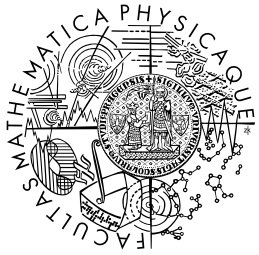


FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

DOCTORAL THESIS

Kristina Mihule

Studies of Higgs boson in its decay to a pair of tau-leptons and search for excited tau-lepton with the ATLAS detector

Institute of Particle and Nuclear Physics

Supervisor of the doctoral thesis: Doc. RNDr. Tomáš Davídek, PhD.

Study programme: Subnuclear Physics

Study branch: P4F9

Prague 2022

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

I thank my supervisor Tomás Davídek. I am grateful for your knowledgeable advice and versatile support throughout PhD years. I am happy to have prepared this project under your wise guidance.
I thank all the people who advised me on scientific and technical subjects. Special thanks to Daniel Scheirich and Vojtěch Pleskot, who generously shared their expertise with me.
Thank you to the colleagues with whom I worked in ATLAS.
I acknowledge the inspiration from people who barely know that they inspired me.
All my gratitude goes to my mother. Thanks to my loved ones, family, and friends.

Title: Studies of Higgs boson in its decay to a pair of tau-leptons and search for excited tau-lepton with the ATLAS detector

Author: Kristina Mihule

Institute: Institute of Particle and Nuclear Physics

Supervisor: Doc. RNDr. Tomáš Davídek, PhD., Institute of Particle and Nuclear Physics

Abstract: The Standard Model (SM) describes known elementary particles and their interactions. The ATLAS experiment is located at the Large Hadron Collider (LHC) in CERN (Geneva). The ATLAS detector measures data from proton-proton collisions at high centre-of-mass energy. The collected data allow physicists to precisely study SM and look for evidence of particles beyond the Standard Model.

The presented thesis describes the analysis of Higgs boson decay to a pair of tau-leptons. The measurement of the Higgs boson production cross section is based on the data set collected by the ATLAS detector in proton-proton (pp) collisions at a centre-of-mass energy of 13 TeV in the years 2015-18 (so-called Run 2). This dissertation presents the update of the algorithm of Higgs boson mass reconstruction for the Run 2 ATLAS conditions. The evaluation of fake tau-lepton background is also discussed.

Further, this dissertation is dedicated to the search for excited tau-leptons in the Run 2 ATLAS data set. Contact interaction of the excited tau-leptons is assumed, which would occur if fermions were composite rather than elementary. The analysed ATLAS data from pp collisions in Run 2 set do not contain signatures of excited tau-leptons production. Within the considered model with the compositeness scale $\Lambda = 10$ TeV, the upper limit on the excited tau-lepton mass is set to 2 TeV.

Keywords: Higgs boson, Higgs boson to tau-leptons, excited tau-lepton, di-tau, Missing Mass Calculator, SM, Standard Model, BSM, Beyond the Standard Model, ATLAS, Large Hadron Collider

Contents

Introduction

The Large Hadron Collider (LHC) in the European Organisation for Nuclear Research (CERN) performs particle collisions at unprecedentedly high energies. Proton-proton collisions result in the production of various types of elementary particles. Interaction between them is described by the Standard Model (SM), which is one of the most precise models constructed by humankind up to now.

The Standard Model triumphed in 2012 when the Higgs boson was discovered by the ATLAS and CMS experiments at the LHC [1, 2]. The first evidence of the Higgs boson was established with the data collected in proton-proton collisions at 7 TeV and 8 TeV centre-of-mass energy. The most significant excess was found in $H \to \gamma\gamma$, $H \to ZZ$, and $H \to WW^*$ decay modes. Since the Higgs boson discovery, several measurements at the LHC confirmed that the Higgs boson couples to other SM particles as predicted. For instance, evidence of Higgs boson decays to two tau-leptons was claimed by the ATLAS experiment based on the analysis of Run 1 ATLAS data set [3].

The second data-taking period at the LHC, Run 2, from 2015 to 2018, gathered a data set of about 150 fb^{-1} delivered luminosity. This data allowed physicists to measure the Higgs boson production cross section in $H \to \tau\tau$ decays. In 2018, the ATLAS Collaboration published the result of the analysis of data from 2015 and 2016 [4]. With the complete Run 2 data set, the analysis team could claim the production cross section with higher significance and smaller uncertainty [5]. One part of this thesis describes the author's participation in measuring the Higgs boson production cross section in its decay to two tau-leptons.

The Standard Model explains the majority of interactions between known particles. There are, however, indications that the Standard Model is not a final theory of fundamental interactions. Among them, there are tensions between the theoretical prediction and experimental finding on electron's and muon's anomalous magnetic momentum [6], [7]. Together with hints at lepton flavour violation as measured in meson decays [8, 9, 10], this experimental observation can be explained by fermion compositeness. The data collected by the ATLAS detector can reveal physics Beyond the Standard Model (BSM). Specifically, lepton compositeness could explain observations. The compositeness could be a result of the contact interactions. The contact interaction hypothetically occurs between four fermions. It could manifest itself with the existence of excited fermion states. Another part of the presented thesis describes the author's work towards the search for excited tau-leptons in the ATLAS Run 2 data.

This thesis is composed as follows. Firstly, I introduce the reader to the Standard Model of particle physics and outline the questions SM theory faces nowadays. Secondly, I describe the physics of proton-proton collisions in high-energy physics, the ATLAS experiment at the LHC, and Monte Carlo simulation in particle physics. Thirdly, I familiarise the reader with the event reconstruction and statistical interpretation of data. In a separate section, I outline the method for evaluating the invariant mass of di-tau. The next chapter informs the reader about my contribution to measuring the Higgs boson production cross section in Higgs boson decay to a pair of tau-leptons. Eventually, the last chapter describes the search for excited tau-leptons in the ATLAS data set collected in 2015-18.

1. Model of elementary particles

1.1 Standard Model

The Standard Model is a commonly accepted and well-established model of elementary particles and the interactions between them. It owes its name to its high accuracy, and the epithet “standard" stands for its acknowledgement by the wide scientific community. The Standard Model has been very successful in predicting particle existence and properties. Up to now, it finds numerous experimental proofs.

The matter around us consists of atoms. An atom is pictured as a nucleus surrounded by a cloud of the electron field. A nucleus is presented as a mixture of protons and neutrons. There are three known forces which govern interactions between particles at the microscopic level. Electromagnetic interactions take place between electrically charged particles and substances with magnetic properties. It is omnipresent in everyday life: lighting, electrical current, voltage transformers, radioactive γ-decays and many more. A subnuclear-level strong interaction holds parts of atomic nuclei together. The interplay between the electromagnetic and strong interactions leads to radioactive α-decay. Another short-distance interaction – a weak interaction – is responsible for permutations of elementary particles within the atoms and nuclei. The weak interaction can be observed in radioactive β-decays and synthesis reactions on the Sun. Massive objects are subject to gravity. The gravitational interaction, among numerous others, adjusts the Solar system together and causes ebbs and flows of the World's Oceans. Gravity is negligible at the level of elementary particles.

