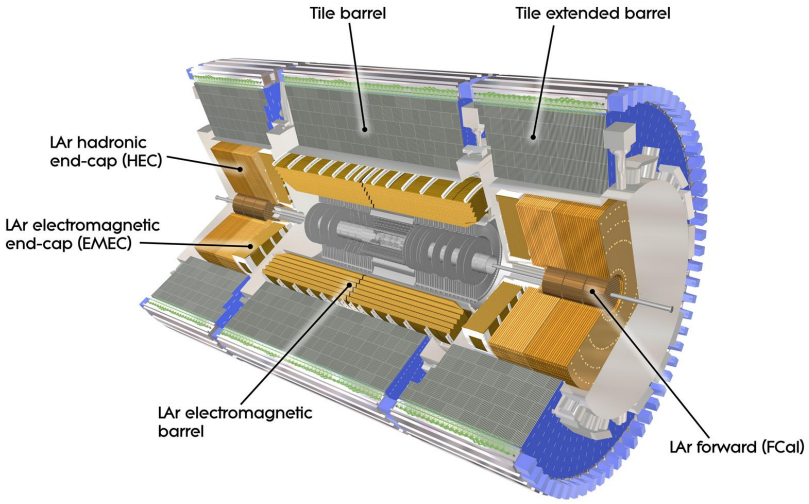


Figure 2.6.: A computer-generated cutaway view of the ATLAS calorimeters [67].

The electromagnetic calorimeter

Electromagnetic (EM) showers are produced by electrons or photons via bremsstrahlung and e^+e^- pair production. EM showers typically develop faster than hadronic showers and

thus have a shorter shower length. For this reason, the EMCal is placed closer to the ID, followed by the hadronic calorimeters. The EMCal has a barrel part that covers $|\eta| < 1.475$ and two end-caps (EMEC) on both sides of the detector that cover $1.375 < |\eta| < 3.2$. All parts of the EMCal use liquid argon (LAr) as the active material and lead plates as absorber. Because of the finer structure of EM showers, the EMCal has a finer granularity than its hadronic counterparts. In most of the covered region, the EMCal has three sampling layers. The first layer has strips of granularity ($\Delta\eta \times \Delta\phi$) 0.0031×0.025 which measures η precisely and a thickness of 4.3 radiation lengths (X_0). The second layer contains most of the EMCal materials. It has a granularity of 0.025×0.025 and a thickness of $16X_0$. The third layer has a coarser granularity of 0.05×0.025 and a thickness of $2X_0$. Other than the three main sampling layers, the EMCal also has a presampler in the region $|\eta| < 1.8$ that corrects for energy lost upstream of the detector. The entire EMCal is around 1.5 nuclear interaction lengths thick, meaning that it is mostly transparent to hadrons.

The hadronic calorimeters

The hadronic calorimeters are designed to induce hadronic showers and measure them. There are three hadronic calorimeters in ATLAS: the scintillator tile calorimeter (TileCal), the LAr hadronic end-cap calorimeter (HEC) and the LAr forward calorimeter (FCal).

The tile calorimeter covers the region $|\eta| < 1.0$ with its barrel and the region $0.8 < |\eta| < 1.7$ with its extended barrels. It uses scintillating tiles as active material and steel as absorber. Both the barrel and the extended barrels have three layers. From the closest to the beam axis to the farthest, the three layers of the barrel are 1.5, 4.1 and 1.8 interaction lengths (λ) thick. For the extended barrels, the three layers are 1.5λ, 2.6λ and 3.3λ thick. The first two layers have a granularity ($\Delta\eta \times \Delta\phi$) of 0.1×0.1 while the last layer has a granularity of 0.2×0.1

The hadronic end-cap calorimeter consists of two wheels installed on each side of the detector directly behind the EMEC. The coverage of the HEC slightly overlaps with that of the TileCal and FCal, spanning from $|\eta| = 1.5$ to $|\eta| = 3.2$. Each end-cap wheel has four layers and uses copper plates as active material. The granularity of the HEC is 0.1×0.1 in $1.5 < |\eta| < 2.5$ and 0.2×0.2 in $2.5 < |\eta| < 3.2$.

The forward calorimeter is located close to the beam line where it can detect particles in the very forward region $3.1 < |\eta| < 4.9$. There are three modules in each end-cap. The module closest to the nominal collision point is made of copper and aims at measuring EM interactions. The remaining two modules are made of tungsten and are devised to measure hadronic interactions. The FCal measures a thickness of approximately 10λ.

2.2.5. The muon spectrometer

The outermost subsystem, which also defines the outline and dimensions of the ATLAS detector, is the muon spectrometer (MS). While most particles typically deposit all of their energy into the calorimeters, and are therefore contained, muons are an exception. At relativistic speed, muons are minimally ionising. As a result, the calorimeters are mostly transparent to muons produced from hard-scattering processes. Therefore, having detectors built outside of the calorimeters are very useful for identifying and reconstructing muons. The MS is built exactly for this reason. The MS consists of two types of chambers: tracking chambers that perform precision tracking and trigger chambers that provide rapid read-out. An overview of the MS is shown in Figure 2.7.

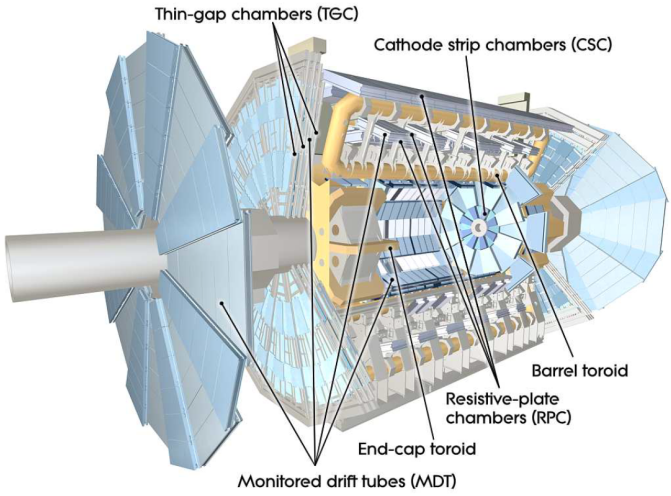


Figure 2.7.: A computer-generated cutaway view of the ATLAS muon spectrometer [67].

Tracking chambers

The tracking chambers enable precision tracking of muons over the range $|\eta| < 2.7$. They make use of monitored drift tubes (MDT) and cathode strip chambers (CSC). The MDTs are pressurised drift tubes filled with 93% argon and 7% carbon dioxide. Tungsten-rhenium wires are used as the anodes of the MDTs. MDTs are used in most part of the tracking chambers. In the barrel region, three layers of MDTs are situated radially 5 m, 7.5 m and 10 m away from the beam axis. In the end-cap region, MDTs are emplaced on four wheels that are located 7.4 m, 10.8 m, 14 m and 21.5 m away from the nominal collision point. The MDTs have a limited counting rate above which they could be damaged. In the forward region where $|\eta| > 2$, the expected particle flux is high enough to exceed such a limit. Therefore, the MDTs on the innermost wheel in the region $|\eta| > 2$ are replaced

by the CSCs. The CSCs are multiwire proportional chambers with wires oriented in the radial direction.

Trigger chambers

The trigger chambers of the MS are resistive plate chambers (RPC) and thin gap chambers (TGC). The fast response and read-out of these chambers allow the trigger system (see Section 2.2.8) to be quickly informed with the multiplicity of muon tracks and their estimated transverse momenta, which are essential for making trigger decisions. The RPCs are used in the barrel to cover the region $|\eta| < 1.05$, while the TGCs are used in the end-caps to cover the region $1.05 < |\eta| < 2.7$. The trigger chambers also complement the tracking chambers by measuring coordinates in the direction orthogonal to that measured by the tracking chambers.

2.2.6. The magnet system

Four superconducting electromagnets are stationed in the ATLAS detector. They create strong magnetic fields that bend charged particle tracks with the Lorentz force. The direction and curvature of the bent tracks can then be used to deduce the charges and momenta of particles. The layout of the four magnets is shown in Figure 2.8.

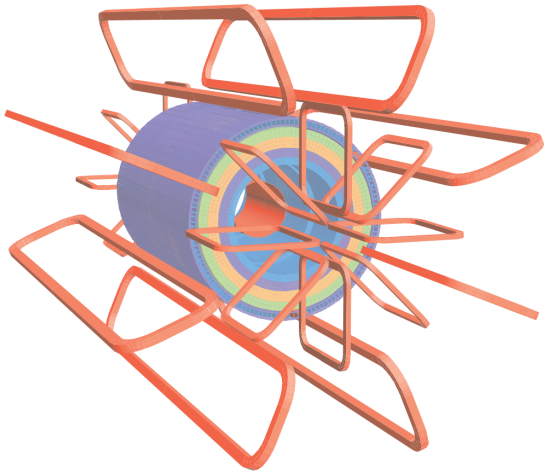


Figure 2.8.: The layout of the ATLAS magnet system [67].

The magnetic field in the inner detector is provided by a solenoid magnet located between the ID and the calorimeters. It has a magnetic field strength of 2 T and is pointing in the z-direction with a high uniformity. As a result, charged particles in the ID follow helical paths with axes parallel to the beam direction.

The remaining magnets are toroidal, which is a defining feature of the ATLAS detector[†]. One of the three toroid magnets is in the barrel region, while the other two are end-cap magnets deployed on both sides of the detector. The toroid magnets are used to supply bending power to the MS. The magnet strengths are approximately 0.5 T in the barrel region and 1.0 T in the end-cap region.

2.2.7. Crack regions

Some regions of the detector are only partially and insufficiently instrumented due to various limitations. These regions are commonly known as "crack" regions. Particles that fall in the crack regions are usually reconstructed with significantly degraded efficiency and quality. One of the crack regions is the transition region between the barrel and the end-cap regions, where detector components are mechanically separated. This corresponds to the region $1.37 < |\eta| < 1.52$. Another crack region is the region $|\eta| < 0.1$. This region is not covered by the MS chambers in order to give way to power and read-out cables for the ID and calorimeters.

2.2.8. The trigger system

During the Run-2 data taking, the LHC operated using the designed parameter of 25 ns spacing between proton bunches. This means that 40 million bunch crossings occurred per second. With the overall average pile-up of $\langle\mu\rangle = 33.7$, a rate of over 1.3 billion collisions per second can be expected. This amounts to a raw data acquisition rate in the order of petabytes per second, which is practically impossible to fully process and store. To resolve this, the ATLAS detector has a trigger system that selects only events that are essential to physics analyses in order to reduce the event rate to a manageable level. Throughout Run 2, a two-level trigger system is used.

The Level-1 (L1) triggers are implemented in specialised hardware installed in the detector and are able to make fast decisions. They are tasked with reducing the event rate from 40 MHz down to 100 kHz. The L1 trigger decisions are made based on information from the calorimeters and the trigger chambers of the MS. Information from the ID is not used because track fitting is a computationally intensive process. The L1 triggers are fired when signatures of muons, photons, electrons, jets or hadronic τ decays are detected, or when an event has large missing transverse momentum or large total transverse energy. Besides a simple binary decision, the L1 trigger system also identifies regions of interests, which are regions of the detector where data are fully read out and further processed. Decisions from the L1 triggers are made within 25 µs of the associated bunch crossing.

Events that are kept by the L1 triggers are further filtered by the software-based High-level triggers (HLT). With a reduced event rate of 100 kHz and defined regions of interests, the HLTs are able to make more elaborate decisions. Physics objects are reconstructed

[†] ATLAS is a genius (but a bit forced in the author's opinion) acronym for A Toroidal LHC ApparatuS.

and tracking information from the ID is included when the HLT decisions are made. This enables the HLTs to take the transverse momenta of reconstructed physics objects as well as the overall topology of the event into account. The event rate is reduced to approximately 1 kHz by the HLTs. The data of the triggered events can then be compressed, transferred and permanently stored.

The trigger selections can usually be characterised by the multiplicity, transverse momentum thresholds and isolation conditions of physics objects. Some trigger selections are prescaled, meaning that only a certain fraction of the events that meet the selection criteria are stored. This allows events which have features that are too common to still be sampled without overloading the system. These are typically events with only low-energy objects.

Chapter 3.

Object reconstruction and identification

In the fields of observation chance favours only the prepared mind.

— Louis Pasteur

The ATLAS detector is an incredibly convoluted machine with many components and literally millions of data readout channels. When particles traverse the detector, the traces they leave are recorded as nothing but a bunch of disjointed digital signals. Without proper processing, these raw data will not be of much use. Therefore, the first step of any physics analysis in the ATLAS experiment is to convert these raw data into meaningful "physics objects". In view that these are common efforts shared by all physics analyses, the reconstruction of SM particles are harmonised within the collaboration. The algorithms that are used to reconstruct and identify physics objects are products of collaborative work, and are developed, maintained and constantly improved by various dedicated working groups.

In this chapter, we will introduce the algorithms for reconstructing and identifying objects that are relevant to the search for LFV $Z \to \ell\tau$ decays. Particularly, attention will be given to the reconstruction and identification of hadronic tau decays, which are crucial for the search and are work that the author has actively and proudly contributed to.

3.1. Jets and flavour tagging

Given that the LHC is a hadron collider, gluons and quarks can be found in almost every collision event. The hadronisations of these gluons and quarks produce showers of particles that travel in directions close to that of the original particles, also known as jets. By reconstructing these jets, we can learn about the kinematic properties of the particles that initiated them, and sometimes even their flavour.

3.1.1. Jet finding and reconstruction

If we were to manually and visually find a jet in an event, it would be rather simple and intuitive – a jet is just a collimated shower of particles. Yet, of course, in order to analyse the billions of recorded events, jet finding must be an automated and objectively defined process. The algorithm used to reconstruct the jets must be able to deliver reproducible results that can be confronted with theories and other experiments. For this, a topological clustering algorithm [71] and a sequential recombination algorithm known as the anti-k_t algorithm [72] are used in the ATLAS experiment.

The reconstruction of a jet begins with the grouping of energy deposits in the calorimeters into so-called "clusters". Calorimeter cells which have a high signal-to-expected-noise ratio (ζ) are used to seed the clusters. Calorimeter cells that are topologically connected to a seed and with ζ passing a certain threshold are joined with the seed to form an initial cluster. The cluster is then grown by repeatedly collecting neighbouring calorimeter cells with ζ above the threshold into the cluster. Each cluster can be interpreted as a massless pseudoparticle which has its own energy and momentum. The energy of a cluster is the signal-weighted sum of the energies in the cells and the direction of a cluster is that of the signal-weighted barycentre of the cells. Furthermore, the shape of a cluster can be quantitatively described by its moments, which are important variables for identifying the origin of a jet. The kinematic properties of clusters are initially calibrated at the electromagnetic scale. Since the calorimeters in the ATLAS detector are non-compensating, i.e. they respond differently to electromagnetic and hadronic particle showers, calibration is needed to correct the scale for hadronic clusters. For that, a local hadronic cluster calibration (LC) scheme is used [73].

In a sequential recombination algorithm, jets are reconstructed by repeatedly combining clusters based on their geometric proximity and energies. For every pair of clusters, the "distance"

$$d_{ij} = \min(p_{\mathrm{T},i}^{2P}, p_{\mathrm{T},j}^{2P}) \frac{(\Delta R_{i,j})^2}{R^2}, \tag{3.1}$$

where i and j are indices for the two clusters, p_{T} is the transverse momentum of the cluster, ΔR is the distance between clusters in the rapidity–azimuth space, and P and R are constant parameters of the algorithm, between them is calculated. In addition, the distance

$$d_{i\mathrm{B}} = p_{\mathrm{T},i}^{2P} \tag{3.2}$$

between every cluster and the beam is also calculated. Then, repeatedly, the smallest of $\{d_{ij}\}$ and $\{d_{i\mathrm{B}}\}$ is identified. If it is a d_{ij}, the two clusters are combined into a "protojet". Otherwise, the cluster/protojet with the smallest $d_{i\mathrm{B}}$ is considered a completely reconstructed jet and is removed in subsequent iterations. The parameter R is usually referred to as the jet radius or cone size and is the maximum angular distance between a cluster in a jet and the central axis of the jet. In ATLAS, R is usually chosen to be 0.4.

When the parameter $P = -1$, the algorithm is called the anti-k_t algorithm, which is the standard algorithm for jet reconstruction in ATLAS. The anti-k_t algorithm prioritises the grouping of low-energy clusters into a close-by high-energy cluster. This avoids reconstructing multiple high-energy jets which are close to each other and results in circular-cone-shaped jets of maximum radius R. The anti-k_t algorithm exhibits a number of desirable properties. By construction, the algorithm is collinear- and infrared-safe. To be collinear-safe means that if particles in a jet are split into multiple particles that travel in the same direction, the reconstructed jet will not change as a result. To be infrared-safe means that soft emissions cannot radically impact the jet reconstructions. Moreover, the anti-k_t algorithm has been shown to be particularly insensitive to underlying events and pile-up. These are very important qualities for a jet reconstruction algorithm as they ensure that the observed jets can faithfully represent the particles from a hard collision event that initiated them. They also make comparisons with theoretical predictions practical as QCD processes at low-energy scales, which are difficult to calculate, cannot significantly alter the reconstruction of energetic jets. On top of all the above mentioned merits, the anti-k_t algorithm, implemented using the FastJet library [74], is also computationally fast.

3.1.2. Flavour tagging

The LHC produces an abundance of top quarks. In the search for $Z \to \ell\tau$ decays, events such as $t\bar{t} \to b\bar{b}W^+W^- \to b\bar{b}\ell^\pm\tau^\mp\nu\bar{\nu}$ are a significant contribution to the SM background if they could not be effectively identified. The top quark decays via weak interaction into a bottom quark and a W boson 99.8% of the time. Therefore, events with top quarks can be effectively identified by recognising b-quark-initiated jets (b-jets), a process also known as b-tagging.

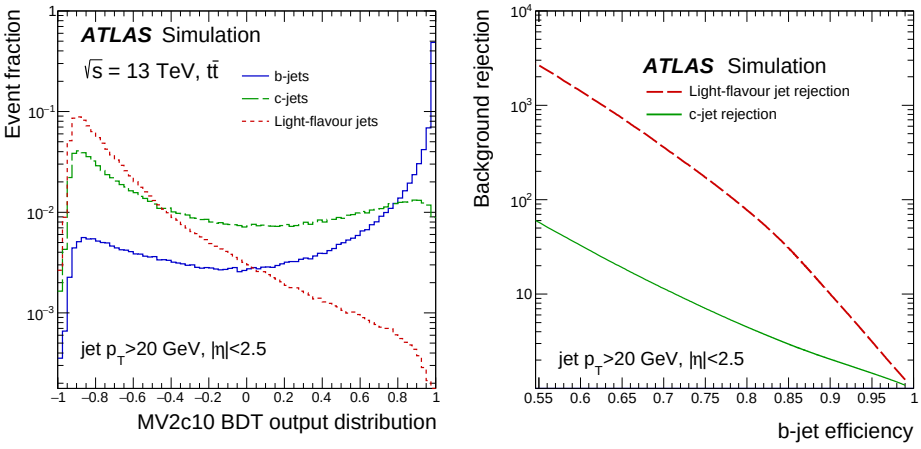
While a b quark always hadronises immediately after its production, the b-hadrons created from the hadronisation have an appreciable lifetime of around 1.5 ps. This relatively long lifetime allows a 50-GeV b-hadron to travel approximately 3 mm on average before it decays. Because of this, a significantly displaced secondary vertex is expected in b-jets, which is the most important feature that is exploited by b-tagging algorithms.

In the ATLAS experiment, b-tagging is performed using several low-level algorithms that provide intermediate outputs which are then fed into a high-level tagging algorithm that creates a final discriminant using multivariate analysis (MVA) [75].

There are four low-level algorithms: IP2D, IP3D, SV1 and JetFitter. The IP2D and IP3D algorithms are likelihood-based algorithms that exploit the large impact parameters of tracks from the b-hadron decay. While the IP2D algorithm only makes use of the transverse impact parameter significance, the IP3D algorithm also takes into account the impact parameter significance in the z-direction. The SV1 and JetFitter algorithms are both vertex reconstruction algorithms. The SV1 algorithm explicitly reconstructs a single displaced secondary vertex in a jet, whereas the JetFitter attempts to reconstruct the full b-hadron decay chain by reconstructing the multiple vertices.

The outputs of the low-level algorithms are combined into a final discriminant by a high-level tagging algorithm. Two high-level tagging algorithms have been developed: the MV2c10 algorithm that uses boosted decision trees (BDT) [76] and the DL1 algorithm that uses a deep feed-forward neural network. For a high target b-tagging efficiency, the MV2c10 and DL1 algorithms provide similar rejection against light-quark- or c-quark-initiated jets.

Figure 3.1 shows the expected output distributions and the background rejection against the signal efficiency of the MV2c10 algorithm.



(a) **(b)**

Figure 3.1.: The (a) expected output distributions and (b) background rejection against the signal efficiency of the MV2c10 algorithm [75].

3.2. Hadronic τ decays

Decays of a τ lepton can be either leptonic ($\mathcal{B} \approx 35.2\%$) or hadronic ($\mathcal{B} \approx 64.8\%$). Either type of decay produces neutrinos that can be partially reconstructed as part of the missing transverse momentum of the event. The remaining products are the visible decay products. The visible decay product of a leptonic τ decay is an electron or muon, which can be reconstructed like any other isolated electrons or muons using the algorithms later described in Sections 3.3 and 3.4. The visible decay products of a hadronic τ decay ($\tau_{\text{had-vis}}$), on the other hand, requires more specialised algorithms, which are detailed in this section. The performance of these algorithms is a crucial factor to the sensitivity of the search for $Z \to \ell\tau$ decays.

The reconstruction and identification of $\tau_{\text{had-vis}}$ are intricate tasks. Due to the complexity, these tasks could often benefit from the use of MVA and machine learning techniques

to automatically take into account the large number of varying factors. During the Run-2 data-taking period, significant improvement in the reconstruction and identification performance has been made by introducing machine learning. In particular, the author has significantly contributed to the development of an MVA-based $\tau_{\text{had-vis}}$ energy calibration algorithm and an electron rejection algorithm, which are both widely used in the collaboration currently. Such developments are valuable to many analyses that consider events with $\tau_{\text{had-vis}}$ in the final states, including, for example, the $H \to \tau\tau$ measurement.

3.2.1. Baseline reconstruction

There are two different approaches for reconstructing a $\tau_{\text{had-vis}}$. The simpler of the two uses mainly information from the calorimeters and considers all the visible decay products within a circular cone as one single collective object [77, 78]. We will refer to this approach as the baseline reconstruction.

As hadrons, $\tau_{\text{had-vis}}$ can almost always be reconstructed as a hadronic jet. Therefore, the natural starting point of the baseline reconstruction is the jets reconstructed by the method mentioned in Section 3.1.1. All reconstructed hadronic jets are initially considered as possible candidates of $\tau_{\text{had-vis}}$.

Tracks measured in the inner detector are associated to the $\tau_{\text{had-vis}}$ candidates if they have matching directions. The direction of a $\tau_{\text{had-vis}}$ candidate is defined by the barycentre of the clusters in the jet, calibrated with the LC scheme and assumed to have zero rest mass. Tracks that are in the "core region" $\Delta R < 0.2$ around the $\tau_{\text{had-vis}}$ direction are the associated tracks. Since a $\tau_{\text{had-vis}}$ consists of one or three charged particles over 99.9% of the time, it can be expected that jets originated from $\tau_{\text{had-vis}}$ have exactly one or three associated tracks. $\tau_{\text{had-vis}}$ candidates are classified into "1-prong" or "3-prong" depending on the number of associated tracks. Tracks that are in the "isolation region" $0.2 < \Delta R < 0.4$ are also considered, but only for identification against quark- or gluon-initiated jets, which will be detailed later.

The momentum of a $\tau_{\text{had-vis}}$ candidate is first calculated by summing the four-momenta of calorimeter clusters in the core region. Since the clusters are calibrated with the LC scheme, the $\tau_{\text{had-vis}}$ momentum obtained at this stage is referred to as the momentum at LC scale (p^{LC}). However, the LC scheme is neither optimised for the cone size $\Delta R = 0.2$, nor the expected hadronic composition of $\tau_{\text{had-vis}}$. This is improved by applying a correction derived using simulations. The correction is a function of the transverse momentum and pseudorapidity at LC scale. The corrected momentum defines a tau energy scale (TES) known as the "calo TES".

The above procedures are the baseline reconstruction for $\tau_{\text{had-vis}}$. It is a relatively simple yet robust method of $\tau_{\text{had-vis}}$ reconstruction. It has an advantage of being relatively easy to model due to the fact that it is only based on averaged properties of calorimeter clusters, and hence does not rely heavily on detailed simulations of jet substructures. It also uses computational resources efficiently by capitalising on the regular jet reconstruction.

3.2.2. Substructure reconstruction

Even though it is simple and robust, the baseline reconstruction is omitting a lot of valuable information about the τ decay that could be used to improve analyses. To recover that information, a more elaborated reconstruction method is used to complement the baseline reconstruction. This method attempts to reconstruct the individual τ decay products instead of treating all of them as a collective object [79]. We will refer to this approach as the substructure reconstruction. A $\tau_{\text{had-vis}}$ that is reconstructed with this approach is sometimes called a "pantau" within the ATLAS collaboration.

Similar to the baseline reconstruction, the substructure reconstruction is also seeded by jets. However, the momentum of $\tau_{\text{had-vis}}$ is not simply calculated by adding the energies and directions of all the clusters in the reconstructed jet cone. Instead, it is calculated as the sum of individually reconstructed charged hadrons ($h^{\pm}$, predominantly pions $\pi^{\pm}$) and neutral pions (π^0). The major challenge of this approach is to disentangle the energy deposits of the $h^{\pm}$'s and π^0's. To achieve that, the hadrons are sequentially reconstructed using an algorithm called Tau Particle Flow. The main idea of Tau Particle Flow is to first make use of the inner detector tracks to reconstruct the $h^{\pm}$'s, and then use that information to approximate the $h^{\pm}$ energy deposits in the EM calorimeter, so that they can be disentangled from the π^0 energy deposits.

In Tau Particle Flow, $h^{\pm}$'s are first identified using the associated tracks. For $\tau_{\text{had-vis}}$ that has an energy around or below 100 GeV, the $h^{\pm}$'s can be correctly identified by the tracks about 98% of the time. Then for each $h^{\pm}$, both the direction and magnitude of the momentum can be reconstructed using the track. Since π^0 showers (dominantly 2γ) very rarely extend beyond the EMCal, the energy deposits in the HCal that matches the direction of the $h^{\pm}$ track can be fully assigned to the $h^{\pm}$. By subtracting the energy deposits in the HCal from the energy calculated using the track, one can then estimate the amount of $h^{\pm}$ energy deposits in the EMCal.

Then, π^0 candidates are reconstructed using clusters in the EMCal. If the π^0 clusters are overlapping with a reconstructed $h^{\pm}$ shower, the estimated energy deposits from the $h^{\pm}$ are subtracted from the clusters. This systematically disentangles the energy deposits of the $h^{\pm}$'s and π^0 and avoids double counting. After that, a BDT classifier is used to identify real π^0's from the π^0 candidates by exploiting the relatively regular shape and size of π^0 showers.

At last, after the $h^{\pm}$'s and π^0's are reconstructed, the full $\tau_{\text{had-vis}}$ can then be reconstructed by treating the $h^{\pm}$'s and π^0's as constituents and summing their momenta. The calculated momentum defines a TES known as the "pantau TES".

By incorporating information from the inner detector, the substructure reconstruction is able to determine the direction of $\tau_{\text{had-vis}}$ more precisely than the baseline reconstruction. For low-p_{T} $\tau_{\text{had-vis}}$ ($p_{\text{T}} \lesssim 100\,\text{GeV}$), the momentum resolution is also superior. Even more importantly, it allows one to determine the actual decay mode of $\tau_{\text{had-vis}}$. A BDT classification algorithm has been developed to achieve that by considering the number of reconstructed $h^{\pm}$'s and π^0's and correcting for potential misreconstructions of π^0's.

The algorithm is capable of classifying a $\tau_{\text{had-vis}}$ into one of the five modes: 1p0n, 1p1n, 1pXn, 3p0n and 3pXn, where the number before 'p' represents the number of $h^{\pm}$'s and the number before 'n' represents the number of π^0's, with 'X' denoting > 1 in 1pXn and ≥ 1 in 3pXn. This opens up the possibility to perform decay-mode-specific calibrations and measurements.

3.2.3. Energy calibration

The calorimeters and the inner detector trackers measure energy/momentum by very different means, which lead to very different performance. The calorimeters measure energy by absorption. They have a principal energy resolution (σ_E) that is independent of the direction and improves with the energy of the initial particle:

$$\frac{\sigma_E}{E} \propto \frac{1}{\sqrt{E}}. \tag{3.3}$$

On the contrary, the inner detector trackers measure momentum by the curvature of magnetically bent track. The momentum resolution (σ_p) of a tracker is at its best when measuring low-energy particles that travel in the transverse direction:

$$\frac{\sigma_p}{p} \propto p_{\text{T}}. \tag{3.4}$$

These differences imply that the calo TES and pantau TES have very different resolution performance. Alone, neither of the calo TES and pantau TES is optimal for the vast range of physics analyses performed in ATLAS. In this regard, an advanced energy calibration that combines the merits of both the calo TES and pantau TES has been developed. The calibration relies on BDT regression algorithm (also known as boosted regression trees, or BRT) [76, 80] and is referred to as the MVA TES [81].

As an intermediate starting point for the machine learning, the calo TES and pantau TES are first combined using a simpler algorithm. A "combined TES" is defined by the inverse-resolution-weighted average of the calo TES and pantau TES. The transverse momentum in the combined TES is calculated as

$$p_{\text{T}}^{\text{comb}} = \frac{\left(\sigma_{\text{p}}^2 - \rho_{\text{c,p}}\sigma_{\text{c}}\sigma_{\text{p}}\right) p_{\text{T}}^{\text{calo}} + \left(\sigma_{\text{c}}^2 - \rho_{\text{c,p}}\sigma_{\text{c}}\sigma_{\text{p}}\right) p_{\text{T}}^{\text{pantau}}}{\sigma_{\text{c}}^2 + \sigma_{\text{p}}^2 - 2\rho_{\text{c,p}}\sigma_{\text{c}}\sigma_{\text{p}}}, \tag{3.5}$$

where σ_{c} and σ_{p} are the resolutions of the calo TES and pantau TES respectively, and $\rho_{\text{c,p}}$ is the correlation coefficient of the two TESs. The resolutions and correlations are estimated as functions of p_{T}, $|\eta|$ and reconstructed decay modes based on simulations.

The combined TES is then further calibrated by a BDT regression algorithm. The major inputs to the BDT are the transverse momenta at different scales, p_{T}^{LC}, $p_{\text{T}}^{\text{calo}}$ and $p_{\text{T}}^{\text{comb}}$. Additionally, the energy-weighted averages of some cluster moments are also fed into the BDT. These variables are found to be correlated with the residues of the calo TES (defined as $p_{\text{T}}^{\text{calo}} - p_{\text{T}}^{\text{truth}}$, where $p_{\text{T}}^{\text{truth}}$ is the true p_{T} of the $\tau_{\text{had-vis}}$) according to simulations.

Table 3.1.: Input variables to the BDT of the MVA TES algorithm. The cluster variables are the energy-weighted averages of all the clusters in the core region. More detailed descriptions of the cluster moments can be found in Reference [73].

Variable	Description
	Event variables
$\langle\mu\rangle$	Average number of interactions per bunch crossing
N_{PV}	Number of pile-up vertices
	$\tau_{\text{had-vis}}$ variables
$p_{\mathrm{T}}^{\mathrm{comb}}$	Transverse momentum at the combined TES
$p_{\mathrm{T}}^{\mathrm{LC}}/p_{\mathrm{T}}^{\mathrm{comb}}$	Transverse momentum at LC scale divided by $p_{\mathrm{T}}^{\mathrm{comb}}$
$p_{\mathrm{T}}^{\mathrm{pantau}}/p_{\mathrm{T}}^{\mathrm{comb}}$	Transverse momentum at the pantau TES divided by $p_{\mathrm{T}}^{\mathrm{comb}}$
η^{pantau}	Pseudorapidity from substructure reconstruction
N_{tracks}	Number of associated tracks
Υ	Relative energy difference between $h^{\pm}$'s and π^0's: $(E(h^{\pm}) - E(\pi^0)) \,/\, (E(h^{\pm}) + E(\pi^0))$
BDT(1p0n vs 1p1n)	BDT classification score for differentiating 1p0n and 1p1n
BDT(1p1n vs 1pXn)	BDT classification score for differentiating 1p1n and 1pXn
BDT(3p0n vs 3pXn)	BDT classification score for differentiating 3p0n and 1pXn
	Cluster variables (energy-weighted average)
$\lambda_{\mathrm{centre}}$	Distance of the shower centre from the calorimeter front face measured along the shower axis
$\langle\lambda^2\rangle$	Second moment in λ, the distance of cells from the shower centre along the shower axis
$\langle\rho\rangle$	First moment in ρ, the energy density
f_{PS}	Energy fraction in the presamplers
P_{EM}	Classification probability that the shower is electromagnetic

The residues and the correlations are believed to be mainly caused by effects of pile-ups and underlying events. Instead of multiple BDTs for $\tau_{\text{had-vis}}$'s in different $|\eta|$ regions or decay modes, only one BDT for all $\tau_{\text{had-vis}}$'s is trained. Such an inclusive approach allows the BDT to make use of continuous variables such as $|\eta|$ and the BDT outputs of the decay mode classifier to dynamically classify the $\tau_{\text{had-vis}}$'s while calibrating the energy. A full list of input variables and their descriptions are shown in Table 3.1.

The BDT is trained to minimise the mean squared error between its output and the training target, which is the ratio $p_{\mathrm{T}}^{\mathrm{truth}}/p_{\mathrm{T}}^{\mathrm{comb}}$. In other words, the output of the BDT

is a correction factor to $p_{\mathrm{T}}^{\mathrm{comb}}$. The predicted transverse momentum $p_{\mathrm{T}}^{\mathrm{MVA}}$ is then used together with the directions (ϕ and η) from the substructure reconstruction to form the final calibrated $\tau_{\text{had-vis}}$ momentum.

The expected distributions of the *response*

$$\mathcal{R} = \frac{p_{\mathrm{T}}^{\mathrm{reco}}}{p_{\mathrm{T}}^{\mathrm{truth}}}, \tag{3.6}$$

where $p_{\mathrm{T}}^{\mathrm{reco}}$ is the reconstructed p_{T} in one of the TESs, are shown in Figure 3.2 for $\tau_{\text{had-vis}}$ in different modes. From the figures, it can be seen that the MVA TES incorporates both the better core resolution of the pantau TES and the better tail behaviour of the calo TES. It also corrects for the biases of the calo TES. As a figure of merit, the relative resolution (of 68% coverage) is defined as

$$\text{Resolution (68\% coverage)} = \frac{Q_{0.84}(\mathcal{R}) - Q_{0.16}(\mathcal{R})}{2}, \tag{3.7}$$

where $Q_x(\mathcal{R})$ is the x^{th} quantile of $\mathcal{R}$. The expected relative resolutions of the calo, pantau, combined, and MVA TESs are shown in Figure 3.3 as functions of $p_{\mathrm{T}}^{\mathrm{truth}}$. At $p_{\mathrm{T}} < 100\,\mathrm{GeV}$, the MVA TES offers a significantly better resolution than both the calo and pantau TESs. The resolution is improved by a factor of two with respect to the calo TES in this region. At $p_{\mathrm{T}} \gg 100\,\mathrm{GeV}$, the resolution of the MVA TES approaches that of the calo TES.

The improved energy resolution provided by the MVA TES directly impacts the sensitivity of the search for $Z \to \ell\tau$ decays as it affects our ability to separate the signal from backgrounds like $Z \to \tau\tau$.

3.2.4. Jet rejection

The reconstruction of $\tau_{\text{had-vis}}$ candidates provides very little rejection against quark- or gluon-initiated jets. Algorithms have been developed to distinguish real $\tau_{\text{had-vis}}$ from other jets. These algorithms use MVA to exploit features of the tracks and clusters in both the core and isolation regions, as well as the reconstructed π^0's. In particular, the most recently developed algorithm uses recurrent neural network (RNN) to fully harness the discriminating power in low-level input variables of individual tracks and clusters. This algorithm, known as the RNN ID [82], doubles the rejection against jets with respect to its predecessor, which uses only high-level variables that collectively describe the tracks and clusters. This significant improvement in jet rejection is one of the important factors that make the sensitivity of the search presented in this thesis achievable.

Isolated $\tau_{\text{had-vis}}$ typically have one or three associated tracks in the core region and very few energetic tracks in the isolation region. Their showers are on average more collimated than jets, especially gluon-initiated ones. $\tau_{\text{had-vis}}$ with π^0's also have distinctive energy deposit patterns in the EMCal. Furthermore, given the relatively long lifetime of τ leptons,

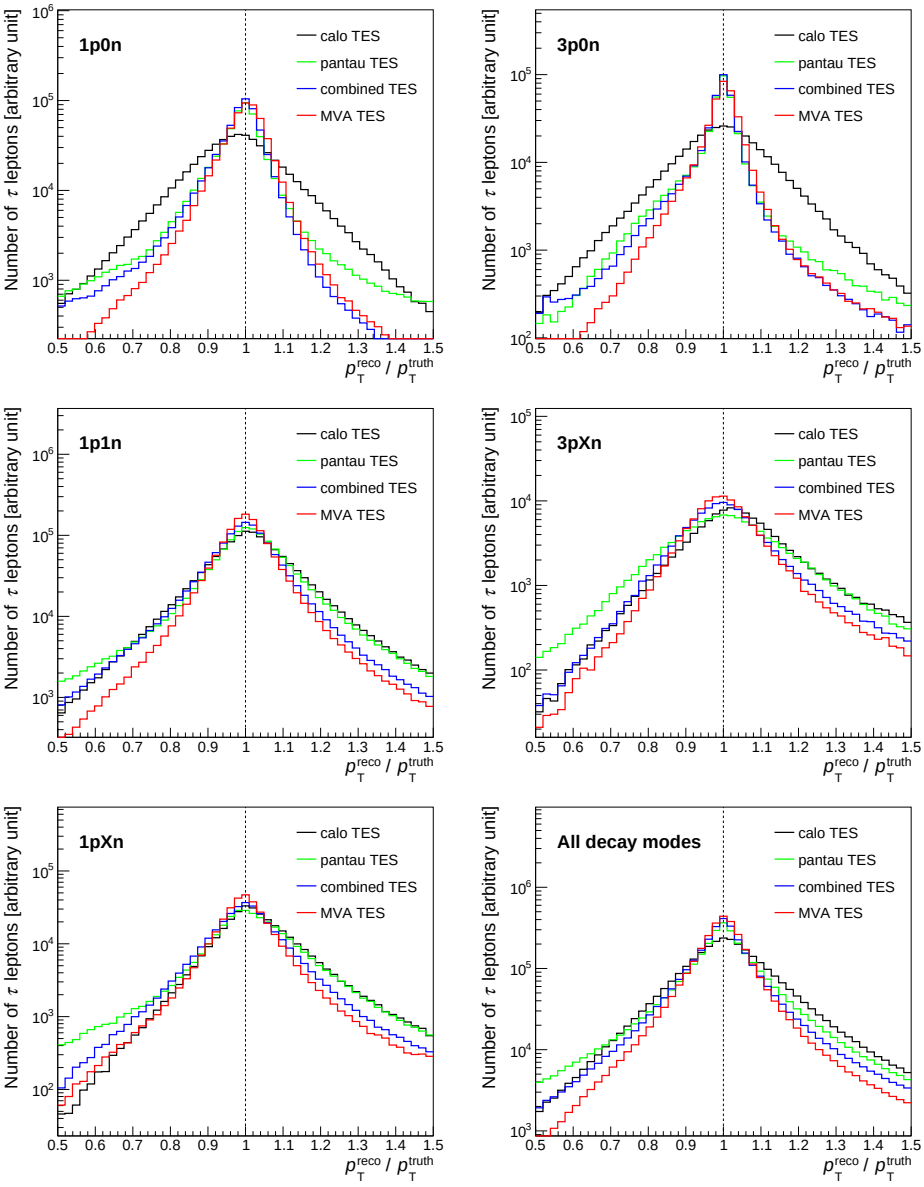


Figure 3.2.: The expected distributions of responses of the calo, pantau, combined and MVA TESs for $\tau_{\text{had-vis}}$'s in different decay modes.