Fermionic matter is described by three generations of leptons and quarks. The Standard Model assumes the existence of seventeen elementary particles (Figure 1.1) [11]. For each matter particle, an antimatter particle exists. Antiparticles have mass identical to particles but have an opposite sign of electric charge or other physical properties (usually quantified by quantum numbers). Fermions which interact electroweakly but not strongly form a group of leptons. Quarks additionally carry colour by means of which they participate in strong interaction.

Electrons, electron neutrinos, up quarks and down quarks make up the first generation of fermions. They are building blocks of the matter around us. Muons, muon neutrinos, strange quarks, and charm quarks are fermions of the second generation. Tau-leptons (τ-leptons), tau-neutrinos, bottom quarks and top quarks are the third-generation fermions. The existence of some second and third-generation particles was predicted by theory. In today's understanding, particles of different generations behave identically in electroweak and strong interactions. However, they differ by mass: i.e. electron is about ~ 200 times lighter than muon and ~ 3500 times lighter than tau-lepton. The hierarchy of masses between generations makes up a question for scientists.

The Standard Model (SM) is built on the idea of local and global symmetry in the elementary particle world [12]. The SM cornerstone idea is the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ gauge symmetry conserved in the strong, weak, and electromagnetic interactions. Flavour-changing currents are mediated by charged

W^+ and W^- intermediate weak bosons. Neutral Z intermediate weak bosons propagate weak neutral currents. Massless photons and gluons mediate the electromagnetic and strong interaction, respectively. The weak and electromagnetic interactions are unified into an electroweak interaction.

According to experimental observations, the strong and electromagnetic interaction preserves its quantum number of colour and electric charge. The non-conservation of parity in weak decays was observed experimentally [13, 14]. However, the properties of weak isospin T and hypercharge Y are not conserved in weak interactions. Only left-handed particles were found to participate in weak processes. Neutrino (antineutrino, an antimatter counterpart of neutrino) has a definitive helicity of -1 (+1) [15]. The SM underpins parity (P) violation by introducing left-handed fermions through $SU(2)$ weak-isospin doublets (L) while right-handed fields remain $SU(1)$ singlets (R):

The strict conservation of the local gauge symmetry prevents the intermediate bosons from acquiring mass. However, the intermediate vector W and Z bosons are massive [16, 17, 18, 19, 20, 21]. Furthermore, the local $SU(2)$ symmetry imposes a zero-mass condition on fermions.

The origin of electroweak bosons and fundamental fermions mass can be explained by their interaction with a complex doublet of scalar fields. This mechanism is called BEH in honour of Rene Brout, François Englert and Peter Higgs. They; and G. Guralnik, C.R. Hagen, and T. Kibble wrote three papers on symmetry breaking [22, 23, 24].

The potential of the 'Higgs" field can be expressed as follows: $V(\phi) = \mu^2\phi^+\phi - \lambda(\phi^+\phi)^2$ where $\lambda > 0$ and μ is a parameter of the Higgs filed self-interaction. If $\mu^2 < 0$, the Higgs potential has its minimum at a non-zero value $\phi\phi^+ = -\mu^2/2\lambda = \upsilon^2/2$ (Figure 1.1). The minimal values of the Higgs potential result in a vacuum expectation value υ.

In this way, the $SU(2)_L \otimes U(1)_Y$ gauge symmetry is spontaneously broken by the vacuum into the $U(1)_{\mathrm{EM}}$ symmetry. The electroweak (EWK) interaction is mediated by a gauge Yang-Mills fields, neutral components of which are mixed. Three intermediate gauge bosons ($W^{\pm}$, Z bosons) acquire mass. The remaining strong $SU(3)_c$ and electromagnetic $U(1)_{\mathrm{EM}}$ gauge symmetries are preserved, which is aligned with the experimental fact of zero gluon's mass and photon's mass. After symmetry breaking, at least one neutral scalar field remains physical according to the Goldstone theorem. A survived scalar is a massive Higgs boson.

The BEH mechanism also explains fermion masses. The interaction between Higgs field H and fermions f is Yukawa-like: $L_{ffH} = g_{ffH}\bar{f}fH$ where $g_{ffH} = -m_f/v$ is a strength of coupling. Therefore, the fermion mass is proportional to its coupling to the Higgs field. In this way, studying third-generation signatures has the potential to achieve higher experimental sensitivity in measuring the Higgs boson properties.

In addition, the introduction of the Higgs field makes the SM theory renormalisable. All processes at a tree level can be described in converged matrix elements. This results in finite cross sections and compliance with unitarity conditions [25].

The evidence of the Higgs boson's existence was observed at the Large Hadron Collider (LHC) in 2012 for the first time [1], [2]. Since that, evidence of the Higgs boson was reaffirmed in multiple channels. The SM Higgs boson properties have been studied with the data collected at the LHC during its Run 2 (2015-18) [5].

This thesis describes the analysis of the Higgs boson production cross section in its decay to tau-leptons.

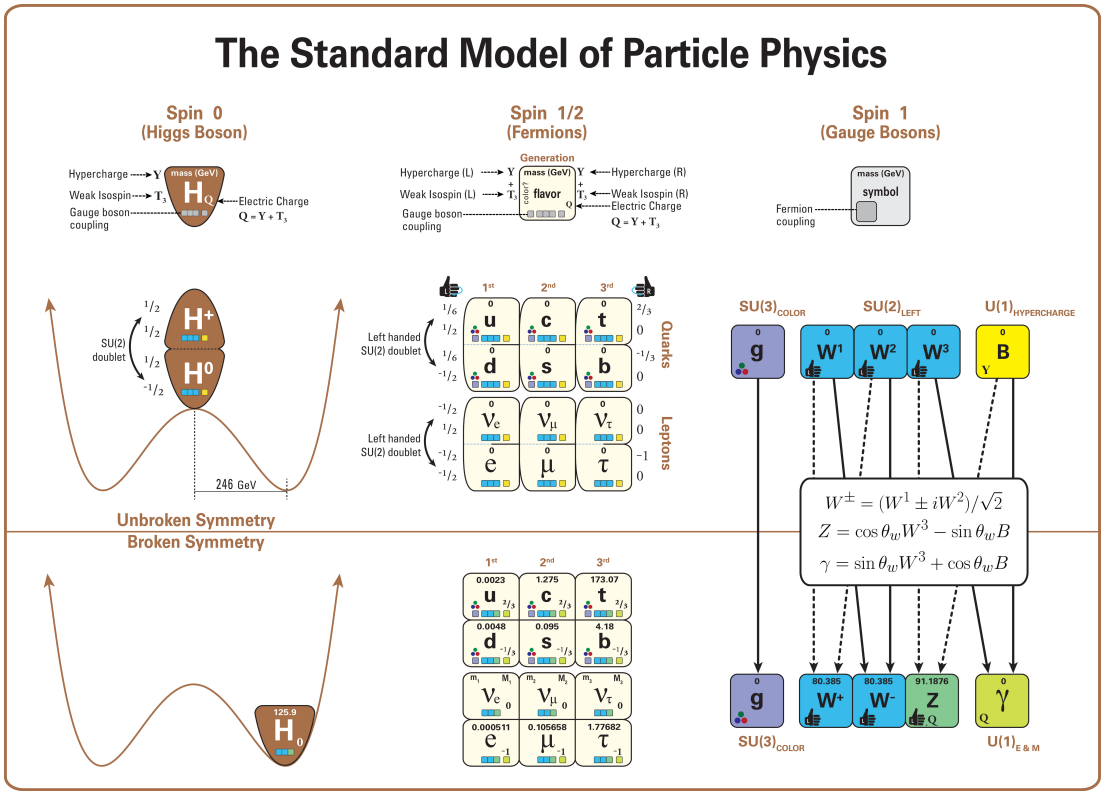


Figure 1.1: The graphical explanation of the Standard Model of elementary particles and the Higgs mechanism [26].