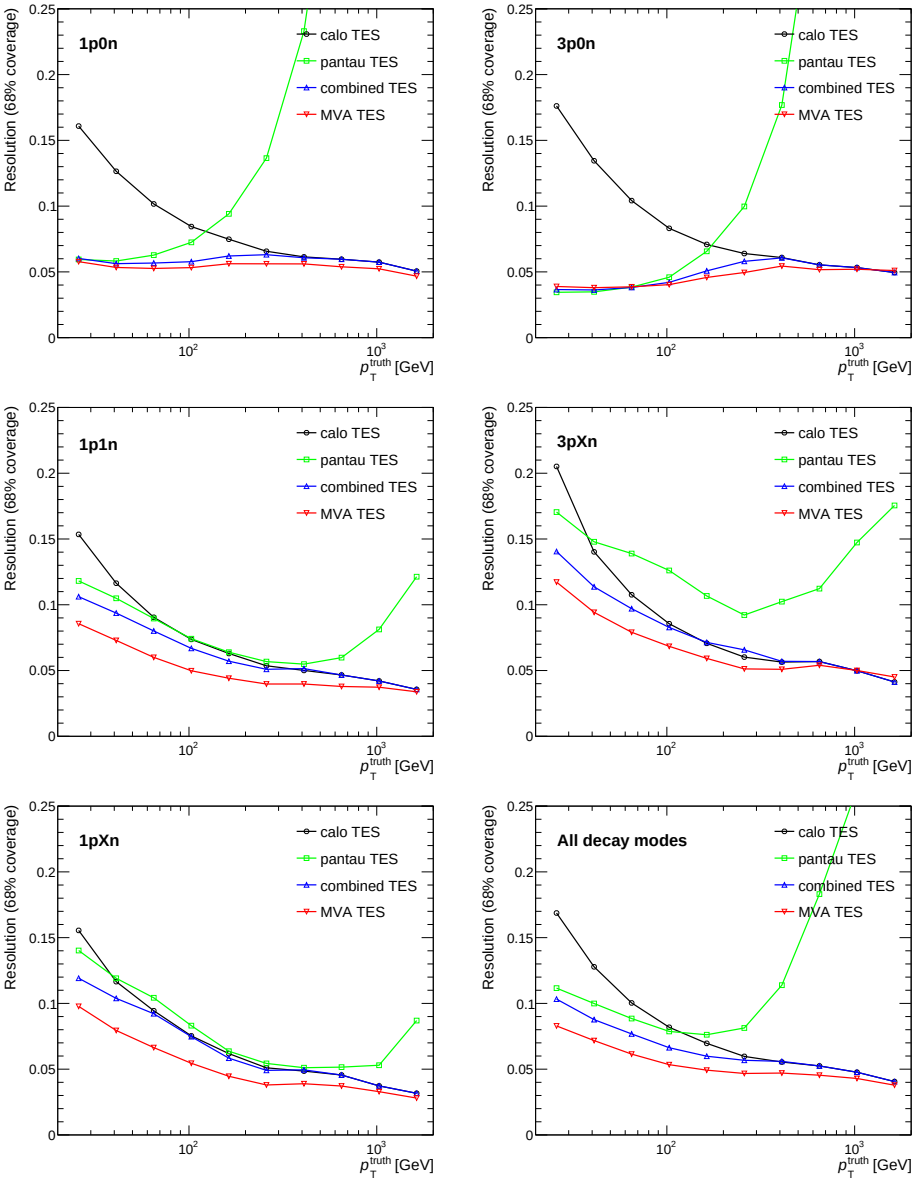


Figure 3.3.: The expected relative resolutions of the calo, pantau, combined and MVA TESs for $\tau_{\text{had-vis}}$'s in different decay modes.

the associated tracks of real $\tau_{\text{had-vis}}$ usually form a displaced secondary vertex and have a large impact parameter with respect to the primary vertex.

Important high-level input variables to the RNN ID include, for example, the central energy fraction (f_{cent}), the invariant mass of the track system (m^{track}) and the transverse flight path significance ($S_{\text{T}}^{\text{flight}}$). f_{cent} quantifies how collimated the shower is by calculating the fraction of energy deposited in the region $\Delta R < 0.1$ to that in the entire core region $\Delta R < 0.2$. m^{track} is defined as the sum of the four-momenta of all the tracks, assuming a pion mass for each track. $S_{\text{T}}^{\text{flight}}$ is the displacement of the secondary vertex in the transverse plane with respect to the primary vertex divided by its estimated uncertainty.

RNNs are neural networks that have internal states which can function as "memory" of the network. The memory can be used to process inputs that are sequences of variable lengths. The RNN ID utilises this feature to take in the variable number of tracks and clusters as low-level inputs. Low-level input variables to the RNN ID include the impact parameters and the number of inner detector hits of the individual tracks, as well as the moments that quantify the longitudinal and radial shapes of the individual clusters.

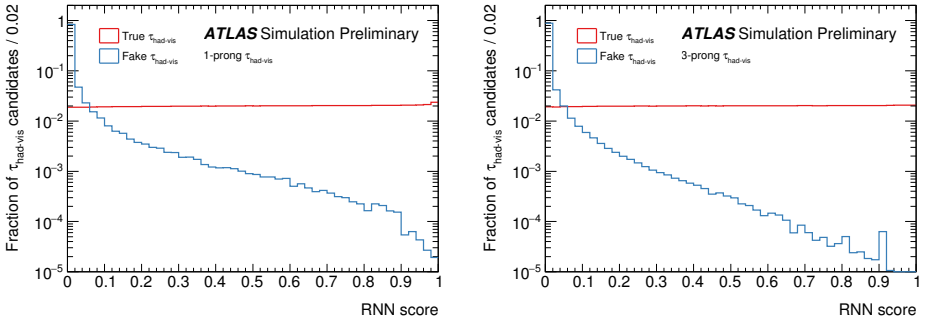
Figure 3.4 shows the expected output distributions and the receiver operating characteristic (ROC) curves of the RNN ID. Four working points (WP), `Tight`, `Medium`, `Loose` and `VeryLoose`, are defined with increasing signal selection efficiencies as indicated in the figure.

3.2.5. Electron rejection

Aside from jets, electrons are also a substantial background to $\tau_{\text{had-vis}}$ identification. The RNN ID, developed specifically for discriminating against jets, does not provide sufficient discriminating power against electrons. For this reason, an e-veto BDT algorithm [77] specialised in telling electrons and $\tau_{\text{had-vis}}$ candidates apart is developed.

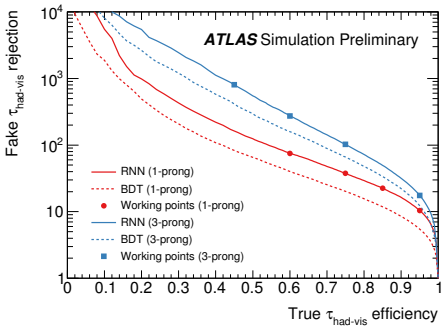
Electrons have features similar to those of 1-prong $\tau_{\text{had-vis}}$, especially for those with one neutral pion. Nonetheless, some differences in the detector response are very effective in differentiating electrons and $\tau_{\text{had-vis}}$. Since electron showers are purely electromagnetic, only a very small amount of the electron energy could leak to the HCal. On the contrary, $\tau_{\text{had-vis}}$ often deposit significant amount of energy in the HCal due to the $h^{\pm}$'s. Another effective discrimination comes from the response of the TRT. Electrons, which are always ultrarelativistic because of their small rest mass, usually leave more hits in the TRT than the heavier $h^{\pm}$'s from $\tau_{\text{had-vis}}$. The e-veto BDT is trained to exploit these differences.

The e-veto BDT was originally developed for analysing Run-1 data. However, due to software and hardware changes, it became deprecated at the beginning of Run 2. It has since been redeveloped to adopt the changes, to which the author has made major contributions. The expected output distributions and ROC curve of the redeveloped e-veto BDT are shown in Figure 3.5. Three working points with increasing efficiencies, `Tight`, `Medium` and `Loose`, are defined. While the algorithm is trained to identify 1-prong $\tau_{\text{had-vis}}$ candidates, it is found to be powerful in identifying 3-prong $\tau_{\text{had-vis}}$ candidates as well.



(a)

(b)



(c)

Figure 3.4.: The expected output distributions for (a) 1-prong and (b) 3-prong $\tau_{\text{had-vis}}$ candidates, and (c) ROC curves of the RNN ID [82]. The markers on the ROC curves indicate the four defined working points, `Tight`, `Medium`, `Loose` and `VeryLoose`, with increasing signal selection efficiencies.

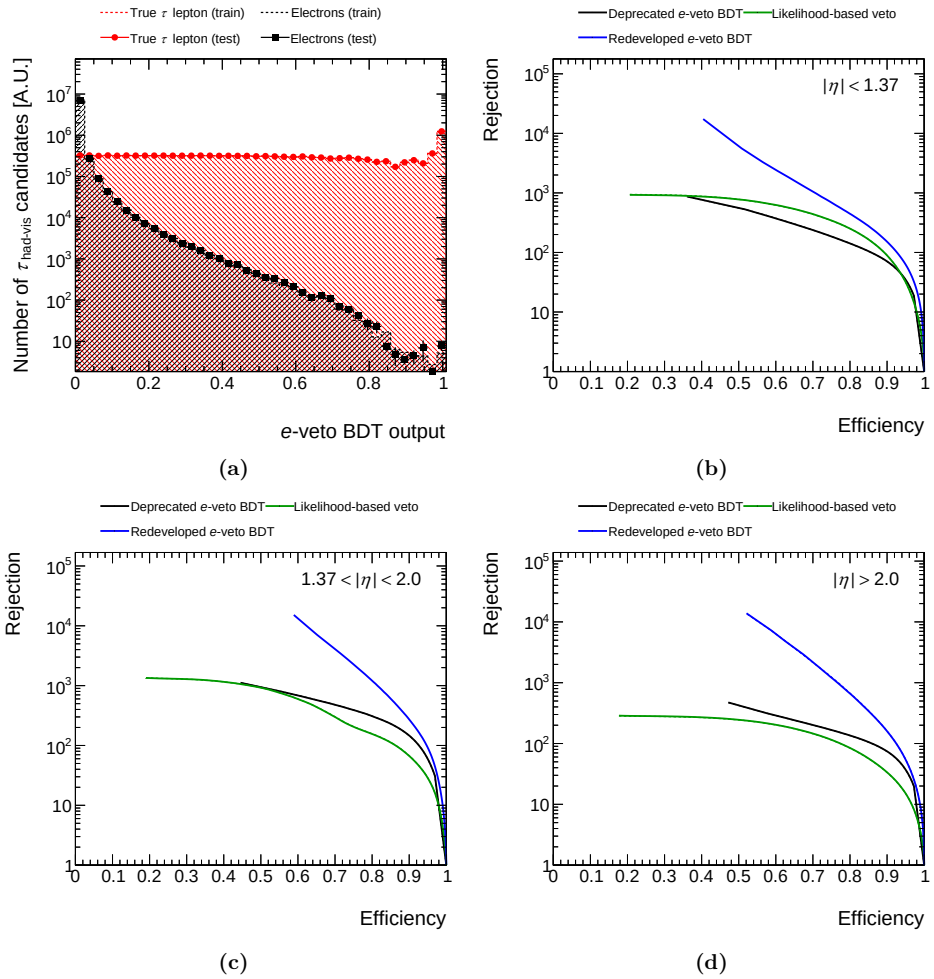


Figure 3.5.: The (a) expected output distributions and (b)–(d) the ROC curves of the *e*-veto BDT. The output distributions evaluated for both the training and testing samples are shown. The ROC curves for $\tau_{\text{had-vis}}$ candidates in different $|\eta|$ regions are shown, alongside the ROC curves of the deprecated *e*-veto BDT used during Run 1 and an alternative likelihood-based vetoing method used in the beginning of Run 2.

3.3. Electrons

In the ATLAS detector, electrons typically appear as a single track in the inner detector that matches geometrically to a distinctive shower in the EMCal [83–85].

The reconstruction of electrons begins with the reconstruction of clusters in the EMCal. A specialised cluster building algorithm is used for electron reconstruction [85]. It creates dynamic, variable-sized clusters, known as superclusters, that are able to recover energy from bremsstrahlung photons or from electrons from photon conversions adaptively. An algorithm is then used to associate tracks to the clusters to form electron candidates. A track is associated if it has $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ and is within an angular distance of $\Delta R = 0.3$ to the centre of a cluster. If multiple tracks are found, only the track that leaves hits in the SCT and is closest to the centre of the cluster are chosen. Since electron and photon showers are basically identical, the presence of an associated track is essential in telling an electron and a photon apart. Therefore, the reconstruction of electrons is restricted to the central region $|\eta| < 2.47$, which is within the coverage of the inner detector. Finally, the energy of the electron candidate, calculated from the energy deposits in the EMCal, is calibrated using studies with Monte Carlo simulations and observed $Z \to ee$ events.

Similar to $\tau_{\text{had-vis}}$, the reconstruction of electrons is susceptible to misreconstructed jets. A multivariate likelihood-based identification is used to improve the purity of reconstructed electron candidates. The identification uses information from both the inner detector and the calorimeters. The information used includes leakage to the HCal, lateral shower width and longitudinal energy distribution in the EMCal, track conditions, TRT response and track-cluster matching. The probability density functions of the relevant quantities are multiplied together to form the likelihood functions for electron signals (L_{S}) and for jet backgrounds (L_{B}). Then, a discriminant

$$d_L = \frac{L_S}{L_S + L_B} \tag{3.8}$$

is calculated. The probability distribution of d_L has a property of peaking sharply at zero for signals and at one for backgrounds. The value of the discriminant is then used to select electron candidates. Three WPs, `Tight`, `Medium` and `Loose`, are defined by cuts on the discriminant that correspond to increasing identification efficiencies.

Another challenge for selecting electron candidates is to draw a distinction between electrons produced directly from hard-scattering vertices or decays of heavy resonances (prompt electron) and those produced from decays of quarks or hadrons within a jet. The main difference between the two types of candidates is in the activity around the reconstructed electron. Two types of isolation variables are calculated. Calorimeter-based isolation variables are calculated based on the total energy of all clusters within a certain distance from the electron candidate, but with the contribution of the candidate itself being subtracted. Track-based isolation variables are constructed based on the total transverse momentum of all tracks around the track associated to the electron candidate, excluding the associated track itself. Various isolation WPs are defined to cater for the different

needs of various analyses. For the search for $Z \to \ell\tau$ decays, where a high jet rejection is desired and a constant efficiency is not required, the `FCTight` WP that is based on fixed cuts on both the calorimeter- and track-based isolation variables is used.

The performance of the electron reconstruction, identification and isolation is summarised in Figure 3.6.

3.4. Muons

The muon is a rather special particle in terms of detection. Muons have a relatively long lifetime[†], and without decaying, their interactions with the calorimeter materials are very limited due to their minimum-ionising nature. It is precisely for this reason that an entire detector subsystem, the muon spectrometer and the toroidal magnets, is dedicated for the detection of muons. The performance of their reconstruction and identification is also quite different from other particles [86].

Muons are reconstructed primarily via tracking. Reconstructions are first carried out separately in the ID and the MS. In the ID, muons are reconstructed as tracks in the same way as any other charged particles. In the MS, segments are first constructed by identifying certain hit patterns. Segments in the different layers of the MS are then fit together to form muon track candidates. For each candidate, a χ^2 fit of the associated hits is performed. The quality of the χ^2 fit is used as a criterion for accepting the candidate.

The information provided by the ID, MS and calorimeters is then combined using a number of algorithms to construct the muon candidates. Four types of reconstructed muons are defined, depending on the algorithm and subdetector information used:

Combined (CB) muons The separately reconstructed tracks in the ID and the MS are matched and their hits are refitted to form combined tracks. The matching is conducted in an outside-in manner, where matching ID tracks are found by extrapolating the MS tracks.

Segment-tagged (ST) muons ID tracks are extrapolated to the MS where matching MS segments are found. If at least one segment in the MDTs or CSCs are matched, the track is considered a muon candidate. ST muons are mostly muons with low p_{T} or in regions with reduced MS acceptance, such that only one layer of the MS is crossed.

Calorimeter-tagged (CT) muons An ID track is also considered a muon candidate if it can be matched to energy deposits in the calorimeters that are compatible with a minimum-ionising particle. CT muons have the highest misreconstruction rate amongst all types of muon candidates, but are useful in recovering acceptance in the crack region of the MS ($|\eta| < 0.1$).

Extrapolated (ME) muons MS tracks that, when extrapolated, are compatible with the nominal interaction point are also reconstructed as a muon candidate. This type of

[†] In the laboratory frame of reference, where they are typically ultrarelativistic.

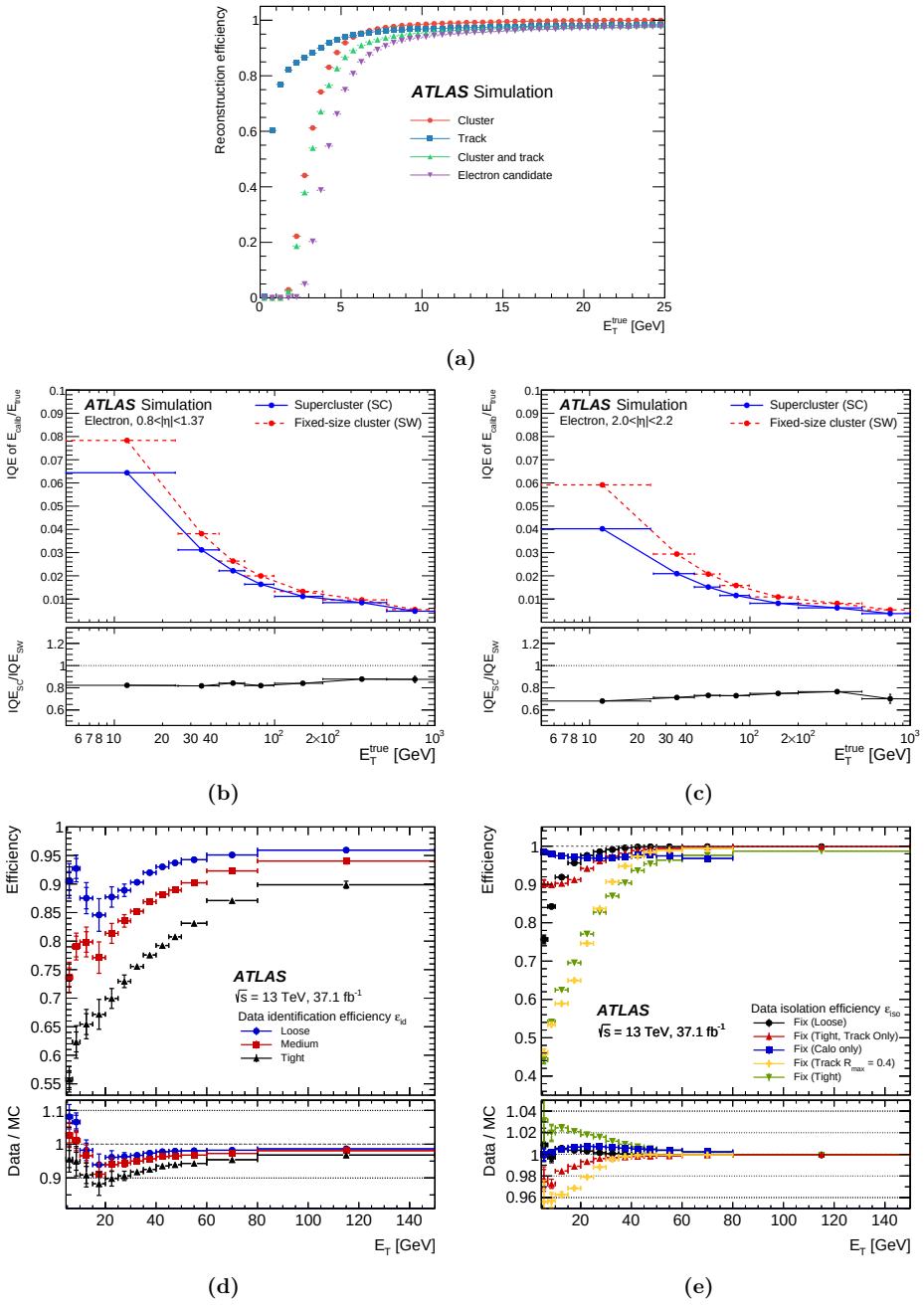


Figure 3.6.: The (a) estimated reconstruction efficiency, (b),(c) estimated energy resolution, (d) measured identification efficiency and (e) measured isolation efficiency of electrons as functions of the electron transverse energy [84, 85]. The energy resolution is given as effective interquartile range (IQE).

reconstruction is mainly to recover acceptance in the region covered by the MS but not by the ID ($2.5 < |\eta| < 2.7$).

In cases where an ID track is shared by multiple muon candidates of different types, preference is first given to CB muons, then to ST muons, and finally to CT muons. MS tracks that are shared by ME muons are given to the candidate with better fit quality and larger hit multiplicity.

Three sets of identification criteria with increasing efficiencies, `Tight`, `Medium` and `Loose`, are defined to identify muons produced directly from hard-scattering vertices or decays of heavy resonances (prompt muon). The major background for prompt muon identification is muons from pion and kaon decays. The identification criteria consider the fact that non-prompt muons typically have tracks with a poor quality of fit and inconsistent momentum measured in the ID and the MS. In addition, an extra set of identification criteria, `High-pT`, is defined specifically for optimised momentum resolution for muons with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$.

Another type of muon identification background is cosmic muons, which are muons from cosmic rays that penetrate the atmosphere and land surface, and reach the detector. Cosmic muons can be effectively rejected by requiring the muon tracks to have small impact parameters with respect to the interaction vertex.

Similar to electrons, isolation WPs are also defined for muons to veto muons originated from jets. These WPs are defined using variables similar to those described in Section 3.3 for electrons.

The performance of the muon reconstruction, identification and isolation is summarised in Figure 3.7.

3.5. Missing transverse momentum

As neutral particles that only interact weakly, neutrinos are never detected by tracking detectors, nor do they deposit energy in the calorimeters. The transverse energy of these invisible particles, nonetheless, can still be inferred. At the LHC, the centre of mass of the colliding protons can be expected to be stationary in the transverse plane. This implies that the total vector sum of the transverse momenta of all outgoing particles must have a vanishing magnitude. That is,

$$\sum_{\text{all particles}} \mathbf{p_T} = \mathbf{0}. \tag{3.9}$$

If some of the particles are invisible to the detector, either because they are neutrinos or some hypothetical inert particles, their total transverse momentum can then be calculated

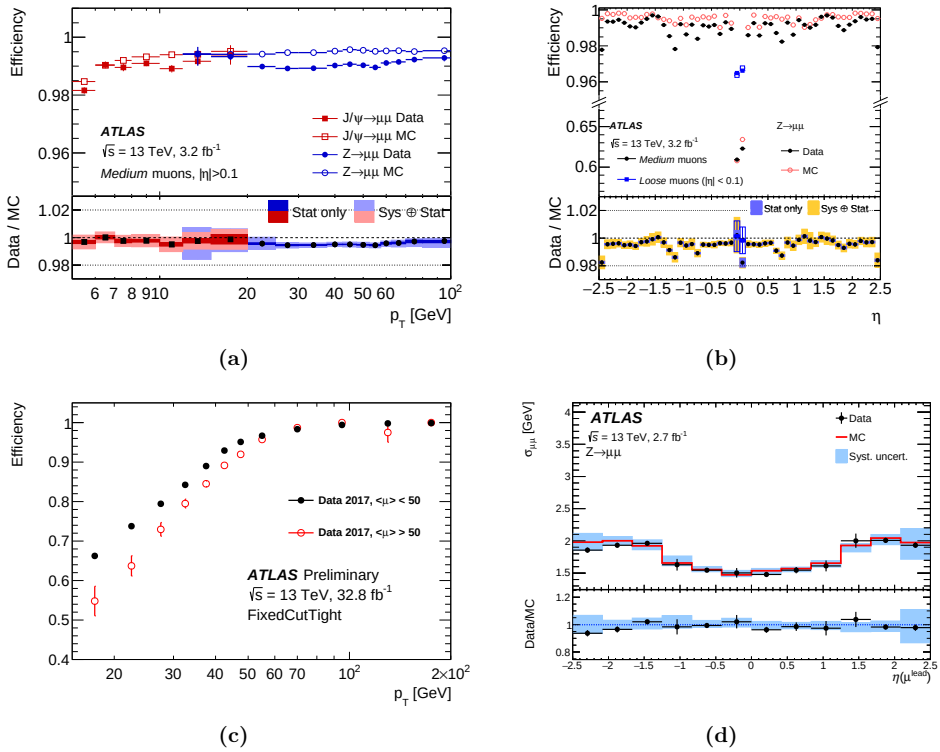


Figure 3.7.: The measured (a),(b) reconstruction and identification efficiency, and (c) isolation efficiency of muons with `Medium` identification and `FCTight` isolation as functions of the muon transverse momentum or pseudorapidity, and (d) the dimuon invariant mass resolution for $Z \to \mu\mu$ events as a function of the leading-p_{T} muon pseudorapidity [86, 87].

as

$$\sum_{\text{invisible}} \mathbf{p_T} = -\sum_{\text{visible}} \mathbf{p_T} \equiv \mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}. \tag{3.10}$$

This is called the missing transverse momentum, with the magnitude and azimuthal angle denoted as $E_{\mathrm{T}}^{\mathrm{miss}}$ and ϕ^{miss} respectively. Thanks to the solid angular coverage of almost 4π, the ATLAS detector is capable of reconstructing $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ accurately in most scenarios.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ of an event is calculated as the sum of two terms, the "hard term" and the "soft term" [88, 89]. The hard term is the negative vector sum of the transverse momenta of all reconstructed physics objects that have passed certain basic quality requirements. To avoid double counting, objects that are geometrically close to another object are removed

in the order: jets, $\tau_{\text{had-vis}}$, photons[†], electrons and muons. The soft term is the negative vector sum of the transverse momenta of all unused ID tracks that are associated to the primary vertex but not to any reconstructed physics objects. This definition of the soft term is called the track-based soft term (TST). By including soft charged particles, the TST reduces the bias and improves the resolution of the reconstructed $E_{\text{T}}^{\text{miss}}$. A more inclusive version of the soft term can be defined by also taking into account the unused clusters in the calorimeters. The advantage of such a definition is that the momenta of soft neutral particles can also be considered. But on the flip side, the inclusion of energy deposits without a clear origin significantly increases the dependence on pile-up, which is generally undesirable.

Depending on the selection criteria of jets in the forward region, four working points for $E_{\text{T}}^{\text{miss}}$ reconstruction are defined: `Loose`, `Tight`, `Tighter` and `Tenacious`. In principle, a tighter selection provides a better resolution, while a looser selection offers possibly smaller bias but is less robust against pile-up.

The resolution of $E_{\text{T}}^{\text{miss}}$ can be measured using processes with no expected $E_{\text{T}}^{\text{miss}}$, such as $Z \to ee$. The measured resolution, as the root-mean-square (RMS) deviation and as a function of pile-up activity, is shown in Figure 3.8.

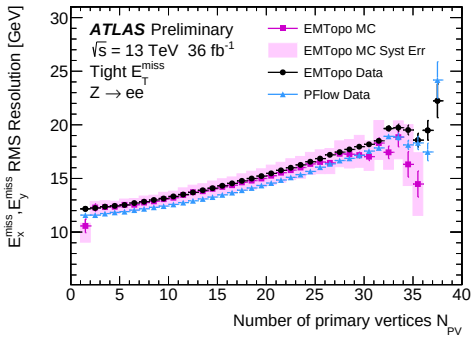


Figure 3.8.: The measured resolution (RMS deviation) of $E_{\text{T}}^{\text{miss}}$ as a function of pile-up activity measured by the number of primary vertices (N_{PV}) [89]. Only the measurement labelled "EMTopo" is relevant to the work presented here.

3.6. Analysis-specific definitions and overlap removal

Object selection criteria that are specific to the search for LFV $Z \to \ell\tau$ decays are listed in Table 3.2. These object definitions ensure good reconstruction qualities, while maintaining a high acceptance and efficiency for the target $Z \to \ell\tau$ signal events.

[†] Photons are reconstructed in a way similar to electrons, but without an associated track. Since they are nowhere used in the work presented in this thesis, details of their reconstruction will be omitted.

Table 3.2.: Physics object selection criteria that are specific to the search for LFV $Z \to \ell\tau$ decays.

Object	Selection criteria
Jet	$p_{\mathrm{T}} > 20\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.5$
b-jet	Passes fixed cut on the MV2c10 output corresponding to a 77% b-tagging efficiency
$\tau_{\text{had-vis}}$	$p_{\mathrm{T}} > 25\,\mathrm{GeV}$ $\lvert\eta\rvert < 1.37$ or $1.52 < \lvert\eta\rvert < 2.5$ `Tight` RNN ID `Medium` (`Loose`) e-veto for 1-prong (3-prong) candidates
Electron	$p_{\mathrm{T}} > 30\,\mathrm{GeV}$ $\lvert\eta\rvert < 1.37$ or $1.52 < \lvert\eta\rvert < 2.47$ `Tight` identification `FCTight` isolation
Muon	$p_{\mathrm{T}} > 30\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.5$ `Tight` identification `FCTight` isolation
$E_{\mathrm{T}}^{\mathrm{miss}}$	`Tight` working point Track-based soft term

The reconstructions of jets, electrons, muons and $\tau_{\text{had-vis}}$ are carried out independently without much harmonisation. Consequently, the same tracks and energy deposits may be attributed to multiple objects, leading to inflated object multiplicities, and incorrectly reconstructed event topologies and kinematics. To resolve this, objects that share the same ID track with or are geometrically close to another object are removed in a specific, well-defined order. This conflict-removing procedure is called overlap removal.

For the purpose of overlap removal, a looser definition for electrons, muons and $\tau_{\text{had-vis}}$ is used, which helps to reduce the misidentification rates. For electrons and muons, the p_{T} threshold is lowered to 25 GeV, the identification requirement is relaxed to `Loose` and no isolation requirement is enforced. For $\tau_{\text{had-vis}}$, the RNN ID requirement is relaxed to `Loose` and no electron rejection is applied.

The criteria for the overlap removal and the order that they are applied are shown in Table 3.3. They are defined mainly to minimise the overall object misidentification rate, hence priority is generally given to candidates with higher purity.

Table 3.3.: The overlap removal criteria listed in the order in which they are applied.

Reject	**Against**	**Criteria**
Electron	Electron	Shared ID track, the electron to be rejected has lower p_{T}
$\tau_{\text{had-vis}}$	Electron	$\Delta R < 0.2$
$\tau_{\text{had-vis}}$	Non-CT muon	$\Delta R < 0.2$, the muon only needs to satisfy $p_{\mathrm{T}} > 2\,\mathrm{GeV}$
$\tau_{\text{had-vis}}$	CT Muon	$\Delta R < 0.2$
CT Muon	Electron	Shared ID track
Electron	Muon	Shared ID track
Jet	Electron	$\Delta R < 0.2$, the jet is not a b-jet
Electron	Jet	$\Delta R < 0.4$
Jet	Muon	$\Delta R < 0.2$ or ghost-associated [90], the jet is not a b-jet and has < 3 associated tracks
Muon	Jet	$\Delta R < 0.4$
Jet	$\tau_{\text{had-vis}}$	$\Delta R < 0.2$

Chapter 4.

Event selection and classification

"Doing or not doing something – they are similar. Both involve an action and sincerity."

— CLAMP, "Yūko Ichihara", *xxxHOLiC*

Throughout the LHC Run 2, tens of billions of pp collision events have been recorded by ATLAS. This enormous amount of data is what provides us the statistical power to search for rare LFV Z decays in the first place. But most of these data are unexciting events from SM processes that are not of our interest. They are the background of the search. At the end of the day, besides the amount of data collected, the sensitivity of a search boils down to two main factors: how well the background can be modelled and how well background events can be rejected while retaining the potential signal events. While the background modelling will be discussed in the next chapter (Chapter 5), focus will be given to the event selection and the discrimination of the $Z \to \ell\tau$ signal from the SM background in this chapter.

4.1. Event selection

The event selection serves two purposes: to select events that are reconstructed with high quality and to segregate events from different processes from each other. Selection criteria are mainly based on the multiplicities and kinematics of reconstructed objects, as well as the topologies of the events. These criteria can either be a simple cut on a certain variable, or make use of MVA, such as neural network classifiers.

4.1.1. Event cleaning

For an apparatus as complex as the ATLAS detector, it would be impractical to assume that every component works flawlessly all the time. During data taking, inevitably,

malfunctions or glitches happen. To ensure that events recorded when the detector is at fault can be identified and discarded in analyses, the conditions of the detector are routinely monitored by human operators and automated algorithms during and after data taking [91]. They are stored as data quality (DQ) information alongside each recorded event. For the search for $Z \to \ell\tau$ decays, selected events are required to be recorded without any critical malfunction in any of the subsystems of the detector[†]. Main sources of data loss include data corruption due to desynchronisation of the read-out electronics, high-voltage trips[‡] due to failures of the power supplies and noise bursts[§]. With the continuous firmware and software development, as well as the replacement of the read-out hardware, these conditions have been improved over the course of the data taking. Overall, the complete Run 2 data set has a combined DQ efficiency of 95.6%.

In addition to the data quality requirement, events are required to have a primary vertex with at least two associated tracks. The primary vertex is selected as the vertex with the largest $\sum p_{\mathrm{T}}^2$, where the sum is over all tracks with transverse momentum $p_{\mathrm{T}} > 0.4\,\mathrm{GeV}$ that are associated with the vertex.

4.1.2. Triggers

In the analysis, only events that passed the unprescaled single-lepton HLTs summarised in Table 4.1 are used. For the data taken in 2016-2018, these triggers require at least one electron (or muon) in the event to have $p_{\mathrm{T}} > 26\,\mathrm{GeV}$ if it is isolated, or $p_{\mathrm{T}} > 60\,\mathrm{GeV}$ ($50\,\mathrm{GeV}$) if it is not isolated. For the data taken in 2015, slightly looser thresholds are employed by the triggers. Since the electron or muon from a $Z \to \ell\tau$ decay can always be expected to be boosted and isolated, the selected triggers have a high efficiency for the signal events.

Table 4.1.: High-level triggers used in the analysis. Detailed definitions and performance of the triggers can be found in Refereces [92, 93].

	2015	**2016-2018**
Single-electron triggers	`HLT_e24_lhmedium_L1EM20VH` `HLT_e60_lhmedium`	`HLT_e26_lhtight_nod0_ivarloose` `HLT_e60_lhmedium_nod0`
Single-muon triggers	`HLT_mu20_iloose_L1MU15` `HLT_mu40`	`HLT_mu26_ivarmedium` `HLT_mu50`

[†] Except events that only encountered errors at the HLT or data-acquisition level, where information can be recovered offline. These events enter the so-called debug stream, and are included in physics analyses to avoid missing potentially interesting events. They constitute a fraction of less than 10^{-7} of all the recorded events in Run 2.

[‡] A sudden drop of high voltage that makes the energy correction unreliable.

[§] A burst of coherent noise over a large region of the detector that typically lasts a microsecond. It mainly affects the LAr end-caps.

Furthermore, depending on the search channel ($e\tau$ or $\mu\tau$), the leading-p_{T} electron or muon in the event must match the object that "fired" the trigger(s). This ensures that the effect of the trigger efficiency can be properly taken into account in the signal and background modelling.

4.1.3. Signal region

The signal $Z \to \ell\tau$ events have a number of unique features that can be exploited to differentiate them from the SM background. To begin with, the event is a resonance, meaning that the ℓ–τ pair, before decaying or radiating, has an invariant mass close to the Z boson rest mass. This also implies that the light lepton and τ lepton are typically boosted due to the massiveness of the parent Z boson. Furthermore, for electric charge and momentum to be conserved, the leptons must carry opposite-sign (OS) charges and travel almost back-to-back in the transverse plane. Since the τ lepton is typically boosted, the $\tau_{\text{had-vis}}$ and neutrino from its decay are expected to travel in almost the same direction in the laboratory frame of reference. Hence, the directions of the reconstructed $\tau_{\text{had-vis}}$ and $E_{\mathrm{T}}^{\text{miss}}$ can be expected to be almost collinear.

Utilising the characteristics of the signal events, a set of criteria is defined to select potential signal events with high efficiency while rejecting background events effectively. The region in the parameter space defined by these selection criteria is the so-called signal region (SR) of the analysis. Maximum-likelihood fits (see Section 6.1) to the measurements in the SR are used to extract possible evidence of $Z \to \ell\tau$ decays.

Events in the SR are required to have exactly one reconstructed light lepton and at least one $\tau_{\text{had-vis}}$ candidate. The light lepton and the $\tau_{\text{had-vis}}$ candidate with leading p_{T} must satisfy the object selection criteria aforementioned in Table 3.2. The flavour of the light lepton determines which search channel will the event be considered in ($e\tau$ or $\mu\tau$). Events with additional τ leptons are not vetoed, since doing so would complicate the estimation of events with misidentified $\tau_{\text{had-vis}}$ candidates, as it mixes per-event and per-object misidentification rates. It has been estimated that only roughly 0.1% of the events in the SR have more than one $\tau_{\text{had-vis}}$ candidate, and the leading-p_{T} $\tau_{\text{had-vis}}$ candidate in signal events is correctly reconstructed from the signal τ lepton 99.8% of the time. For the sake of convenience, from this point onwards, "*the* $\tau_{\text{had-vis}}$ candidate" of an event would always refer to the leading-p_{T} $\tau_{\text{had-vis}}$ in the event unless the context clearly implies otherwise.

The light lepton and the $\tau_{\text{had-vis}}$ candidate are required to carry OS charges. The visible invariant mass,

$$m_{\text{vis}}(\ell,\tau) \equiv m(\ell,\tau_{\text{had-vis}}) = \sqrt{2p_{\mathrm{T}}(\ell)p_{\mathrm{T}}(\tau_{\text{had-vis}})\left[\cosh\Delta\eta(\ell,\tau_{\text{had-vis}}) - \cos\Delta\phi(\ell,\tau_{\text{had-vis}})\right]}, \tag{4.1}$$

is required to be larger than 60 GeV to reject background events with ℓ–$\tau_{\text{had-vis}}$ pairs that are unlikely to originate from $Z \to \ell\tau$ decays. While an upper cut on $m_{\text{vis}}(\ell,\tau)$ can also reject background events, it has been found that keeping events with high $m_{\text{vis}}(\ell,\tau)$ are

actually beneficial as they are useful for constraining the predicted background yields in the maximum-likelihood fits. Events with identified b-jets are vetoed in order to suppress events from $t\bar{t}$, single-top or Wt productions (collectively known as the "top background").

Additionally, in the $\mu\tau$ channel, events with 1-prong $\tau_{\text{had-vis}}$ candidates reconstructed in the crack region of the muon spectrometer ($|\eta| < 0.1$) are rejected. This is because of the large $\mu \to \tau_{\text{had-vis}}$ misidentification rate due to the inefficacy of the μ–$\tau_{\text{had-vis}}$ overlap removal in the region.

With the above mentioned selection criteria, the remaining major background processes are $Z \to \tau\tau$, where one of the τ leptons decays leptonically and the other hadronically; $Z \to \ell\ell$, where one of the light leptons is misidentified as the $\tau_{\text{had-vis}}$ candidate ($\ell \to \tau_{\text{had-vis}}$ fakes); and events with a quark- or gluon-initiated jet misidentified as the $\tau_{\text{had-vis}}$ candidate (jet $\to \tau_{\text{had-vis}}$ fakes, or simply "fakes" without further clarification), which are dominantly W+jets and multijet events. These background events can be effectively separated from the signal events by the transverse mass variables and outputs of neural network classifiers.

The transverse mass, m_{T}, of a system of any two objects, X and Y, is defined by

$$\begin{aligned} m_{\text{T}}^2(X,Y) &\equiv \left(E_{\text{T}}(X) + E_{\text{T}}(Y)\right)^2 - \left(\mathbf{p_T}(X) + \mathbf{p_T}(Y)\right)^2 \\ &\equiv m^2(X) + m^2(Y) + 2\left(E_{\text{T}}(X)E_{\text{T}}(Y) - \mathbf{p_T}(X)\cdot\mathbf{p_T}(Y)\right). \end{aligned} \tag{4.2}$$

Its value is invariant under Lorentz boosts in the z-direction, making it a useful quantity when one of the considered objects is the missing transverse momentum, which has an unknown z-component. In the case where X and Y are both approximately massless (i.e. highly relativistic), the transverse mass is approximately

$$m_{\text{T}}(X,Y) = \sqrt{2p_{\text{T}}(X)p_{\text{T}}(X)\left(1 - \cos\Delta\phi(X,Y)\right)}, \tag{4.3}$$

where $\Delta\phi(X,Y)$ is the angle between the transverse momenta of X and Y. When X and Y are collinear in the transverse plane, $m_{\text{T}}(X,Y)$ vanishes. On the other hand, when they are back-to-back, $m_{\text{T}}(X,Y)$ is proportional to the magnitudes of their transverse momenta. This property is exploited in the event selection for the SR.

Figure 4.1 shows the typical topology of a signal $Z \to \ell\tau$ event in the transverse plane, compared to that of a $Z \to \tau\tau$ or W+jets event. For a signal event, the neutrino from the single τ decay is expected to be the sole contributor to the $E_{\text{T}}^{\text{miss}}$ and travel in a direction close to that of the $\tau_{\text{had-vis}}$ but back-to-back to that of the light lepton. This can be reflected by a relatively small $m_{\text{T}}(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})$ but large $m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}})$. In contrast, the neutrinos from both the τ decays in a $Z \to \tau\tau$ event contribute to the total $E_{\text{T}}^{\text{miss}}$, resulting in a $E_{\text{T}}^{\text{miss}}$ that could either line up with the $\tau_{\text{had-vis}}$ or the light lepton. Therefore, a large fraction of $Z \to \tau\tau$ events have a small $m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}})$ but large $m_{\text{T}}(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})$. In a W+jets event, the scenario is again different. The $\tau_{\text{had-vis}}$ candidate, which is a misidentified jet, is generally not aligned with the $E_{\text{T}}^{\text{miss}}$, nor the light lepton. This results in both $m_{\text{T}}(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})$ and $m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}})$ being large in general.