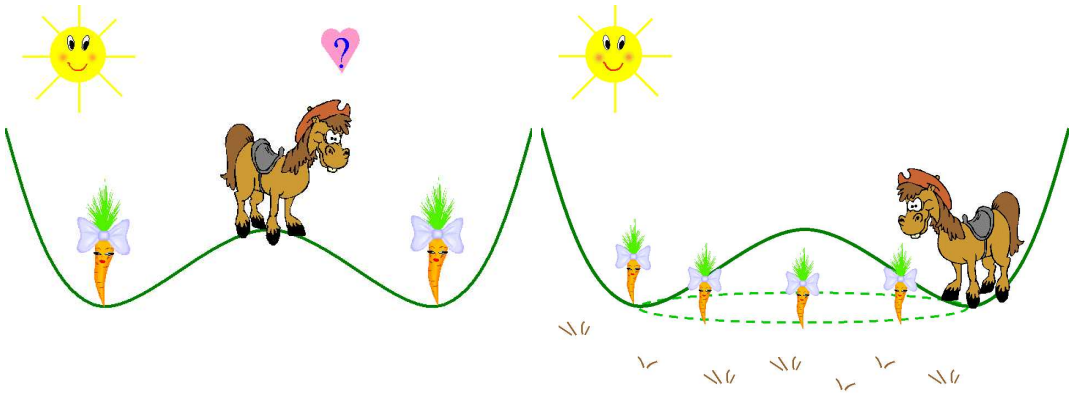


Figure 1.2: Although Nicolás likes the symmetric food configuration, he must break the symmetry to decide which carrot is more appealing. In three dimensions, there is a continuous valley where Nicolás can move from one carrot to the next without effort. Courtesy of Antonio Pich [27].

1.2 Physics Beyond the Standard Model

Despite versatile corroborations, the Standard Model does not find an explanation for a range of experimental observations. Among other findings, there are hints

of lepton flavour violation in heavy meson decays [8, 9, 10] and tension between experimental values of electron and muon anomalous magnetic momentum and their prediction by the SM [6, 28].

Numerous theoretical models aim to eliminate SM shortcomings, to unite all fundamental forces like the Grand Unification Theory (GUT), to explain the mechanism of lepton and baryon number conservation, to describe the physics of the combined charge-parity (CP) violation, to find the reason for the mass difference between generations (models of leptoquarks and composite fermions, supersymmetric models, etc.). The presence of not yet discovered forces would reveal itself in the existence of exotic particles. Altogether, the models describing non-SM processes or extending SM are called **B**eyond the **S**tandard **M**odel (BSM) or new physics.

As mentioned, the existence of the three generations of particles with the mass hierarchy questions the minds of physicists. What is the reason for the mass difference between the generations? Is there a gauge-mediated interaction carrying a lepton number? SM does not answer these questions.

The idea that leptons are composite follows from the observation of the matryoshka pattern: molecule $\rightarrow$ atoms $\rightarrow$ nucleus $\rightarrow$ proton, neutron $\rightarrow$ quark. For that, the search for composite fermions, in particular, is performed with the pp collision data collected by the ATLAS experiment at the LHC. This thesis presents the effort towards finding excited tau-leptons (ETL, τ^*) in the ATLAS Run 2 data set.

2. Proton-proton collisions at the Large Hadron Collider

2.1 The Large Hadron Collider

The Large Hadron collider (LHC) at the European centre of nuclear research (Geneva, Switzerland) is the largest laboratory for particle physics. It provides unique conditions for studying elementary particles and interactions between them with unprecedented accuracy.

The LHC is a circular colliding beam accelerator (Figure 2.1). It realizes high-energy proton-proton (pp) as well as heavy-ion collisions. Physics is probed by four major detectors: multi-purpose ATLAS and CMS experiments, ALICE, and a B-factory LHCb. Besides them, numerous other experiments (TOTEM, LHCf, MoEDAL, and others) are designed for various measurements. Several test beam facilities (such as the North Area and the East Area) provide a crucial resource for the validation of detectors and technologies in real-life situations. The experiments are constructed to study the Standard Model and to search for BSM physics. The data collected at the LHC are used for precision measurements of the SM parameters and for testing various exotics and other new physics theories. Much recent high-energy physics (HEP) measurements are based on the LHC data.

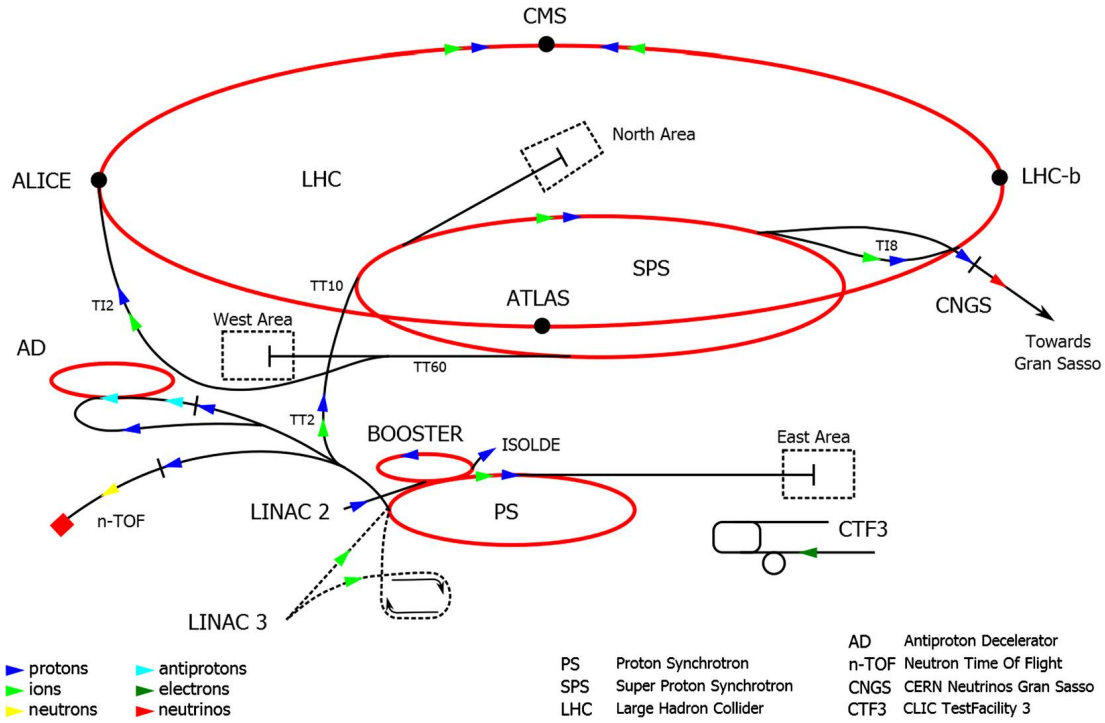


Figure 2.1: The complex of the LHC accelerator and related facilities [29].

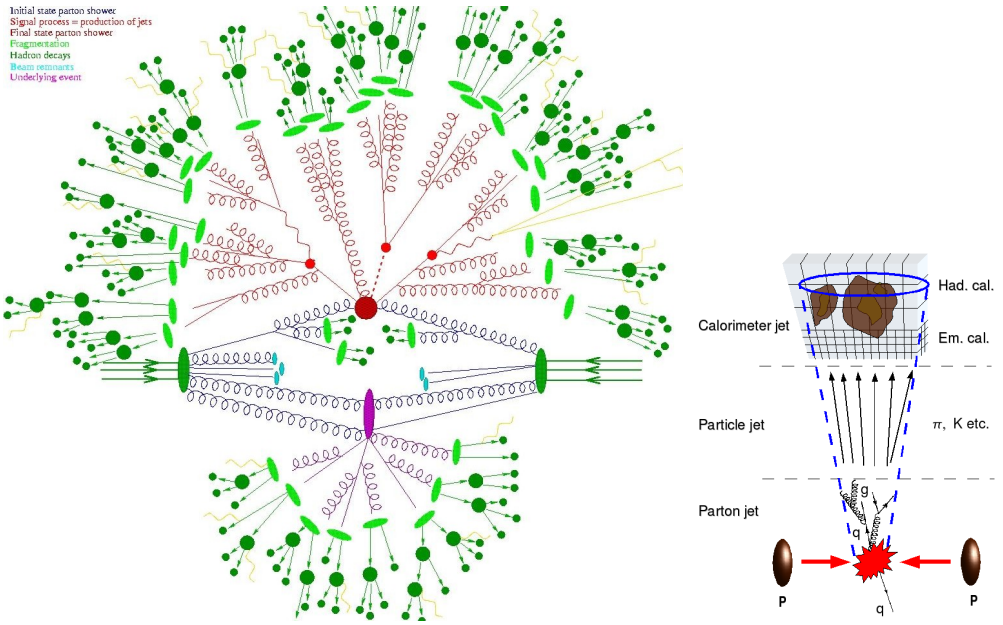
Protons are produced in duoplasmotrons where a high electric field dissociates atoms of hydrogen into electrons and protons. Next, high electromagnetic fields accelerate protons. They are accelerated to 50 MeV energy in **Lin**ear **Ac**celerator

(Linac) by using alternating electric potentials. The Proton Synchrotron Booster gives further momentum to the proton beam and injects a 1.4 GeV beam into the Proton Synchrotron for acceleration up to 26 GeV. The Super Proton Synchrotron serves as a final accelerator (to 450 GeV) before injection of the proton beam into the LHC. The LHC complex consists of a ring with superconductive magnets used for particle beam accelerating, bending, and focusing. In the LHC synchrotron, the proton beam achieves unprecedentedly high energies up to 7 TeV. The LHC provides a unique opportunity to study the structure of matter in collisions at the $\sqrt{s} = 14$ TeV centre-of-mass energy.

During the second campaign of data taking, Run 2 (2015-2018), the proton beams collided at $\sqrt{s} = 13$ TeV. The facility provided densely populated proton bunches with an instantaneous luminosity of up to 10^{34} cm^{-2} s^{-1} with a 25 ns interval between proton beam crossings. The amount of the data recorded by the ATLAS experiment at the LHC during Run 2 allows to perform statistically significant measurements.

2.2 Physics of proton-proton collisions

The proton beam at the LHC consists of a group of bunches, squeezed packets of protons. Intersections of bunches cause interaction between proton constituents – partons: quarks, antiquarks, and gluons. This physical activity is recorded as an event in an output data set.



(a) (b)

Figure 2.2: A sketch of a proton-proton collision at high energies [30] (a). Deep inelastic scattering in pp collisions, jet emergence and reconstruction [31] (b).

The majority of interactions between protons result in non-elastic scattering, which can be hard or soft depending on the energy transferred between partons upon collision. In the deep inelastic scattering (DIS), partons undergo a very inelastic interaction with a large momentum transfer of several GeV (Figure 2.2a).

In a perturbative phase, partons emit gluons. Gluons, in their turn, can emit quark-antiquark pairs. Due to the confinement, colourful quarks and gluons are hadronised. As a result of a non-perturbative phase of the hadronisation, final-state bound states of quarks are formed. The formed hadrons are colourless. Based on a quark composition, mesons and baryons are distinguished in hadrons. The energetic collimated sprays of the hadrons produced upon fragmentation of the initially interacting partons are called jets (Figure 2.2b).

Proton fragments and other beam constituents can interact as well. These interactions are referred to as underlying events and multi-parton interactions. Apart from hard interactions, multiple soft collisions usually accompany bunch crossing. Soft inelastic collisions produce hadrons of lower energies.

The probability of a given process to occur in parton hard scattering can be quantified by a cross section. The cross section evaluates the probability of production and decay of a given particle. In pp collisions, it depends on the parton density function (PDF) and matrix element (ME). The parton density function $f_a^h(x, \mu_F)$ describes the density of parton of a type a carrying x fraction of momentum in a proton h which interacts at the scale μ_F. The matrix element $\hat{\sigma}_{ab\to F}$ quantifies the probability of the transition from an initial state (partons a and b) to a final state (F). The perturbative and non-perturbative phases are independent. Therefore, the cross section is factorised out [32] in the following way:

$$\frac{d\sigma}{d\mathcal{O}} = \sum_{a,b} \int_i f_a^{h_1}(x, \mu_F) f_b^{h_2}(x, \mu_F) \frac{d\hat{\sigma}_{ab\to F}}{d\hat{\mathcal{O}}} D_F(\hat{\mathcal{O}} \to \mathcal{O}, \mu_F) \tag{2.1}$$

where $\mathcal{O}$ is a hadronic observable and $\hat{\mathcal{O}}$ is a parton-level variable. The indices a and b run over parton types in colliding protons, and the indices h_1 and h_2 label two colliding hadrons. The term $D_F(\hat{\mathcal{O}} \to \mathcal{O}, \mu_F)$ stands for the probability of the transition from the partonic-level variable $\hat{\mathcal{O}}$ to the hadronic observable $\mathcal{O}$ after hadronisation.