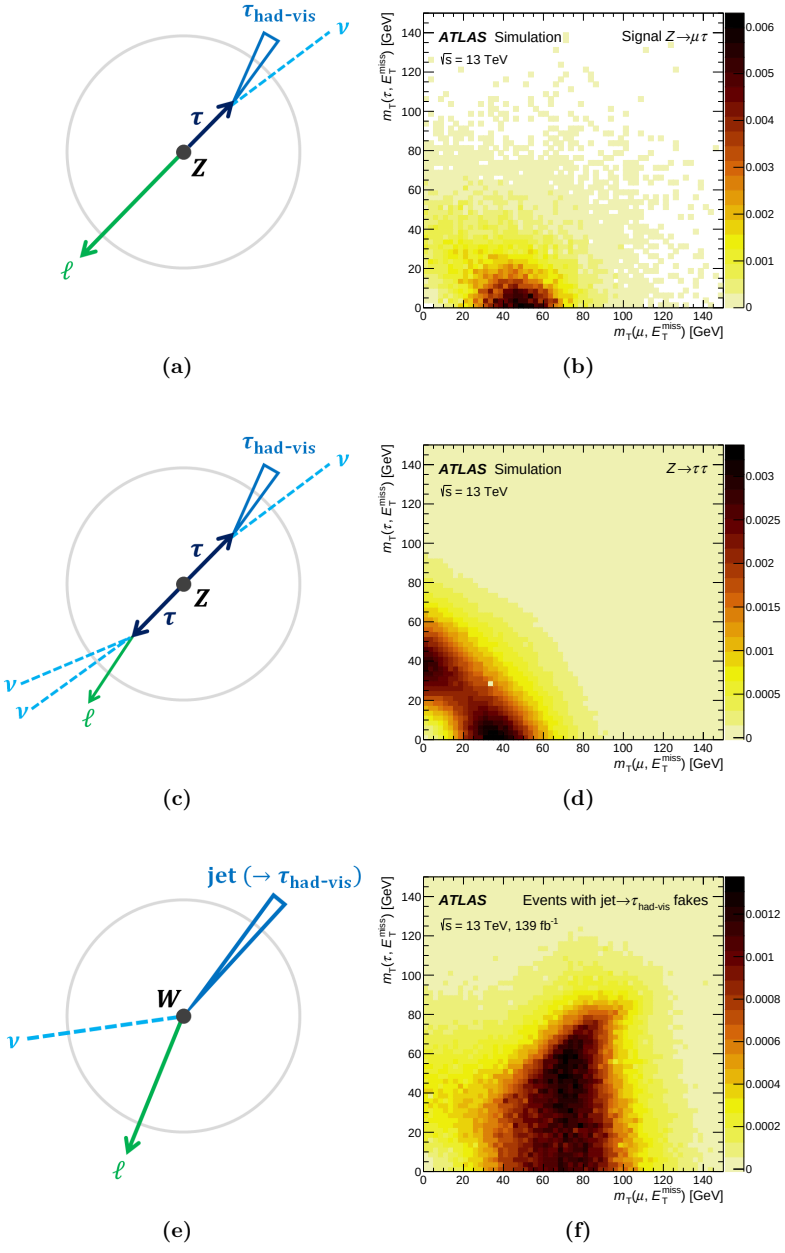


Figure 4.1.: A schematic representation of the typical topology of a (a) signal $Z \to \ell\tau$, (c) $Z \to \tau\tau$ or (e) W+jets event, as seen in the transverse plane. The two-dimensional histograms show the distributions of $m_{\mathrm{T}}(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\mathrm{miss}})$ versus $m_{\mathrm{T}}(\mu, E_{\mathrm{T}}^{\mathrm{miss}})$ of (b) simulated $Z \to \mu\tau$ events, (d) simulated $Z \to \tau\tau$ events and (f) estimated jet$\to$ $\tau_{\text{had-vis}}$ fakes in the SR without the cuts on $m_{\mathrm{T}}(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\mathrm{miss}})$ and $m_{\mathrm{T}}(\mu, E_{\mathrm{T}}^{\mathrm{miss}})$. The colours represent the fraction of events in each bin.

Using these properties to our advantage, events in the SR are required to have $m_{\mathrm{T}}(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\text{miss}}) < 35\,\text{GeV}$ to remove the majority of $Z \to \tau\tau$ and W+jets events. The events that are not selected are split by a cut on $m_{\mathrm{T}}(\ell, E_{\mathrm{T}}^{\text{miss}})$ into two different regions, each enriched in either $Z \to \tau\tau$ or W+jets events. These regions are used to control the $Z \to \tau\tau$ background and estimate fakes from W+jets events, and will be detailed in the next subsection (Section 4.1.4).

Finally, outputs of neural network classifiers, which will be described later in Section 4.2, are used to refine the SR selection. Events that are classified as the most background-like are rejected.

Due to the difference in background compositions, the SR (and regions that will later be defined) is further split into two regions depending on the number of tracks associated to the $\tau_{\text{had-vis}}$ candidates. The region with 1-prong (3-prong) $\tau_{\text{had-vis}}$ candidates is denoted with "1P" ("3P").

Table 4.2 summarises the selection criteria for events in the SR.

Table 4.2.: Selection criteria for events in the signal region. The last two criteria involve outputs of neural network classifiers, which are described in Section 4.2.

Selection criteria	
$e\tau$ channel	**$\mu\tau$ channel**
At least one $\tau_{\text{had-vis}}$ candidate	
One electron and no muon	One muon and no electron
Opposite-sign charged e–$\tau_{\text{had-vis}}$ pair	Opposite-sign charged μ–$\tau_{\text{had-vis}}$ pair
$m_{\text{vis}}(e, \tau) > 60\,\text{GeV}$	$m_{\text{vis}}(\mu, \tau) > 60\,\text{GeV}$
	$\lvert\eta(\text{1-prong } \tau_{\text{had-vis}})\rvert > 0.1$
$m_{\mathrm{T}}(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\text{miss}}) < 35\,\text{GeV}$	
No b-jets	
NN_{comb} output > 0.1 (0.2) for 1P (3P) region	
$\text{NN}_{\text{Z}\ell\ell}$ output > 0.2	

4.1.4. Control regions and fakes-enriched regions

Events that are rejected by the SR selection contain almost no potential signal events. Despite that, they are still useful in helping to constrain the background modelling. These events are split into several regions that are each enriched in events from a particular background process. These regions are the so-called control regions (CR).

To help constrain the overall yield of the $Z \to \tau\tau$ background and reduce systematic uncertainties related to the modelling of real $\ell\tau_{\text{had-vis}}$ signal states, a control region enriched

in $Z \to \tau\tau \to \ell\tau_{\text{had-vis}} + 3\nu$ events, CRZ$\tau\tau$, is defined. Events in CRZ$\tau\tau$ are required to pass the same selection criteria as those for the SR, except the cut on $m_{\text{T}}(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})$. Events are required to have $m_{\text{T}}(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}}) > 35\,\text{GeV}$ and $m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}}) < 40\,\text{GeV}$ instead, which can be justified by the expected m_{T} distributions as shown in Figure 4.1. Additionally, the so-called collinear mass, m_{coll} (ℓ,τ), of the event must be between 70 GeV and 110 GeV for the event to be considered. The collinear mass is the invariant mass of the ℓ–$\tau_{\text{had-vis}}$–$E_{\text{T}}^{\text{miss}}$ system, assuming that the $E_{\text{T}}^{\text{miss}}$ has a virtual z-component such that its pseudorapidity is that same as that of the $\tau_{\text{had-vis}}$ candidate. In the case of a signal event, the collinear mass is a good proxy for the true invariant mass of the ℓ–$\tau_{\text{had-vis}}$–ν system, which is expected to be the Z boson mass.

Another control region, CRZ$\ell\ell$, is defined for measuring corrections to the modelled $\ell \to \tau_{\text{had-vis}}$ misidentification rates. The region consists of dominantly $Z \to \ell\ell$ events and is defined using outputs of the neural network classifiers. The event selection is the same as that for the SR, except that the last criterion in Table 4.2 is inverted and both the outputs of the $\text{NN}_{\text{Z}\tau\tau}$ and NN_{Wjets} classifiers (see Section 4.2) must be larger than 0.8. These cuts select events that are classified by the neural networks to be likely a $Z \to \ell\ell$ event but not likely a $Z \to \tau\tau$ or W+jets event.

Furthermore, four regions known as the fakes-enriched regions (FR) are defined. The FRs are a special kind of control region. These regions are enriched in jet$\to \tau_{\text{had-vis}}$ fakes and are essential for deriving the data-driven prediction of fakes in other regions, including the SR (see Section 5.5). The four FRs are:

FRW Dominated by $W(\to \tau\nu)$+jets events, which contribute the most to the total amount of fakes in the SR. The region is defined by inverting the cut on $m_{\text{T}}(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})$ for the SR and requiring $m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}}) > 40\,\text{GeV}$.

FRQ Dominated by QCD multijet events, which are the second largest contribution to all the fakes in the SR. This region requires events to satisfy the SR selection but with an inverted isolation requirement for the light lepton. Events with $m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}}) > 40\,\text{GeV}$ are excluded to suppress the W+jets background. Moreover, the ℓ–$\tau_{\text{had-vis}}$ pair is required to carry same-sign (SS) electric charges instead of OS charges to reduce contamination from $Z \to \tau\tau$ events.

FRZ Dominated by $Z(\to \ell\ell)$+jets events. The event selection for this region is the same as that for the SR, except that exactly two same-flavour, opposite-sign-charged light leptons are required instead of just one. The invariant mass of the light lepton pair must fall within the interval (81 GeV, 101 GeV). The event is still required to have a $\tau_{\text{had-vis}}$ candidate.

FRT Dominated by $t\bar{t}$ events. Events in the region must satisfy the SR selection but with two or more b-jets identified[†].

[†] Events from single-top or Wt productions also contribute to the total fakes background. However, the contribution from these events is smaller than that from $t\bar{t}$ events by orders of magnitude in the SR. Therefore, the FRT is only optimised to include $t\bar{t}$ events and requires two or more b-jets instead of at least one b-jet.

Table 4.3.: Purities of the fakes-enriched regions.

Fakes-enriched region	Purity [%]	
	$e\tau$ channel	$\mu\tau$ channel
FRW	95	94
FRQ	90	89
FRZ	99	99
FRT	96	96

The purity of an FR in the targeted process is defined as the expected relative contribution of the targeted process to the total amount of fakes in the region. The estimated purities of the FRs are shown in Table 4.3, which are generally high.

Table 4.4 summarises the event selection criteria for CRZ$\tau\tau$, CRZ$\ell\ell$ and the FRs. The purpose and usage of these regions will be discussed in more details in the following chapters.

Additionally, several regions are defined particularly for validating the modelling of fakes. They include a validation region, VRSS, that has the same event selection criteria as the SR except that SS instead of OS charged ℓ–$\tau_{\text{had-vis}}$ pairs are required. Three special FRs are also defined, they are defined in the same way as FRW, FRZ and FRT, but again requires SS instead of OS charged ℓ–$\tau_{\text{had-vis}}$ pairs. They are referred to as the SS FRW, SS FRZ and SS FRT respectively, and are collectively referred to as the SS FRs together with FRQ. These regions are all dominated by fakes.

4.2. Neural network classifiers

Neural network (NN) classifiers[†] are used to optimise the discrimination between the signal and different background events. The utilisation of NN classifiers is at the heart of the analysis. Not only do they allow us to identify the most signal-like events, but they also provide separation between different background processes, which is valuable for improving the precision and accuracy of the background modelling.

Multiple binary classifiers are trained for each channel to discriminate signal events from the three major backgrounds: $Z \to \tau\tau$, $W(\to \ell\nu)$+jets and $Z \to \ell\ell$ events. They are labelled $\text{NN}_{\text{Z}\tau\tau}$, NN_{Wjets} and $\text{NN}_{\text{Z}\ell\ell}$ respectively. The output scores of these classifiers are then combined into one final discriminant, the distribution of which is used in binned maximum-likelihood fits in the SR to extract evidence of potential signals or to set upper limits on the LFV branching fractions.

[†] See Appendix A for a brief introduction to neural network classification.

Table 4.4.: Event selection for CRZ$\tau\tau$, CRZ$\ell\ell$ and the FRs, listing only the criteria that are different or additional to those for the SR. The selection criteria for CRZ$\ell\ell$ involve outputs of neural network classifiers, which will be described in Section 4.2.

Region	**Selection criteria**	
	***e*τ channel**	**$\mu\tau$ channel**
CRZ$\tau\tau$	$m_{\mathrm{T}}(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\mathrm{miss}}) > 35\,\mathrm{GeV}$	
	$m_{\mathrm{T}}(e, E_{\mathrm{T}}^{\mathrm{miss}}) < 40\,\mathrm{GeV}$	$m_{\mathrm{T}}(\mu, E_{\mathrm{T}}^{\mathrm{miss}}) < 40\,\mathrm{GeV}$
	$70\,\mathrm{GeV} < m_{\mathrm{coll}}(e, \tau) < 110\,\mathrm{GeV}$	$70\,\mathrm{GeV} < m_{\mathrm{coll}}(\mu, \tau) < 110\,\mathrm{GeV}$
CRZ$\ell\ell$	$\mathrm{NN}_{\mathrm{Z}\ell\ell}$ output < 0.2	
	$\mathrm{NN}_{\mathrm{Z}\tau\tau}$ output > 0.8	
	$\mathrm{NN}_{\mathrm{Wjets}}$ output > 0.8	
FRW	$m_{\mathrm{T}}(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\mathrm{miss}}) > 35\,\mathrm{GeV}$	
	$m_{\mathrm{T}}(e, E_{\mathrm{T}}^{\mathrm{miss}}) > 40\,\mathrm{GeV}$	$m_{\mathrm{T}}(\mu, E_{\mathrm{T}}^{\mathrm{miss}}) > 40\,\mathrm{GeV}$
FRQ	Non-isolated light lepton candidate	
	$m_{\mathrm{T}}(e, E_{\mathrm{T}}^{\mathrm{miss}}) < 40\,\mathrm{GeV}$	$m_{\mathrm{T}}(\mu, E_{\mathrm{T}}^{\mathrm{miss}}) < 40\,\mathrm{GeV}$
	SS charged e–$\tau_{\text{had-vis}}$ pair	SS charged μ–$\tau_{\text{had-vis}}$ pair
FRZ	Two electrons and no muons	Two muons and no electrons
	$81\,\mathrm{GeV} < m(e_1, e_2) < 101\,\mathrm{GeV}$	$81\,\mathrm{GeV} < m(\mu_1, \mu_2) < 101\,\mathrm{GeV}$
FRT	Two or more b-jets	

4.2.1. Training samples

The NNs are trained using MC samples (see Section 5.1) selected by the same event selection criteria for the SR, except that the cuts on $m_{\mathrm{vis}}(\ell, \tau)$ and the NN outputs are omitted. In addition, real $\tau_{\text{had-vis}}$ candidates from $Z \to \ell\tau$ and $Z \to \tau\tau$ events are only required to pass the `Loose` instead of the `Tight` identification in order to increase the size of the training samples. For the $Z \to \ell\ell$ sample, events are selected only if the true origin of the $\tau_{\text{had-vis}}$ candidate is a misidentified light lepton.

Different NNs are separately trained for the $e\tau$ and $\mu\tau$ channels and for the 1P and 3P regions. To increase the signal sample size, the $Z \to e\tau$ and $Z \to \mu\tau$ samples are combined and used for training in both channels, assuming equivalent event topology when exchanging e and μ. Due to the small contribution of $Z \to \ell\ell$ events to the total background in the 3P regions, no classifier is trained to discriminate against them. The sizes of the training samples used for the NN training are shown in Table 4.5.

Table 4.5.: Numbers of simulated events used in the NN training.

Process	$e\tau$ 1P	$e\tau$ 3P	$\mu\tau$ 1P	$\mu\tau$ 3P
$Z \to \ell\tau$	31643	11927	31643	11927
$Z \to \tau\tau$	3515217	1379724	3344264	1373453
W+jets	100204	39213	100289	40273
$Z \to \ell\ell$	36751	not used	13953	not used

The training samples are randomly (but reproducibly) split into two sets of equal size. For each classifier, a pair of NNs are constructed and trained using the different sets of events. For each NN, the set of events that is not used in the training is considered the "evaluation set". The performance of the NN is estimated using the evaluation set. This ensures that the NNs are evaluated using unseen samples to avoid bias and allows cross validation of the two NNs for each classifier. The evaluation set is further split into two halves, one of which is used to optimise the hyperparameters for the NN architecture and training, while the other half is used to make other optimisation such as adding or dropping input variables.

4.2.2. Input variables

The input to the NNs is a mix of low-level and high-level kinematic variables.

The low-level variables are the four-momenta of the reconstructed ℓ, $\tau_{\text{had-vis}}$ and $E_{\text{T}}^{\text{miss}\dagger}$ in each event. In order to remove known spatial symmetries for optimal training, the low-level variables are first transformed before they are fed into the NNs. The transformation consists of the following steps in the listed order:

1. The ℓ–$\tau_{\text{had-vis}}$–$E_{\text{T}}^{\text{miss}}$ system is boosted in a direction in the transverse plane such that the total transverse momentum of the system becomes zero.
2. The system is rotated about the z-axis such that $E_{\text{T}}^{\text{miss}}$ is aligned with the x-axis.
3. If the $\tau_{\text{had-vis}}$ momentum has a negative z-component, the entire system is rotated about the (new) x-axis by π.

To preserve Lorentz invariance, no boosting in the z-direction or rotation about the x- or y-axis is involved, given that the true momentum of undetected particles in the z-direction is unknown to us. By removing known spatial symmetries, we spare the NNs the extra effort to "learn" the equivalence between events that are boosted and/or rotated with respect to each other but are otherwise the same. This essentially increases the effective training sample size. After the transformation, only six independent non-vanishing components are

† The four-momentum of $E_{\text{T}}^{\text{miss}}$ is defined with a zero mass and a zero z-component (i.e. $\eta = 0$).

left for the ℓ–$\tau_{\text{had-vis}}$–$E_{\text{T}}^{\text{miss}}$ system[†]: $\hat{p}_z(\ell)$, $\hat{E}(\ell)$, $\hat{p}_x(\tau_{\text{had-vis}})$, $\hat{p}_z(\tau_{\text{had-vis}})$, $\hat{E}(\tau_{\text{had-vis}})$ and $\hat{E}_{\text{T}}^{\text{miss}}$ (the hats indicate that these are variables measured in the transformed frame of reference). They are direct inputs to the NNs.

The high-level variables include the visible mass $m_{\text{vis}}(\ell, \tau)$, the collinear mass $m_{\text{coll}}(\ell, \tau)$, the track-based invariant mass $m(\ell, \tau \text{ track})$, and a kinematic discriminant known as $\Delta\alpha$ [61].

The track-based invariant mass is similar to the visible mass, except that the momentum of the $\tau_{\text{had-vis}}$ candidate is replaced by that of the leading-p_{T} track associated to the $\tau_{\text{had-vis}}$ candidate. This variable carries especially useful information for identifying $Z \to \ell\ell$ events with ℓ misidentified as $\tau_{\text{had-vis}}$. For $Z \to ee$ events with $e \to \tau_{\text{had-vis}}$ fakes, it is common for the electron to emit bremsstrahlung before entering or in the inner detector. In such cases, the momentum of the inner-detector track would be smaller than the original momentum of the electron before radiation. However, the momentum of the $\tau_{\text{had-vis}}$ candidate, which is reconstructed with energy deposits in the calorimeters taken into account, would likely recover the original momentum of the electron as the calorimeters are able to capture the energy carried away by the bremsstrahlung. Therefore, for these events, one can expect $m_{\text{vis}}(\ell, \tau_{\text{had-vis}}) \approx m_Z$ but $m(\ell, \tau \text{ track}) \not\approx m_Z$, where m_Z is the rest mass of the Z boson. On the other hand, for $Z \to \mu\mu$ events with $\mu \to \tau_{\text{had-vis}}$ fakes, the situation could be different. Since muons do not usually deposit much energy in the calorimeters, they are usually misidentified as $\tau_{\text{had-vis}}$ candidate only when the inner-detector track is wrongly matched with some random energy deposits in the calorimeters from other sources. For these events, one can instead expect $m_{\text{vis}}(\ell, \tau_{\text{had-vis}}) \not\approx m_Z$ but $m(\ell, \tau \text{ track}) \approx m_Z$. In any case, the difference between $m_{\text{vis}}(\ell, \tau_{\text{had-vis}})$ and $m(\ell, \tau \text{ track})$ is a piece of valuable information for the $\text{NN}_{\text{Z}\ell\ell}$ classifier.

The kinematic discriminant $\Delta\alpha(\ell, \tau)$ is defined by the equation

$$\Delta\alpha(\ell, \tau) = \frac{m_Z^2 - m_\tau^2}{2p(\ell) \cdot p(\tau_{\text{had-vis}})} - \frac{p_{\text{T}}(\ell)}{p_{\text{T}}(\tau_{\text{had-vis}})}, \tag{4.4}$$

where m_Z and m_τ are the nominal rest masses of the Z boson and τ lepton respectively, and $p(\ell)$ and $p(\tau_{\text{had-vis}})$ are the measured four-momenta of the ℓ and $\tau_{\text{had-vis}}$ candidates respectively. In the event of a signal $Z \to \ell\tau$ decay, one can reasonably assume that the momentum of the τ lepton before it decays ($p(\tau)$) is collinear with the $\tau_{\text{had-vis}}$ momentum ($p(\tau_{\text{had-vis}})$), i.e.

$$p(\tau) = \alpha\, p(\tau_{\text{had-vis}}), \tag{4.5}$$

where α is some constant. The value of α can be estimated with two different approaches. First, since the ℓ and τ lepton are the products of a Z boson decay,

$$\begin{aligned} m_Z^2 - m_\tau^2 &= 2\, p(\ell) \cdot p(\tau) \\ &= 2\, \alpha\, p(\ell) \cdot p(\tau_{\text{had-vis}}). \end{aligned} \tag{4.6}$$

[†] The $\tau_{\text{had-vis}}$ is assumed to have a zero rest mass

Second, assuming that the transverse momentum of the Z boson is small, the ℓ and τ lepton can be expected to have transverse momenta that are equal in magnitude (but opposite in direction). In other words,

$$\begin{aligned} p_{\mathrm{T}}(\ell) &= p_{\mathrm{T}}(\tau) \\ &= \alpha\, p_{\mathrm{T}}(\tau_{\text{had-vis}}). \end{aligned} \tag{4.7}$$

The difference between the values of α estimated by Equations (4.6) and (4.7) gives us $\Delta\alpha$ in Equation (4.4). For the signal events, where the assumptions needed to estimate α can be justified, $\Delta\alpha$ is expected to be close to zero. In contrast, for background events, where the assumptions generally do not hold true, $\Delta\alpha$ is expected to deviate from zero significantly. This makes $\Delta\alpha$ a powerful discriminant.

Including both low-level and high-level kinematic variables as input is found to be beneficial to the performance of the NN classifiers. Since the information provided by the variables is redundant, one might expect the addition of high-level variables to be unnecessary, and that the low-level variables alone would suffice. Although that might be true in theory, the finite training sample size and computing resources limit the actual ability of the NNs to fully explore the correlations between the low-level variables. In this regard, the addition of high-level variables is able to help the NNs to converge faster, while they continue to exploit the residual correlations between the low-level variables.

The input variables to the NN classifiers are listed in Table 4.6. The expected background and signal distributions of all the input variables in the SR can be found displayed in Appendix C.

4.2.3. Software, architecture and optimiser

The NN training and optimisation are implemented using the open-source software package Keras [94] with TensorFlow [95] backend.

All of the NNs used in the analysis share the same architecture. Each NN is a feed-forward network consisting of an input layer, two hidden layers of 20 neurons each, and an output layer with a single node. Each layer is fully connected to the neighbouring layers. Low-level and high-level variables are treated equally in the input layer, with one input node per variable. The hidden-layer neurons are rectified linear units, while the activation of the output node is a standard logistic sigmoid function. The number of hidden layers and the number of neurons per layer has been chosen by maximising the area under the ROC curve. The optimisation is done semi-manually with a coarse grid search.

The NNs are trained using the Adam algorithm [96] to minimise the binary cross-entropy. All NNs are trained with a batch size of 256 and 200 epochs. Optimised together with the number of hidden layers and the number of neurons per layer, the learning rate of the optimiser is set to 5×10^{-5} for the $\mathrm{NN}_{\mathrm{Z}\tau\tau}$ classifier for the 1P regions, 2.5×10^{-4} for the $\mathrm{NN}_{\mathrm{Z}\ell\ell}$ classifier, and 1×10^{-4} for the other classifiers. There is no indication of over-training, as the loss of the NNs on the evaluation sample sets stabilises and does not

Table 4.6.: Input to the NN classifiers. The list is common between the classifiers for the $e\tau$ and $\mu\tau$ channels. The top six variables are the low-level variables and are measured in the transformed frame of reference as described in text. The bottom four variables are the high-level variables and are measured in the laboratory frame of reference.

Variable	1P region			3P region	
	$\mathbf{NN_{Z\tau\tau}}$	$\mathbf{NN_{Wjets}}$	$\mathbf{NN_{Z\ell\ell}}$	$\mathbf{NN_{Z\tau\tau}}$	$\mathbf{NN_{Wjets}}$
$\hat{p}_z(\ell)$	•	•	•	•	•
$\hat{E}(\ell)$	•	•	•	•	•
$\hat{p}_x(\tau_{\text{had-vis}})$	•	•	•	•	•
$\hat{p}_z(\tau_{\text{had-vis}})$	•	•	•	•	•
$\hat{E}(\tau_{\text{had-vis}})$	•	•	•	•	•
$\hat{E}_{\text{T}}^{\text{miss}}$	•	•	•	•	•
$\Delta\alpha(\ell,\tau)$	•	•	•	•	•
$m_{\text{vis}}(\ell,\tau)$	•	•	•	•	•
$m_{\text{coll}}(\ell,\tau)$	•	•	•	•	•
$m(\ell,\tau$ track)			•		

have a tendency to increase after a certain epoch (see Appendix B). In view of that, no regularisation or dropout is added. Other unspecified configurations or hyperparameters have not been optimised explicitly and the default values from Keras 1.1.0 are used.

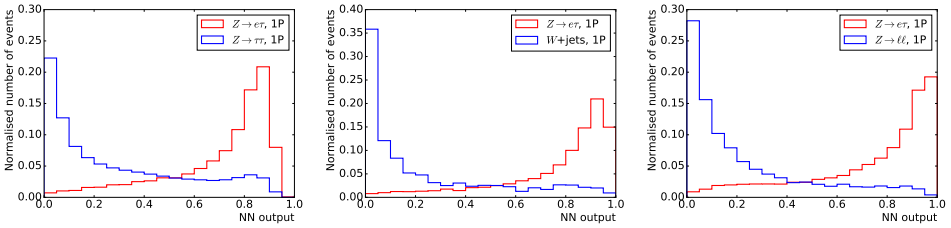
The configurations and hyperparameters of the NN architecture and optimiser are summarised in Table 4.7.

Figure 4.2 shows the expected output distributions of the targeted background and signal of the different NN classifiers.

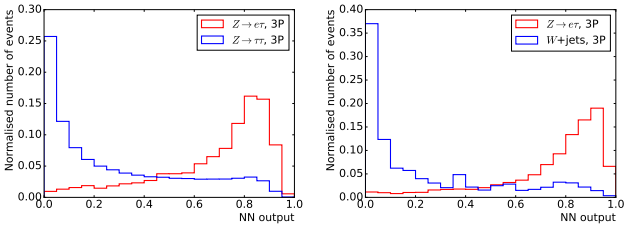
4.2.4. Combined output

For every event, each NN classifier would output a score ranging from zero to one. The score reflects the likelihood of an event to be a signal event. The most background-like events would be given a score close to zero, while the most signal-like events would receive a score close to one. For each channel ($e\tau$ or $\mu\tau$, and 1P or 3P) separately, the output scores from the different classifiers are combined into one final discriminant, called the "combined NN output" or "NN_{comb} output", using the formula

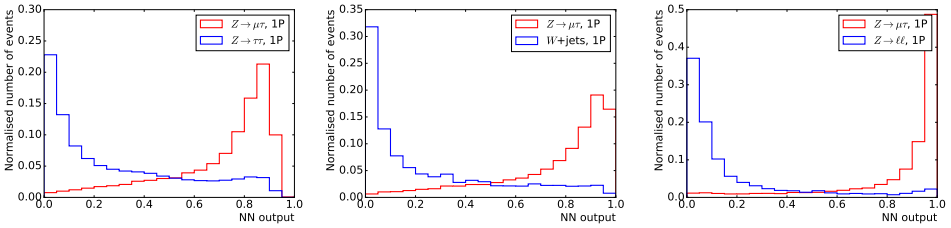
$$\text{combined NN output} = 1 - \sqrt{\frac{\sum_b w_b \ (1 - \text{NN}_b \text{ output})^2}{\sum_b w_b}}, \tag{4.8}$$



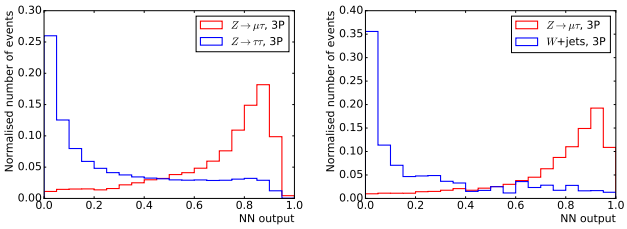
(a) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $e\tau$ 1P **(b)** $\mathrm{NN}_{\mathrm{Wjets}}$, $e\tau$ 1P **(c)** $\mathrm{NN}_{\mathrm{Z}\ell\ell}$, $e\tau$ 1P



(d) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $e\tau$ 1P **(e)** $\mathrm{NN}_{\mathrm{Wjets}}$, $e\tau$ 1P



(f) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $\mu\tau$ 1P **(g)** $\mathrm{NN}_{\mathrm{Wjets}}$, $\mu\tau$ 1P **(h)** $\mathrm{NN}_{\mathrm{Z}\ell\ell}$, $\mu\tau$ 1P



(i) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $\mu\tau$ 1P **(j)** $\mathrm{NN}_{\mathrm{Wjets}}$, $\mu\tau$ 1P

Figure 4.2.: Expected output distributions of the targeted background and signal of the different NN classifiers.

Table 4.7.: Configurations and hyperparameters of the NN architecture and optimiser. Definitions of β_1, β_2 and the learning rate can be found in Reference [96].

Configuration/hyperparmeter	**Value**
Architecture	
Number of hidden layer	2
Number of neurons per hidden layer	20
Connection between layers	Fully connected (Dense)
Hidden-layer activation	Rectified linear unit (ReLU)
Output-layer activation	Standard logistic sigmoid
Optimiser	
Loss function	Binary cross-entropy
Training epochs	200
Batch size	256
Algorithm	Adam (non-AMSGrad variant)
Learning rate ($\mathrm{NN}_{\mathrm{Z}\tau\tau}$, 1P)	5×10^{-5}
Learning rate ($\mathrm{NN}_{\mathrm{Z}\ell\ell}$, 1P)	2.5×10^{-4}
Learning rate (others)	1×10^{-4}
β_1	0.9
β_2	0.999
Regularisation	None
Dropout	None

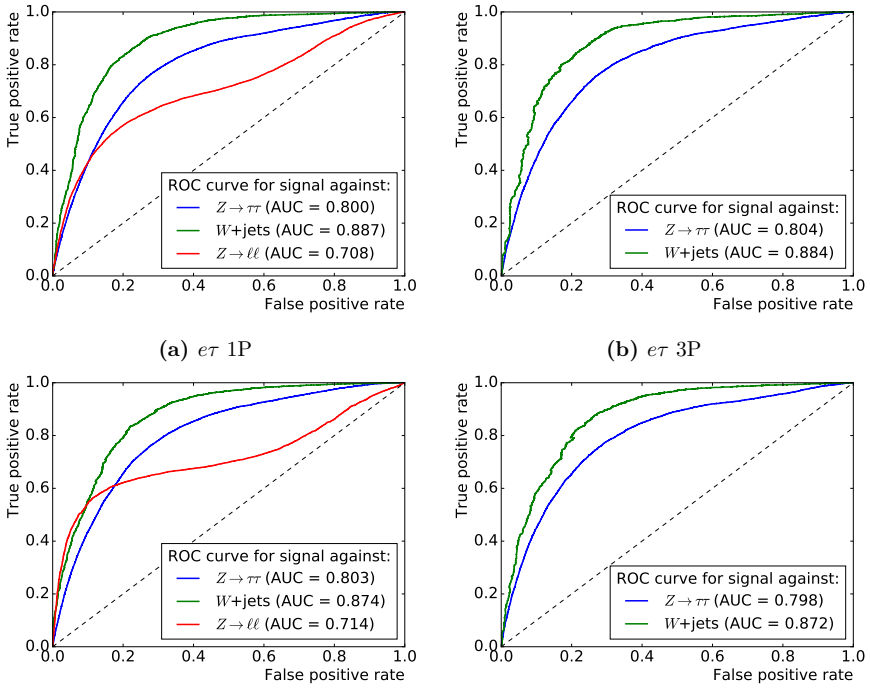
where the summation is over $b = \mathrm{Z}\tau\tau, \mathrm{Wjets}, \mathrm{Z}\ell\ell$ ($b = \mathrm{Z}\tau\tau, \mathrm{Wjets}$) for events with 1-prong (3-prong) $\tau_{\text{had-vis}}$ candidates, and $\{w_b\}$ are constant weights assigned to the corresponding individual classifiers and are free parameters of the combination formula. By construction, the combined NN output also ranges from zero to one, with one again representing the most signal-like event and zero representing the most generally background-like event.

With Equation (4.8), we have the freedom to choose the values of $\{w_b\}$. If $w_{\mathrm{Z}\tau\tau} = w_{\mathrm{Wjets}} = w_{\mathrm{Z}\ell\ell}$, for events in the 1P region, the combined NN output simply represents one minus the normalised distance in a three-dimensional space between the point ($\mathrm{NN}_{\mathrm{Z}\tau\tau}$ output, $\mathrm{NN}_{\mathrm{Wjets}}$ output, $\mathrm{NN}_{\mathrm{Z}\ell\ell}$ output) for the considered event and the point $(1, 1, 1)$ that corresponds to the theoretically most signal-like event possible (or similarly in a two-dimensional space for events in the 3P region). If different values for $\{w_b\}$ are chosen, the distance along certain directions in the $\mathrm{NN}_{\mathrm{Z}\tau\tau}$-output–$\mathrm{NN}_{\mathrm{Wjets}}$-output–$\mathrm{NN}_{\mathrm{Z}\ell\ell}$-output space would be amplified, thus giving the output of certain individual classifiers more importance. The choice of the values of $\{w_b\}$ has an impact on the sensitivity of the analysis,

as they affect the separation of the different background processes along the combined NN output, which in turn affects our ability to constrain the background modelling using observed data in statistical fits. To optimise the values of $\{w_b\}$, a grid search was performed and the values that yield the best expected upper limit on the LFV branching fractions were chosen. The optimised values have the ratio $w_{\mathrm{Ztt}} : w_{\mathrm{Wjets}} : w_{\mathrm{Zll}} = 1.0 : 1.5 : 0.33$, common to both the $e\tau$ and $\mu\tau$ channels.

Figure 4.3 shows the expected ROC curves for the combined NN output in the different channels, while Figure 4.4 shows the expected signal and major background distributions in the SR. As shown in the figures, the distributions of different background processes are clearly separated from each other, as expected from the optimisation of $\{w_b\}$. The distribution of the combined NN output in the SR is used in binned maximum-likelihood fits to extract evidence of signal or set upper limits on the LFV branching fractions (see Section 6.1). It is also used to reject the most background-like events in the SR. These events are kinematically very different from the signal events.

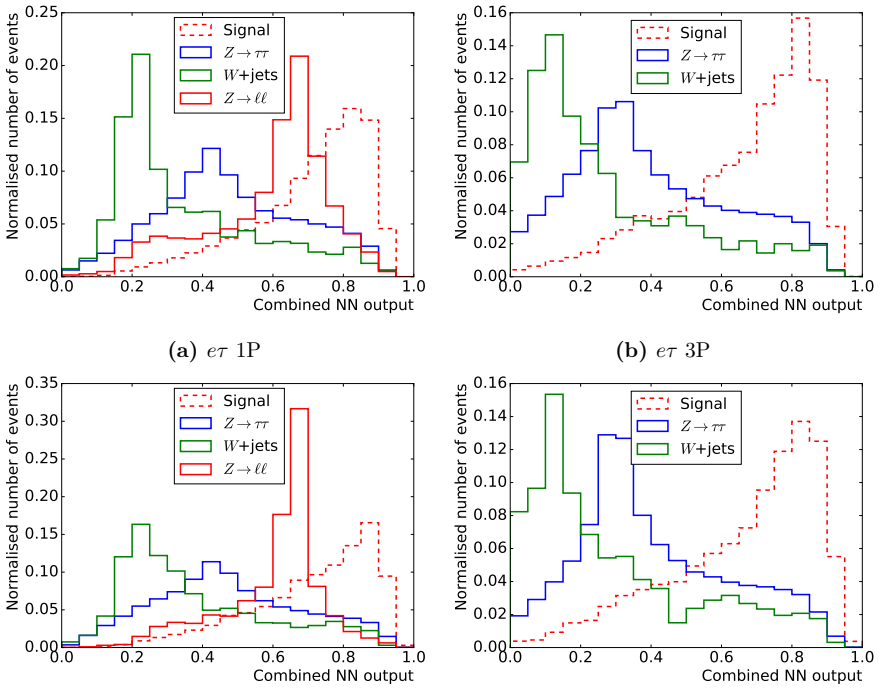
Other alternative ways of creating a one-dimensional final discriminant have also been experimented. That includes what will be referred to as the "combined-background classifier approach" and the "multiclass classifier approach". The combined-background classifier approach uses a single binary classifier trained with a combined set of background samples instead of having multiple, individual classifiers for the different backgrounds. The background samples are weighted according to the relative importance of the backgrounds in the SR. The multiclass classifier approach uses a neural network with (number of major backgrounds + 1) output nodes to classify the signal and the major background events at the same time. The multiclass classifier uses a softmax activation [97] and is trained to minimise the categorical cross-entropy. The performance of the final discriminants created with the considered alternative approaches are found to be worse, in terms of expected sensitivity, than what the current chosen approach could offer. This is not unreasonable as in the alternative approaches, NN classifiers that directly output the final discriminant could only be optimised based on some proxy figures of merit, such as cross-entropies or the area under ROC curves. However, these figures of merit might not faithfully reflect the actual sensitivity of the analysis. Whereas in the chosen approach, the parameters for the combination ($\{w_b\}$) can be optimised with the actual procedure of maximum-likelihood fitting and the systematic uncertainties taken into account.



(a) $e\tau$ 1P **(b)** $e\tau$ 3P

(c) $\mu\tau$ 1P **(d)** $\mu\tau$ 3P

Figure 4.3.: Expected ROC curves for the combined NN output in the 1P and 3P SR of the $e\tau$ and $\mu\tau$ channels. The ROC curves for classifying signal events against $Z \to \tau\tau$, W+jets or $Z \to \ell\ell$ events are separately shown. The area under the curves (AUC) are shown as a figure of merit. The rather peculiar shape of the ROC curves for the classification of $Z \to \ell\ell$ events is mainly due to the low weight given to the $\mathrm{NN}_{\mathrm{Z}\ell\ell}$ output in the combination.



(a) $e\tau$ 1P **(b)** $e\tau$ 3P

(c) $\mu\tau$ 1P **(d)** $\mu\tau$ 3P

Figure 4.4.: Expected distributions of the combined NN output in the 1P and 3P SR of the $e\tau$ and $\mu\tau$ channels. The distributions for the signal, $Z \to \tau\tau$, W+jets and $Z \to \ell\ell$ events are separately shown.

Chapter 5.

Signal and background modelling

> *"The fake is of far greater value. In its deliberate attempt to be real, it's more real than the real thing."*
>
> — Ishin Nishio, "Deshū Kaiki", *Nisemonogatari*

To be able to compare the observed data with predictions, one must be able to model the probability distributions of all relevant observables according to the SM or the hypotheses that are to be tested. However, given the many coupled degrees of freedom, it is virtually impossible to calculate the distributions analytically. In this chapter, we will discuss the methods used to model the LFV $Z \to \ell\tau$ signal and the SM background events. These methods are partly based on MC simulations and partly based on measurements in regions that have little or no signal contamination expected (data-driven). The signal and the contributions to the SM background can be classified into three main categories:

1. Events with a single Z-boson decay. These are the signal, $Z \to \tau\tau$ and $Z \to \ell\ell$ events.
2. Events with a quark- or gluon-initiated jet misidentified as the $\tau_{\text{had-vis}}$ candidate (fakes). They are dominantly $W(\to \ell\nu)$+jets and multijet events, with little contribution from $Z(\to \ell\ell)$+jets and $t\bar{t}$ events.
3. Other events that are not single Z-boson decays or fakes. These include $t\bar{t}$, diboson, $H \to \tau\tau$, $H \to WW$ and $W(\to \tau\nu)$+jets events.

Events in the first category are modelled using MC simulations with data-driven corrections for the production cross section and p_{T} spectrum of the Z bosons, and for the $\ell \to \tau_{\text{had-vis}}$ fake rate for $Z \to \ell\ell$ events. The generation of the MC samples will be described in Section 5.1 while the data-driven corrections will be described in Sections 5.3 and 5.4. Events in the second category are predicted using a data-driven method known as the fake-factor method, which will be detailed in Section 5.5. Events in the third category, which have only minor contributions to the total SM background, are purely modelled by MC simulations.

5.1. Monte Carlo simulations

Different Monte Carlo samples are generated for different physical processes. These processes have either different final states or very little interference with each other, meaning that their contributions to the total event yields can simply be summed together (represented by histograms stacked on top of each other in plots).

The signal $Z \to e\tau$ and $Z \to \mu\tau$ samples are simulated Z+jets events where the Z boson undergoes a two-body decay into a $e\tau$ or $\mu\tau$ final state. These events are generated using Pythia 8.2 [98] with matrix elements calculated at leading order (LO) in the strong coupling constant α_s. Parameters for initial-state radiations, multiparton interactions and beam remnants are set according to the A14 set of tuned parameters (tune) [99] with the NNPDF2.3 LO parton distribution function (PDF) set [100]. Nominal signal samples are generated with unpolarised τ leptons, which correspond to the specific scenario where the effective $Z\ell\tau$ interaction vertex is parity-conserving. Scenarios with polarised τ leptons and parity-violating $Z\ell\tau$ vertices are considered by reweighting the nominal samples according to the probability of occurrence of each generated events under different assumptions. The reweighting method will be described in Section 5.2.

The $Z(\to \tau\tau)$+jets and W+jets events are generated using Sherpa 2.2.1 [101] with matrix elements calculated using the Comix [102] and OpenLoops [103–105] libraries at next-to-leading-order (NLO) in α_s for up to two partons, and at LO for up to four partons. They are matched with the Sherpa parton shower [106] using the MEPS@NLO prescription [107–110] with the default Sherpa tune.

The $Z(\to \ell\ell)$+jets events are generated using Powheg-Box [111] for calculating the matrix elements and Pythia 8 for modelling the parton showers, hadronisation, and underlying-event activity (Powheg+Pythia 8). The matrix elements are calculated at NLO in α_s. The CT10 PDF set [112] is used for the hard-scattering processes, whereas the CTEQ6L1 PDF set [113] and the AZNLO tune [114] are used for the parton showers.