Usually, the final state of interest is an ensemble of reconstructed objects and event-level signatures characteristic of certain processes. Knowledge of expected properties allows physicists to select objects and events. This narrows down the full ATLAS record to a set of events under investigation. A rate of registering of a given type of event is defined as a ratio of the process cross section and the collider instantaneous luminosity: $\text{rate}_{ab\to D} = \mathcal{L}/\sigma_{ab\to D}$. At the LHC, the instantaneous luminosity is defined through the number of particles per bunch N_{bunch} by the equation

$$\mathcal{L} = \frac{n N_{\text{bunch}}^2 f_{\text{LHC}} \gamma}{\beta \varepsilon F} \tag{2.2}$$

in which n is the number of bunches in the beam, f_{LHC} is the revolution frequency (collision rate) at the LHC, γ is a relativistic factor, ε is a normalised transverse beam emittance, β is a beam β-function, and F is an acceptance factor [33]. The γ, β, and ε parameters define the transverse beam size: $S = \sqrt{\varepsilon\beta}$ [34]. The size of the data set delivered by the accelerator or collected by the detectors is measured in units of an integrated luminosity. The integrated luminosity is integral to the instantaneous luminosity throughout data taking.

The designed parameters of the LHC take into account small cross sections. For inelastic processes cross section achieves around $\sigma_{\text{inel}} \approx 80$ mb. The elastic

cross section is about $\sigma_{\text{el}} \approx 31$ mb at $\sqrt{s} = 13$ TeV (Figure 2.3). Yet smaller cross sections characterise exclusive processes. For example, the estimated cross section is about 15.6 fb for contact-interaction production of hypothesised excited tau-leptons with mass $m_{\tau*} = 1$ TeV at the compositeness scale $\Lambda = 10$ TeV. The total Higgs boson production cross section $\sigma(pp \to H + X)$ reaches ~ 50 pb at $\sqrt{s} = 13$ TeV. To increase or open up the probability of registering events of interest, physicists perform high-rate collisions of high-luminosity beams at high energies over extended periods.

The large luminosity and the high collision rate lead to multiple interactions per bunch crossing, and their number μ is defined as follows:

$$\mu = \frac{\mathcal{L}\sigma_{\text{inel}}}{N_{\text{bunch}} f_{\text{LHC}}} \tag{2.3}$$

The effect of simultaneously happening or simultaneously registered events is known as a pile-up. The concurrent events are measured simultaneously (so-called in-time pile-up). As a detector's readout time window is larger than the bunch spacing of 25 ns, the signal from neighbouring bunch crossings can be sampled with the current one (so-called out-of-time pile-up).

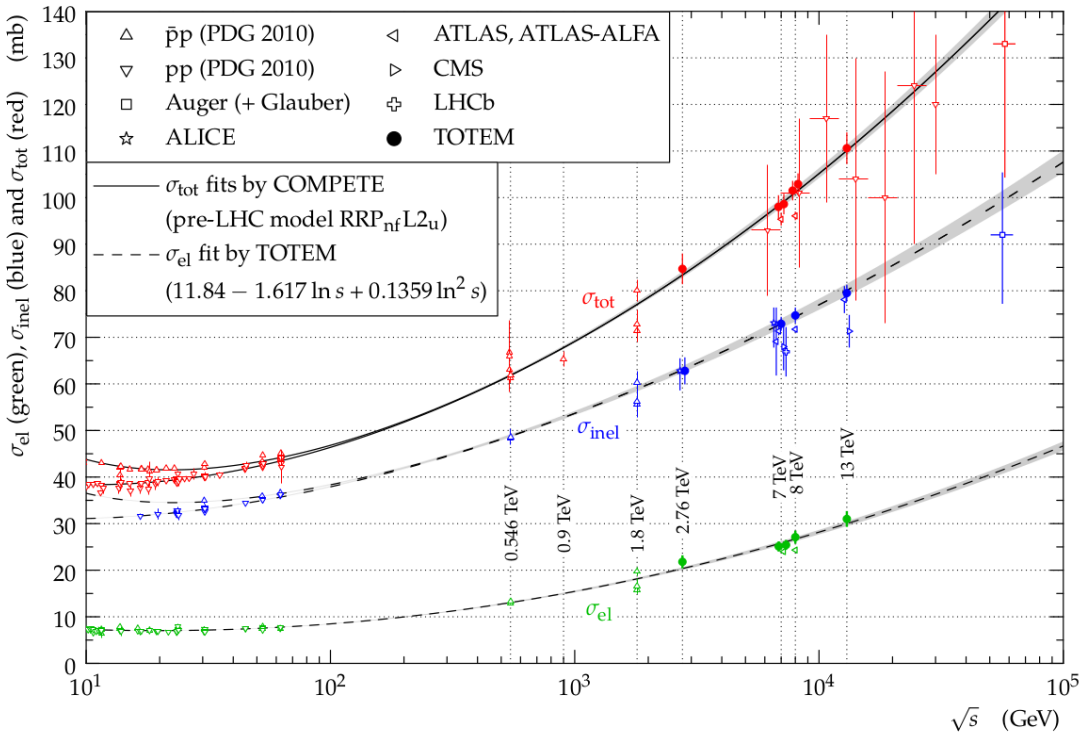


Figure 2.3: The elastic, inelastic, and total cross section on interactions in pp and $p\bar{p}$ as a function $\sqrt{s}$ [35].

There were about 34 interactions per bunch crossing under Run 2 data-taking conditions (Figure 2.4). This pile-up is the price physicists pay to maximise the integrated luminosity. The ultimate goal is to measure rare events with the highest statistically allowed accuracy. In Run 2, the ATLAS detector collected the set of 140 fb^{-1} data ready for physics analyses.

The ATLAS records data based on the decision of a set of two-level triggers [37]. In Run 2, it included the Level 1 (L1) system, based on customised

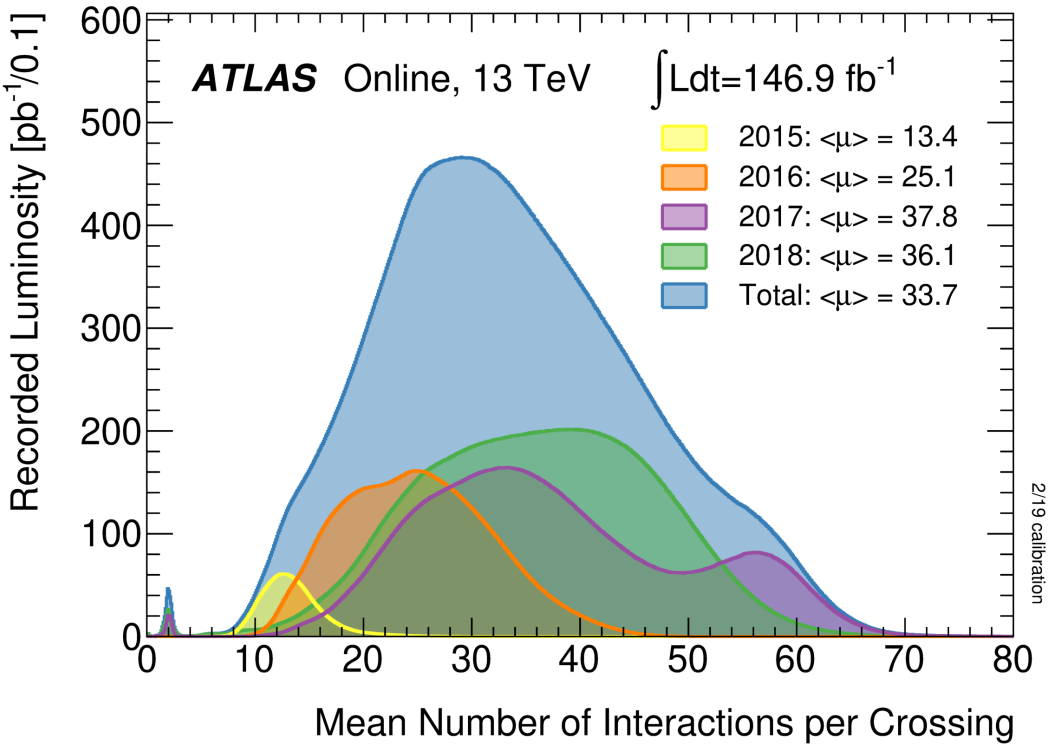


Figure 2.4: The mean number of interactions per bunch crossing in proton-proton collisions observed per year of ATLAS data-taking during Run 2 [36].