The overall prefit (see Section 6.1) yield of each of the abovementioned Z-decay samples is normalised to match the measured Z-production cross section $\sigma(Z) = 1.981\,\text{nb}$ [62].

The $t\bar{t}$, single-top and Wt events are generated using Powheg+Pythia 8 with the NNPDF3.0 NLO PDF set [115] and the A14 tune.

The diboson events are generated using Sherpa 2.2.1 for purely leptonic decays and Sherpa 2.2.2 for semileptonic and semihadronic decays with the NNPDF3.0 next-to-next-to-leading-order (NNLO) PDF and a set-up for parton showers modelling similar to that for the $Z \to \tau\tau$ and W+jets samples.

The $H \to \tau\tau$ and $H \to WW$ events are generated using Powheg+Pythia 8 with the NNPDF3.0 NNLO PDF set and a set-up for parton showers modelling similar to that for the $Z \to \ell\ell$ samples. Samples are generated separately for gluon-fusion and vector-boson-fusion productions. For gluon-fusion productions, the NNLOPS program [116, 117] is used to yield a fully exclusive, hadron-level description of the final state. No

samples are generated for top-fusion productions, which have negligible contribution to the total background in the relevant regions. The overall yields of the Higgs boson samples are normalised using the predicted cross sections and Higgs-decay branching fractions recommended by the LHC Higgs Cross Section Working Group in Reference [118]. The cross sections are calculated at next-to-next-to-next-to-leading-order (N3LO) in α_s and NLO in the electroweak coupling constants.

After being generated, all simulated events are set to go through a detailed simulation of the ATLAS detector implemented using GEANT 4 [119]. Furthermore, simulated inelastic pp collisions, generated using PYTHIA 8 with the NNPDF2.3 LO PDF set and the A3 tune [120], are overlaid with the simulated hard-scattering events to model pile-up. The events are reweighted to match the pile-up conditions measured in the different data taking periods. Finally, all simulated events are processed using the same reconstruction algorithms as the ones used for the actual data, which have been introduced in Chapter 3.

5.2. τ polarisation reweighting

As mentioned earlier, the simulated signal events are generated with unpolarised τ leptons (nominal signal), which corresponds to the assumption of a parity-conserving $Z\ell\tau$ interaction vertex. However, depending on the mechanism behind the LFV coupling, the effective $Z\ell\tau$ vertex could be parity-violating and the τ leptons from the Z-boson decays could be polarised.

The τ polarisation state determines the decay topology of the τ lepton, which in turn affects the kinematic distributions of the reconstructed $\tau_{\text{had-vis}}$ and $E_{\text{T}}^{\text{miss}}$ in a signal event. Therefore, the assumption for the τ polarisation has a significant impact on the acceptance of $\tau_{\text{had-vis}}$ candidates and the NN output distributions for signal events in the SR. This ultimately impacts the signal sensitivity of the search.

In order to generalise the analysis, the nominal signal samples are reweighted using the software package TAUSPINNER [121] to consider as well the scenarios in which the effective $Z\ell\tau$ vertex is maximally parity-violating and the τ leptons from the LFV Z-boson decays are either all left-handed or all right-handed. Different constraints on the LFV branching fraction can then be set for the different considered scenarios in case of no discovery.

For each generated signal event, TAUSPINNER uses the truth event records to identify the τ decay mode and the kinematics of the final-state particles. It then calculates the probabilities of occurrence of the event, averaged over all possible initial states according to the assumed parton distribution (NNPDF2.3 LO), under the assumptions that the τ lepton is left-handed or unpolarised. By dividing the probability under the left-handed τ assumption by the probability under the unpolarised τ assumption, TAUSPINNER obtains and outputs an event weight $w(p = -1)$ that can be used to reweight the nominal signal sample to simulate the scenario where only left-handed τ leptons are produced.

Generally, the event weight can be interpreted as

$$w(p) = 1 + \mathbf{s} \cdot \mathbf{h} = 1 + p \cdot h_z, \tag{5.1}$$

where $\mathbf{s}$ is the τ polarisation vector, $\mathbf{h}$ is the polarimetric vector constructed using the hadronic currents, p is the polarisation of the τ lepton (equals -1 for left-handed and 1 for right-handed τ leptons) and h_z is the component of $\mathbf{h}$ in the travelling direction of the τ lepton. Therefore, the weight $w(-1)$ provided by TauSpinner can also be used to derive weights for simulating scenarios with any assumed τ polarisation states:

$$w(p) = 1 - p\,(w(-1) - 1)\,. \tag{5.2}$$

This means that by applying the weight $2 - w(-1)$ to the nominal signal sample, the scenario where only right-handed τ leptons ($p = 1$) are produced can also be simulated.

Since a nominal signal sample is generated with unpolarised τ leptons and can therefore be viewed as an ensemble of events with 50% left-handed and 50% right-handed τ leptons, it is able to represent scenarios with any assumed τ polarisation states with a sufficient sample size. For this reason, the neural network classifiers described in Section 4.2 are only trained with the nominal signal sample.

Figure 5.1 shows the combined NN output distributions of the signal events in the SR before and after the τ polarisation reweighting. As can be seen, the assumption for the τ polarisation has a rather significant impact on combined NN output distributions in the 1P regions.

5.3. Corrections to the simulated Z-boson production

Events with Z-boson decays ($Z \to \tau\tau$, $Z \to \ell\ell$ and the signal $Z \to \ell\tau$) are modelled based on MC simulations. These simulations come with uncertainties in the theoretical predictions (theory uncertainties). The Z-boson decays themselves are purely electroweak processes and can be predicted with negligible uncertainties. Nonetheless, the cross section, $\sigma(Z)$, and transverse momentum, $p_{\mathrm{T}}(Z)$, of the simulated Z bosons can only be predicted with much larger theory uncertainties, as they also depend on the simulation of QCD activities. These theory uncertainties could have a significant negative impact on the sensitivity of the analysis. In order to avoid that, data-driven corrections are applied to the overall prefit normalisations and the simulated $p_{\mathrm{T}}(Z)$ spectra of the $Z \to \tau\tau$, $Z \to \ell\ell$ and signal MC samples.

The prefit normalisations of all the Z-decay MC samples are first corrected to match the measured Z-boson production cross section $\sigma(Z) = 1.981\,\mathrm{nb}$ [62]. Then for the $Z \to \tau\tau$ and signal samples, their actual normalisations are determined in a maximum-likelihood fit to data in the SR and CRZ$\tau\tau$ (see Section 6.1). This allows us to make use of the large amount of data, especially in CRZ$\tau\tau$, to strongly constrain $\sigma(Z)$, along with the systematic uncertainties in the simulation of real $\ell\tau_{\text{had-vis}}$ final states. For the $Z \to \ell\ell$

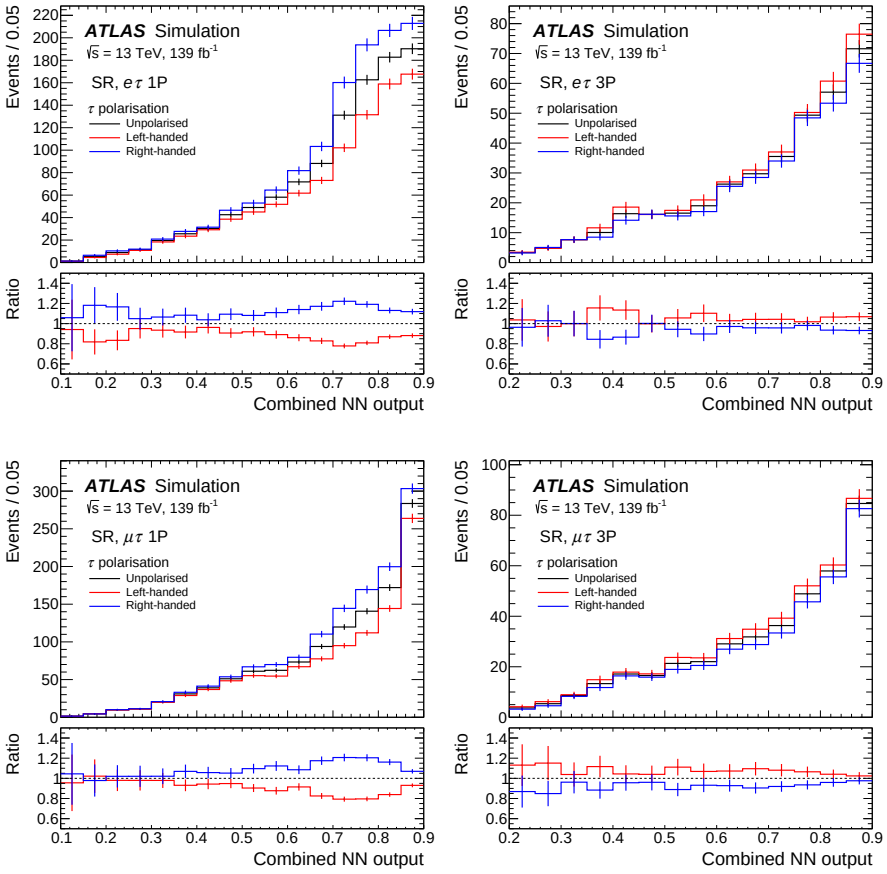


Figure 5.1.: Combined NN output distributions in the SR before and after the τ polarisation reweighting. The ratios of the reweighted distributions to the original distributions (corresponding to the scenario with unpolarised τ leptons) are shown in the bottom panels. The signal is normalised to $\mathcal{B}(Z \to \ell\tau) = 10^{-5}$. The error bars represent the MC statistical uncertainties.

samples, which only account for a relatively small fraction of background events in the SR, their normalisations are fixed at the prefit value in the maximum-likelihood fit, with the uncertainty in the cross section measurement, $\pm 0.057\,\text{nb}$, considered as a systematic uncertainty.

The $p_{\text{T}}(Z)$ distributions for each of the the Z-decay MC samples is corrected by reweighting the simulated events to match the measured $p_{\text{T}}(Z)$ spectrum [122]. The measurement was performed with pp collision data at $\sqrt{s} = 13\,\text{TeV}$, where events with dielectron or dimuon final states are analysed in a fiducial phase space that is close to the detector acceptance for leptons in $p_{\text{T}}(\ell)$, $\eta(\ell)$ and $m(\ell,\ell)$. The unfolded normalised $p_{\text{T}}(\ell,\ell)$ spectrum after correcting for the detector effects was reported at Born level (before final-state radiations).

Since the $Z \to \tau\tau$, $Z \to \ell\ell$ and signal samples are generated with different MC generators, PDF sets and tunes, correction scale factors are derived separately for each of them. For each simulated event in the fiducial region where the $p_{\text{T}}(Z)$ spectrum was measured, the true value of $p_{\text{T}}(Z)$ is retrieved or recalculated from the truth event records. The obtained $p_{\text{T}}(Z)$ distributions are then compared to the measured $p_{\text{T}}(Z)$ spectrum to derive scale factors as functions of $p_{\text{T}}(Z)$. The uncertainties in the scale factors, which are dominated by MC statistical errors, are considered as a systematic uncertainty in maximum-likelihood fits.

Figure 5.2 shows the simulated $p_{\text{T}}(Z)$ distributions in comparison with the measured distribution, and derived correction scale factors and their uncertainties. The obtained values of the correction scale factors are consistent with the data-model comparisons reported in Reference [122].

5.4. Corrections to simulated events with $\ell \to \tau_{\text{had-vis}}$ misidentification

As mentioned, $Z \to \ell\ell$ events, where an electron or muon is misidentified as a $\tau_{\text{had-vis}}$ candidate, are modelled by simulations. However, the simulations are not perfect in modelling the rate of $\ell \to \tau_{\text{had-vis}}$ misidentification. To improve the accuracy of the prediction, Corrections are derived using data in CRZ$\ell\ell$ and applied to simulated $Z \to \ell\ell$ events in the other regions. Correction scale factors, binned in $p_{\text{T}}(\tau)$ and $|\eta(\tau)|$, are calculated as the ratio of the observed number of events minus the predicted number of non-$Z \to \ell\ell$ SM events to the predicted number of $Z \to \ell\ell$ events. Given the low misidentification rate of 3-prong $\tau_{\text{had-vis}}$ candidates, the corrections are only derived for events with 1-prong $\tau_{\text{had-vis}}$ candidates. Separate scale factors are derived for $Z \to ee$ and $Z \to \mu\mu$ events. The binning, values and errors of the derived scale factors are shown in Figure 5.3. The statistical uncertainties in the scale factors are considered as a systematic uncertainty in maximum-likelihood fits. The systematic uncertainties in the derivation of the scale factors due to the subtraction of other SM background events are negligible compared to the data and MC statistical uncertainties.

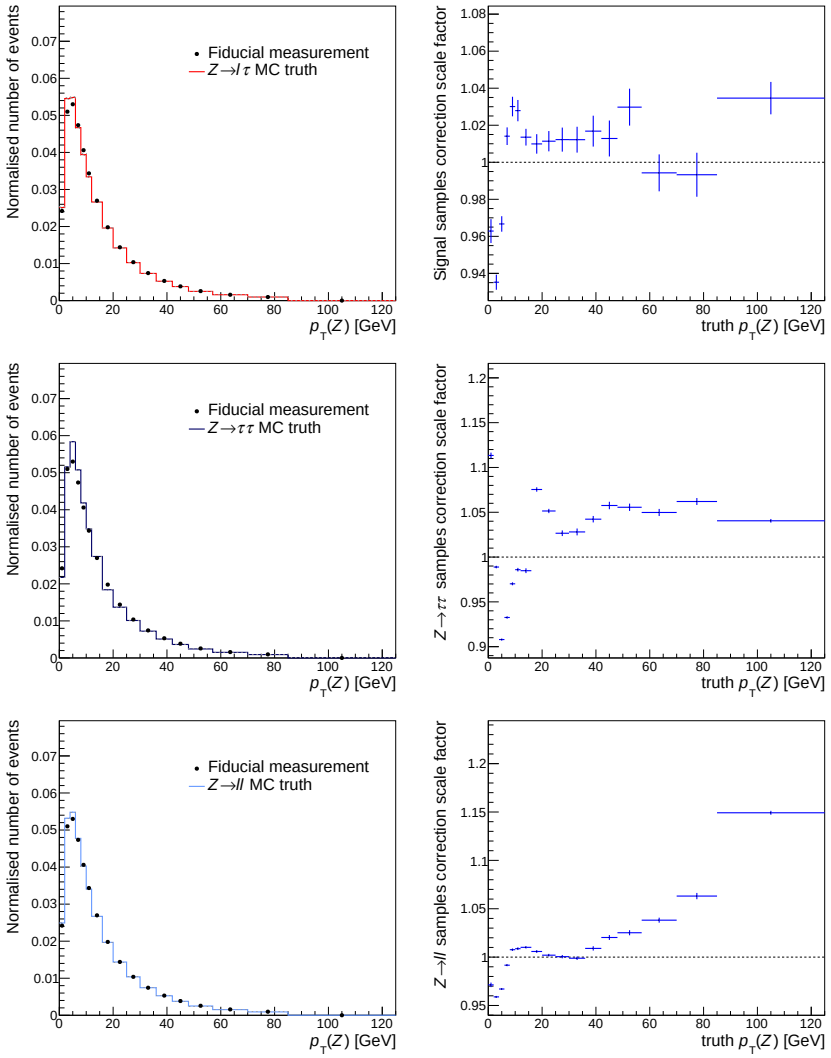
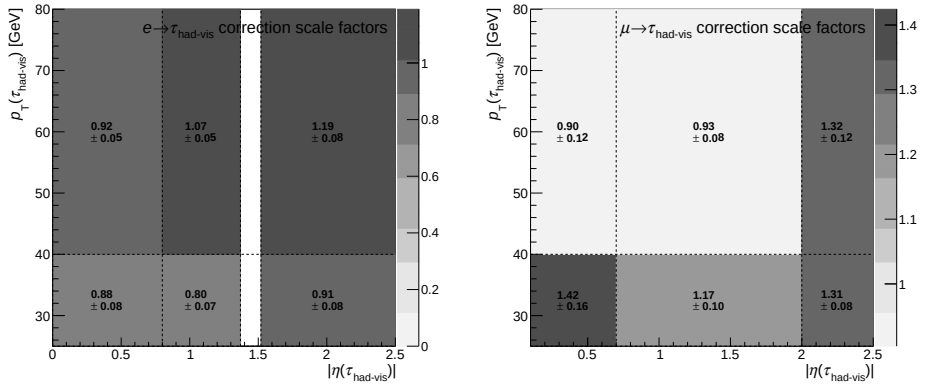


Figure 5.2.: Normalised simulated $p_{\mathrm{T}}(Z)$ distributions (left column) for the signal (top row), $Z \to \tau\tau$ (middle row) and $Z \to \ell\ell$ (bottom row) samples compared to the measured distribution, and the corresponding $p_{\mathrm{T}}(Z)$ correction scale factors (right column). For the $p_{\mathrm{T}}(Z)$ distributions, the bin contents are divided by the bin widths in units of GeV. The rightmost bins include overflow events up to $p_{\mathrm{T}}(Z) = 2500\,\mathrm{GeV}$. The errors for both the predicted and measured distributions are not visible in the displayed scale, while the errors for the correction scale factors are represented by the error bars.



(a) **(b)**

Figure 5.3.: The (a) $e \to \tau_{\text{had-vis}}$ and (b) $\mu \to \tau_{\text{had-vis}}$ correction scale factors and their uncertainties. While the plot only shows $p_{\text{T}}(\tau_{\text{had-vis}})$ up to 80 GeV, the actual bin boundaries are at infinity.

The observed and expected distributions of several kinematic variables in CRZ$\ell\ell$ after the correction are shown in Figures 5.4 and 5.5 for the $e\tau$ and $\mu\tau$ channels respectively. The data and the corrected predictions agree well with each other within uncertainties.

5.5. Modelling of events with jet $\to \tau_{\text{had-vis}}$ misidentifcation

Modelling events with jet $\to \tau_{\text{had-vis}}$ fakes purely by MC simulations is extremely challenging. This is because it requires detailed and accurate modelling of the hadronisation of quarks or gluons and the interactions between the particle showers and the materials in the calorimeters. Adding to the challenge, there is simply no robust MC simulation for purely QCD multijet events because of the difficulties in calculating QCD matrix elements at low-energy scales. Therefore, a data-driven method broadly known as the fake-factor (FF) method is used to model such background processes. The method also partially addresses jet $\to \ell$ fakes indirectly for the cases where these fakes come from multijet events. This section describes in details the implementation of the FF method in our analysis.

5.5.1. Concept and definitions

In this section, the following definitions of subscripts or superscripts are used:

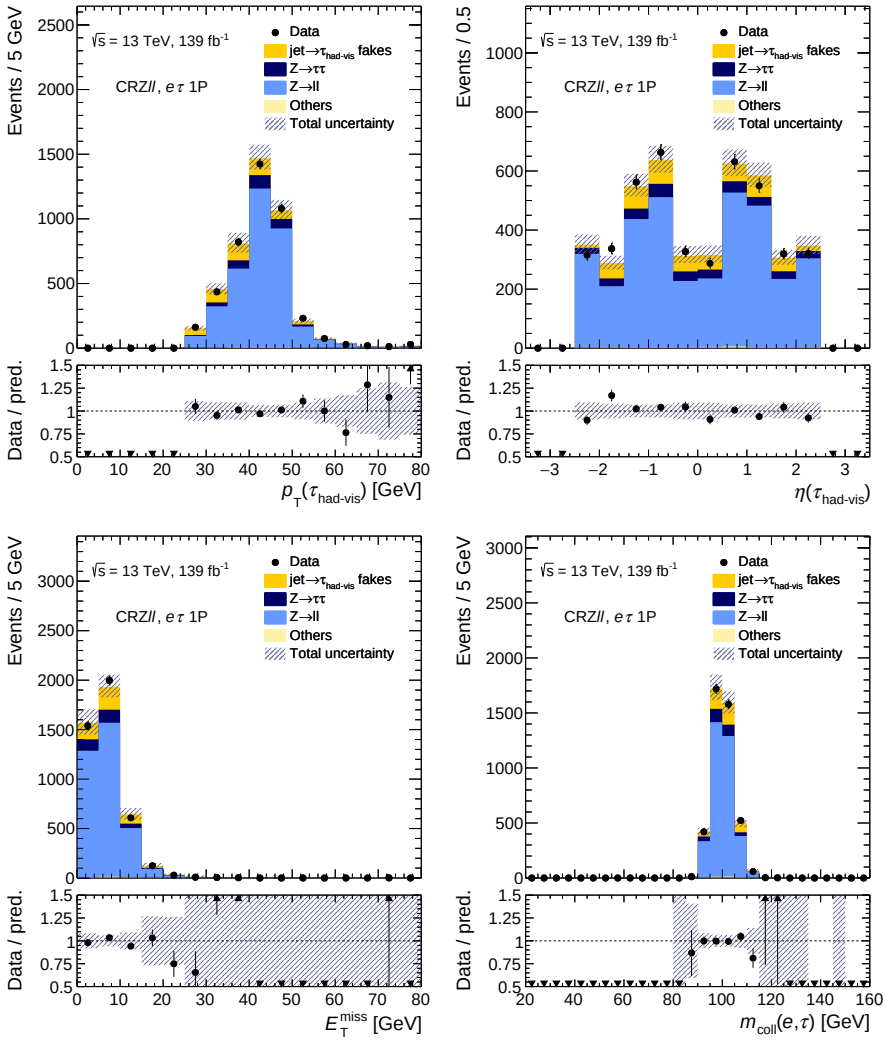


Figure 5.4.: Expected and observed distributions of $p_{\mathrm{T}}(\tau_{\text{had-vis}})$, $\eta(\tau_{\text{had-vis}})$, $E_{\mathrm{T}}^{\mathrm{miss}}$ and $m_{\mathrm{coll}}(e,\tau)$ in the CRZ$\ell\ell$ of the $e\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties.

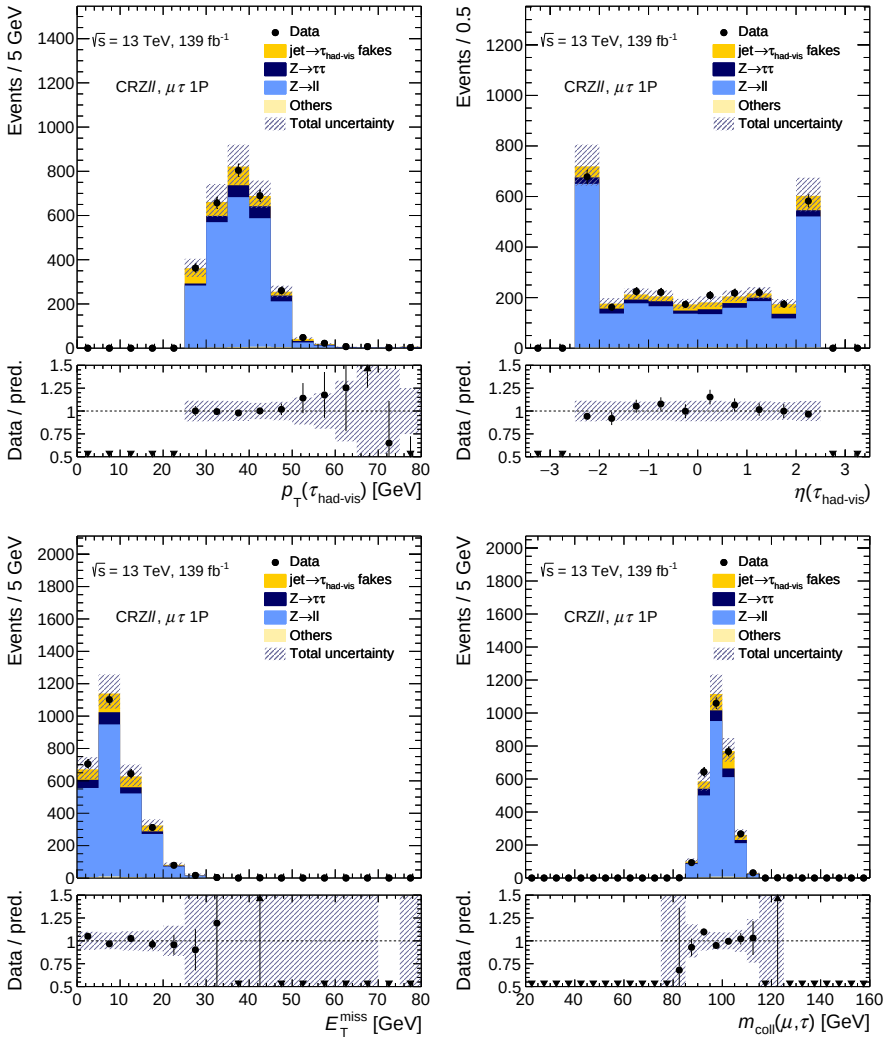


Figure 5.5.: Expected and observed distributions of $p_{\mathrm{T}}(\tau_{\text{had-vis}})$, $\eta(\tau_{\text{had-vis}})$, $E_{\mathrm{T}}^{\text{miss}}$ and $m_{\text{coll}}(\mu,\tau)$ in the CRZ$\ell\ell$ of the $\mu\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties.

pass Events of which the leading-p_{T} $\tau_{\text{had-vis}}$ candidate passed the `Tight` RNN ID requirement.

fail Events of which the leading-p_{T} $\tau_{\text{had-vis}}$ candidate failed the `Tight` but passed the `Loose` RNN ID requirements.

fake Simulated events of which the leading-p_{T} $\tau_{\text{had-vis}}$ candidate originates from a misidentified gluon- or quark-initiated jet.

real Simulated events of which the leading-p_{T} $\tau_{\text{had-vis}}$ candidate originates from a real τ-lepton or a misidentified light lepton.

data Observed events in the collected data.

MC Simulated events from any SM processes (except multijet events).

MC$_p$ Simulated events from a specific process p.

MC$_{\text{not }p}$ Simulated events from any SM processes except from a specific process p (and except multijet events).

With the fake-factor method, the yield of fakes in a given region r[†] is estimated by

$$N^{\text{pass}}_{r,\text{fake}} = \left(N^{\text{fail}}_{r,\text{data}} - N^{\text{fail}}_{r,\text{MC,real}}\right) \times F_r. \tag{5.3}$$

The (region-specific) fake factor F_r for any region r is defined as a weighted average of the process-specific fake factors F^p:

$$F_r = \sum_p R^p_r F^p = R^{\mathrm{W}}_r F^{\mathrm{W}} + R^{\mathrm{Z}}_r F^{\mathrm{Z}} + R^{\text{top}}_r F^{\text{top}} + R^{\text{QCD}}_r F^{\text{QCD}}, \tag{5.4}$$

where $p = \mathrm{W}, \mathrm{Z}, \text{top}, \text{QCD}$ are labels representing the SM processes $W(\to \ell\nu)$+jets, $Z(\to \ell\ell)$+jets, $t\bar{t}$ and QCD multijet respectively.

The process-specific fake factors are derived from data in the corresponding fakes-enriched regions FR_p that have been defined in Section 4.1.4: $\mathrm{FR}_{\mathrm{W}} = \mathrm{FRW}$, $\mathrm{FR}_{\mathrm{Z}} = \mathrm{FRZ}$, $\mathrm{FR}_{\text{top}} = \mathrm{FRT}$ and $\mathrm{FR}_{\text{QCD}} = \mathrm{FRQ}$. They are defined as the ratio of the number of events with fakes passing the `Tight` RNN ID requirement to the number of those failing it but passing the `Loose` RNN ID requirement:

$$F^p = \frac{N^{\text{pass}}_{\mathrm{FR}_p,\text{data}} - N^{\text{pass}}_{\mathrm{FR}_p,\mathrm{MC}_{\text{not }p}} - N^{\text{pass}}_{\mathrm{FR}_p,\mathrm{MC}_{p,\text{real}}}}{N^{\text{fail}}_{\mathrm{FR}_p,\text{data}} - N^{\text{fail}}_{\mathrm{FR}_p,\mathrm{MC}_{\text{not }p}} - N^{\text{fail}}_{\mathrm{FR}_p,\mathrm{MC}_{p,\text{real}}}}. \tag{5.5}$$

The second terms in both the denominator and numerator are there to subtract the contributions from SM other processes (except multijet), while the third terms are there to subtract contributions from events that are not with jet $\to \tau_{\text{had-vis}}$ fakes. Since there is no reliable MC simulations for multijet events, they cannot be subtracted when p is not QCD. Still and all, because the FRs are enriched in events from the process that they are

[†] This could refer to a subregion, or a single bin in any variables, in any of the regions defined in Section 4.1.

designed for, the impact of not having simulated multijet events or that of the potential mismodelling in the subtracted simulated events on the derived fake factors is expected to be negligible.

The weights R_r^p in Equation (5.4) are the relative contributions of events from the process p to the total yield of events with fakes in region r. They are estimated using the formulae:

$$R_r^p = \frac{N^{\text{fail}}_{r,\text{MC}_p,\text{fake}}}{N^{\text{fail}}_{r,\text{data}} - N^{\text{fail}}_{r,\text{MC,real}}} \tag{5.6}$$

for $p = \text{W}, \text{Z}, \text{top}$, and

$$R_r^{\text{QCD}} = 1 - \sum_{p \neq \text{QCD}} R_r^p = 1 - R_r^{\text{W}} - R_r^{\text{Z}} - R_r^{\text{top}} \tag{5.7}$$

for $p = \text{QCD}$.

Since Equations (5.5) and (5.6) rely (partially) on the fakes estimated by MC simulations, offset can be introduced by virtue of the fact that these estimates are not scaled to data. Given that the FRs are enriched in the processes they are dedicated for, they can be used to extract normalisations for the MC estimates of the corresponding processes by subtracting other processes and truth-matched events from data. The correction factors

$$k^p_{\text{pass}} = \frac{N^{\text{pass}}_{\text{FR}_p,\text{data}} - N^{\text{pass}}_{\text{FR}_p,\text{MC}_{\text{not } p}} - N^{\text{pass}}_{\text{FR}_p,\text{MC}_p,\text{real}}}{N^{\text{pass}}_{\text{FR}_p,\text{MC}_p,\text{fake}}} \quad \text{and} \tag{5.8}$$

$$k^p_{\text{fail}} = \frac{N^{\text{fail}}_{\text{FR}_p,\text{data}} - N^{\text{fail}}_{\text{FR}_p,\text{MC}_{\text{not } p}} - N^{\text{fail}}_{\text{FR}_p,\text{MC}_p,\text{real}}}{N^{\text{fail}}_{\text{FR}_p,\text{MC}_p,\text{fake}}} \tag{5.9}$$

are derived for each process p that has a corresponding FR (except $p = \text{QCD}$). The simulated fakes can then be corrected by multiplying predicted yields with the correction factors. Expressed more explicitly, with the corrections, Equations (5.5) and (5.6) would become:

$$F^p = \frac{N^{\text{pass}}_{\text{FR}_p,\text{data}} - \left(N^{\text{pass}}_{\text{FR}_p,\text{MC}_{\text{not } p,\text{real}}} + k^p_{\text{pass}} \cdot N^{\text{pass}}_{\text{FR}_p,\text{MC}_{\text{not } p,\text{fake}}}\right) - N^{\text{pass}}_{\text{FR}_p,\text{MC}_p,\text{real}}}{N^{\text{fail}}_{\text{FR}_p,\text{data}} - \left(N^{\text{fail}}_{\text{FR}_p,\text{MC}_{\text{not } p,\text{real}}} + k^p_{\text{fail}} \cdot N^{\text{fail}}_{\text{FR}_p,\text{MC}_{\text{not } p,\text{fake}}}\right) - N^{\text{fail}}_{\text{FR}_p,\text{MC}_p,\text{real}}}, \tag{5.10}$$

$$R_r^p = \frac{k^p_{\text{fail}} \cdot N^{\text{fail}}_{r,\text{MC}_p,\text{fake}}}{N^{\text{fail}}_{r,\text{data}} - N^{\text{fail}}_{r,\text{MC,real}}}. \tag{5.11}$$

The described method relies on the assumption that for the same background process, the quark/gluon compositions are the same or similar in the FRs and in the region where the fakes need to be estimated. Such an assumption can be justified: The FRs, with the exception of FRQ, have the same charge correlation requirement as the SR and CRZ$\tau\tau$, i.e. they all require the ℓ–$\tau_{\text{had-vis}}$ pair to carry OS charges. This ensures that the quark/gluon composition is similar in the FRs and the SR for processes like W+jets, which is indeed

expected to show charge correlations between the fake $\tau_{\text{had-vis}}$ candidate and the light lepton. And while FRQ is defined with a SS charged ℓ–$\tau_{\text{had-vis}}$ pair requirement, the charge correlation is not a concern there since multijet events are expected to be largely gg events with no charge correlations. To add more confidence to the validity of the assumption, it has been estimated using MC samples that the quark/gluon compositions of jet $\rightarrow \tau_{\text{had-vis}}$ fakes in W+jets events (which is the dominant contribution to the total fakes) are indeed similar in the SR and FRW. Table 5.1 shows the estimated quark/gluon compositions in the SR, FRW, VRSS (SR but with SS ℓ–$\tau_{\text{had-vis}}$ pairs) and SS FRW (FRW but with SS ℓ–$\tau_{\text{had-vis}}$ pairs).

Table 5.1.: The quark/gluon compositions of jet $\rightarrow \tau_{\text{had-vis}}$ fakes in W+jets events in the SR, FRW, VRSS and SS FRW estimated using MC samples. A light quark is an up, down or strange quark.

Region	**Composition [%]**			
	Light quark	**Charm quark**	**Bottom quark**	**Gluon**
SR ($e\tau$)	71.5	17.4	1.2	9.9
FRW ($e\tau$)	75.3	14.4	0.8	9.5
SR ($\mu\tau$)	71.4	18.4	1.2	8.9
FRW ($\mu\tau$)	76.2	13.7	1.1	9.0
VRSS ($e\tau$)	54.3	15.7	3.6	26.3
SS FRW ($e\tau$)	59.3	11.5	2.6	26.7
VRSS ($\mu\tau$)	53.9	15.6	3.9	26.7
SS FRW ($\mu\tau$)	60.0	11.0	3.2	25.9

5.5.2. Measurement and sources of uncertainties

The fake factors are measured separately in the $e\tau$ and $\mu\tau$ channels. They are measured in bins of $p_{\text{T}}(\tau_{\text{had-vis}})$ and $p_{\text{T}}(\tau \text{ track})$ in 1P regions and in bins of $p_{\text{T}}(\tau_{\text{had-vis}})$ in 3P regions. Binning in $p_{\text{T}}(\tau \text{ track})$ in 1P regions is necessary for the modelling of the variable $m(\ell, \tau \text{ track})$, which is used as an input to the neural network classifiers. The measured process-specific fake factors, F^p, and the estimated relative contribution of each process to the total fakes in the SR, R^p_{SR}, are summarised in Appendix D.

There are three major sources of uncertainties in the estimation of fakes via the FF method. They are:

1. The statistical errors in $\left(N^{\text{fail}}_{r,\text{data}} - N^{\text{fail}}_{r,\text{MC,real}}\right)$ (the observed yields that are multiplied by the fake factors). These errors are combined bin-by-bin with the statistical errors in the MC samples for the other backgrounds in the binned maximum-likelihood fits.

2. The statistical errors in the measured F^p. The error in each $p_{\mathrm{T}}(\tau_{\text{had-vis}})$ and $p_{\mathrm{T}}(\tau \text{ track})$ bin is propagated to the final fake factor, F_r, in the same bin and combined with the error in the estimation of k^p and R_r^p. The combined errors are introduced as constrained nuisance parameters in the binned maximum-likelihood fits (one independent parameter per bin).

3. The statistical errors in the k^p and R_r^p estimations. As described above, these errors are propagated to F_r and combined with the errors in F^p.

In theory, there are also uncertainties associated to the extrapolations from the FRs to the SR. However, given the excellent data-model agreement observed in the same-sign region test (described in the next section), which validated the extrapolation and showed that there are enough degrees of freedom for the maximum-likelihood fits to describe the fakes accurately, no extra uncertainties for the extrapolations are needed to be considered.

5.5.3. The FR closure test and same-sign region test

To validate the modelling of fakes, a closure test is performed, where the FF method is used to predict kinematic distributions in the four FRs. By definition, the predicted $p_{\mathrm{T}}(\tau_{\text{had-vis}})$ and $|\eta(\tau_{\text{had-vis}})|$ distributions should always agree with what is observed from data, barring the differences due to binning. Furthermore, assuming that the jet $\rightarrow \tau_{\text{had-vis}}$ misidentification rate depends only on $p_{\mathrm{T}}(\tau_{\text{had-vis}})$ and $|\eta(\tau_{\text{had-vis}})|$, which is to a large extent a good approximation, the predicted distributions in other variables can also be expected to agree with the observed distributions within statistical and systematic uncertainties. Closure is observed for all the input and output variables of the NN classifiers. As examples, the predicted and observed $m_{\text{vis}}(\ell, \tau)$ distributions are shown in Figures 5.6, 5.7, 5.8 and 5.9.

Another validation for the FF method is the same-sign region test. The entire procedure of deriving and applying the fake factors are repeated in the SS FRs and the VRSS, instead of the FRs and the SR. The VRSS is dominated by fakes but is kinematically similar to the SR, which allows us to independently validate the fakes modelling in the SR. The predicted and observed distributions of the combined NN output are shown in Figure 5.10. The excellent data-model agreement observed in the VRSS shows that the difference in the compositions of fakes in the FRs and the SR is indeed correctly accounted for by the FF method. It also validated the extrapolation of the process-specific fake factors from the FRs to the SR.

5.6. Summary

To summarise this chapter, the methods used to model each of the signal and background processes are listed in Table 5.2.

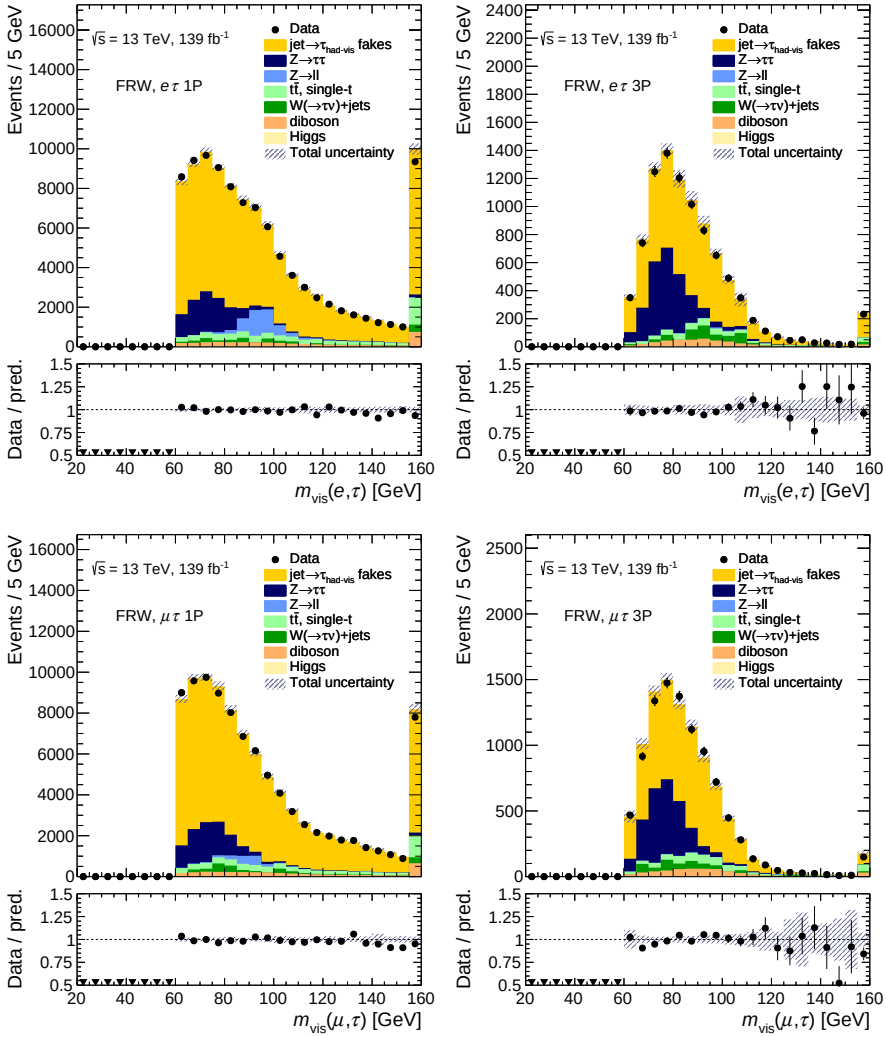


Figure 5.6.: Expected and observed distributions of $m_{\text{vis}}(\ell, \tau)$ in the FRW. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

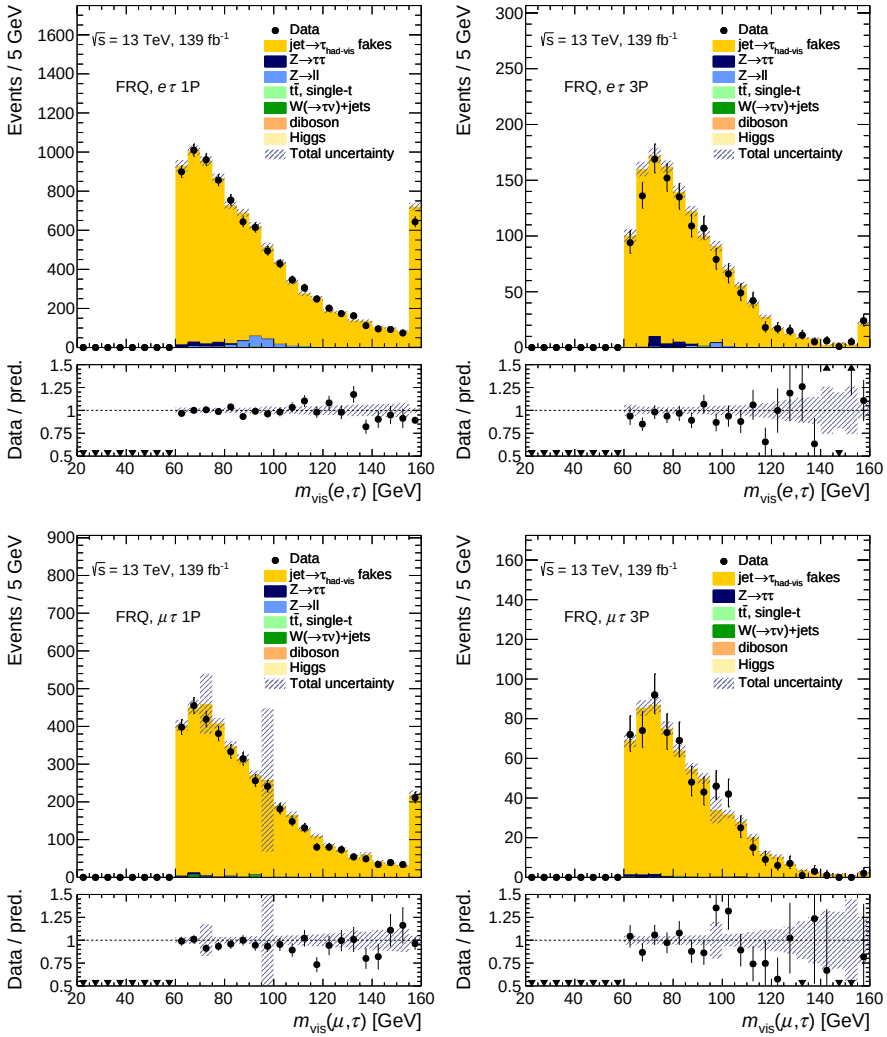


Figure 5.7.: Expected and observed distributions of $m_{\text{vis}}(\ell,\tau)$ in FRQ. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The unusually large uncertainties in certain bins are due to MC samples with occasionally large event weights being used in the calculation of R^p_{FRQ} or k^p. The last bin in each plot includes overflow events.

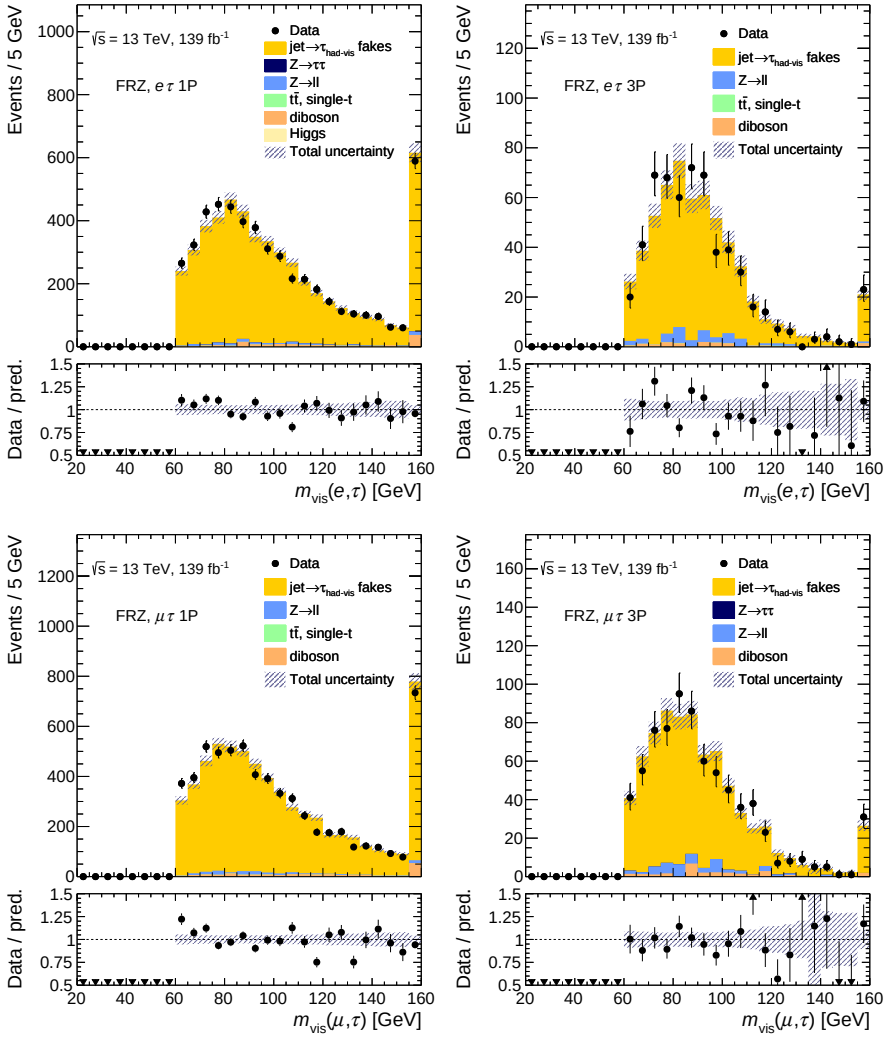


Figure 5.8.: Expected and observed distributions of $m_{\text{vis}}(\ell, \tau)$ in FRZ. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

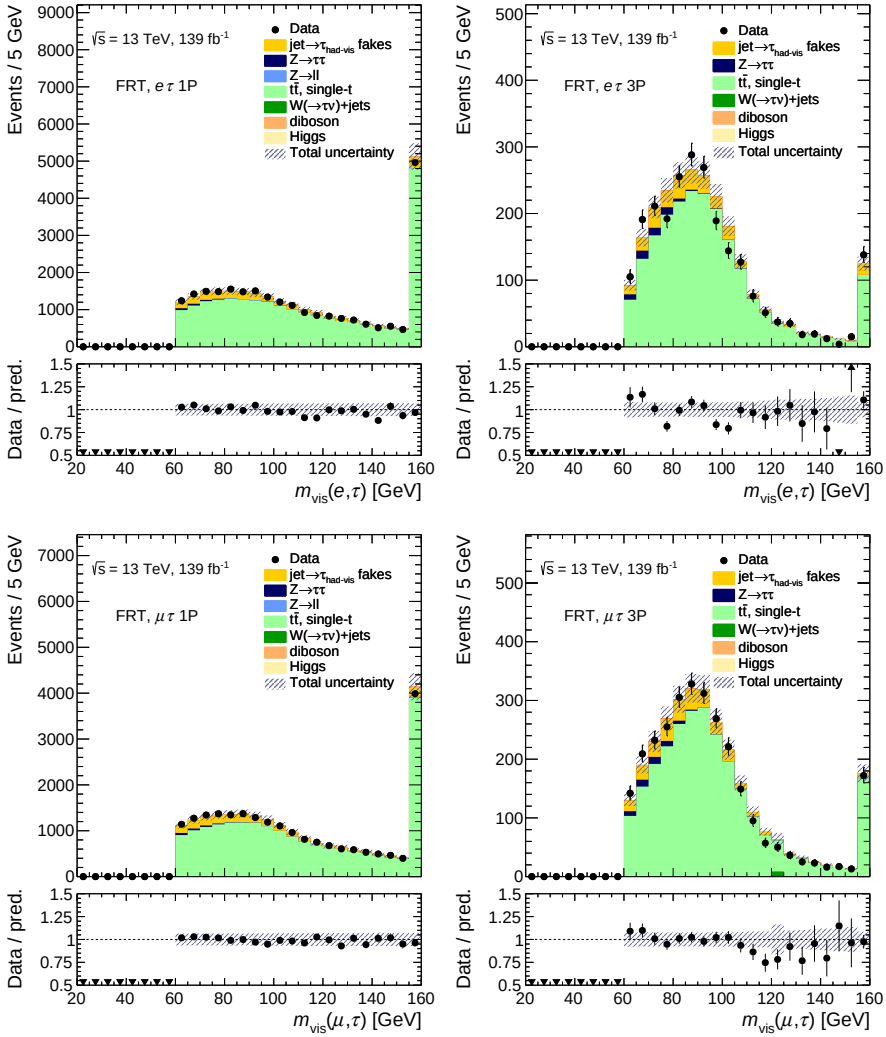


Figure 5.9.: Expected and observed distributions of $m_{\text{vis}}(\ell, \tau)$ in FRT. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

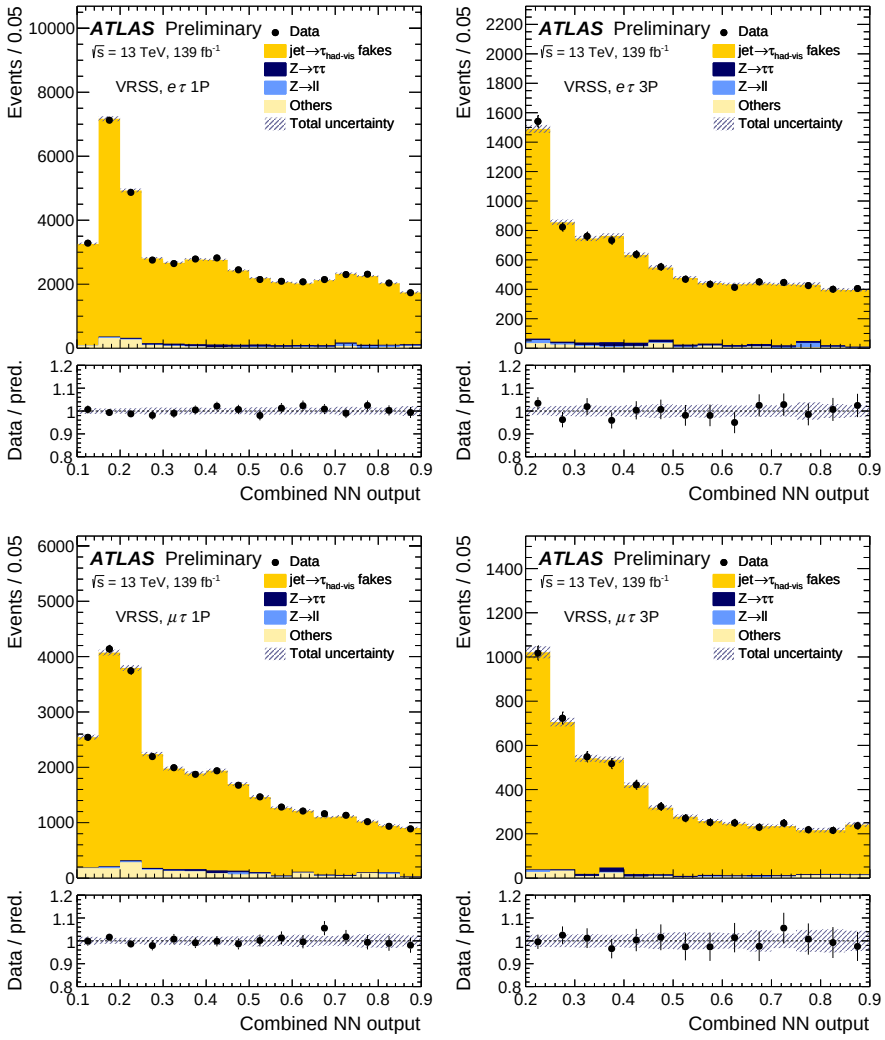


Figure 5.10.: Expected and observed distributions of the combined NN output in the VRSS of the $e\tau$ (top row) and $\mu\tau$ (bottom row) channels. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

Table 5.2.: Summary of the modelling methods used to model each of the signal and background processes.

Process	**Method**	**Normalisation**	**Corrections**	**Enriched region**
$Z \to \ell\tau$	MC simulations	Fit to data	$p_{\mathrm{T}}(Z)$ spectrum	SR (high NN output)
$Z \to \tau\tau$	MC simulations	Fit to data	$p_{\mathrm{T}}(Z)$ spectrum	CRZ$\tau\tau$ and SR (low NN output)
$Z \to \ell\ell$	MC simulations	Fixed to measurement	$p_{\mathrm{T}}(Z)$ spectrum, $\ell \to \tau_{\text{had-vis}}$ misidentification rate	CRZ$\ell\ell$
Fakes	FF method	Fit to data	N/A	FRs, VRSS and SR (low NN output)
Others	MC simulations	Fixed to theory predictions	N/A	N/A

Chapter 6.

Statistical interpretation and results

Nothing in the world causes so much misery as uncertainty.

— Martin Luther

To interpret the observed data, binned maximum-likelihood fits are used to compare observations with predictions and to extract evidence of possible signal events. The significance of any possible excess in data is assessed, and the background-only (null) and background-plus-signal (alternative) hypotheses are tested. In case no significant excess in data is observed, exclusion upper limits on the LFV branching fractions are set by inverting the hypothesis tests. This chapter details the methods and results of the statistical interpretation for the analysis.

6.1. Maximum-likelihood fit

For each of the $e\tau$ and $\mu\tau$ channels, a binned maximum-likelihood fit to data is used to compare the expected distributions of the combined NN output in the SR and the collinear mass in CRZ$\tau\tau$ with the observed distributions, and to extract evidence of possible signal events. The high-NN-output region of the SR is sensitive to the signal, which allows the fit to determine the potential signal yield precisely. Meanwhile, the low-NN-output region of the SR is enriched in background events and can be used by the fit to constrain the background predictions. Furthermore, the $m_{\text{coll}}(\ell, \tau)$ distribution in CRZ$\tau\tau$ is utilised by the fit to improve the precision of the determined $Z \to \tau\tau$ yields, as well as the constraints on the TES related uncertainties, which are both factors that significantly impact the signal sensitivity. Due to the differences in acceptances, efficiencies and background compositions, 1P and 3P regions are considered separately in the fit.

The probability of compatibility between the data and the background-only or background-plus-signal hypothesis is assessed using the modified frequentist CL_s method [123], and exclusion upper limits on $\mathcal{B}(Z \to \ell\tau)$ are set by the inversion of these hypothesis tests in case of no significant discovery.

The fits are implemented through the use of the statistical analysis packages RooFit [124], RooStats [125] and HistFitter [126].

6.1.1. Likelihood function and fit parameters

The parameter of interest (POI) in the analysis is the signal strength modifier μ_s, which quantifies the size of the LFV Z-decay branching fraction $\mathcal{B}(Z \to \ell\tau) = \mu_s \times \mathcal{B}_{\text{prefit}}(Z \to \ell\tau)$ ($\mathcal{B}_{\text{prefit}}$ can be set to any arbitrary value without affecting the physical interpretation of the results).

The likelihood function L that is being maximised in the fits can be expressed as:

$$L(\mu_s, \mathbf{b}, \boldsymbol{\theta}|\mathbf{n}) = \prod_{i \in \text{bin}} \text{Pois}(n_i|\lambda_i(\mu_s, \mathbf{b}, \boldsymbol{\theta})) \times \prod_{j \in \text{syst}} C_j(\theta_j^0, \theta_j). \tag{6.1}$$

The first product in the equation corresponds to the Poisson measurements of $\mathbf{n} = \{n_i\}$, the observed number of events in each bin in the SR and CRZ$\tau\tau$. The Poisson expectation values λ_i are functions of the predicted yields $\mathbf{b}$ for the different background contributions, the nuisance parameters (NP) $\boldsymbol{\theta}$ that parameterise the effects of systematic uncertainties in the modelling, and the POI μ_s. The second product in the equation is the constraints on the NPs $\boldsymbol{\theta}$.

The background predictions $\mathbf{b}$ depend on three unconstrained NPs, also known as normalisation factors (NF): μ_Z, $\mu_{\text{1P fakes}}$ and $\mu_{\text{3P fakes}}$. They determine the overall yields of $Z \to \tau\tau$, fakes in 1P regions and fakes in 3P regions respectively. The values of the NFs are mainly constrained by the data in CRZ$\tau\tau$ and the low-NN-output region of the SR, which are dominantly $Z \to \tau\tau$ events and fakes. In the background-plus-signal model, the signal yield is determined jointly by μ_Z and μ_s. By fitting the predicted $Z \to \tau\tau$ yield to data, the Z-boson production cross section $\sigma(Z)$, the acceptance $A(Z \to \tau\tau \to \ell\tau_{\text{had-vis}})$, and the combined trigger, reconstruction, identification and isolation efficiency $\varepsilon(Z \to \tau\tau \to \ell\tau_{\text{had-vis}})$ for the $Z \to \tau\tau$ events are determined. The determined value of $\sigma(Z) \times A(Z \to \tau\tau \to \ell\tau_{\text{had-vis}}) \times \varepsilon(Z \to \tau\tau \to \ell\tau_{\text{had-vis}})$ is then reflected by the value of μ_Z, which equals to the ratio of the total postfit $Z \to \tau\tau$ yield to the total prefit $Z \to \tau\tau$ yield. Given that the signal and $Z \to \tau\tau$ samples are both normalised to the same measured $\sigma(Z)$ before the fit, and that the $A \times \varepsilon$ for $Z \to \tau\tau$ events can be expected to be similar to that for the signal events in the SR, the value of μ_Z determined by fitting $Z \to \tau\tau$ events can also be interpreted as $\sigma(Z) \times A(Z \to \ell\tau \to \ell\tau_{\text{had-vis}}) \times \varepsilon(Z \to \ell\tau \to \ell\tau_{\text{had-vis}})$ for the signal events. Therefore, by normalising the signal sample with $\mu_Z \times \mu_s$, the POI μ_s is effectively decoupled from $\sigma(Z)$, $A(Z \to \ell\tau \to \ell\tau_{\text{had-vis}})$ and $\varepsilon(Z \to \ell\tau \to \ell\tau_{\text{had-vis}})$, and represents unequivocally the ratio $\mathcal{B}(Z \to \ell\tau)/\mathcal{B}_{\text{prefit}}(Z \to \ell\tau)$.

The NPs $\boldsymbol{\theta}$ are constrained by the probability density functions $C_j(\theta_j^0, \theta_j)$ for each systematic uncertainty j. Depending on the nature of the uncertainty, the constraint C_j is either a Poisson or a unit-variance Gaussian distribution with a central value θ_j^0 around which θ_j can be varied. The total statistical uncertainties in the MC samples in each bin are modelled by NPs with Poisson constraints and are considered to be independent of

each other. Other systematic uncertainties are modelled by NPs with Gaussian constraints, where bin-to-bin correlations are preserved in the parameterisation.

6.1.2. Test statistics and hypothesis tests

The significance of an excess is determined in p-value using the modified frequentist $\mathrm{CL_s}$ method. The test statistic employed is the log-likelihood ratio q_{μ_s}, which is modified into q_0 for testing the background-only (null) hypothesis.

q_{μ_s} is constructed based on the likelihood ratio $\lambda(\mu_\mathrm{s})$:

$$\lambda(\mu_\mathrm{s}) = \begin{cases} \frac{L\left(\mu_\mathrm{s}, \hat{\hat{\mathbf{b}}}(\mu_\mathrm{s}), \hat{\hat{\boldsymbol{\theta}}}(\mu_\mathrm{s})\right)}{L\left(\hat{\mu}_\mathrm{s}, \hat{\mathbf{b}}, \hat{\boldsymbol{\theta}}\right)} & \text{if } \hat{\mu}_\mathrm{s} \geq 0, \\ \frac{L\left(\mu_\mathrm{s}, \hat{\hat{\mathbf{b}}}(\mu_\mathrm{s}), \hat{\hat{\boldsymbol{\theta}}}(\mu_\mathrm{s})\right)}{L\left(0, \hat{\hat{\mathbf{b}}}(0), \hat{\hat{\boldsymbol{\theta}}}(0)\right)} & \text{if } \hat{\mu}_\mathrm{s} < 0, \end{cases} \tag{6.2}$$

where the single-hatted quantities are the best-fit values and the doubled-hatted quantities are values that maximise the likelihood while given a fixed value of μ_s.

For quantifying the significance of a potential discovery, the test statistic used is

$$q_0 = \begin{cases} -2\ln\lambda(\mu_\mathrm{s}=1) & \text{if } \hat{\mu}_\mathrm{s} \geq 0, \\ 0 & \text{if } \hat{\mu}_\mathrm{s} < 0. \end{cases} \tag{6.3}$$

When testing a specific background-plus-signal model, the test statistic is given by

$$q_{\mu_\mathrm{s}} = \begin{cases} -2\ln\lambda(\mu_\mathrm{s}) & \text{if } \hat{\mu}_\mathrm{s} \leq \mu_\mathrm{s}, \\ 0 & \text{if } \hat{\mu}_\mathrm{s} > \mu_\mathrm{s}. \end{cases} \tag{6.4}$$

This choice of test statistics ensures that q_0 is zero when the best-fit value of μ_s is smaller than zero, i.e. the number of events is below the SM expectation. It also ensures that no exclusion limit will be set below $\hat{\mu}_\mathrm{s}$.

The level of (dis)agreement between data and the null hypothesis is quantified by the p-value

$$p_0 = \int_{q_\mathrm{obs}}^{\infty} f(q_0|0)\,\mathrm{d}q_0, \tag{6.5}$$

where q_obs is the value of the test statistic observed from data and f is the probability distribution of the test statistic. For testing the alternative hypotheses and setting exclusion limits, the $\mathrm{CL_s}$ value

$$\mathrm{CL_s} = \frac{p_{\mu_\mathrm{s}}}{1-p_0} \tag{6.6}$$

is used instead of the simple p-value

$$p_{\mu_s} = \int_{q_{\text{obs}}}^{\infty} f(q_{\mu_s}|\mu_s)\,\mathrm{d}q_{\mu_s} \tag{6.7}$$

alone. The CL_s value conservatively quantifies the strength of statistical evidence for the alternative hypothesis as against the null hypothesis. The probability distributions of the test statistics are approximated using the asymptotic formulae described in Reference [127].

6.2. Uncertainties

As for any scientific research, careful consideration of possible uncertainties is what makes a truthful and trustworthy result. In the following, the sources and handling of uncertainties considered in the maximum-likelihood fits will be discussed.

6.2.1. Prefit uncertainty estimations

In the likelihood function, systematic uncertainties are parameterised by Gaussian-constrained NPs (denoted as α). Their prefit values and uncertainties are set to be 0 ± 1.

Reconstructed objects

Systematic uncertainties associated to the reconstructed objects are estimated by the respective ATLAS working groups based on various measurements. These include the electron-related uncertainties (resolution, scale, and trigger, reconstruction, isolation and identification efficiencies) [84, 85], muon-related uncertainties (resolution, scale, and trigger, reconstruction, isolation and track-to-vertex-association efficiencies) [86, 87], $\tau_{\text{had-vis}}$-related uncertainties (energy scale, and reconstruction and identification efficiencies) [81, 82], $E_{\text{T}}^{\text{miss}}$-related uncertainties (track soft term resolution and scale) [88, 89], jet-related uncertainties (energy resolution and scale) [128] and flavour-tagging-related uncertainties (b-tagging efficiencies) [75], The effect of these uncertainties are modelled by varying the corresponding kinematics or efficiency scale factors by ± 1 standard deviations (σ) and interpolating/extrapolating their effects on event yields in each bin.

Tau energy scale

Since both the $m_{\text{coll}}(\ell, \tau)$ distribution in CRZ$\tau\tau$ and the combined NN output distribution in the SR are sensitive to the reconstructed $\tau_{\text{had-vis}}$ energy, uncertainties in the TES are one of the dominant sources of systematic uncertainties and deserve careful considerations. These uncertainties mainly affect the $Z \to \tau\tau$ background modelling.

The prefit TES and its uncertainties are estimated using measurements in so-called "tag-and-probe" analyses, where simulated $Z \to \tau\tau$ events are compared to observation from data [81]. However, these estimations are not entirely applicable to this analysis for several reasons: the measurements in the tag-and-probe analyses are performed based on $Z \to \tau\tau$ MC samples generated using Powheg+Pythia with a vastly different set-up than what is used for the $Z \to \tau\tau$ MC samples in this analysis; the measurements in the tag-and-probe analyses are performed inclusively for 1-prong and 3-prong $\tau_{\text{had-vis}}$ candidates, while distinctions are made between 1-prong and 3-prong $\tau_{\text{had-vis}}$ candidates in this analysis; the estimated TES uncertainties in Reference [81] are only based on the measured peak in the m_{vis} (ℓ,τ) distribution, but in this analysis, the entire m_{coll} (ℓ,τ) spectrum in CRZ$\tau\tau$ is used to constrain the TES. Furthermore, by considering all the events in the SR and in CRZ$\tau\tau$, this analysis makes use of more data than what is used in the tag-and-probe analyses. In other words, the maximum-likelihood fit in this analysis is basically remeasuring the TES and its uncertainties to a higher precision. Because of this, the postfit values of the TES NPs can be expected to be different from the prefit values and more constrained.

Simulated Z-boson production

As mentioned in Section 5.3 in the previous chapter, the dominant source of theory uncertainties in the simulated signal, $Z \to \tau\tau$ and $Z \to \ell\ell$ events is the uncertainties in the predicted $\sigma(Z)$ and $p_{\text{T}}(Z)$. However, by applying corrections derived from measurements, the impact of these uncertainties are reduced.

The statistical uncertainties in the derived $p_{\text{T}}(Z)$ correction scale factors are assigned to the $Z \to \ell\ell$, $Z \to \tau\tau$ and signal samples. The systematic uncertainty in the measured Z production cross section (2.0%) [62] is assigned to the overall normalisation of the $Z \to \ell\ell$ sample. Since the normalisation of the signal and $Z \to \tau\tau$ samples are determined by a maximum-likelihood fit to data, no associated uncertainties are assigned.

$\ell \to \tau_{\text{had-vis}}$ fakes

For the $Z \to \ell\ell$ events with $\ell \to$ 1-prong $\tau_{\text{had-vis}}$ misidentification, the uncertainties in the correction scale factors derived from CRZ$\ell\ell$ (see Section 5.4) are assigned to the $Z \to \ell\ell$ samples in the maximum-likelihood fit.

For the events with $\ell \to$ 3-prong $\tau_{\text{had-vis}}$ misidentification, there is no corrections derived or applied. Instead, extra uncertainties are assigned to the corresponding samples. The uncertainty in the efficiency of correctly vetoing $e \to$ 3-prong $\tau_{\text{had-vis}}$ events using the e-veto BDT algorithm is considered by assigning a conservative variation of $\pm 50\%$ at $\pm 1\sigma$ on the expected yield of $Z \to ee$ events. On the other hand, The uncertainty in the (in)efficiency of incorrectly vetoing $e \to$ 3-prong $\tau_{\text{had-vis}}$ events is considered by varying the expected yields of $Z \to \tau\tau$ and signal events by $\pm 50\%$ of the inefficiency of applying the

e-veto BDT algorithm on true $\tau_{\text{had-vis}}$ candidates at $\pm 1\sigma$, which is equivalent to 1.8% of the overall normalisations of the samples in the 3P SR.

Jet $\rightarrow \tau_{\text{had-vis}}$ fakes

As described in Section 5.5.2, the statistical uncertainties in the measured fake factors, as well as the statistical uncertainties in $\left(N_{r,\text{data}}^{\text{fail}} - N_{r,\text{MC,real}}^{\text{fail}}\right)$ are considered.

Luminosity

The relative uncertainty in the measured integrated luminosity (1.7%) [129] is assigned to the overall normalisations of all the background MC samples except for the $Z \rightarrow \tau\tau$ sample, which is normalised by fitting a free parameter.

Minor backgrounds

For the minor backgrounds, theory uncertainties in the overall yields of the backgrounds are considered. These uncertainties are conservatively estimated and they include the overall uncertainties in the PDF, the strong coupling constant α_s and the scale variations in the production of Higgs bosons, top quarks, dibosons and $W(\rightarrow \tau\nu)$+jets. For the $H \rightarrow \tau\tau$ samples, a relative uncertainty of 4.1% in the total yield is assigned, which, aside from the Higgs production cross section, also accounts for the theory uncertainty in the $H \rightarrow \tau\tau$ branching fraction [118]. For the top-quark and diboson samples, relative uncertainties of 5% and 10% are assigned respectively [130, 131]. For the $W(\rightarrow \tau\nu)$+jets samples, which have extremely minor contribution, a conservative 20% relative uncertainty is assigned.

6.2.2. Pruning and symmetrisation

Some of the MC samples have a relatively small sample size in the SR or CRZ$\tau\tau$. For these samples, the estimated variations due to uncertainties might be dominated by statistical noise, which could introduce numerical instabilities in the fits. Moreover, uncertainties with little impact on the fit results can contribute to the complexity of the fits and lead to longer computation time. These uncertainties can be safely omitted without significantly affecting the final fit results. The removal of these uncertainties is also known as pruning.

The following pruning and symmetrisation procedures are employed in order to stabilise and speed up the fit. They are performed on each uncertainty for each MC sample in each region.

Samples with extremely low MC sample size Uncertainties of samples that have extremely small MC sample size are completely removed from the fit. Having extremely

small MC sample size is defined as having an overall relative statistical error larger than 10% in the concerned region. In this analysis, samples that have extremely small MC sample size include $W(\to \tau\nu)$+jets in the both the 1P and 3P SR and CRZ$\tau\tau$, and $Z \to \ell\ell$ in the 3P regions.

Removal of normalisation effects The effect of an uncertainty on the overall normalisation of a sample is kept only when its impact at $\pm 1\sigma$ is larger than 0.1% of the total nominal yield.

Removal of shape effects In order to test if the effect of an uncertainty on the shape of the expected distribution is significant compared to the MC statistical uncertainties, a χ^2 test is performed between the normalised nominal distribution and the varied distribution at $\pm 1\sigma$. Only if the corresponding p-value from the χ^2 test is smaller than 0.9[†]. The effect of an uncertainty on the distribution shapes is always kept for uncertainties in efficiency scale factors.

Symmetrisation Sometimes, due to statistical fluctuation, variations due to uncertainties could be one-sided in certain bins, i.e. the $+1\sigma$ and -1σ variations of the expected yield are both larger than or both smaller than the nominal expected yield. These variations are symmetrised by inverting the direction of the smaller variation.

6.2.3. Impact on the best-fit LFV branching fraction

A summary of the postfit uncertainties and their impact on the best-fit LFV branching fraction is given in Table 6.1. It shows that the primary source of uncertainty in the analysis is statistical. This is mainly due to the fact that the fakes modelling and the normalisation of the $Z \to \tau\tau$ background are both determined using data. This heavily data-driven modelling approach is what allows the analysis to make full use of the data, without being heavily limited by uncertainties in theoretical calculations and simulations. The remaining dominant uncertainties are those in the TES and in the $e \to \tau_{\text{had-vis}}$ misidentification modelling.

6.3. Results

The observed and the best-fit expected distributions of the combined NN output in the SR and the collinear mass in CRZ$\tau\tau$ are shown in Figures 6.1 and 6.2 for the $e\tau$ and $\mu\tau$ channels respectively.

The corresponding best-fit POIs and NFs are listed in Table 6.2. Almost no signal is fitted in the $e\tau$ channel, while a small positive signal is fitted in the $\mu\tau$ channel. The precision of the determined yields of $Z \to \tau\tau$ events and fakes ranges from 2% to 4%.

[†] Note that a smaller p-value in a χ^2 test represents more confidence for rejecting the hypothesis of identity.

Table 6.1.: A summary of the postfit uncertainties and their impacts on the measured signal branching fraction. The statistical uncertainties include those in the determination of the total yields from fakes and $Z \to \tau\tau$ or $Z \to \ell\tau$ decays. The uncertainties in the Z-boson modelling include those in the correction of the simulated transverse momentum and the measured production cross section of the Z boson.

Uncertainty	**Impact on the signal branching fraction [10^{-6}]**	
	$e\tau$	$\mu\tau$
Statistical	± 3.5	± 2.8
Systematic	± 2.3	± 1.6
Tau	± 1.9	± 1.5
Energy calibration	± 1.3	± 1.4
Jet rejection	± 0.3	± 0.3
Electron rejection	± 1.3	
Light leptons	± 0.4	± 0.1
$E_{\mathrm{T}}^{\mathrm{miss}}$, jets and flavour tagging	± 0.6	± 0.5
Z-boson modelling	± 0.7	± 0.3
Luminosity and other minor backgrounds	± 0.8	± 0.3
Total	± 4.1	± 3.2

Table 6.2.: Best-fit values and uncertainties of $\mathcal{B}(Z \to \ell\tau)$ and the background normalisation factors. The uncertainties include both the statistical and systematic contributions.

	$e\tau$	$\mu\tau$
$\mathcal{B}(Z \to \ell\tau)$ [10^{-6}]	-0.1 ± 4.1	4.3 ± 3.3
μ_Z	1.02 ± 0.02	0.97 ± 0.02
$\mu_{\text{1P fakes}}$	0.98 ± 0.02	1.00 ± 0.02
$\mu_{\text{3P fakes}}$	1.03 ± 0.04	1.14 ± 0.04

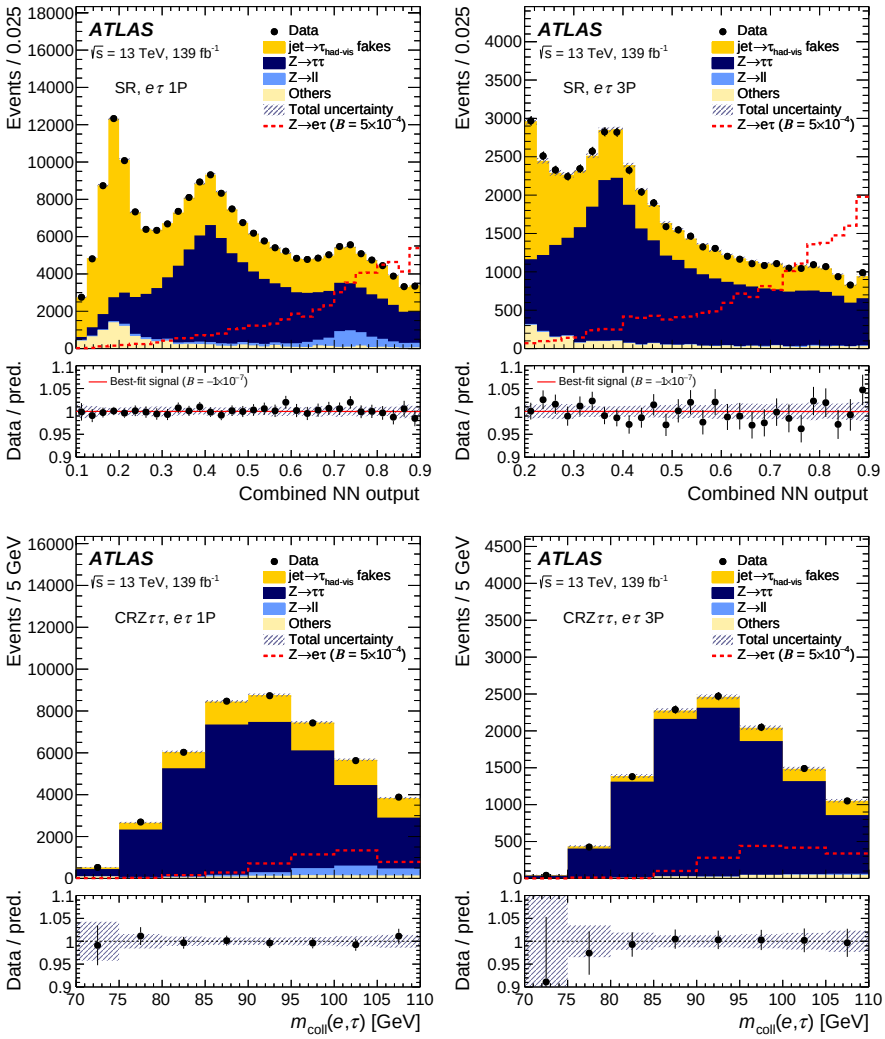


Figure 6.1.: Best-fit expected and observed distributions of the combined NN output in the SR and $m_{\text{coll}}(e,\tau)$ in CRZ$\tau\tau$ of the $e\tau$ channel. In the lower panel of each plot, the ratios of the observed yields (dots) and the best-fit background-plus-signal yields (solid red line) to the best-fit background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin of combined NN output in the SR includes overflow events.

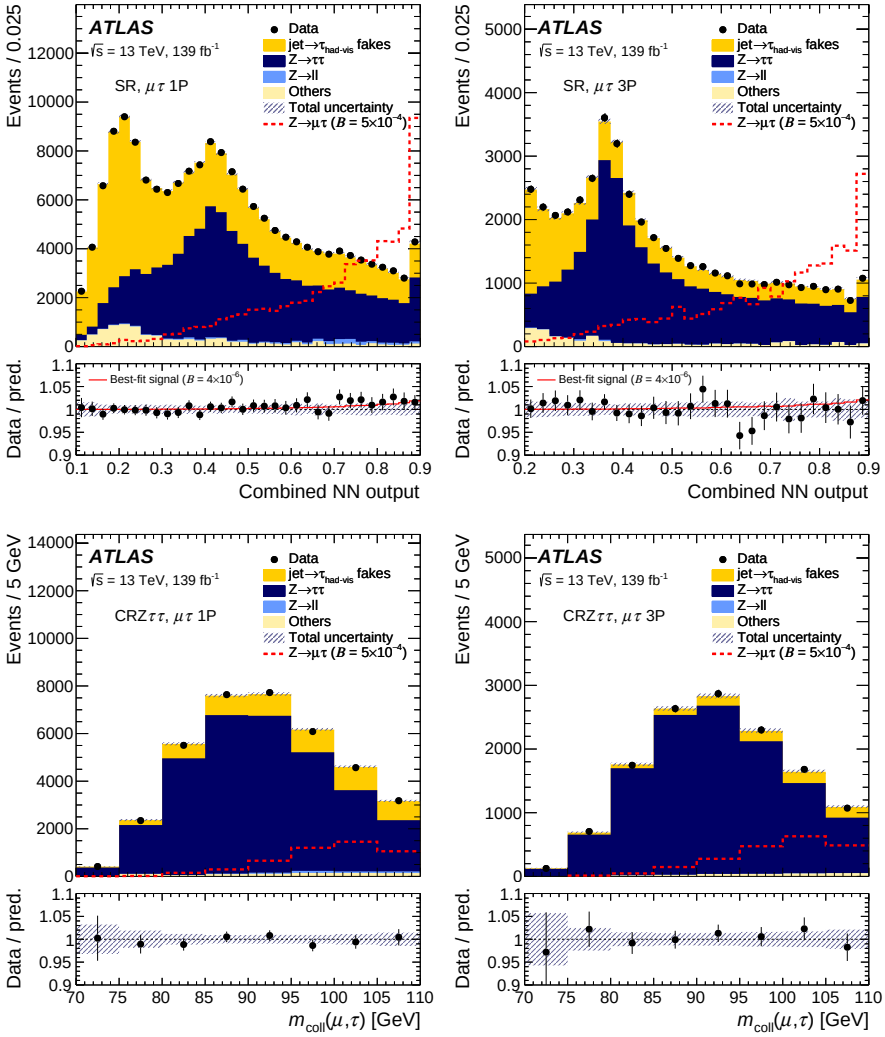


Figure 6.2.: Best-fit expected and observed distributions of the combined NN output in the SR and $m_{\text{coll}}(\mu, \tau)$ in CRZ$\tau\tau$ of the $\mu\tau$ channel. In the lower panel of each plot, the ratios of the observed yields (dots) and the best-fit background-plus-signal yields (solid red line) to the best-fit background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin of combined NN output in the SR includes overflow events.

Table 6.3.: The expected (median) and observed upper limits on the signal branching fraction at 95% CL, under different assumptions for the τ-polarisation state. The difference between the observed and expected limits are due to the non-zero best-fit signal branching fractions.

Experiment, τ-polarisation state assumption	**Observed (expected) upper limit on $\mathcal{B}(Z \to \ell\tau)$ [10^{-6}]**	
	$e\tau$	$\mu\tau$
ATLAS Run 2, unpolarised	8.1 (8.1)	9.9 (6.3)
ATLAS Run 2, left-handed	8.2 (8.6)	9.5 (6.7)
ATLAS Run 2, right-handed	7.8 (7.6)	10 (5.8)
ATLAS Run 1, unpolarised [132]		17 (26)
ATLAS Run 1 and Run 2, unpolarised (see Section 6.4)		9.5 (6.1)
OPAL, unpolarised [58]	9.8	17
DELPHI, unpolarised [60]	22	12

The best-fit NFs and NPs for systematic uncertainties, ranked in their estimated impact on the best-fit POI, are summarised in Figures 6.3 and 6.4 for the $e\tau$ and $\mu\tau$ channels respectively. Only the highest ranked 40 parameters are shown. As expected, uncertainties related to the TES and fakes modelling have the highest impact on the POI. Additionally, in the $e\tau$ channel, the $e \to \tau_{\text{had-vis}}$ fakes correction scale factors are also a dominant source of uncertainty.

Some of the TES NPs appear to be pulled by the fit away from their prefit values, especially in the $\mu\tau$ channel. This is reasonable as the prefit estimated values and uncertainties for these NPs are not entirely applicable to this analysis, as already discussed in Section 6.2.

The p-values (significance) for the null hypotheses in the $e\tau$ and $\mu\tau$ channels are 0.99 (-2.17σ) and 0.18 (0.90σ) respectively. Since no significant excess in data is observed, exclusion upper limits on the LFV signal branching fractions are set. For all the three considered scenarios with different assumed τ-polarisation states and $Z\ell\tau$ interaction vertices, the corresponding upper limits at 95% confidence level (CL) are shown in Table 6.3. For comparison, existing upper limits from the LEP experiments and from previous measurement by ATLAS are also shown in the same table. The observed upper limits on $\mathcal{B}(Z \to e\tau)$ and $\mathcal{B}(Z \to \mu\tau)$ at 95% CL under the assumption of parity-conserving $Z\ell\tau$ vertices are 8.1×10^{-6} and 9.9×10^{-6} respectively.