hardware, and the High-Level Trigger (HLT) system, implemented in software. A trigger decision is positive if requirements on objects' energy, position, identification, isolation and specific criteria on event topology are satisfied. The geometrical match between trigger-level (L1) objects and final-state reconstructed objects is needed. The di-tau trigger is discussed in detail in Appendix A. For instance, Figure A.1 presents the composition of the τ_{had} tau-lepton trigger chain.

During data taking, cross section estimates depend on data acquisition lifetime, trigger prescales, and operational status of detector parts. Data taking at the LHC is performed in luminosity blocks, each lasting about a minute. A continuous period of data taking is called a run. Runs usually last for hours which corresponds to a single fill of the LHC. Runs are grouped into sub-periods and periods based on the similarity of data-taking conditions. Eventually, a term of the LHC Run covers all the data-taking campaigns between long technical shutdowns. During these shutdowns, the system undergoes technical maintenance and upgrades.

3. ATLAS experiment at the Large Hadron Collider

3.1 Particle interaction with matter

The particles produced immediately after the hard scattering typically have large energies (orders of GeV). A pp interaction typically results in the production of quarks, gluons, electrons, muons, tau-leptons, neutrinos, intermediate vector bosons, and Higgs boson.[1] The detector registers these final-state particles. Short-lived particles – such as W, Z, or Higgs bosons – instantly decay instantly to more stable ones. Hadronised quarks are observed as jets and hadrons. Besides elusive neutrinos, emerged particles interact with detector material. There are numerous mechanisms for how particles interact with matter at the atomic, nuclear and subnuclear levels.

Electrons and positrons with an energy of MeV and above are decelerated in the electromagnetic field of atomic nuclei. As a result, Bremsstrahlung radiation is emitted. A fast charged particle can produce secondary electrons. Electrons and other charged particles of lower energies preferably lose their energy by ionisation. When a positron and an atomic electron annihilate, two photons emerge.

Photons with energy above double electron mass $2m_e$ can form a pair of an electron and a positron. Energetic photons interact with atomic electrons and nuclei through Compton scattering. Low-energy photons (below ~ 10 keV) have a higher probability of interacting through a photoelectronic effect, which is accompanied by electron emission.

Mutual conversion between electrons, positrons and photons results in a chain reaction which is called an electromagnetic shower. The electromagnetic cascade is sustained with the $e^- - e^+$ pair production by photons and Bremsstrahlung radiation by electrons and positrons. The EM shower development is schematized in Figure 3.1. To register EM cascades throughout their evolution, a dedicated detector should be appropriately segmented. The electromagnetic calorimeter is required to have a width and length of several radiation lengths X_0. The such size ensures that the calorimeter optimally accepts and efficiently registers EM showers.

Incoming high-energy (above 1 GeV) hadrons undergo inelastic scattering with nuclei. As a result, secondary charged and neutral hadrons (pions, kaons, protons, neutrons) are produced. The secondary hadron can further participate in inelastic interaction which leads to the development of a hadronic shower (Figure 3.2). Neutral pions decay to photon pairs, which initiate subsequent electromagnetic shower. The decay of charged pions and other hadrons can produce muons.

Neutrons scatter multiple times elastically and inelastically until they slow down to thermal energies (0.025 eV). The nucleon-free path length λ_{nuclear} defines the size of a hadronic calorimeter. A hadronic calorimeter's length and width correspond to the expected longitudinal and lateral span of hadronic showers.

[1] Emergence of anti-particles goes without saying. Electrons e^- and positrons e^+, muons μ^- and μ^+, tau-leptons τ^- and τ^+, quarks d, u, s, c, b, t and anti-quarks $\bar{d}, \bar{u}, \bar{s}, \bar{c}, \bar{b}, \bar{t}$, neutrinos ν_e, ν_μ, ν_τ and anti-neutrinos $\tilde{\nu_e}, \tilde{\nu_\mu}, \tilde{\nu_\tau}$ can be produced.

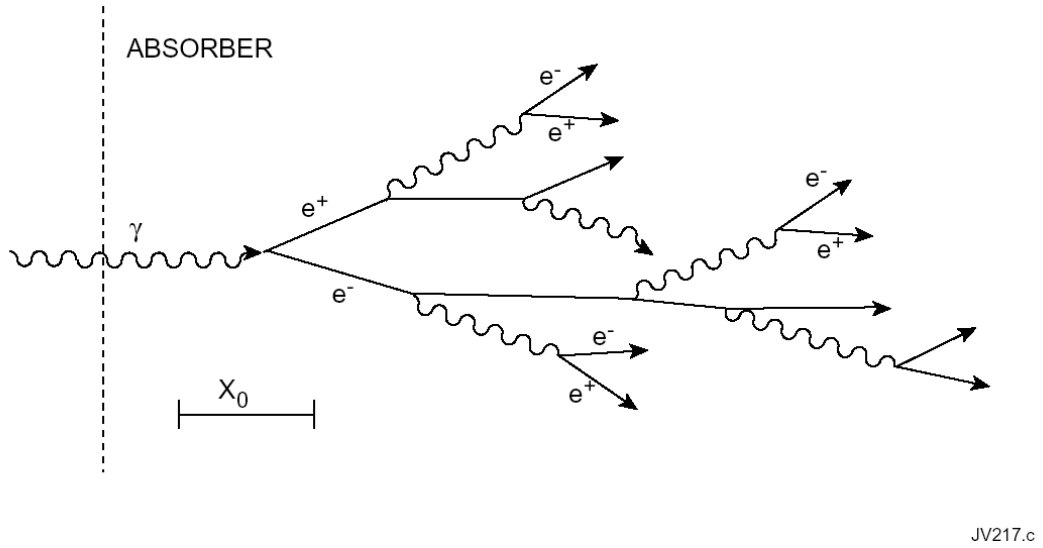


Figure 3.1: The development of an electromagnetic shower in matter [38].