As can be seen from the table, the observed upper limits set by this analysis superseded those set by the LEP experiments, which are otherwise the most stringent limits established, for the first time in over two decades. Since it is shown that the sensitivity of this analysis is

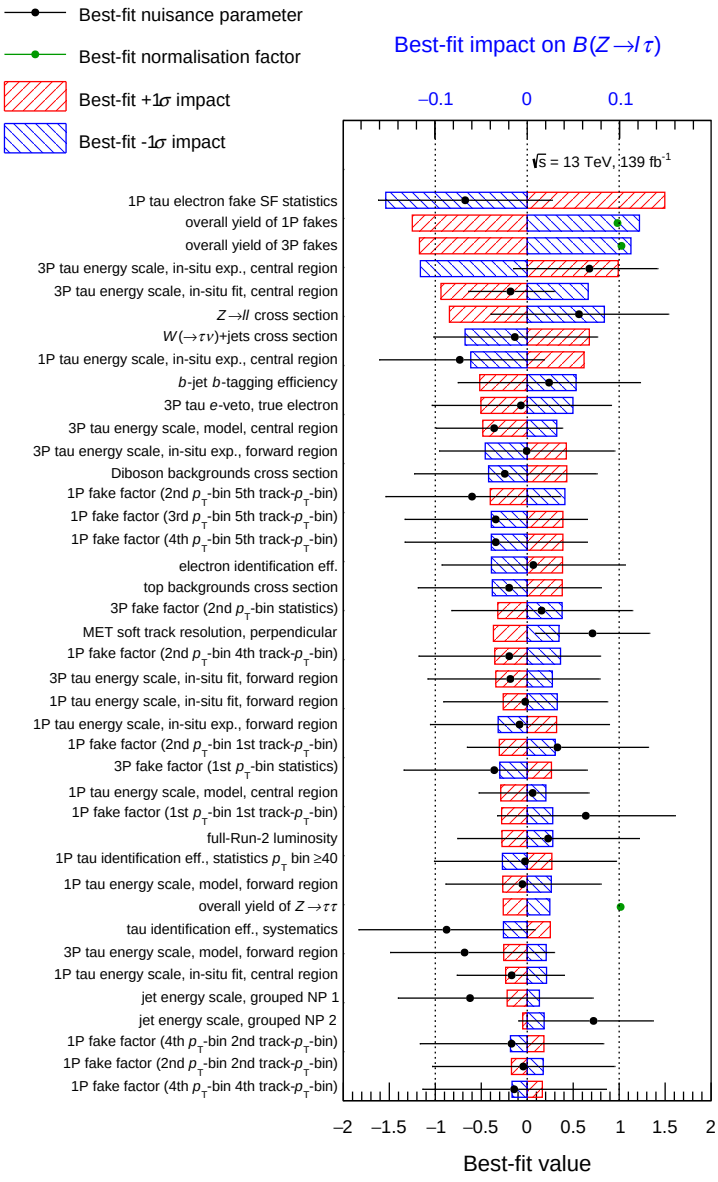


Figure 6.3.: The best-fit values of the nuisance parameter ranked in their estimated impact on μ_s for the $e\tau$ channel.

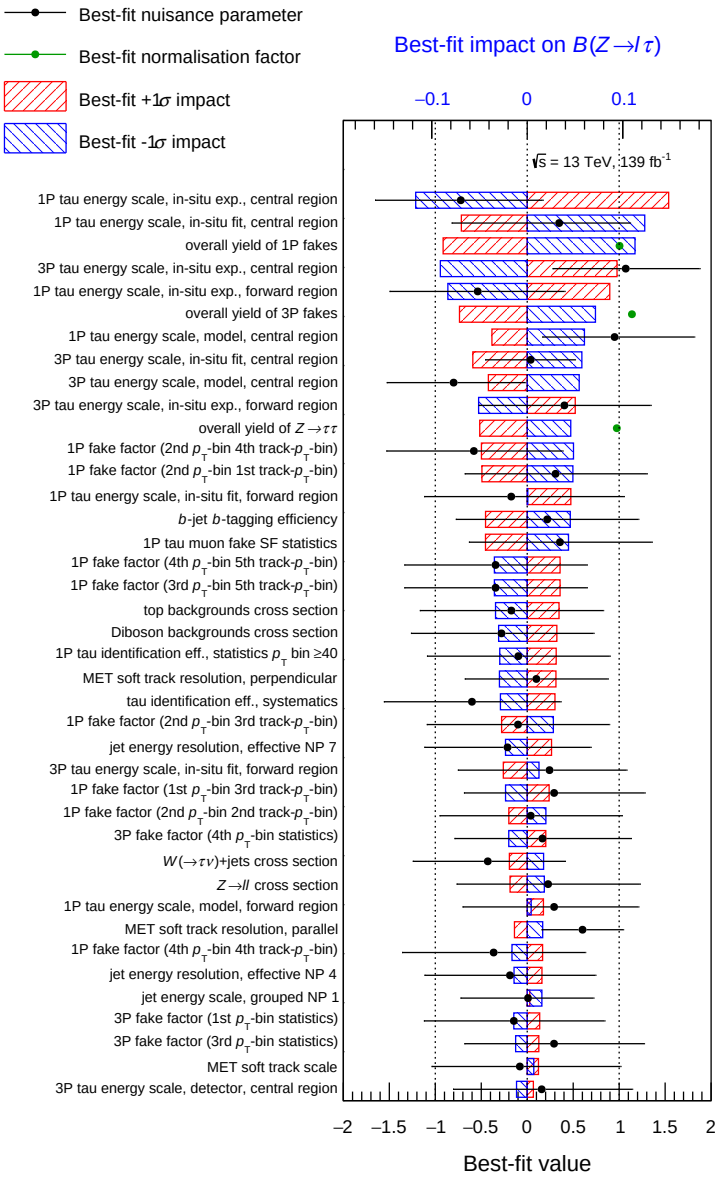


Figure 6.4.: The best-fit values of the nuisance parameter ranked in their estimated impact on μ_s for the $\mu\tau$ channel.

still primarily limited by statistical uncertainties, as more and more data will be collected in future runs of the LHC and the High-Luminosity LHC, more and more stringent constraints on the LFV $Z \to \ell\tau$ decays, or otherwise a discovery, can be expected.

6.4. Combination with existing measurement

An observed (expected) upper limit on the $Z \to \mu\tau$ branching fraction at 95% CL, $\mathcal{B}(Z \to \mu\tau) < 1.7(2.6) \times 10^{-5}$, has previously been set using $20.3\,\text{fb}^{-1}$ of $\sqrt{s} = 8\,\text{TeV}$ TeV pp collision data collected by ATLAS during the LHC Run 1 [132]. This Run-1 measurement is combined with the measurement in the current analysis to set an even more stringent upper limit on $\mathcal{B}(Z \to \mu\tau)$.

In the Run-1 analysis, the resonant mass $m_{\mu\tau}^{\text{MMC}}$ of the μ–τ pair reconstructed using the Missing Mass Calculator [133] was used as the final discriminant for statistical interpretations. Four regions were simultaneously fitted to data: two signal regions (SR1 and SR2) binned in $m_{\mu\tau}^{\text{MMC}}$ and two one-bin control regions (TCR and WCR) enriched in top-quark and W+jets background events. Same as the current analysis, the parameter of interest was the LFV branching fraction $\mathcal{B}(Z \to \mu\tau)$. The normalisations of the major backgrounds, $Z \to \tau\tau$ and W+jets, were free floating parameters in the fit. Other minor backgrounds were normalised to their theoretically predicted cross sections. A different data-driven method (the "same-sign method") was used to estimate fakes instead of the FF method used in the current analysis. The $Z \to \tau\tau$ background was estimated using so-called embedded events, which are observed $Z \to \mu\mu$ events where muons are replaced by simulated τ leptons [134]. τ leptons in the signal events are assumed to be unpolarised. The likelihood function for the measurement and the signal and background models were preserved in a RooFit workspace, which makes combination with the current analysis technically possible.

When combining the likelihood functions from the two measurements, a correlation scheme is employed where no parameters except the parameters of interest are considered correlated between the two fit models. This decision can be justified by the following reasoning.

In both fit models, the parameters of interest carry the same physical meaning as a modifier to $\mathcal{B}(Z \to \ell\tau)$. They are also both decoupled from $\sigma(Z)$, $A(Z \to \ell\tau \to \ell\tau_{\text{had-vis}})$ and $\varepsilon(Z \to \ell\tau \to \ell\tau_{\text{had-vis}})$. Moreover, the signal MC samples in both analyses are generated using the same MC generator. Therefore, in the combined fit, the parameters of interest from the two models are set to be 100% correlated.

NPs for statistical uncertainties in the two analyses should naturally be uncorrelated, given that the two measurements are based on completely different sets of data. Systematic uncertainties related to the $Z \to \tau\tau$ background estimations are also uncorrelated in the two analyses, given that embedded events were used in the Run-1 analysis while MC generated events are used in the current analysis. Furthermore, both analyses use data-driven methods to estimate fakes and the methods used are very different, implying that

Table 6.4.: Best-fit values and uncertainties of $\mathcal{B}(Z \to \mu\tau)$ and the normalisation factors, and the expected (median) and observed upper limits on $\mathcal{B}(Z \to \mu\tau)$ at 95% CL. Results from standalone fits to the Run 1 or Run 2 data set are shown alongside the results from the combined fit for comparison. The uncertainties include both the statistical and systematic contributions.

	Run 1	**Run 2**	**Combined**
$\mathcal{B}(Z \to \mu\tau)$ $[10^{-6}]$	$-15.8\ ^{+12.1}_{-11.5}$	4.3 ± 3.2	3.1 ± 3.1
μ_Z (Run 2)	–	0.97 ± 0.02	0.97 ± 0.02
$\mu_{\text{1P fakes}}$ (Run 2)	–	1.00 ± 0.02	1.01 ± 0.02
$\mu_{\text{3P fakes}}$ (Run 2)	–	1.14 ± 0.04	1.14 ± 0.04
μ_Z (Run 1)	$1.02\ ^{+0.07}_{-0.06}$	–	$1.02\ ^{+0.07}_{-0.06}$
μ_W (Run 1)	$1.10\ ^{+0.09}_{-0.08}$	–	1.09 ± 0.07
Expected upper limit at 95% CL $[10^{-6}]$	25.7	6.3	6.1
Observed upper limit at 95% CL $[10^{-6}]$	16.9	9.9	9.5

statistical and systematic uncertainties related to fakes estimation are also uncorrelated in the two analyses. Most of the object reconstruction and identification algorithms have been significantly changed in Run 2 compared to Run 1. This is especially true for $\tau_{\text{had-vis}}$ algorithms – the MVA TES, RNN ID and e-veto BDT are all newly developed algorithms that differ from the algorithms used in Run 1. Performance measurements, which are based on data, are also statistically uncorrelated in the two runs. In view of all the above, it appears reasonable to consider all the NPs in the two fit models to be uncorrelated.

The best-fit values and uncertainties of the POI and NFs after a fit to the Run 1 and Run 2 combined data set, as well as the expected and observed upper limits at 95% CL, are shown in Table 6.4, alongside the results from standalone fits to the Run 1 or Run 2 data set for comparison. The observed upper limit on $\mathcal{B}(Z \to \mu\tau)$ from the combined measurement is 9.5×10^{-6}.

Conclusion and outlook

"Would you tell me, please, which way I ought to go from here?"
"That depends a good deal on where you want to get to."
"I don't much care where —"
"Then it doesn't matter which way you go."
"— so long as I get SOMEWHERE"
"Oh, you're sure to do that, if you only walk long enough."

— Lewis Carroll, "Alice" and "the Cheshire Cat",
Alice's Adventures in Wonderland

In this thesis, a search for lepton-flavour-violating $Z \to \ell\tau$ decays is presented. The search is motivated by the fact that lepton flavour symmetry in the Standard Model is only accidental and approximate, and that any observation of $Z \to \ell\tau$ decays would be an unambiguous signal of physics beyond the Standard Model. The large amount of high-energy proton–proton collision data collected by the LHC and the ATLAS detector, with approximately eight billion Z bosons produced in the second operational run, created the opportunity to search for these rare hypothetical decays with an unprecedented sensitivity. Although no significant evidence of $Z \to \ell\tau$ decays has been found, stringent upper limits on the LFV branching fractions $\mathcal{B}(Z \to \ell\tau)$ have been set, which superseded the otherwise most stringent limits set by the LEP experiments more than two decades ago. Under the assumption that the LFV $Z\ell\tau$ interaction is parity-conserving, the following upper limits at 95% confidence level are set: $\mathcal{B}(Z \to e\tau) < 8.1 \times 10^{-6}$ and $\mathcal{B}(Z \to \mu\tau) < 9.5 \times 10^{-6}$. Similar constraints are also set under different scenarios where the $Z\ell\tau$ interaction is assumed to be parity-violating, which, to the best of the author's knowledge, has never been done before in similar experiments.

The high sensitivity of the presented search does not only rely on the availability of a large data set, but also hinges on the analysis techniques. In the heart of the search are the neural network classifiers. While using neural network for event classification is not a novel idea in and of itself, it is rather unconventional for a rare-decay search like the one presented here[†] to be heavily based on neural network classification. In the analysis, neural networks are used to exploit correlations in low-level kinematic variables, which are often neglected or not fully explored in conventional "cut-based" analyses or analyses that utilise boosted decision trees. An original approach of having multiple classifiers,

[†] which is similar to a precision measurement

each trained to discriminate against a particular background contribution, has also been employed. The outputs of these classifiers are combined into a powerful discriminant in a flexible and transparent way. This allows us to optimise the separation between the different background events, which in turn allows the maximum-likelihood fits to more precisely constrain the background models. It is estimated that the use of neural network classification has improved the sensitivity of the search by roughly 50% with respect to a conventional cut-based analysis[†].

Another important factor to the sensitivity is the precision of the signal and background predictions. For this, data-driven techniques are utilised to reduce the impact of uncertainties related to theories and simulations. For events where quark- or gluon-initiated jets are misidentified as $\tau_{\text{had-vis}}$ candidates, predictions are made using the data-driven fake-factor method. For the signal and the $Z \to \tau\tau$ and $Z \to \ell\ell$ backgrounds, corrections to the production cross section and transverse momentum of the Z bosons are derived from data and applied to the simulated events. Furthermore, the misidentification rate of light leptons as $\tau_{\text{had-vis}}$ candidates in $Z \to \ell\ell$ events is also corrected using observed data. Owing to these data-driven estimations and corrections, the search is able to make full use of the available data, without being heavily limited by systematic uncertainties.

The result of this thesis marks an important point of transition for lepton flavour violation searches. It demonstrated that searches for rare LFV phenomena at the LHC and with the ATLAS detector are not only possible, but also comparable to the best searches in the past. Better yet, it has been shown that the sensitivity of the search is still primarily limited by statistical uncertainties, thanks to the data-driven approach employed. This implies that with even more data that will be collected in the future runs of the LHC, further improvement on the sensitivity can be expected.

Besides directly benefiting from the increase of data size, there are also other improvements left to explore. An obvious one is the inclusion of leptonically decaying τ leptons (τ_{lep}) in the search. In this thesis, only $Z \to \ell\tau$ events where the τ lepton subsequently decays into hadrons (τ_{had}) are considered. However, by doing so, roughly 35% of all the possible signal events are being omitted. Although less signal events can be expected in the $\ell\tau_{\text{lep}}$ channels compared to the $\ell\tau_{\text{had}}$ channels, significantly less background events can also be expected, especially for events with misidentified jets. In fact, extensive efforts have already been made within the ATLAS collaboration to explore these potential new channels. Preliminary studies show that the expected sensitivity of the $\ell\tau_{\text{lep}}$ channels could be comparable to that of the $\ell\tau_{\text{had}}$ channels considered in this thesis. Since the measurements in the $\ell\tau_{\text{lep}}$ and $\ell\tau_{\text{had}}$ channels are complementary to each other, combining the measurements can significantly improve the sensitivity of the search.

To conclude, we are living in exciting times, where the LHC and the ATLAS experiment have opened up new opportunities for lepton flavour violation searches. The result of this thesis is just the first of many more exciting results to come.

[†] The estimation was done at an early stage of the study, where both the neural network classification and the conventional cut-based analysis in the comparison had not been explored to their full extent. However, the result still shows undeniably the superiority of the neural network classification.

Appendix A.

Introduction to neural network classification

This appendix provides a minimal introduction to neural network (NN) classification, with the aim of providing the readers with the necessary vocabulary.

Generally speaking, an artificial neural network (or simply called neural network) is any simulated collection of connected units known as (artificial) neurons, or nodes, where each neuron outputs a certain signal when given a set of input signals. Inputs to a neuron can either be an external signal, or an output signal from another neuron in the network. Similarly, the output from a neuron can either be an input to another neuron, or an external output. A neural network is therefore ultimately a mapping from a set of external input signals to a set of external output signals. Such a mapping is determined by the way each individual neuron responds to its inputs and the way that the neurons are connected to each other. The mapping can be highly non-linear and, with enough neurons in a network, can approximate any continuous functions to an arbitrary accuracy [135]. In other words, NNs are universal approximators.

The function that defines the response of a neuron when given an input or a set of inputs is called the activation of the neuron. Common activations for neurons with a single input (x) include the standard logistic sigmoid function

$$f(x) = \frac{1}{1 + e^{-x}}, \tag{A.1}$$

and the rectified linear unit (ReLU) function

$$f(x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } x \geq 0. \end{cases} \tag{A.2}$$

For neurons with multiple inputs ($\mathbf{x}$), a linear combination (L) of the inputs can be used instead as the input to the above functions:

$$x = L(\mathbf{x}) = \sum_{x_i \in \mathbf{x}} w_i x_i + b, \tag{A.3}$$

where $\{w_i\}$ and b are parameters commonly referred to as the weights and bias. When a neural network is being optimised, the weights and biases are usually the "trainable parameters" that are being varied.

The way that neurons are connected to each other is referred to as the architecture of the NN. One of the simplest and most common types of architectures is the feed-forward NN. A neural network is feed-forward when the connections between neurons do not form cycles. Neurons in a feed-forward NN can often be arranged into layers, where neurons in a certain layer are only connected to neurons in the previous or the next layer. The first layer of a feed-forward NN is the input layer in which the neurons receive external inputs (usually one neuron per input). The last layer is the output layer that provides external output(s). Any layers in between are called hidden layers. Two layers are fully connected if every neuron in one of the layers is connected to every neuron in the other layer. A feed-forward NN consisting of only fully connected layers is called a dense NN.

Neural networks can be trained for various purposes, and there are different methods to train them. Here, we will only discuss neural networks as classifiers trained via supervised learning. To train a neural network for classification, training samples are given to the NN. The true classes of the training samples are known and each class is associated to a certain ideal output of the NN. For example, for a binary classification between signal and background, the signal can be associated with an output of "1" from a single output neuron, while the background can be associated with an output of "0". The actual output of a neural network can then be compared with the expected output to calculate a value known as the loss (or cost). The loss is defined such that it decreases as the accuracy of the classifier increases. The parameters (often weights and biases) of the NN are then varied to minimise the loss. A commonly chosen definition for the loss of a binary classifier is the binary cross-entropy:

$$H(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log (1 - \hat{y}) \,, \tag{A.4}$$

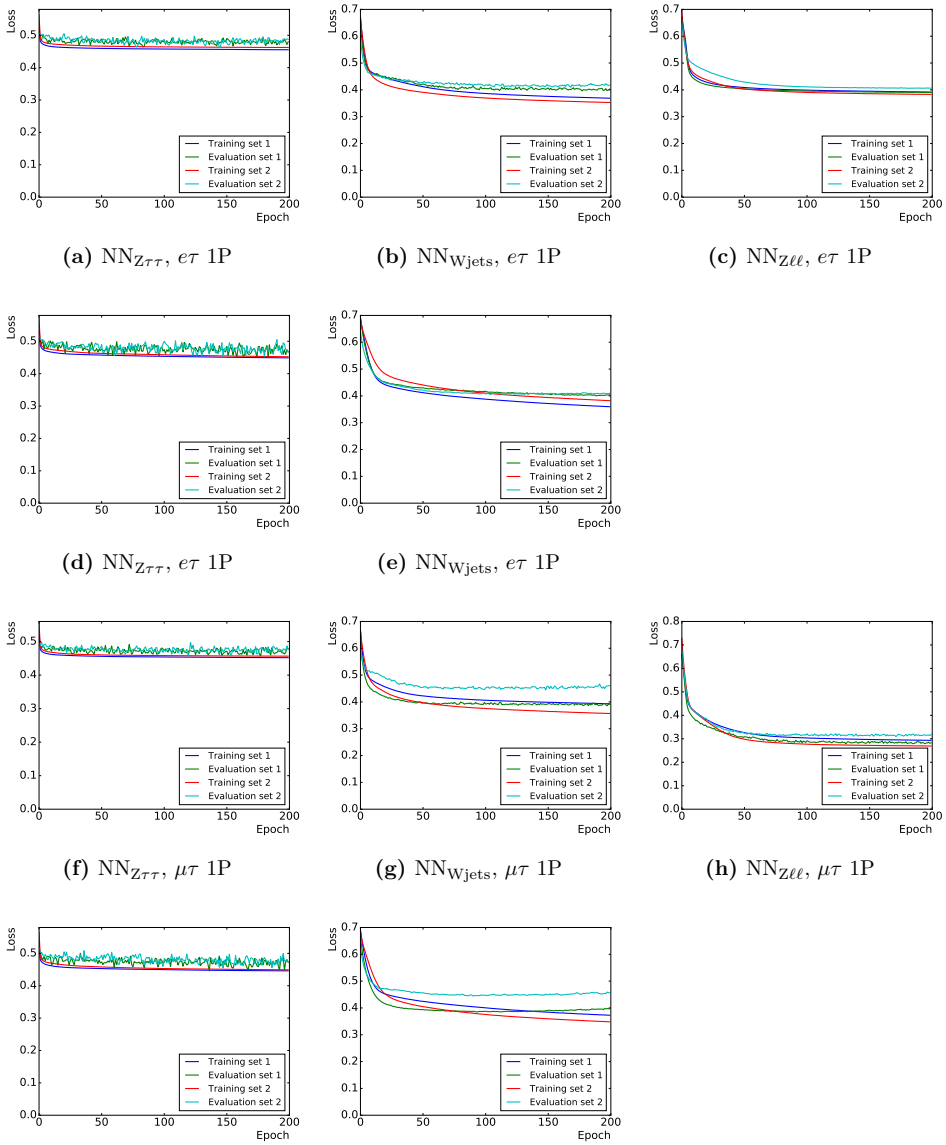
where y is the target output and $\hat{y}$ is the actual output of the classifier. Using this definition, the output of a trained classifier would estimate the Bayesian *a posteriori* probability [136].

Iterative methods are often used for the optimisation of NNs (minimisation of the loss). In these methods, training samples are usually randomly partitioned into smaller batches. For each iteration, the parameters of the NN are varied to reduce the loss of the NN evaluated on a single batch of samples. After iterating through each batch in the full training sample set, the training samples can be shuffled and randomly partitioned again into different batches for more iterations. Each cycle through the full training sample set is known as an epoch.

Appendix B.

Training history of the neural network classifiers

The following figures show the loss and accuracy of the neural network classifiers as functions of the training epoch.



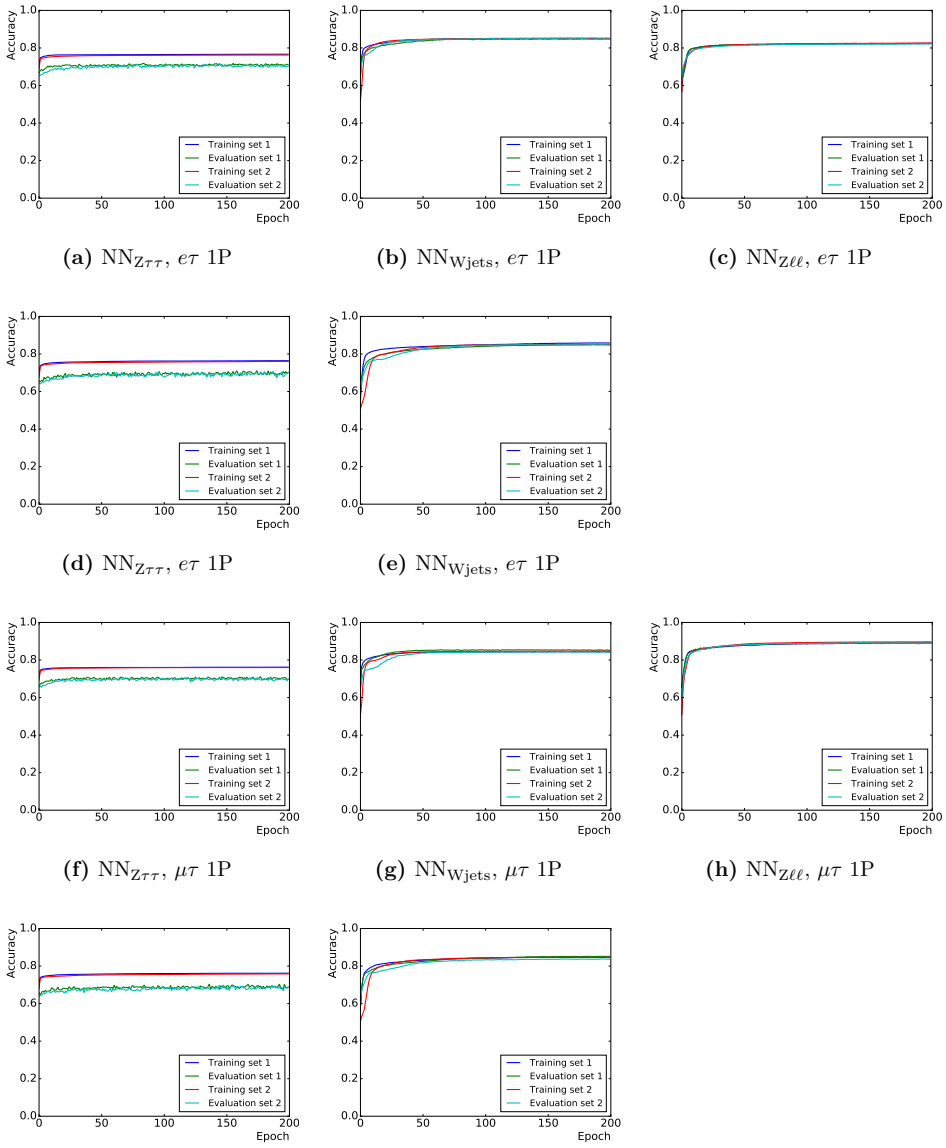
(a) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $e\tau$ 1P **(b)** $\mathrm{NN}_{\mathrm{Wjets}}$, $e\tau$ 1P **(c)** $\mathrm{NN}_{\mathrm{Z}\ell\ell}$, $e\tau$ 1P

(d) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $e\tau$ 1P **(e)** $\mathrm{NN}_{\mathrm{Wjets}}$, $e\tau$ 1P

(f) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $\mu\tau$ 1P **(g)** $\mathrm{NN}_{\mathrm{Wjets}}$, $\mu\tau$ 1P **(h)** $\mathrm{NN}_{\mathrm{Z}\ell\ell}$, $\mu\tau$ 1P

(i) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $\mu\tau$ 1P **(j)** $\mathrm{NN}_{\mathrm{Wjets}}$, $\mu\tau$ 1P

Figure B.1.: Binary cross-entropy (loss) versus training epoch for the different NN classifiers. Each classifier consists of two independent NNs, one of which is trained with samples in "Training set 1" and tested with independent samples in "Evaluation set 1", and the other with "Training set 2" and "Evaluation set 2".



(a) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $e\tau$ 1P **(b)** $\mathrm{NN}_{\mathrm{Wjets}}$, $e\tau$ 1P **(c)** $\mathrm{NN}_{\mathrm{Z}\ell\ell}$, $e\tau$ 1P

(d) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $e\tau$ 1P **(e)** $\mathrm{NN}_{\mathrm{Wjets}}$, $e\tau$ 1P

(f) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $\mu\tau$ 1P **(g)** $\mathrm{NN}_{\mathrm{Wjets}}$, $\mu\tau$ 1P **(h)** $\mathrm{NN}_{\mathrm{Z}\ell\ell}$, $\mu\tau$ 1P

(i) $\mathrm{NN}_{\mathrm{Z}\tau\tau}$, $\mu\tau$ 1P **(j)** $\mathrm{NN}_{\mathrm{Wjets}}$, $\mu\tau$ 1P

Figure B.2.: Classifying accuracy versus training epoch for the different NN classifiers. Each classifier consists of two independent NNs, one of which is trained with samples in "Training set 1" and tested with independent samples in "Evaluation set 1", and the other with "Training set 2" and "Evaluation set 2".

Appendix C.

Distributions and modelling of the neural network input variables

The following figures show the expected and observed distributions of all the neural network input variables in the SR.

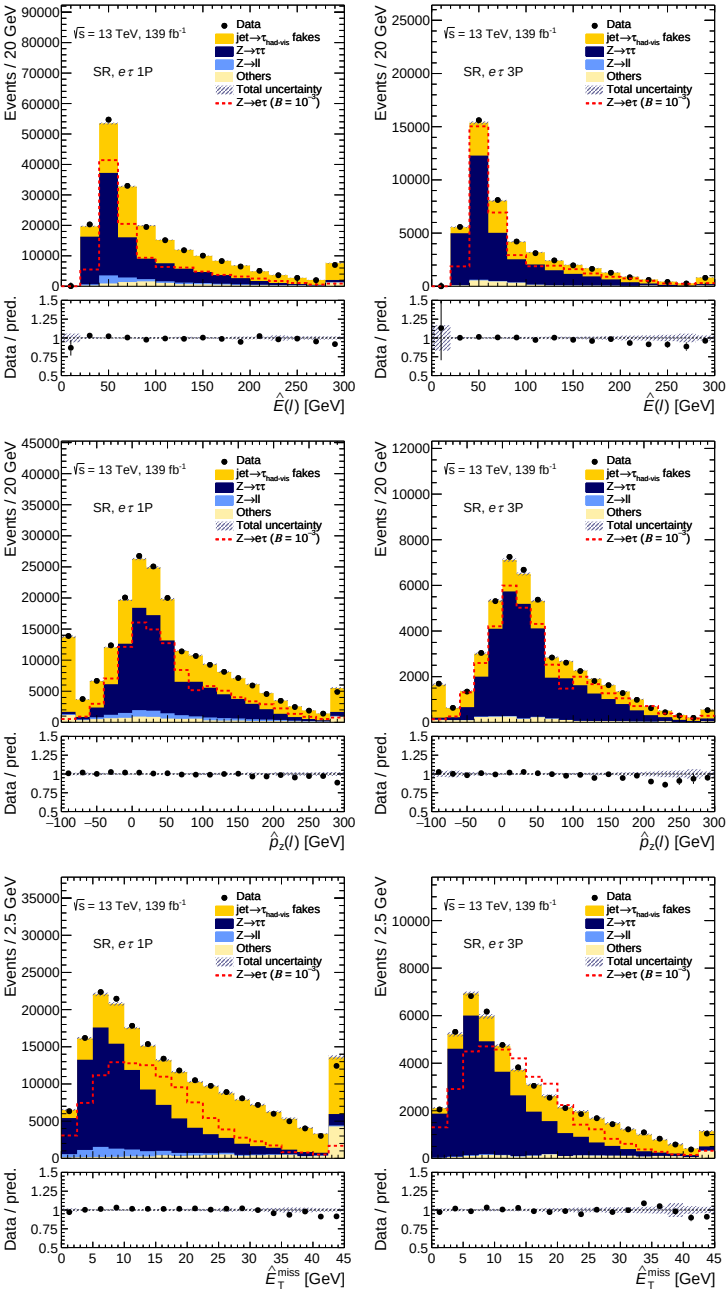


Figure C.1.: Expected and observed distributions of $\hat{E}(\ell)$, $\hat{p}_z(\ell)$ and $\hat{E}_{\mathrm{T}}^{\mathrm{miss}}$ in the SR of the $e\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

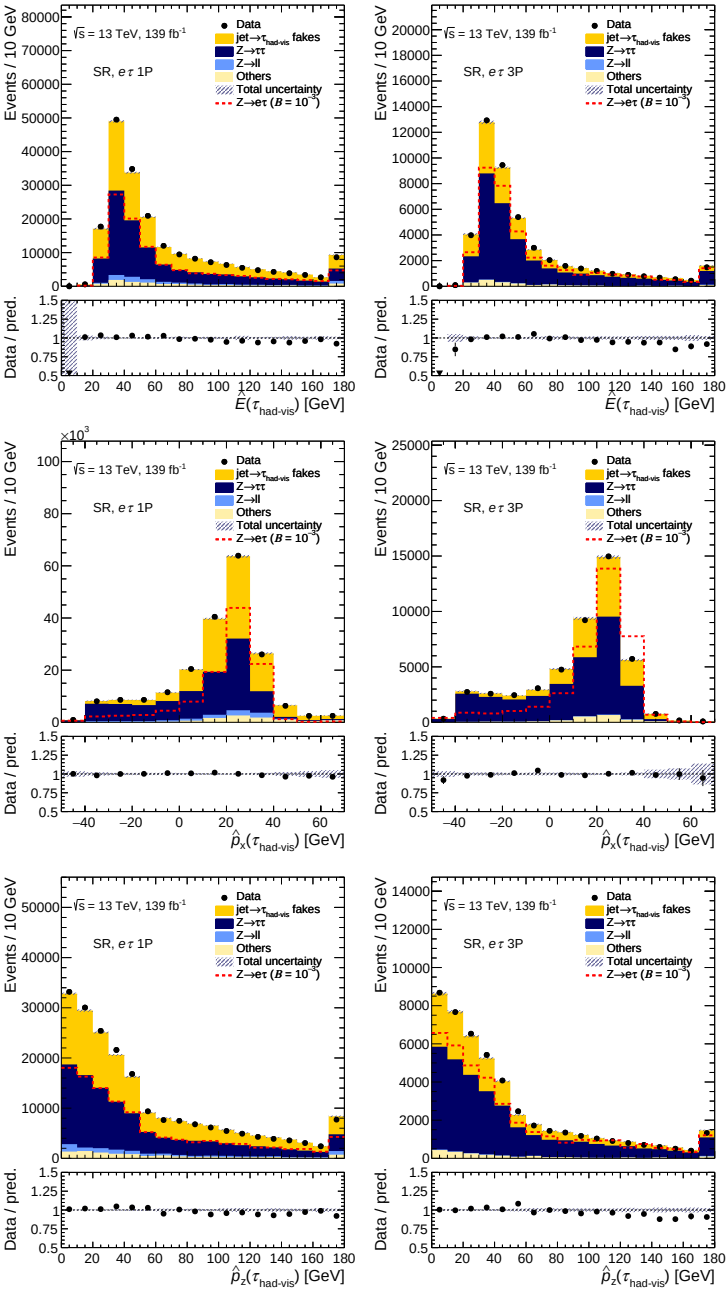


Figure C.2.: Expected and observed distributions of $\hat{E}(\tau_{\text{had-vis}})$, $\hat{p}_x(\tau_{\text{had-vis}})$ and $\hat{p}_z(\tau_{\text{had-vis}})$ in the SR of the $e\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

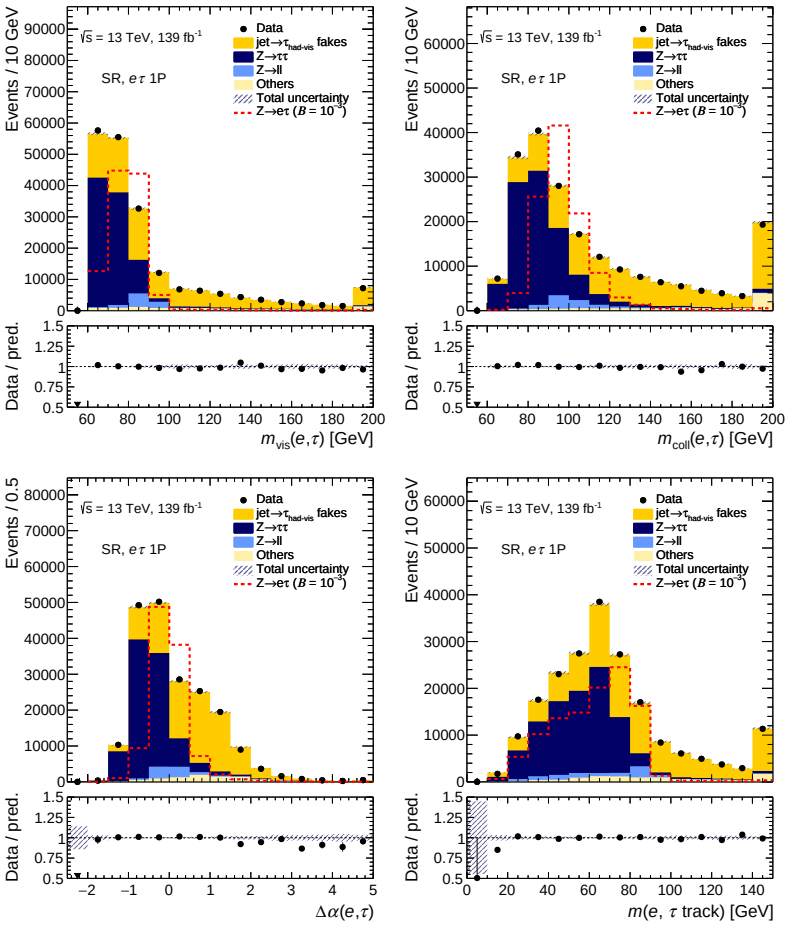


Figure C.3.: Expected and observed distributions of $m_{\text{vis}}(e,\tau)$, $m_{\text{coll}}(e,\tau)$, $\Delta\alpha(e,\tau)$ and $m(e,\tau$ track) in the 1P SR of the $e\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

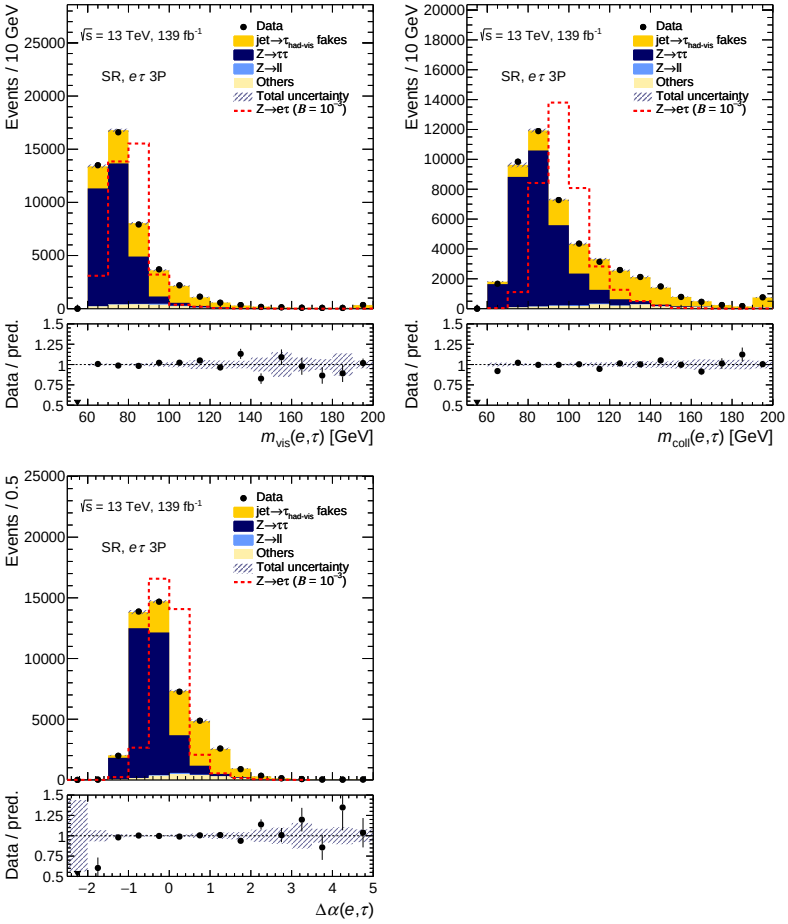


Figure C.4.: Expected and observed distributions of $m_{\text{vis}}(e,\tau)$, $m_{\text{coll}}(e,\tau)$, $\Delta\alpha(e,\tau)$ and $m(e,\tau$ track) in the 3P SR of the $e\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

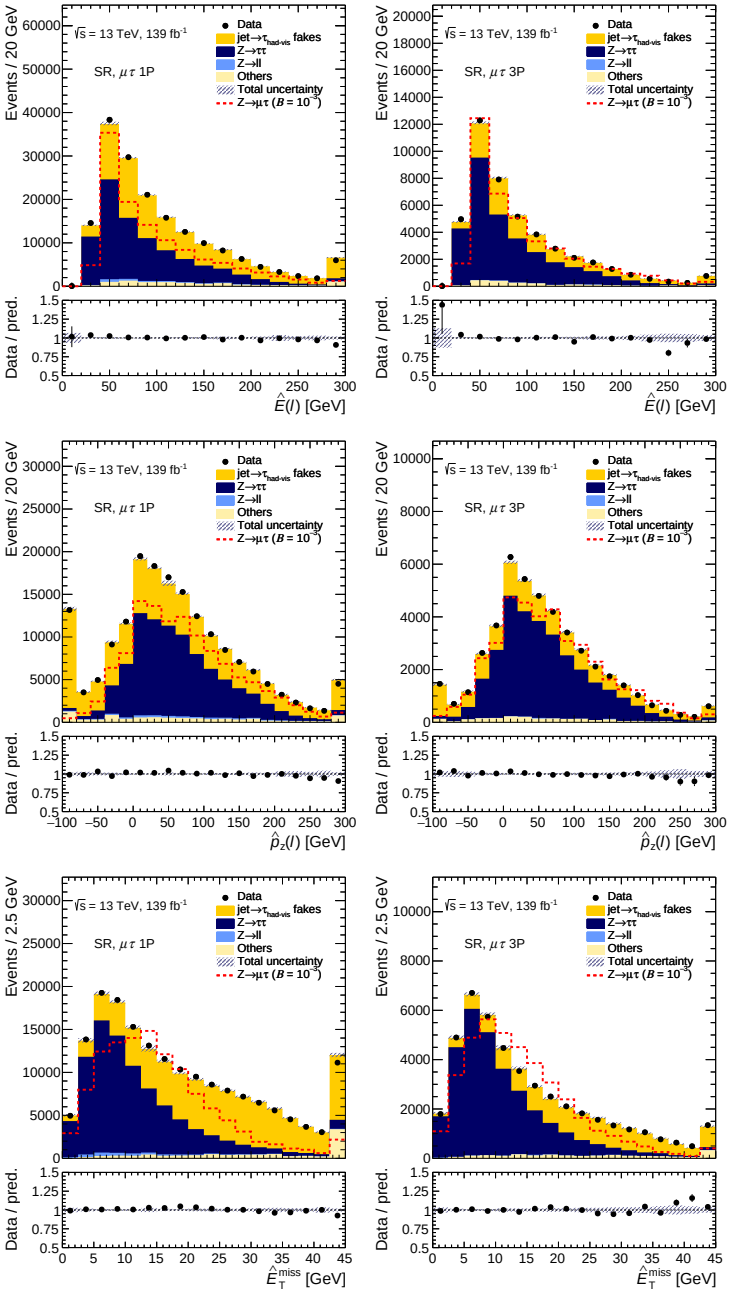


Figure C.5.: Expected and observed distributions of $\hat{E}(\ell)$, $\hat{p}_z(\ell)$ and $\hat{E}_{\mathrm{T}}^{\mathrm{miss}}$ in the SR of the $\mu\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

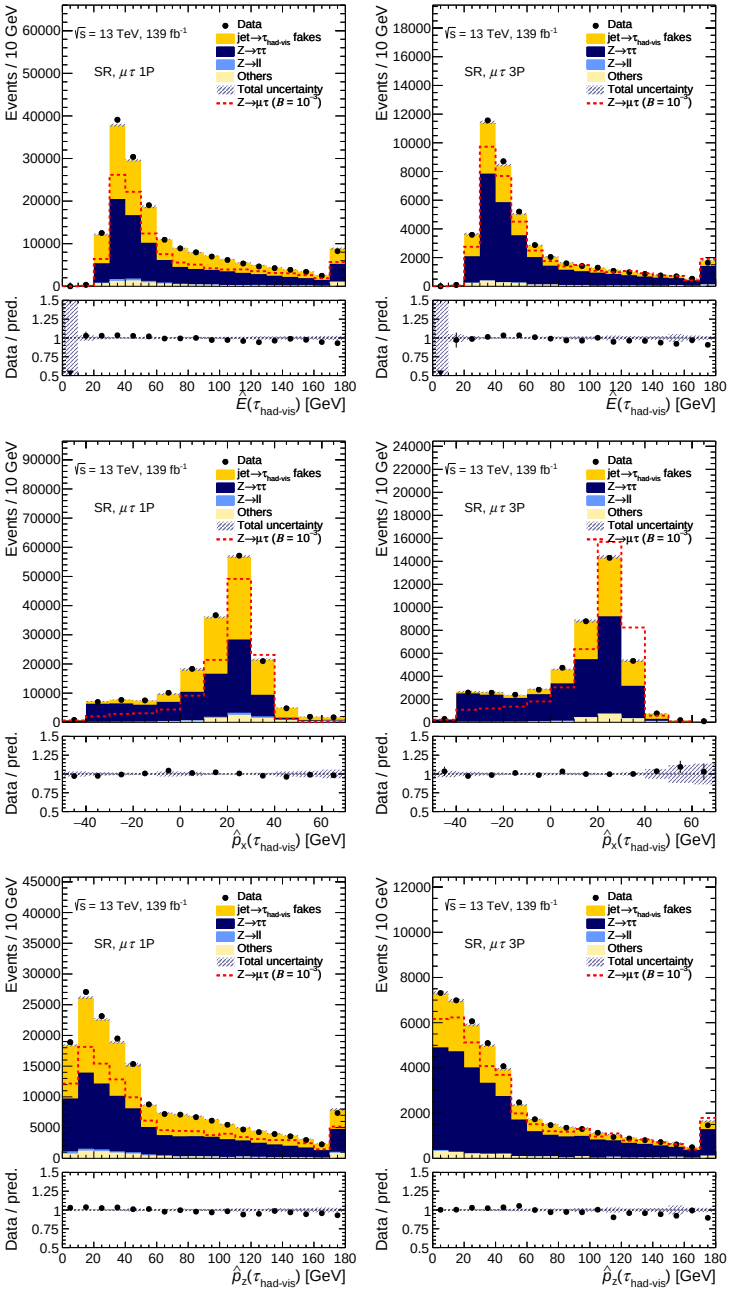


Figure C.6.: Expected and observed distributions of $\hat{E}(\tau_{\text{had-vis}})$, $\hat{p}_x(\tau_{\text{had-vis}})$ and $\hat{p}_z(\tau_{\text{had-vis}})$ in the SR of the $\mu\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

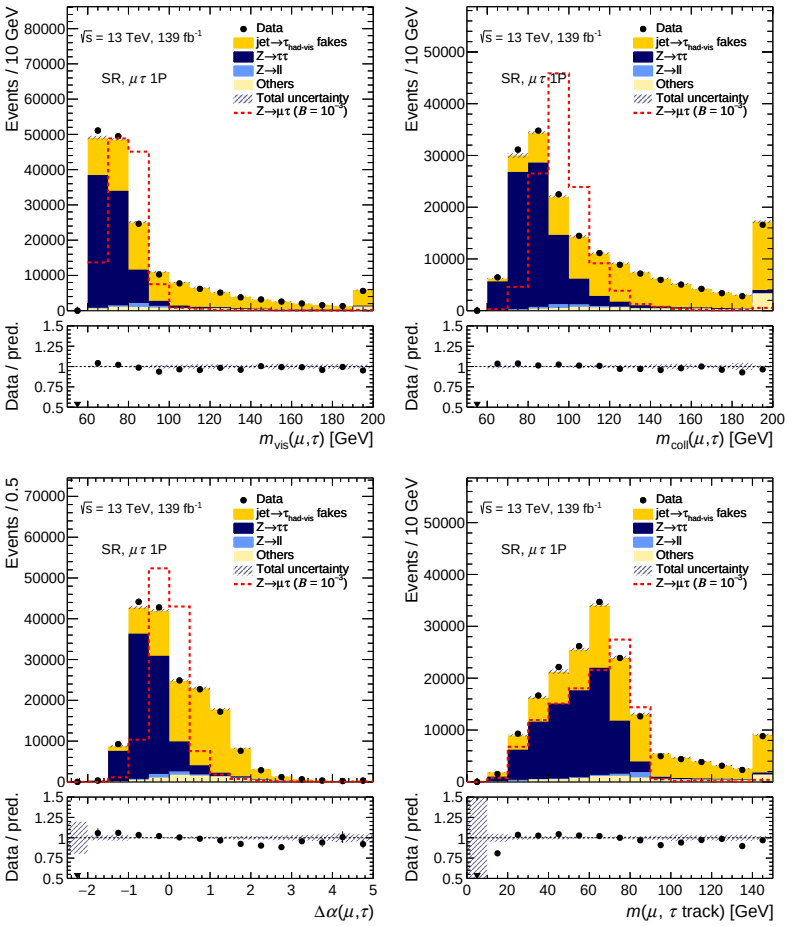


Figure C.7.: Expected and observed distributions of $m_{\text{vis}}(\mu,\tau)$, $m_{\text{coll}}(\mu,\tau)$, $\Delta\alpha(\mu,\tau)$ and $m(\mu,\tau$ track) in the 1P SR of the $\mu\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

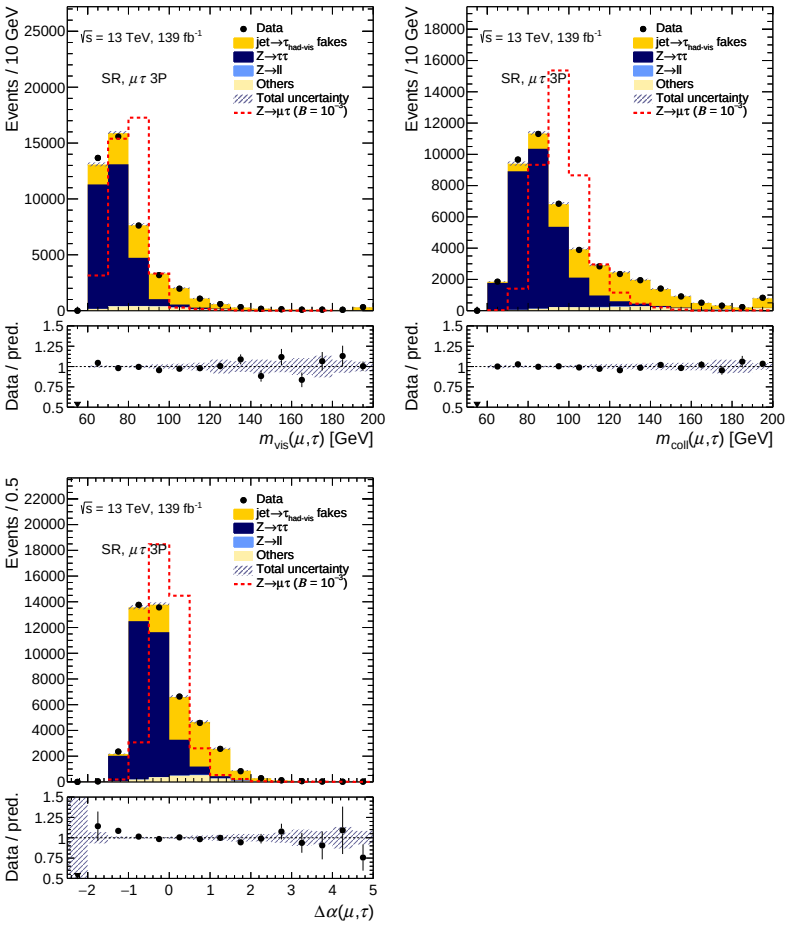


Figure C.8.: Expected and observed distributions of $m_{\text{vis}}(\mu,\tau)$, $m_{\text{coll}}(\mu,\tau)$, $\Delta\alpha(\mu,\tau)$ and $m(\mu,\tau$ track) in the 3P SR of the $\mu\tau$ channel. In the lower panel of each plot, the ratios of the observed yields to the predicted background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

Appendix D.

Measured fake factors

The following figures show the measured process-specific fake factors, F^p, as well as the relative contributions of each process to the total yield of fakes in the SR, R^p_{SR}.

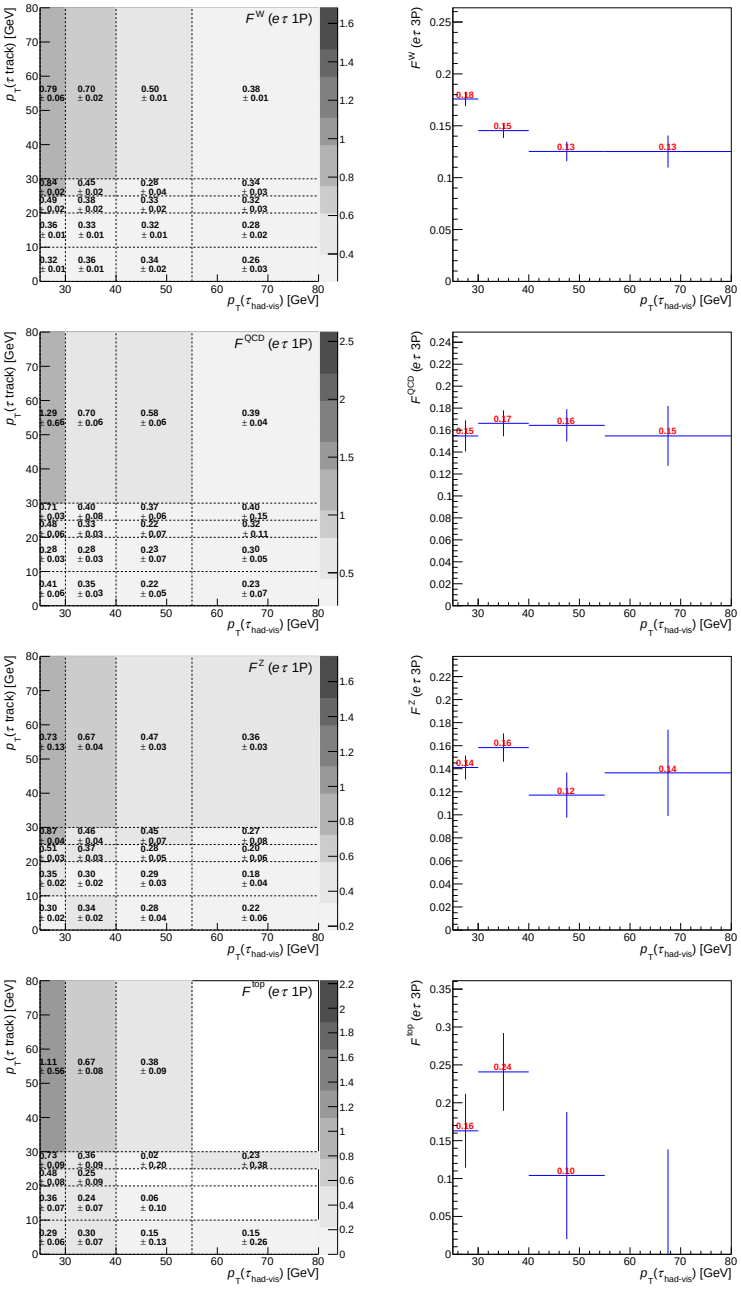


Figure D.1.: The measured process-specific fake factors, F^p, in the $e\tau$ channel. While the plot only shows $p_T(\tau_{\text{had-vis}})$ and $p_T(\tau$ track) up to 80 GeV, the actual bin boundaries are at infinity.

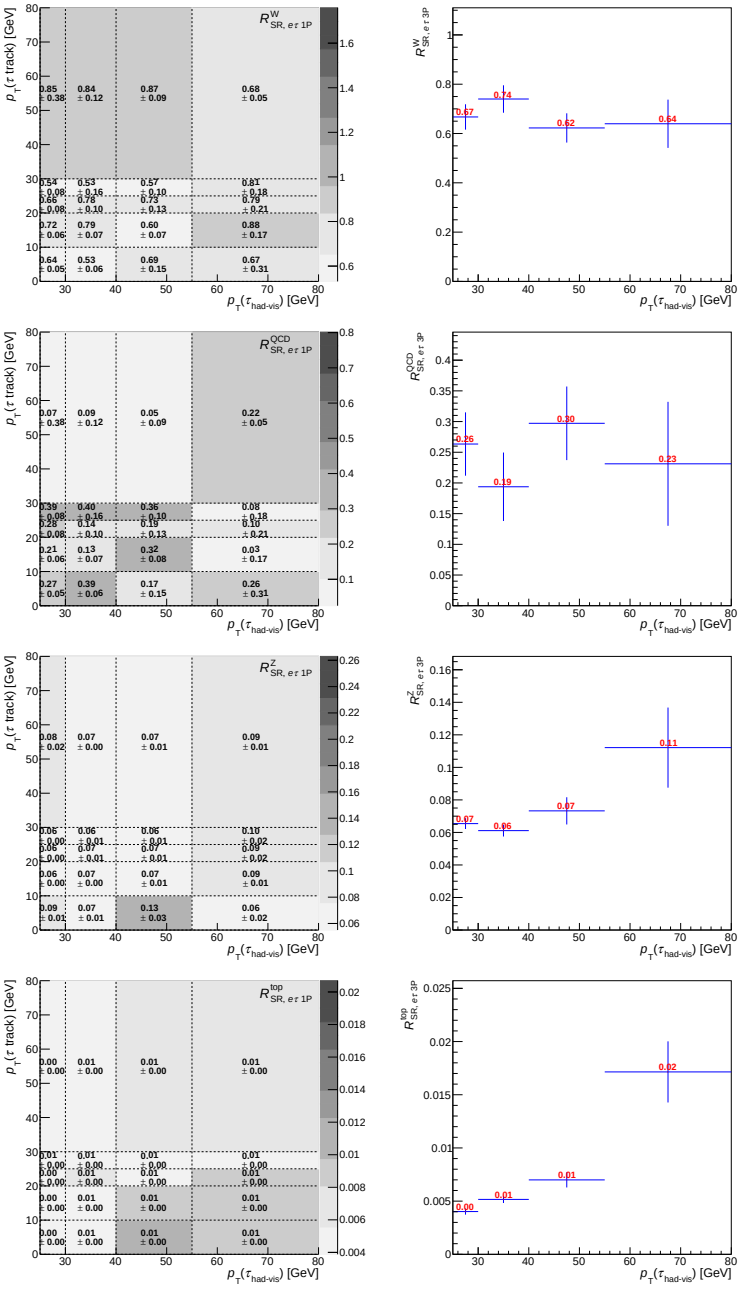


Figure D.2.: The estimated relative contributions of each process to the total yield of fakes in the SR, R^{p}_{SR}, in the $e\tau$ channel. While the plot only shows $p_{\mathrm{T}}(\tau_{\text{had-vis}})$ and $p_{\mathrm{T}}(\tau$ track) up to 80 GeV, the actual bin boundaries are at infinity.

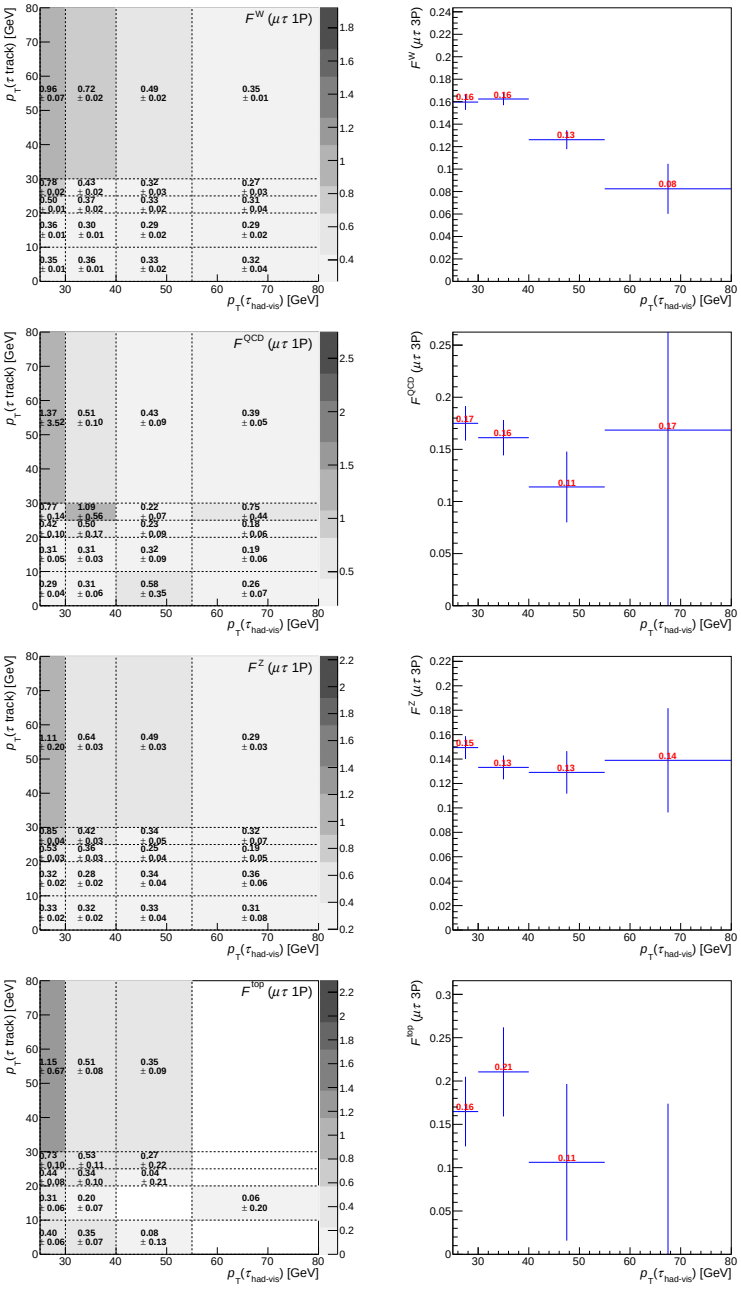


Figure D.3.: The measured process-specific fake factors, F^p, in the $\mu\tau$ channel. While the plot only shows $p_T(\tau_{\text{had-vis}})$ and $p_T(\tau$ track) up to 80 GeV, the actual bin boundaries are at infinity.

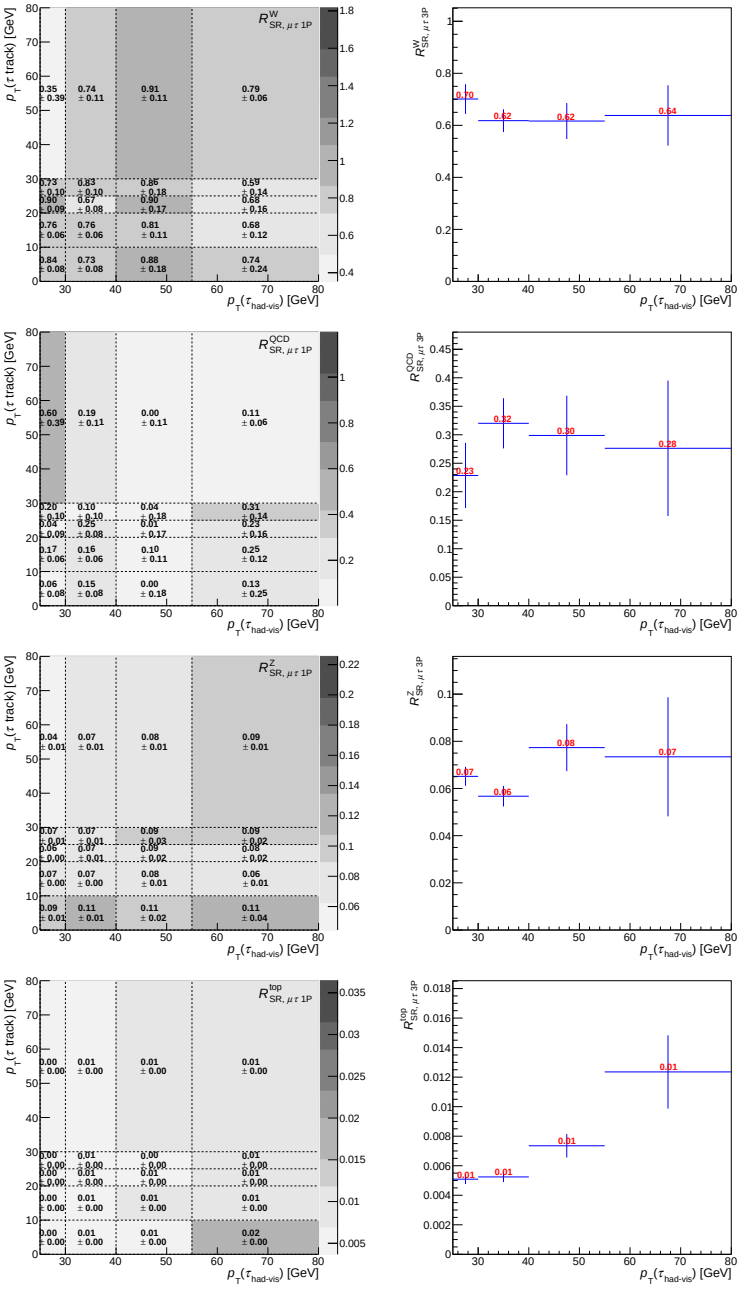


Figure D.4.: The estimated relative contributions of each process to the total yield of fakes in the SR, R^{p}_{SR}, in the $\mu\tau$ channel. While the plot only shows $p_{\mathrm{T}}(\tau_{\text{had-vis}})$ and $p_{\mathrm{T}}(\tau$ track) up to 80 GeV, the actual bin boundaries are at infinity.

Bibliography

[1] ATLAS Collaboration, *Charged-lepton-flavour violation at the LHC: a search for $Z \to e\tau/\mu\tau$ decays with the ATLAS detector* (2020), arXiv: 2010.02566 [hep-ex] (cit. on p. 2).

[2] S. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579, DOI: 10.1016/0029-5582(61)90469-2 (cit. on pp. 4, 11).

[3] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264, DOI: 10.1103/PhysRevLett.19.1264 (cit. on pp. 4, 11).

[4] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367, DOI: 10.1142/9789812795915_0034 (cit. on pp. 4, 11).

[5] G. 't Hooft and M. Veltman, *Regularization and renormalization of gauge fields*, Nucl. Phys. B **44** (1972) 189, DOI: https://doi.org/10.1016/0550-3213(72)90279-9 (cit. on p. 4).

[6] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001, DOI: 10.1103/PhysRevD.98.030001 (cit. on pp. 8–10, 16, 25).

[7] National Institute of Standards and Technology, *2018 CODATA recommended values*, URL: https://physics.nist.gov/cuu/Constants/bibliography.html (visited on 06/03/2020) (cit. on p. 9).

[8] P. A. M. Dirac, *The Quantum Theory of the Emission and Absorption of Radiation*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **114** (1927) 243, DOI: 10.1098/rspa.1927.0039 (cit. on p. 10).

[9] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, *Tenth-Order QED Contribution to the Electron $g-2$ and an Improved Value of the Fine Structure Constant*, Phys. Rev. Lett. **109** (2012) 111807, DOI: 10.1103/PhysRevLett.109.111807 (cit. on p. 10).

[10] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321, DOI: 10.1103/PhysRevLett.13.321 (cit. on p. 11).

[11] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132, DOI: 10.1016/0031-9163(64)91136-9 (cit. on p. 11).

[12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508, DOI: 10.1103/PhysRevLett.13.508 (cit. on p. 11).

[13] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585, DOI: 10.1103/PhysRevLett.13.585 (cit. on p. 11).

[14] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156, DOI: 10.1103/PhysRev.145.1156 (cit. on p. 11).

[15] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (1967) 1554, DOI: 10.1103/PhysRev.155.1554 (cit. on p. 11).

[16] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, DOI: 10.1103/PhysRevLett.114.191803, arXiv: 1503.07589 [hep-ex] (cit. on p. 11).

[17] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (1970) 1285, DOI: 10.1103/PhysRevD.2.1285 (cit. on p. 11).

[18] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics*, Phys. Rev. **76** (1949) 769, DOI: 10.1103/PhysRev.76.769 (cit. on p. 13).

[19] E. Noether, *Invariante Variationsprobleme*, ger, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918** (1918) 235 (cit. on p. 14).

[20] Super-Kamiokande Collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, DOI: 10.1103/PhysRevLett.81.1562 (cit. on p. 16).

[21] SNO Collaboration, *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301, DOI: 10.1103/PhysRevLett.89.011301 (cit. on p. 16).

[22] P. de Salas, D. Forero, C. Ternes, M. Tórtola, and J. Valle, *Status of neutrino oscillations 2018: 3σ hint for normal mass ordering and improved CP sensitivity*, Phys. Lett. B **782** (2018) 633, DOI: 10.1016/j.physletb.2018.06.019 (cit. on p. 17).

[23] T. Asaka and M. Shaposhnikov, *The nuMSM, dark matter and baryon asymmetry of the universe*, Phys. Lett. B **620** (2005) 17, DOI: 10.1016/j.physletb.2005.06.020 (cit. on p. 17).

[24] R. P. Feynman, *Feynman lectures on gravitation*, Addison-Wesley, 1995, ISBN: 978-0201627343 (cit. on p. 18).

[25] A. Bosma, *The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types* (1978) (cit. on p. 18).

[26] D. Maity and P. Saha, *Connecting CMB anisotropy and cold dark matter phenomenology via reheating*, Phys. Rev. D **98** (2018) 103525, DOI: 10.1103/PhysRevD.98.103525, arXiv: 1801.03059 [hep-ph] (cit. on p. 18).

[27] J. Einasto, *Dark matter and large scale structure*, ASP Conf. Ser. **252** (2001) 85, arXiv: astro-ph/0012161 (cit. on p. 18).

[28] A. D. Sakharov, *Violation of CP in variance, C asymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34** (1991) 392, DOI: 10.1070/PU1991v034n05ABEH002497 (cit. on p. 18).

[29] BaBar Collaboration, *Evidence for an Excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ Decays*, Phys. Rev. Lett. **109** (2012) 101802, DOI: 10.1103/PhysRevLett.109.101802 (cit. on p. 19).

[30] Belle Collaboration, *Measurement of the branching ratio of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ relative to $\overline{B} \to D^{(*)}\ell^-\overline{\nu}_\ell$ decays with hadronic tagging at Belle*, Phys. Rev. D **92** (2015) 072014, DOI: 10.1103/PhysRevD.92.072014 (cit. on p. 19).

[31] L. Collaboration, *Measurement of the Ratio of Branching Fractions $\mathcal{B}(\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \to D^{*+}\mu^-\overline{\nu}_\mu)$*, Phys. Rev. Lett. **115** (2015) 111803, DOI: 10.1103/PhysRevLett.115.111803 (cit. on p. 19).

[32] L. Collaboration, *Measurement of the Ratio of the $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ Branching Fractions Using Three-Prong τ-Lepton Decays*, Phys. Rev. Lett. **120** (2018) 171802, DOI: 10.1103/PhysRevLett.120.171802 (cit. on p. 19).

[33] L. Collaboration, *Test of lepton flavor universality by the measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays*, Phys. Rev. D **97** (2018) 072013, DOI: 10.1103/PhysRevD.97.072013 (cit. on p. 19).

[34] L. Collaboration, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, DOI: 10.1007/JHEP08(2017)055, arXiv: 1705.05802 [hep-ex] (cit. on p. 19).

[35] HFLAV group, *Average of $R(D)$ and $R(D^{*+})$ for spring 2019*, URL: https://hflav-eos.web.cern.ch/hflav-eos/semi/spring19/html/RDsDsstar/RDRDs.html (visited on 10/09/2020) (cit. on p. 19).

[36] A. de Gouvêa and P. Vogel, *Lepton flavor and number conservation, and physics beyond the standard model*, Prog. Part. Nucl. Phys. **71** (2013), Fundamental Symmetries in the Era of the LHC 75, DOI: https://doi.org/10.1016/j.ppnp.2013.03.006 (cit. on p. 19).

[37] J. I. Illana, M. Jack, and T. Riemann, "Predictions for $Z \to \mu\tau$ and related reactions", *2nd Workshop of the 2nd Joint ECFA / DESY Study on Physics and Detectors for a Linear Electron Positron Collider*, 1999, pp. 490–524, arXiv: hep-ph/0001273 (cit. on p. 19).

[38] F. F. Deppisch, P. S. B. Dev, and A. Pilaftsis, *Neutrinos and collider physics*, New J. Phys. **17** (2015) 075019, DOI: 10.1088/1367-2630/17/7/075019 (cit. on p. 20).

[39] J. I. Illana and T. Riemann, *Charged lepton flavor violation from massive neutrinos in Z decays*, Phys. Rev. D **63** (2001) 053004, DOI: 10.1103/PhysRevD.63.053004 (cit. on p. 21).

[40] V. De Romeri, M. J. Herrero, X. Marcano, and F. Scarcella, *Lepton flavor violating Z decays: A promising window to low scale seesaw neutrinos*, Phys. Rev. D **95** (2017) 075028, DOI: 10.1103/PhysRevD.95.075028 (cit. on p. 21).

[41] G. L. Kane and M. A. Shifman, *Introduction to 'the supersymmetric world: The Beginnings of the theory'* (2001), arXiv: hep-ph/0102298 (cit. on p. 22).

[42] S. P. Martin, "A Supersymmetry primer", *Perspectives on Supersymmetry II*, 2010, pp. 1–153, DOI: 10.1142/9789812839657_0001, arXiv: hep-ph/9709356 (cit. on p. 22).

[43] J. M. Yang, *Lepton flavor violating Z-boson decays at GigaZ as a probe of supersymmetry*, Sci. China Phys. Mech. Astron. **53** (2010) 1949, DOI: 10.1007/s11433-010-4146-3, arXiv: 1006.2594 [hep-ph] (cit. on p. 22).

[44] J. Cao, Z. Xiong, and J. M. Yang, *Lepton flavor violating Z decays in supersymmetric seesaw model*, Eur. Phys. J. C **32** (2004) 245, DOI: 10.1140/epjc/s2003-01391-1, arXiv: hep-ph/0307126 (cit. on p. 22).

[45] T. Lee, *A Theory of Spontaneous T Violation*, Phys. Rev. D **8** (1973), ed. by G. Feinberg 1226, DOI: 10.1103/PhysRevD.8.1226 (cit. on p. 22).

[46] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rep. **516** (2012), Theory and phenomenology of two-Higgs-doublet models 1, DOI: 10.1016/j.physrep.2012.02.002 (cit. on p. 23).

[47] R. Benbrik, C.-H. Chen, and T. Nomura, $h, Z \to \ell_i \bar{\ell}_j$, Δa_μ, $\tau \to (3\mu, \mu\gamma)$ *in generic two-Higgs-doublet models*, Phys. Rev. D **93** (2016) 095004, DOI: 10.1103/PhysRevD.93.095004, arXiv: 1511.08544 [hep-ph] (cit. on p. 23).

[48] MEG Collaboration, *Search for the lepton flavour violating decay* $\mu^+ \to e^+\gamma$ *with the full dataset of the MEG experiment*, Eur. Phys. J. C **76** (2016) 434, DOI: 10.1140/epjc/s10052-016-4271-x, arXiv: 1605.05081 [hep-ex] (cit. on p. 25).

[49] SINDRUM Collaboration, *Search for the Decay* $\mu^+ \to e^+e^+e^-$, Nucl. Phys. B **299** (1988) 1, DOI: 10.1016/0550-3213(88)90462-2 (cit. on p. 25).

[50] KTeV Collaboration, *Search for lepton flavor violating decays of the neutral kaon*, Phys. Rev. Lett. **100** (2008) 131803, DOI: 10.1103/PhysRevLett.100.131803, arXiv: 0711.3472 [hep-ex] (cit. on p. 25).

[51] BESIII Collaboration, *Search for the lepton flavor violation process* $J/\psi \to e\mu$ *at BESIII*, Phys. Rev. D **87** (2013) 112007, DOI: 10.1103/PhysRevD.87.112007, arXiv: 1304.3205 [hep-ex] (cit. on p. 25).

[52] CMS Collaboration, *Search for lepton flavour violating decays of the Higgs boson to* $e\tau$ *and* $e\mu$ *in proton–proton collisions at* $\sqrt{s} = 8$ *TeV*, Phys. Lett. B **763** (2016) 472, DOI: 10.1016/j.physletb.2016.09.062, arXiv: 1607.03561 [hep-ex] (cit. on p. 25).

[53] CMS Collaboration, *Search for Lepton Flavor Violation in Z decays in pp collisions at sqrt(s)=8 TeV*, tech. rep. CMS-PAS-EXO-13-005, CERN, 2015, URL: https://cds.cern.ch/record/2019863 (cit. on p. 25).

[54] BaBar Collaboration, *Searches for Lepton Flavor Violation in the Decays* $\tau^\pm \to e^\pm\gamma$ *and* $\tau^\pm \to \mu^\pm\gamma$, Phys. Rev. Lett. **104** (2010) 021802, DOI: 10.1103/PhysRevLett.104.021802 (cit. on p. 25).

[55] Belle Collaboration, *Search for lepton-flavor-violating* τ *decays into three leptons with 719 million produced* $\tau^+\tau^-$ *pairs*, Phys. Lett. B **687** (2010) 139, DOI: 10.1016/j.physletb.2010.03.037 (cit. on p. 25).

[56] BES Collaboration, *Search for the lepton flavor violation processes* $J/\psi \to \mu\tau$ *and* $e\tau$, Phys. Lett. B **598** (2004) 172, DOI: 10.1016/j.physletb.2004.08.005, arXiv: hep-ex/0406018 (cit. on p. 25).

[57] ATLAS Collaboration, *Searches for lepton-flavour-violating decays of the Higgs boson in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, Phys. Lett. B **800** (2020) 135069, DOI: 10.1016/j.physletb.2019.135069, arXiv: 1907.06131 [hep-ex] (cit. on p. 25).

[58] OPAL Collaboration, *A Search for lepton flavor violating Z^0 decays*, Z. Phys. C **67** (1995) 555, DOI: 10.1007/BF01553981 (cit. on pp. 25, 26, 115).

[59] CMS Collaboration, *Search for lepton flavour violating decays of the Higgs boson to $\mu\tau$ and $e\tau$ in proton–proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **06** (2018) 001, DOI: 10.1007/JHEP06(2018)001, arXiv: 1712.07173 [hep-ex] (cit. on p. 25).

[60] DELPHI Collaboration, *Search for lepton flavor number violating Z^0 decays*, Z. Phys. C **73** (1997) 243, DOI: 10.1007/s002880050313 (cit. on pp. 25, 26, 115).

[61] S. Davidson, S. Lacroix, and P. Verdier, *LHC sensitivity to lepton flavour violating Z boson decays*, JHEP **09** (2012) 092, DOI: 10.1007/JHEP09(2012)092, arXiv: 1207.4894 [hep-ph] (cit. on pp. 24, 77).

[62] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **759** (2016) 601, DOI: 10.1016/j.physletb.2016.06.023, arXiv: 1603.09222 [hep-ex] (cit. on pp. 24, 86, 88, 109).

[63] LIGO Scientific Collaboration and Virgo Collaboration, *GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs*, Phys. Rev. X **9** (2019) 031040, DOI: 10.1103/PhysRevX.9.031040 (cit. on p. 27).

[64] LIGO Scientific Collaboration and Virgo Collaboration, *Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1*, Phys. Rev. D **100** (2019) 104036, DOI: 10.1103/PhysRevD.100.104036 (cit. on p. 27).

[65] E. Mobs, "The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018", General Photo, 2018, URL: http://cds.cern.ch/record/2636343 (cit. on p. 29).

[66] ATLAS Collaboration, *Luminosity Public Results Run 2*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on 04/16/2020) (cit. on p. 31).

[67] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003, DOI: 10.1088/1748-0221/3/08/S08003 (cit. on pp. 32, 36, 38, 40, 41).

[68] J. Pequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: https://cds.cern.ch/record/1095924 (cit. on p. 34).

[69] J. Pequenao and P. Schaffner, "How ATLAS detects particles: diagram of particle paths in the detector", 2013, URL: https://cds.cern.ch/record/1505342 (cit. on p. 35).

[70] K. Potamianos, *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*, tech. rep. ATL-PHYS-PROC-2016-104, 15 pages, EPS-HEP 2015 Proceedings: CERN, 2016, URL: https://cds.cern.ch/record/2209070 (cit. on p. 36).

[71] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, DOI: 10.1140/epjc/s10052-017-5004-5, arXiv: 1603.02934 [hep-ex] (cit. on p. 46).

[72] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, DOI: 10.1088/1126-6708/2008/04/063, arXiv: 0802.1189 [hep-ph] (cit. on p. 46).

[73] ATLAS Collaboration, *Local Hadronic Calibration*, ATL-LARG-PUB-2009-001-2, 2008, URL: https://cds.cern.ch/record/1112035 (cit. on pp. 46, 52).

[74] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, DOI: 10.1140/epjc/s10052-012-1896-2, arXiv: 1111.6097 [hep-ph] (cit. on p. 47).

[75] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, JHEP **08** (2018) 089, DOI: 10.1007/JHEP08(2018)089, arXiv: 1805.01845 [hep-ex] (cit. on pp. 47, 48, 108).

[76] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*, Chapman & Hall, 1993, ISBN: 0412048418 (cit. on pp. 48, 51).

[77] ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. C **75** (2015) 303, DOI: 10.1140/epjc/s10052-015-3500-z, arXiv: 1412.7086 [hep-ex] (cit. on pp. 49, 56).

[78] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, 2015, URL: https://cds.cern.ch/record/2064383 (cit. on p. 49).

[79] ATLAS Collaboration, *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, Eur. Phys. J. C **76** (2016) 295, DOI: 10.1140/epjc/s10052-016-4110-0, arXiv: 1512.05955 [hep-ex] (cit. on p. 50).

[80] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide*, tech. rep. physics/0703039, CERN, 2007, URL: https://cds.cern.ch/record/1019880 (cit. on p. 51).

[81] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2017-029, 2017, URL: https://cds.cern.ch/record/2261772 (cit. on pp. 51, 108, 109).

[82] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, 2019, URL: https://cds.cern.ch/record/2688062 (cit. on pp. 53, 57, 108).

[83] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, JINST **14** (2019) P03017, DOI: 10.1088/1748-0221/14/03/P03017, arXiv: 1812.03848 [hep-ex] (cit. on p. 59).

[84] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} =$ 13 *TeV*, Eur. Phys. J. C **79** (2019) 639, DOI: 10.1140/epjc/s10052-019-7140-6, arXiv: 1902.04655 [hep-ex] (cit. on pp. 59, 61, 108).

[85] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, DOI: 10.1088/1748-0221/14/12/P12006, arXiv: 1908.00005 [hep-ex] (cit. on pp. 59, 61, 108).

[86] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **76** (2016) 292, DOI: 10.1140/epjc/s10052-016-4120-y, arXiv: 1603.05598 [hep-ex] (cit. on pp. 60, 63, 108).

[87] ATLAS Collaboration, *High-*μ *Muon Reconstruction and Isolation Studies on 2017 Data*, 2017, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/MUON-2017-005/index.html (visited on 07/11/2020) (cit. on pp. 63, 108).

[88] 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, DOI: 10.1140/epjc/s10052-018-6288-9, arXiv: 1802.08168 [hep-ex] (cit. on pp. 63, 108).

[89] ATLAS Collaboration, E_T^{miss} *performance in the ATLAS detector using 2015–2016 LHC pp collisions*, ATLAS-CONF-2018-023, 2018, URL: https://cds.cern.ch/record/2625233 (cit. on pp. 63, 64, 108).

[90] M. Cacciari, G. P. Salam, and G. Soyez, *The catchment area of jets*, JHEP **2008** (2008) 005, DOI: 10.1088/1126-6708/2008/04/005 (cit. on p. 66).

[91] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003, DOI: 10.1088/1748-0221/15/04/P04003, arXiv: 1911.04632 [physics.ins-det] (cit. on p. 68).

[92] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47, DOI: 10.1140/epjc/s10052-019-7500-2, arXiv: 1909.00761 [hep-ex] (cit. on p. 68).

[93] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2* (2020), arXiv: 2004.13447 [hep-ex] (cit. on p. 68).

[94] F. Chollet et al., *Keras*, 2015, URL: https://keras.io (cit. on p. 78).

[95] M. Abadi et al., *TensorFlow: large-acale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015, URL: https://www.tensorflow.org/ (cit. on p. 78).

[96] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014, arXiv: 1412.6980 [cs.LG] (cit. on pp. 78, 81).

[97] R. A. Dunne and N. A. Campbell, "On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function", *Proc. 8th Aust. Conf. on the Neural Networks, Melbourne*, 1997, pp. 181–185 (cit. on p. 82).