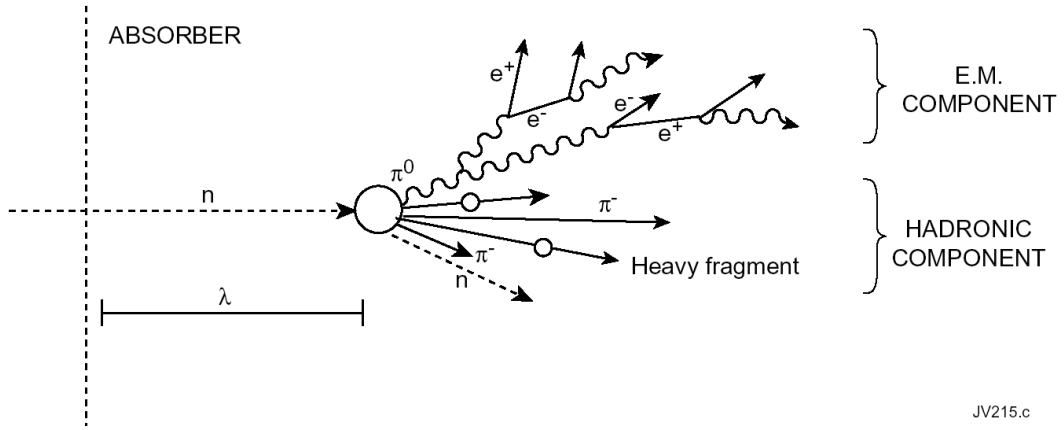


Figure 12: Schematic of development of hadronic showers.

Figure 3.2: The development of an hadronic shower in matter [38].

Muons can be produced primarily in a hard-scatter process and in the decay of secondary particles (vector bosons, tau-leptons, charged pions, and others). For muons with energy above several hundred GeV, radiative losses are significant. In a wide energy range below 0.5 TeV ionization losses are a dominant principle for detecting muons. Radiative losses become significant for highly energetic muons (with energy above several hundred GeV).

3.1.1 Registration principles of high-energy detectors

The evolution of electromagnetic and hadronic avalanches is stopped when the secondary particles do not have enough energy to create new particles. The final daughter particles formed in showers are subsequently thermalised, absorbed, or captured. At this stage, charged particles can be detected due to their ionization losses. Semiconductor and silicon detectors, gaseous detectors and counters exploit materials with a low ionization potential to gain higher sensitivity.

Scintillator-based detectors use the material's property to be excited by ion-

izing particles. This mechanism is employed in electromagnetic calorimeters to measure the energy of electrons and photons. Scintillators are also used in hadronic calorimeters, which register hadrons and jets.

The drift tubes and multiwire proportional chambers, resistive plate and thin gap chambers are exploited for muons registering. These detectors deploy gases with lower ionization potential (e.g. argon or pentane). The muon tracks are also detected by silicon-based semiconductor detectors based on the muon ionization losses.

Besides, detectors of Cherenkov and transition radiation are widely used in experimental particle physics for particle detection, identification and momentum measurement.

3.2 The ATLAS experiment

The ATLAS detector (**A** **T**oroidal **L**HC **A**pparatu**S**) [39, 40, 41] is a multi-purpose detector designed to observe particles produced in pp and heavy-ion collisions at the LHC. The ATLAS experiment is dedicated to measuring SM parameters and assisting in searches for new physics.

The ATLAS detector surrounds the main interaction point where the proton bunches cross (Figure 3.4). Several subdetectors are located at different distances from the main interaction point. Figure 3.3 illustrates how the signal from passing through particles is reconstructed and interpreted.

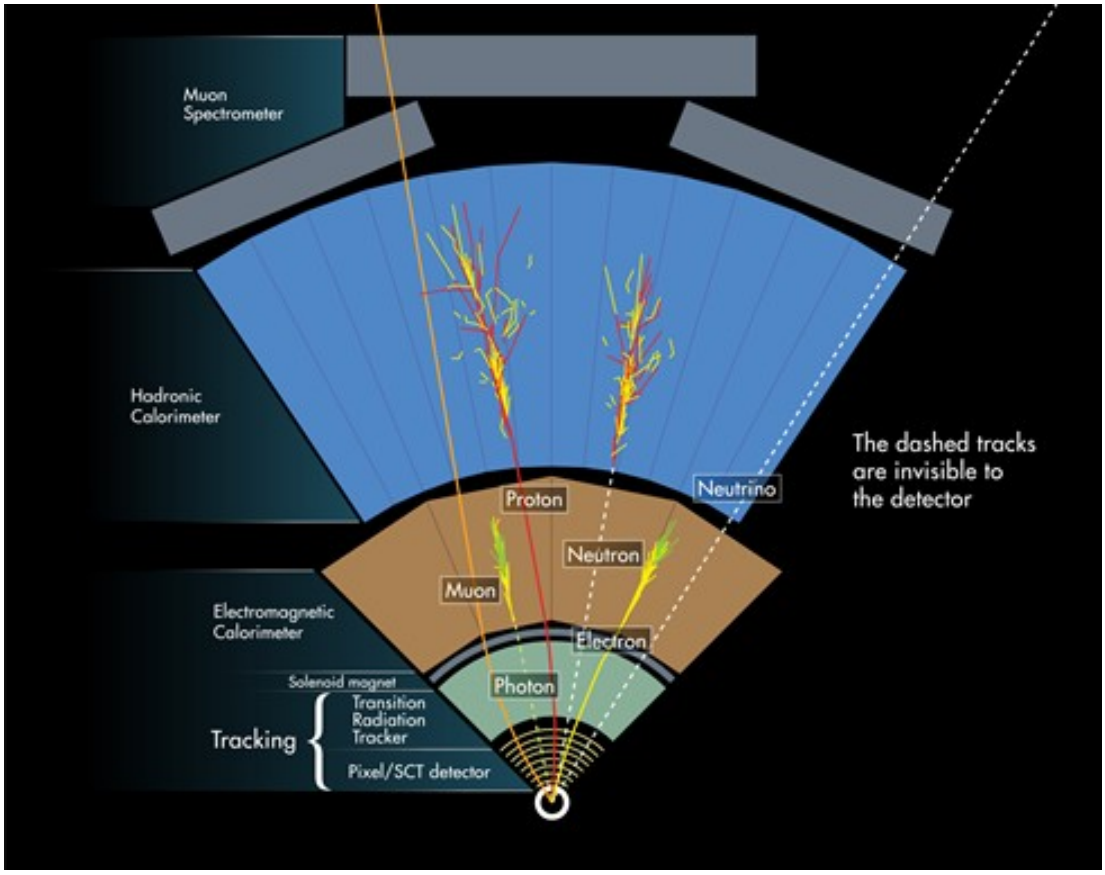


Figure 3.3: The diagram of the particle signatures in the ATLAS detector [42].

They enable particle identification and measurement of their four-momenta. Energy and location of passing through particles are usually defined in terms of transverse momentum p_{T}, pseudorapidity η, and azimuthal angle ϕ.[2] The detector symmetrically covers the main interaction point in the range of pseudorapidity up to $|\eta| < 4.9$ with almost a full solid angle.

The ATLAS detector is designed to register particles which are produced in pp collisions and subsequent particle evolution: electrons, photons, muons, hadronically decaying tau-lepton, charged and neutral pions, jets. The energy carried by elusive neutrinos is commonly estimated as an energy disbalance in the plane transverse to the beam axis.