[98] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, DOI: 10.1016/j.cpc.2015.01.024, arXiv: 1410.3012 [hep-ph] (cit. on p. 86).

[99] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on p. 86).

[100] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, DOI: 10.1016/j.nuclphysb.2012.10.003, arXiv: 1207.1303 [hep-ph] (cit. on p. 86).

[101] E. Bothmann et al., *Event Generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, DOI: 10.21468/SciPostPhys.7.3.034, arXiv: 1905.09127 [hep-ph] (cit. on p. 86).

[102] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, DOI: 10.1088/1126-6708/2008/12/039, arXiv: 0808.3674 [hep-ph] (cit. on p. 86).

[103] F. Buccioni et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866, DOI: 10.1140/epjc/s10052-019-7306-2, arXiv: 1907.13071 [hep-ph] (cit. on p. 86).

[104] F. Cascioli, P. Maierhöfer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, DOI: 10.1103/PhysRevLett.108.111601, arXiv: 1111.5206 [hep-ph] (cit. on p. 86).

[105] A. Denner, S. Dittmaier, and L. Hofer, *Collier: A fortran-based complex one-loop library in extended regularizations*, Comput. Phys. Commun. **212** (2017) 220, DOI: 10.1016/j.cpc.2016.10.013, arXiv: 1604.06792 [hep-ph] (cit. on p. 86).

[106] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, JHEP **03** (2008) 038, DOI: 10.1088/1126-6708/2008/03/038, arXiv: 0709.1027 [hep-ph] (cit. on p. 86).

[107] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *A critical appraisal of NLO+PS matching methods*, JHEP **09** (2012) 049, DOI: 10.1007/JHEP09(2012)049, arXiv: 1111.1220 [hep-ph] (cit. on p. 86).

[108] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, JHEP **04** (2013) 027, DOI: 10.1007/JHEP04(2013)027, arXiv: 1207.5030 [hep-ph] (cit. on p. 86).

[109] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *QCD Matrix Elements + Parton Showers*, JHEP **11** (2001) 063, DOI: 10.1088/1126-6708/2001/11/063, arXiv: hep-ph/0109231 (cit. on p. 86).

[110] S. Höche, F. Krauss, S. Schumann, and F. Siegert, *QCD matrix elements and truncated showers*, JHEP **05** (2009) 053, DOI: 10.1088/1126-6708/2009/05/053, arXiv: 0903.1219 [hep-ph] (cit. on p. 86).

[111] 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, DOI: 10.1007/JHEP06(2010)043, arXiv: 1002.2581 [hep-ph] (cit. on p. 86).

[112] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, DOI: 10.1103/PhysRevD.82.074024, arXiv: 1007.2241 [hep-ph] (cit. on p. 86).

[113] J. Pumplin et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, JHEP **07** (2002) 012, DOI: 10.1088/1126-6708/2002/07/012, arXiv: hep-ph/0201195 (cit. on p. 86).

[114] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, DOI: 10.1007/JHEP09(2014)145, arXiv: 1406.3660 [hep-ex] (cit. on p. 86).

[115] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, DOI: 10.1007/JHEP04(2015)040, arXiv: 1410.8849 [hep-ph] (cit. on p. 86).

[116] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013) 222, DOI: 10.1007/JHEP10(2013)222, arXiv: 1309.0017 [hep-ph] (cit. on p. 86).

[117] K. Hamilton, P. Nason, and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05** (2015) 140, DOI: 10.1007/JHEP05(2015)140, arXiv: 1501.04637 [hep-ph] (cit. on p. 86).

[118] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector* (2016), DOI: 10.23731/CYRM-2017-002, arXiv: 1610.07922 [hep-ph] (cit. on pp. 87, 110).

[119] S. Agostinelli et al., GEANT4 – *a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250, DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 87).

[120] 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 (cit. on p. 87).

[121] T. Przedzinski, E. Richter-Was, and Z. Was, *Documentation of TauSpinner algorithms: program for simulating spin effects in τ -lepton production at LHC*, Eur. Phys. J. C **79** (2019) 91, DOI: 10.1140/epjc/s10052-018-6527-0, arXiv: 1802.05459 [hep-ph] (cit. on p. 87).

[122] ATLAS Collaboration, *Measurement of the transverse momentum distribution of Drell-Yan lepton pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 616, DOI: 10.1140/epjc/s10052-020-8001-z, arXiv: 1912.02844 [hep-ex] (cit. on p. 90).

[123] A. L. Read, *Presentation of search results: the CL_s technique*, J. Phys. G **28** (2002) 2693, DOI: 10.1088/0954-3899/28/10/313 (cit. on p. 105).

[124] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, 2003, arXiv: physics/0306116 [physics.data-an] (cit. on p. 106).

[125] L. Moneta et al., *The RooStats Project*, PoS **ACAT2010** (2010), ed. by T. Speer et al. 057, DOI: 10.22323/1.093.0057, arXiv: 1009.1003 [physics.data-an] (cit. on p. 106).

[126] M. Baak et al., *HistFitter software framework for statistical data analysis*, Eur. Phys. J. C **75** (2015) 153, DOI: 10.1140/epjc/s10052-015-3327-7, arXiv: 1410.1280 [hep-ex] (cit. on p. 106).

[127] 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, DOI: 10.1140/epjc/s10052-011-1554-0, arXiv: 1007.1727 [physics.data-an] (cit. on p. 108).

[128] ATLAS Collaboration, *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector* (2020), arXiv: 2007.02645 [hep-ex] (cit. on p. 108).

[129] 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 (cit. on p. 110).

[130] LHC TOP Physics Working Group, *NNLO+NNLL top-quark-pair cross sections*, 2015, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO#Top_quark_pair_cross_sections_at (visited on 06/23/2020) (cit. on p. 110).

[131] LHC TOP Physics Working Group, *NLO single-top channel cross sections*, 2017, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec#Predictions_at_7_8_13_and_14_TeV (visited on 06/23/2020) (cit. on p. 110).

[132] ATLAS Collaboration, *Search for lepton-flavour-violating decays of the Higgs and Z bosons with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 70, DOI: 10.1140/epjc/s10052-017-4624-0, arXiv: 1604.07730 [hep-ex] (cit. on pp. 115, 118).

[133] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. A **654** (2011) 481, DOI: 10.1016/j.nima.2011.07.009, arXiv: 1012.4686 [hep-ex] (cit. on p. 118).

[134] ATLAS Collaboration, *Modelling $Z \to \tau\tau$ processes in ATLAS with τ-embedded $Z \to \mu\mu$ data*, JINST **10** (2015) P09018, DOI: 10.1088/1748-0221/10/09/P09018, arXiv: 1506.05623 [hep-ex] (cit. on p. 118).

[135] K. Hornik, M. Stinchcombe, and H. White, *Multilayer feedforward networks are universal approximators*, Neural Networks **2** (1989) 359, ISSN: 0893-6080, DOI: 10.1016/0893-6080(89)90020-8 (cit. on p. 125).

[136] M. D. Richard and R. P. Lippmann, *Neural network classifiers estimate Bayesian a posteriori probabilities*, Neural computation **3** (1991) 461, DOI: 10.1162/neco.1991.3.4.461 (cit. on p. 126).

Summary

"Mysterious things – when you were a child you used to think: 'They exist for sure!', but all of a sudden it changes to, 'It would be nice if they existed.' I wonder when and why you stop believing."

— Kozue Amano, "Akari Mizunashi", *ARIA*

Nature's mysteries

Since the beginning of history, humans have always been fascinated by Nature and trying to solve all its mysteries. We make observations, find patterns in them, and then create theories that we hope could describe and predict Nature. But more often than not, solving a mystery would only lead us to discover even more mysteries. Yet, this seemingly never ending dialectic cycle of theory – problem – solution – new theory has always been how we advanced our understanding of the world. A problem is but an opportunity for us to break through and expand our knowledge.

Today, probably the best theory we have to describe the physical world is the Standard Model (SM) of particle physics. The SM is a theory that postulates a set of elementary particles (Figure S.1) and uses their interactions to accurately describe many observed phenomena. It has withstood the scrutiny of countless experiments to date. Nevertheless, there are still many unsolved mysteries of Nature that the SM could not explain, including:

- Why is there so much more matter than antimatter in the universe? (the matter-antimatter asymmetry problem)
- What is dark matter – the invisible matter in our universe that can only be "observed" indirectly through gravitational phenomena? (the dark matter problem)
- Within the same theoretical framework that the SM is based on (quantum field theory), what is gravity? (the quantum gravity problem)
- Why is gravity so much weaker than the other fundamental forces? Or related to this, why is the Higgs boson so light compared to the Planck mass? (the hierarchy problem)

- Why are the neutrino masses so much lighter than the other matter particles in the SM? (the neutrino mass problem)

Unable to address these problems, the SM simply could not be a complete theory in describing Nature. Therefore, physicists have been trying to extend the SM to create the so-called beyond-the-Standard-Model (BSM) theories. There are many different ways that the SM can be extended, and in order to know whether we are going in the right direction, we need experimental data to confirm or constrain our many BSM theories.

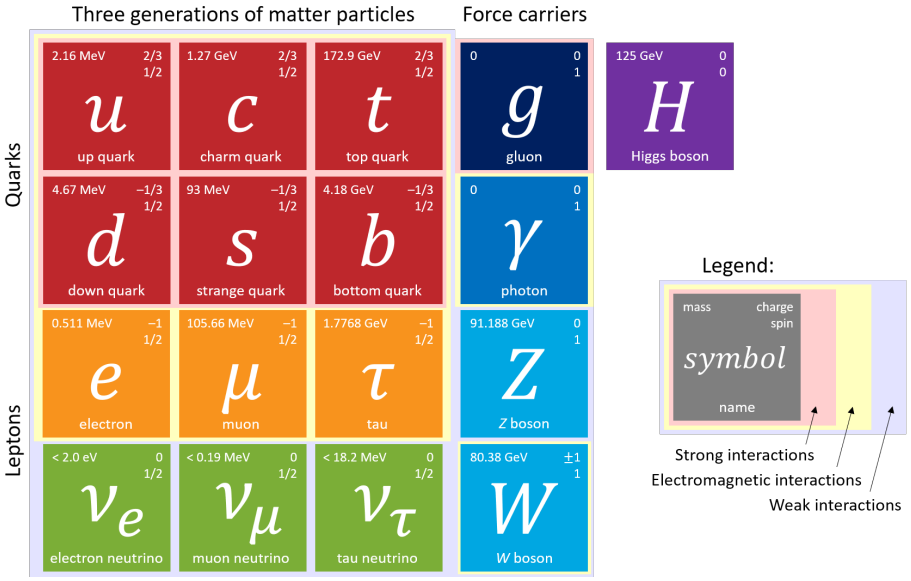


Figure S.1.: Elementary particles in the Standard Model of particle physics.

Lepton flavour violation

In the SM, there are three generations of matter particles. In particular, leptons (a class of particles in the SM) come in three different flavours: the electron (e), the muon (μ) and the tau (τ) lepton. According to the SM, the number of leptons in each flavour remains the same before and after any interactions. For example, a Z boson (a massive neutral particle) can decay into a muon and an antimuon ($Z \to \mu\mu$), but never into a muon and an anti-τ lepton ($Z \to \mu\tau$) (Figure S.2). Violation of this rule is known as lepton flavour violation. Any observation of lepton flavour violation would certainly constitute an exciting sign of BSM physics.

There are many BSM theories that permit lepton flavour violation. These theories are usually motivated by the unsolved mysteries mentioned above. For example, theories with heavy neutrinos are able to explain the smallness of the SM neutrino masses, while

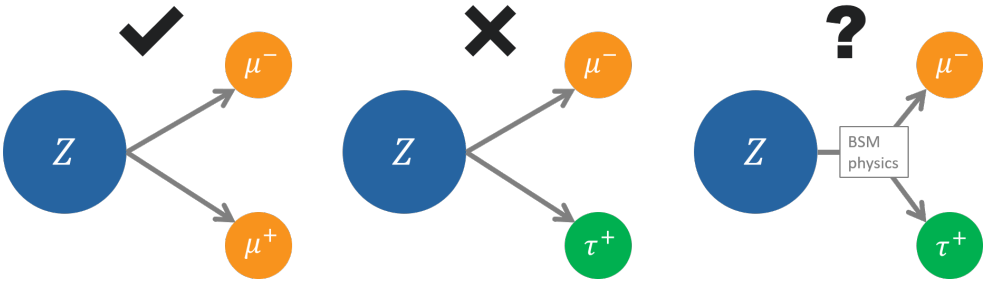


Figure S.2.: Examples of processes that are allowed (left) or forbidden (middle and right) in the Standard Model of particle physics.

predicting an observable occurring probability of lepton-flavour-violating (LFV) processes. Therefore, by searching for LFV processes in experiments, we can either verify or constrain these theories, and guide ourselves towards finding the right missing pieces of New Physics.

The search for $Z \to \ell\tau$ decays with the ATLAS detector

In this thesis, a search for LFV $Z \to \ell\tau$ decays ($\ell = e$ or μ) using data collected by the ATLAS detector at the Large Hadron Collider (LHC) is presented. At the LHC, protons are accelerated to nearly the speed of light and made to collide with each other. These high-energy proton–proton collisions are able to produce heavy particles that are not seen in everyday life. These particles decay very quickly into lighter particles, which are then detected by the ATLAS detector. The Z boson is one of the heavy particles that are produced. During the data-taking periods in 2015–2018, approximately eight billion Z bosons have been produced by the LHC at the ATLAS detector. Such a large sample size allows us to search for LFV $Z \to \ell\tau$ decays even if they only occur very rarely.

The ATLAS detector is a general-purpose particle detector. It allows physicists to reconstruct and identify most of the SM particles (except neutrinos) from the collision data it collects. Using the reconstructed energies and momenta of the identified particles, and by exploiting expected differences between $Z \to \ell\tau$ decays (signal) and events from SM processes (background), potential signal events are selected and separated from the background events. Events reconstructed with an electron or muon and a τ lepton that decayed into hadrons are selected for analysis. These selected events are then classified by neural network classifiers based on how similar their kinematic properties are to a possible signal event.

The selected observed events are then compared with predictions from the background-only or background-plus-signal models. These predictions are based on simulations as well as data of observed background-like events. Maximum-likelihood fits to the observed neural network output distributions are performed to determine the overall yields of the

signal events and major background events in the models. Finally, the compatibility between the data and the models are assessed with hypothesis tests using a likelihood ratio as test statistic.

Results

To the author's slight disappointment, no statistically significant evidence of LFV $Z \to \ell\tau$ decays is found. The observed and expected neural network output distributions are shown in Figure S.3. Nonetheless, the analysis is still able to pose stringent constraints on the occurring probability of $Z \to \ell\tau$ decays, which are meaningful in constraining BSM theory candidates. Upper limits on the LFV branching fraction (fraction of Z bosons which decay by an LFV decay mode with respect to the total number of Z bosons) are set to be 8.1×10^{-6} and 9.5×10^{-6} for the $Z \to e\tau$ and $Z \to \mu\tau$ decays, respectively, at 95% confidence level. These limits are currently the most stringent experimental limits on the decays, which superseded the formerly most stringent limits set by experiments at the Large Electron-Positron Collider (LEP) more than two decades ago.

The success of the analysis is founded on both the amount of collected data and the analysis techniques used. An especially important analysis technique is the use of multiple neural network classifiers to optimally separate the signal and different backgrounds, which is a relatively novel approach that has not been exploited in similar searches before.

Outlook

Looking into the future, the result presented in this thesis is likely just the first of many more exciting results to come.

With the current analysis techniques, the sensitivity of the search of $Z \to \ell\tau$ decays is still primarily limited by statistical uncertainties due to the limited amount of data, instead of systematic uncertainties in the predictions of signal and background events. This implies that with more data collected in the future by the LHC and the ATLAS detector, we will be able to search for LFV $Z \to \ell\tau$ decays with an even higher, unprecedented sensitivity.

Moreover, in the present analysis, only events where the τ lepton decays into hadrons are analysed. It is possible that the analysis can be further improved by considering events where the τ lepton decays into lighter leptons as well. In fact, there are already ongoing efforts within the ATLAS collaboration to try to search for $Z \to \ell\tau$ decays in these events. Significant improvements in sensitivity can be expected when these searches are combined with the present analysis.

To conclude, we are now at the beginning of a new era, where the LHC and the ATLAS experiment have opened up new opportunities for lepton flavour violation searches. While this thesis has proudly set new stringent limits on the interesting LFV $Z \to \ell\tau$ decays, even more exciting discoveries might be just ahead of us.

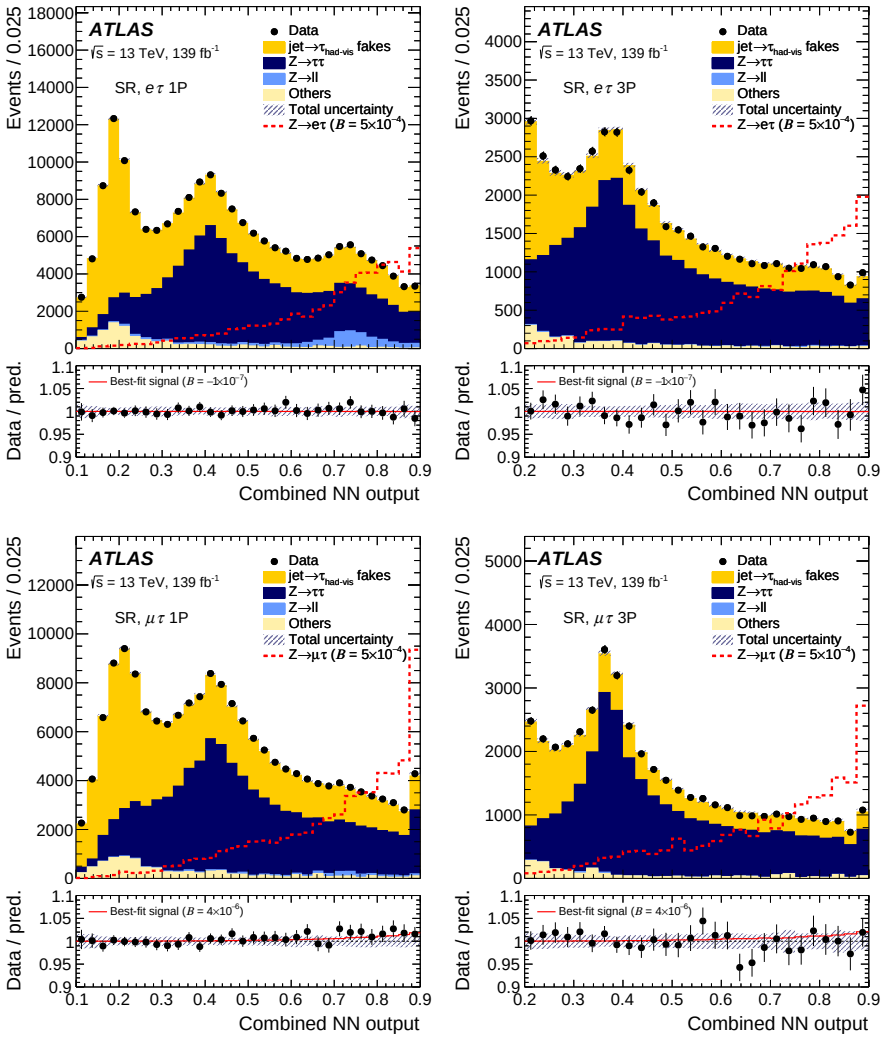


Figure S.3.: Best-fit expected and observed distributions of the combined neural network (NN) output in the signal regions (SR) of the $e\tau$ (top row) and $\mu\tau$ (bottom row) channels. In the lower panel of each plot, the ratios of the observed yields (dots) and the best-fit background-plus-signal yields (solid red line) to the best-fit background yields are shown. The hatched error bands represent the combined statistical and systematic uncertainties. The last bin in each plot includes overflow events.

Samenvatting

"Mysterieuze dingen – toen u een kind was, dacht u altijd: 'Ze bestaan zeker!', maar plotseling verandert het in, 'Het zou mooi zijn als ze zouden bestaan.' Ik vraag me af wanneer en waarom je stopt met geloven."

— Kozue Amano, "Akari Mizunashi", *ARIA*

De raadsels van de Natuur

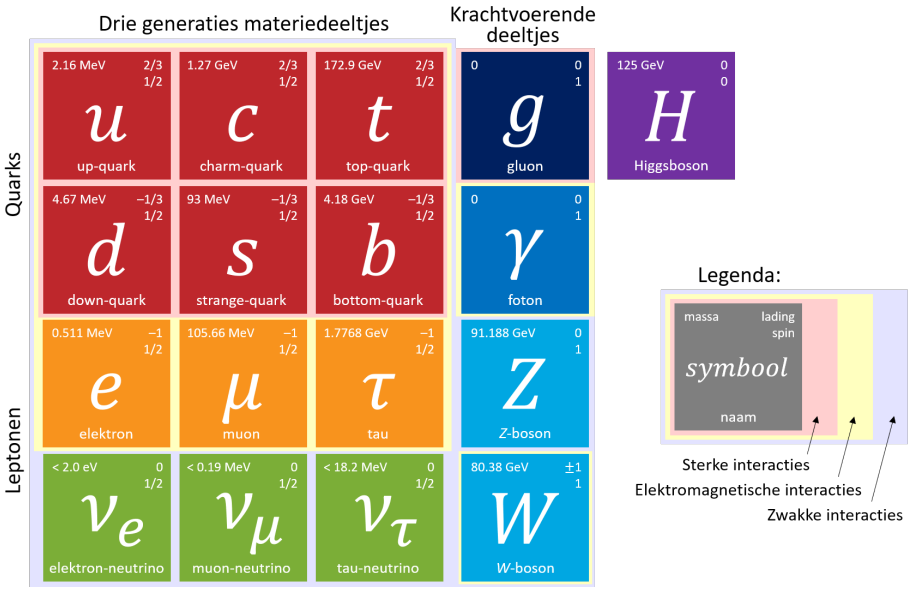
Sinds het begin van de geschiedenis is de mensheid al gefascineerd met de Natuur en het oplossen van haar raadsels. We observeren, ontdekken patronen, en creëren vervolgens theorieën die hopelijk de natuurwetten kunnen beschrijven en voorspellen. Maar vaker dan niet leidt het onthullen van het eerste mysterie slechts tot nieuwere mysteries. En toch is deze dialectische cyclus van theorie – probleem – oplossing – nieuwe theorie altijd de manier geweest hoe wij onze kennis van de aarde ontwikkelen. Elk probleem is slechts een kans voor een nieuwe doorbraak die onze kennis verbreedt.

De beste theorie voor het beschrijven van de fysische wereld is momenteel het Standaardmodel (SM) van de deeltjesfysica. Het SM postuleert een set van elementaire deeltjes (Figuur S.1) en gebruikt hun interacties om vele geobserveerde fenomenen nauwkeurig te kunnen beschrijven. De theorie is gevalideerd door talloze experimenten. Toch blijven er onopgeloste raadsels in de Natuur die het SM nog niet heeft kunnen verklaren, zoals:

- Waarom is er zoveel meer materie dan antimaterie in het universum? (*the matter-antimatter asymmetry problem*)
- Wat is donkere materie – de onzichtbare materie in ons universum dat alleen indirect "geobserveerd" kan worden middels de zwaartekracht? (*the dark matter problem*)
- Binnen hetzelfde theoretisch raamwerk waarop het SM is gebaseerd (kwantumveldentheorie), wat is zwaartekracht? (*the quantum gravity problem*)
- Waarom is zwaartekracht zoveel zwakker dan andere fundamentele krachten? Of gerelateerd, waarom is het Higgsboson zo licht in vergelijking tot de Planckmassa? (*the hierarchy problem*)

- Waarom zijn neutrinos zoveel lichter dan de andere massieve deeltjes in het SM? (*the neutrino mass problem*)

Omdat zij niet in staat is deze problemen op te lossen, kan het SM simpelweg niet een complete theorie zijn voor het beschrijven van de Natuur. Daarom proberen natuurkundigen nu het SM uit te breiden om zogeheten theorien voor fysica buiten het Standaardmodel (BSM) te creëren. Er zijn veel verschillende manieren waarop het SM kan worden uitgebreid, en om te weten of we in de juiste richting gaan is experimentele data nodig of deze BSM theorieën te bevestigen of uit te sluiten.

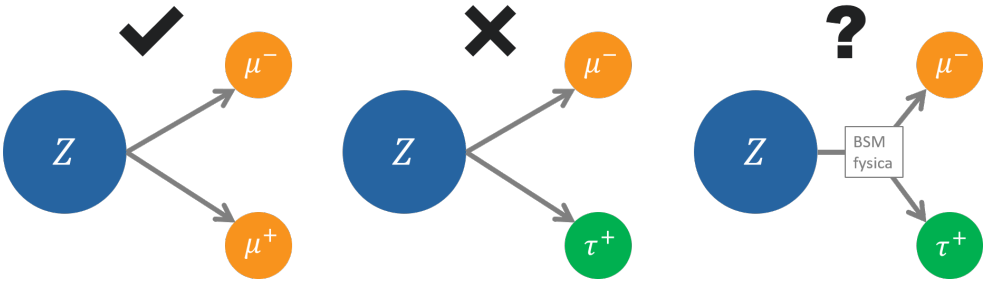


Figuur S.1.: Elementaire deeltjes in het Standaardmodel van de deeltjesfysica.

Lepton flavour violation

In het SM zijn er drie generaties van materiedeeltjes. Leptonen (een klasse van deeltjes in het SM) bestaan in drie verschillende smaken: het elektron (e), het muon (μ) en het tau (τ) lepton. Volgens het SM blijft het aantal leptonen van elke smaak constant voor en na iedere interactie. Een Z boson (een massief ongeladen deeltje) kan bijvoorbeeld vervallen in een muon en een antimuon ($Z \to \mu\mu$), maar nooit in een muon en een anti-τ lepton ($Z \to \mu\tau$) (Figuur S.2). Overtreding van deze regel staat bekend als *lepton flavour violation* (lepton smaak schending). Iedere observatie van lepton flavour violation zou met zekerheid een teken zijn van BSM fysica.

Er zijn vele BSM theorieën die lepton flavour violation toelaten. Deze theorieën zijn normaliter gemotiveerd door de onopgeloste raadsels die eerder zijn genoemd. Als



Figuur S.2.: Voorbeelden van processen die zijn toegestaan (links) of verboden (midden en rechts) in het Standaardmodel van de deeltjesfysica.

voorbeeld: theorieën met zware neutrinos zijn in staat om de lage massas van de SM neutrinos te verklaren, en voorspellen tegelijkertijd een meetbare waarschijnlijkheid van lepton-flavour-violating (LFV) processen. Dus door het zoeken naar LFV processen middels experimenten is het mogelijk om deze theorieën te bevestigen of uit te sluiten en onszelf te gidsen naar de juiste missende puzzelstukjes van de Nieuwe Fysica.

De zoektocht naar $Z \to \ell\tau$ verval met de ATLAS detector

In did thesis is een zoektocht daar LFV $Z \to \ell\tau$ verval ($\ell = e$ of μ) in data verzameld door de ATLAS detector bij de Large Hadron Collider (LHC) gepresenteerd. Bij de LHC worden protonen versneld tot bijna de lichtsnelheid om vervolgens op elkaar te botsen. Deze hoog-energetische proton–proton botsingen zijn in staat om zware deeltjes te produceren die niet in het gewone leven gezien worden. Deze deeltjes vervallen erg snel in lichtere deeltjes, die vervolgens gedetecteerd worden door de ATLAS detector. Het Z-boson is een van de zwaarste deeltjes dat geproduceerd kan worden. Gedurende de data-opname periode in 2015–2018 zijn ongeveer acht miljard Z-bosonen geproduceerd door de LHC bij de ATLAS detector. Een dergelijke dataset kan worden benut om te zoeken naar LFV $Z \to \ell\tau$ verval zelfs als deze maar zeer zeldzaam optreedt.

De ATLAS detector is een deeltjesdetector voor algemeen gebruik, in staat om de deeltjes van het SM (met uitzondering van muonen) te reconstrueren en identificeren door middel van de botsingdata die wordt verzameld. Met de gereconstrueerde energieën en impulsen van de geïdentificeerde deeltjes, en door de verschillen tussen $Z \to \ell\tau$ verval (signaal) en processen van SM processen (achtergrond) te benutten kunnen potentiële signaalevenementen geselecteerd en gescheiden worden van achtergrond. Evenementen gereconstrueerd met een elektron of muon en een τ lepton dat hadronisch vervalt zijn geselecteerd voor deze analyse. Deze geselecteerde evenementen worden dan geclassificeerd door een neuraal netwerk gebaseerd op hoe overeenkomend hun kinematische eigenschappen zijn ten opzichte van een mogelijk signaalevenement.

De geselecteerde geobserveerde evenementen worden dan vergeleken met voorspellingen uitgaande van modellen met slechts achtergrond danwel achtergrond plus signaal. Deze voorspellingen worden gebaseerd op simulaties en op data van geobserveerde achtergrondachtige evenementen. Fits van de meest aannemelijke schatter, fittend op de distributies geproduceerd door het gebruikte neuraal netwerk bepalen de totale hoeveelheid aan signaalevementen en achtergrondevenementen in de modellen. Ten slotte wordt de compatibiliteit tussen de data en de modellen bestudeerd middels het testen van hypotheses, met als schatter de ratio van de meest aannemelijke schatters van de twee modellen.

Resultaten

Tot de milde teleurstelling van de auteur is geen statistisch significant bewijs gevonden voor LFV $Z \to \ell\tau$ verval. De verwachte en geobserveerde distributies geproduceerd door het neuraal netwerk zijn getoond in Figuur S.3. Echter is de analyse wel in staat om sterke limieten te stellen aan de waarschijnlijkheid van het optreden van $Z \to \ell\tau$ vervallen, wat betekenisvol bijdraagt aan het uitsluiten van BSM theorie-kandidaten. Bovengrenzen op de LFV vertakkingsfractie (fractie van de Z-bosonen dat vervalt via een LFV vervalsmodus tegen het totaal aantal Z-bosonen) zijn gezet op 8.1×10^{-6} en 9.5×10^{-6} voor de $Z \to e\tau$ en $Z \to \mu\tau$ vervallen respectievelijk, met een 95% betrouwbaarheidsinterval. Deze limieten zijn momenteel de sterkste experimentele limieten op deze vervallen, en zijn een verbetering op de vroegere meest sterke limieten gesteld door de Large Electron-Positron Collider (LEP) meer dan twee decennia geleden.

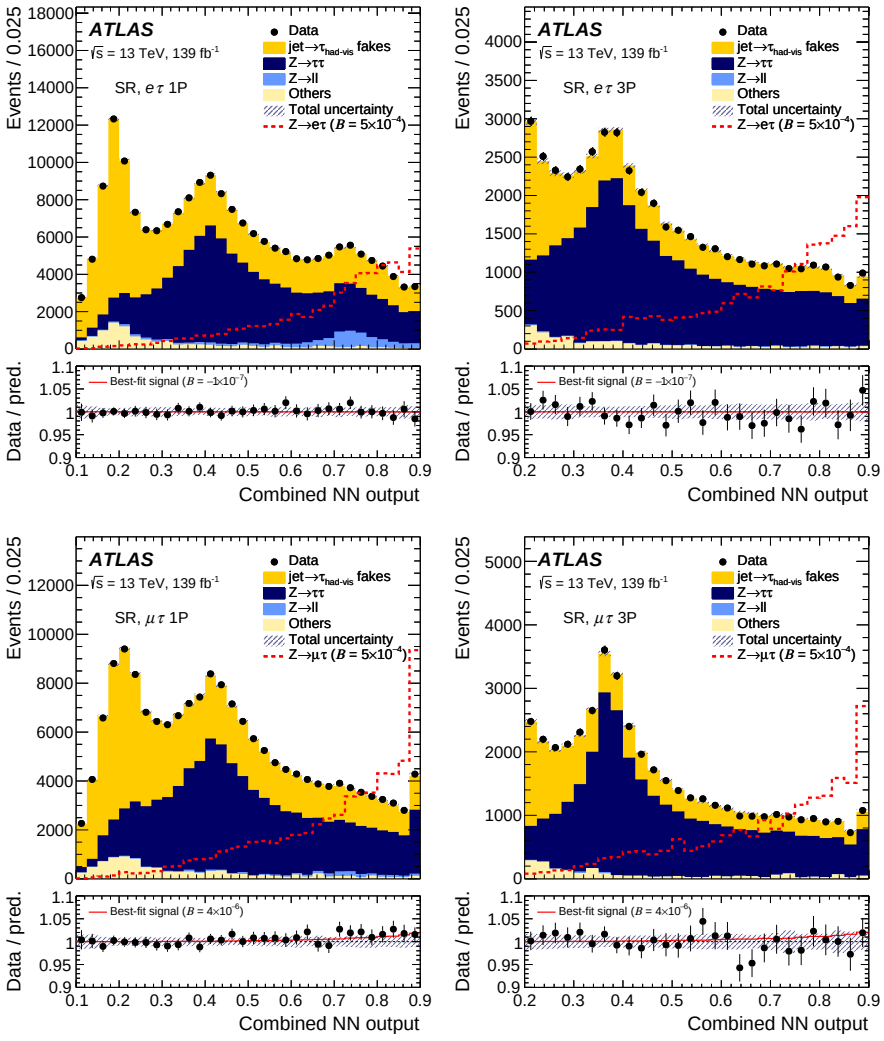
Het success van de analyse rust op zowel de hoeveelheid verzamelde data en de analysetechnieken die zijn gebruikt. Een belangrijke analysetechniek is het benutten van meerdere neurale netwerkclassificeerders om optimaal het signaal van de verschillende achtergronden te scheiden, hetgeen een relatief nieuwe aanpak is die nog niet in vergelijkbare zoektochten is gebruikt.

Uitzicht

Vooruit kijkend is het resultaat gepresenteerd in dit thesis is waarschijnlijk slechts de eerste van vele interessante resultaten die nog zullen volgen.

Met de huidige analysetechnieken is de gevoeligheid van de zoektocht daar $Z \to \ell\tau$ vervallen vooral gelimiteerd door statistische onzekerheden volgend uit de gelimiteerde beschikbaarheid van data, en niet door systematische onzekerheden op de voorspellingen van het aantal signaal- en achtergrondevenementen. Dit impliceert dat met het verzamelen van meer data in de toekomst van de LHC en de ATLAS detector, het mogelijk wordt om te zoeken naar LFV $Z \to \ell\tau$ vervallen met een nog grotere, ongeëvenaarde gevoeligheid.

Ook is in de huidige analyse alleen aandacht besteed aan evenementen waar het τ lepton vervalt naar hadronen. Het is mogelijk dat de analyse verder verbeterd kan worden door ook evenementen mee te nemen waar het τ lepton vervalt naar lichtere leptonen. En



Figuur S.3.: Beste fit op de verwachte en geobserveerde distributies van de gecombineerde neuraal netwerk (NN) resultaten in de signaalgebieden van de $e\tau$ (boven) en $\mu\tau$ (onder) kanalen. De panelen onder elke plot tonene de ratios van de geobserveerde aantallen (punten) en de beste-fit van de achtergrond plus signaal (rode lijn) ten opzichte van de beste fit op de achtergrond. The gestreepte foutbanden tonen de combinatie van statistische en systematische onzekerheden. De laatste bin in elke lot bevat de overflow-evenementen.

er zijn inderdaad binnen de ATLAS collaboratie al projecten op weg om de zoektocht naar $Z \to \ell\tau$ vervallen in deze evenementen voort te zetten. Significante verbeteringen in gevoeligheid kunnen verwacht worden wanneer deze zoektochten gecombineerd worden met de huidige analyse.

Ter conclusie: we bevinden ons in het begin van een nieuw tijdperk, waar de LHC en het ATLAS experiment nieuwe mogelijkheden hebben geopend om lepton flavour violation te zoeken. Alhoewel dit thesis met trots nieuwe nauwkeurige limieten aan de interessante LFV $Z \to \ell\tau$ vervallen heeft gesteld, zullen nog veel meer fascinerende ontdekkingen in het verschiet liggen.

Acknowledgements

If you were an ice cream flavour, you would be my favourite one.

— Mindy Gledhill, *All About Your Heart*

So, this is it. This is the end of the journey of Terry Wing Sheung Chan as a physics PhD student. It was an amazing adventure full of new experiences and challenges, and I genuinely enjoyed every moment of it. Studies have shown that people perceive time to be slower when they experience new things, which is probably why the beginning of my four-year study seems so distant to me now. Along the journey, I have met awesome people and was given a lot of support and joy. Here, I would like to thank everyone who has made this journey so extraordinary and unforgettable.

The first person that I must thank was the person who brought me to Nikhef to start my PhD life in the first place. Olga, thank you for admitting me and giving me this awesome research topic. Sorry that I wasn't able to show you this thesis and its results earlier. You had taught me many, many things, not only about physics research, but also about life. Your enthusiasm, positivity, perseverance, braveness and love for life will forever have my admiration and respect.

Another special thanks goes to my promoter and the manuscript committee. Nicolo, thanks for being a responsible supervisor in the last year of my study. Prof. Ronald Kleiss, Prof. Stan Bentvelsen, Prof. Raymond Snellings, Dr Sascha Caron and Dr Mengqing Wu, thanks for reading through my thesis critically and providing me with constructive feedbacks to improve it.

I also want to thank the wonderful team without which I could not have delivered the analysis in such quality. Stefania, thank you for taking on the role of being my copromoter and daily supervisor, and helping to push the paper towards approval and publication. Daniele, thanks for all the assists and supports you have given me. I have learned a lot from you. Ann-Kathrin, thank you for taking care of the leptonic tau channels. Sorry that I had to let you do all the work alone using the messy analysis framework I left you with. I am still very excited to see what we can achieve with these channels.

I must also express my gratitude to the people at CERN and in the ATLAS collaboration who have given the many assists I needed. Many thanks to Noam Hod, Luca Fiorini and Antonio De Maria for being the editorial board of my paper and critically reviewing my

analysis and writings. Huge thanks to Bertrand Martin dit Latour, Will Davey, Pier-Olivier Deviveiros and Christian Grefe for providing all the help to the analysis and also my projects in the ATLAS tau working group.

There are many persons from Nikhef that I would like to thank. Nikhef is truly an extraordinary institute. Everyone I met is brilliant, enthusiastic, friendly, open-minded and inspiring. I never regretted coming to Nikhef from far away because of every one of you:

Matteo, there are a lot I have to thank you for. Thanks for asking me to be your paranymph - I feel really honoured. Thanks for often inviting me to your house parties. And when you were still at Nikhef, thanks for always inviting me to coffee breaks and telling me interesting stories of yours during them.

Pepijn, I enjoyed the time I shared with you in all the summer schools and workshops we have been to together. Sorry that I am not a talkative person, and I know sometimes it must be difficult to stand the uncomfortable silence with me, but thanks for still sticking with me.

Birgit, thanks for being my office mate for my last two year at Nikhef (excluding the last months when we had to work from home due to the pandemic). Also thank you for inviting me to your sinterklaas parties. They are super nice and a ton of fun.

Let me also thank my first-year office and foosball mates: Broos, Marc, Kees, Milo, Daan, Wouter. It was so much fun playing foosball and chatting with you guys (or listening to you guys chat) in all the lunch and coffee breaks.

Alice, Michiel, Rahul, Ashley, Jordy, Peter, Federica, Anamika, Brian, although we didn't spent much time together, I enjoyed being with you guys a lot. And for those of you who were at Birgit's place when we played the drawing game (Scrawl), I will remember all the stupid drawings and the great laughs that we had.

Edwin, I am very glad that you, another Hongkonger, has decided to join the Nikhef ATLAS group. It feels much more at home with someone in the group I can chat with using my mother tongue. Thanks for going to dinners with me sometimes when I needed to unwind and vent out my frustrations after some long, stressful days at work. Let's go discover more (Japanese) restaurants in Amsterdam after the pandemic.

Ka Wa, thanks for coming to Nikhef together with me, even though you deserted high-energy physics and joined the gravitational-wave group instead. Also thank you for carrying me (while trolling Edwin) in our League of Legends games. I am grateful to have met you during our Master's. I hope our paths will cross once again in the future.

To my family in Hong Kong, thanks for supporting my choice to come to Nikhef and pursue my passions, even though that means we have to be separated and cannot see each other every day. Dad, you worried about my thesis progress more than anyone else, including my supervisors and myself. But here you are, finally, you are holding it in your hands. Thanks for the time you came all the way to Amsterdam just to cook some nice meals for me. Mum, thanks for taking care of all the annoying little things for me that I couldn't do from abroad, and sorry for not contacting you more often. I know my decision to leave Hong Kong was a rather sudden one and might have shocked you a bit, but you were still very supportive. Thanks for always worrying and caring about me. Cindy,

thanks for visiting me during New Year. I loved our trip to Austria. To all of you, just know that I will always be fine wherever, and will always love you.

My dear friends from Hong Kong, thanks for doing everything you can to protect my homeland, while I was away focusing on my research. I wish I could have done more. There were some very difficult times, when I was genuinely depressed, could not sleep well and could not concentrate at work after hearing all the saddening news. But chatting with you guys had always made me feel better and a bit more positive. You and the future of Hong Kong will always be in my prayers. Hold on and take care.

Sira, Ira, Jeffrey, Kit, Earl, Peter, Dennis, thanks for welcoming me every time I return to Hong Kong, and finding time to entertain me when I craved for some board game gatherings.

Jason, thanks for delaying my thesis writing progress by inviting me to play RO in the most critical times. You are the best bad company.

Siu Ming, thanks for visiting me in Amsterdam. Travelling with you is fun. Let's go visit somewhere else together someday.

A very special thanks to Prof. MC Chu, who brought me into the field while I was still a Master's student. I would not have had the opportunity to pursue a PhD in high-energy physics without you. You have my full respect for your efforts in promoting fundamental science in Hong Kong, which I know is in no way an easy task. It is unfortunate that I missed all the opportunities to meet with you every time you visited Amsterdam or Geneva. I will let you strangle me next time I am back in Hong Kong.

Online e-book:

https://www.publicatie-online.nl/publications/wing-sheung-chan/