The layout for Run 1 (2012-2015) and Run 2 of the ATLAS detectors was designed in the 90s and 00s. The detectors, their readout, and related electronics components were extensively tested during numerous test beam campaigns at the Prevessin in the CERN North Area. The North Area facility allows validation equipment with particle beams of up to 400 GeV energy. The detector response to hadrons (pions, kaons, protons), muons, and electrons was extensively studied.

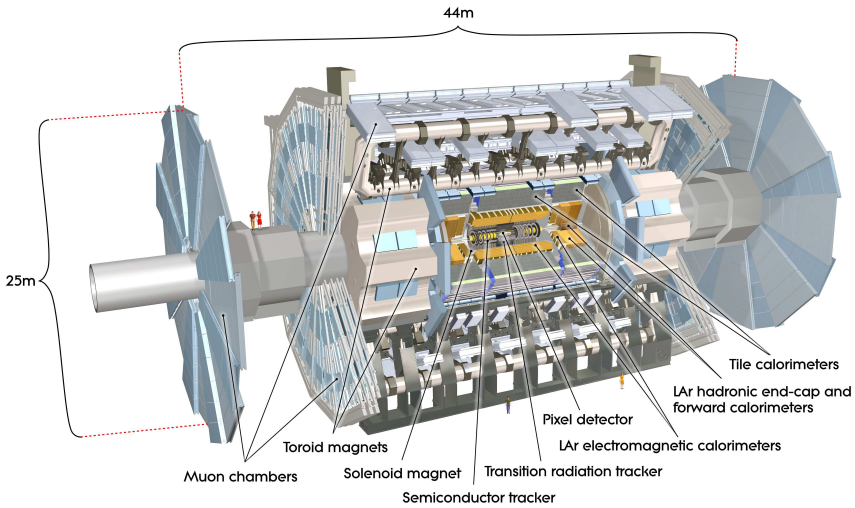


Figure 3.4: A cut-away view of the ATLAS experiment [39].

3.2.1 Inner detector

The Inner Detector (ID) accomplishes the challenging task of tracking multiple charged particles produced in pp collisions in a compact volume [43]. The ATLAS Inner Detector is represented by finely granular semiconductor and transition radiation detectors with high sensitivity and fast readout. The Inner detector is required to be radiation-hard due to high radiation exposure. Thus, the innermost

[2]The ATLAS coordinate system is a right-handed Cartesian one. The centre of the coordinate system is placed at the main interaction point where two beams are colliding. The z-axis is directed along the beam line. The x-axis passes through the interaction point and the centre of the LHC ring, and the y-axis is pointed perpendicularly upwards. The cylindrical system of the $(r,\ \phi)$ coordinates is introduced in the transverse plane, where r is parallel to the x-axis and ϕ is the azimuthal angle raised from the z-axis. In terms of a polar coordinate θ in the longitudinal $(x,\ y)$ plane, pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$.

layer of the Pixel system is engineered to withstand up to 2×10^{15} 1-MeV neutron equivalent per cm^{-2} after 10 years of operation (2008-18) [44].

The Inner Detector is situated in 2 T axial magnetic field generated by a superconducting solenoid magnet. The Inner Detector has a total radius of about 1.15 m and comprises four subdetectors:

- The Insertable B-Layer (IBL) with 12 million readout channels introduced in 2014 and employing 3D pixel sensors 230 μm thick [45],
- The Pixel Detector comprising 80 million tiny (50 μm $\times$ (250 $\div$ 400) μm) silicon pixels for accurate registration of vertices,
- The Semiconductor Tracker (SCT) with dense readout of 6.3 million channels for precise detection of tracks, and
- The Transition Radiation Tracker (TRT), a gaseous ($Xe - CO_2 - O_2$ or $Ar - CO_2 - O_2$) detector made of 300,000 straw drift tubes, which assists in particle identification (e.g. electron/pion separation).

The SCT detector operates at $\sim 0°$ being cooled with a carbon dioxide CO_2 and fluorocarbon C_3H_8 mixture. The TRT is kept at room temperature [46]. An ATLAS Inner Detector End Cap is housed in a LAr cryostat to avoid thermal noise. For Run 2, the Inner Detector **D**ata **Ac**q**u**isition (DAQ) hardware was improved to cope with the high-luminosity challenges.

The efficiency and resolution of the Inner Detector trigger were studied in several processes. The Inner Detector tracking achieved nearly unit efficiency for the 24 GeV muon and the 25 GeV one-prong tau-lepton trigger. The response is uniform for offline muons with $p_{\text{T}} > 10$ GeV and approximately stable at the various pile-up level [47]. The processing time of the tau-lepton trigger signatures was reduced by applying a two-stage approach [48]. The fast track finding takes 45 ms and the precision tracking is done within 5 ms [47].

3.2.2 Calorimeters

Electromagnetic and hadron calorimeters form the next layer of the ATLAS detector. Development of electromagnetic (EM) and hadronic showers is defined by the radiation length X_0 and the nuclear interaction length λ_{nuclear}. The nuclear interaction length is typically one to two orders larger than the radiation length as the cross section of interaction between incident hadrons and nuclei in the detector matter is lower than the cross section of the electroweak interaction. Therefore, the electromagnetic calorimeter is surrounded by the hadronic calorimeter in the ATLAS. The majority of the particles in electromagnetic showers – electrons and photons – are registered by electromagnetic calorimeters. Hadronic showers develop in the EM and hadronic calorimeters. The ATLAS calorimeters are non-compensative. It means that they respond to electrons and isolated hadrons of the same energy with pulses of different values. The estimated e/h (electron-to-hadron) non-compensation was 1.74 (1.36) for an electromagnetic (hadronic) compartment of the combined calorimeter [49, 50].

The electromagnetic Liquid Argon Calorimeter. The ATLAS electromagnetic (EM) calorimeter is the Liquid Argon (LAr) calorimeter [51]. It is composed of the barrel in the central $|\eta| < 1.45$ region and the End Caps (EC) in the high-η region ($1.375 < |\eta| < 3.2$) as shown in Figure 3.5. The total length of $> 24X_0$ in the barrel and $> 26X_0$ in the EC ensures high detector acceptance and EM shower containment.

The lead absorber plates are interspaced with $\sim$ 182,000 electrodes with the LAr gaps, 2 mm each. Electromagnetic cascade is initiated in the high-Z lead by incoming particles (namely, electrons, and photons). Liquid Argon is chosen as active material due to its low excitation and ionization energetic thresholds. LAr also provides a linear and stable response. The molecules of the cooled LAr are ionized by passing through charged particles of the electromagnetic shower. The produced ionization current is read out, and the energy and position of electrons and photons are measured.

The Liquid Argon Calorimeter reached its designed energy measurement parameters in Run 2. The electron energy resolution was probed in $Z \to ee$ events and the constant term did not exceed 1% for the majority of the LAr barrel channels [52]. The energy scale uncertainty defined with corrections from Z boson mass stayed within 4 ‰ [53]. The LAr time resolution was below 0.3 ns for energies above 30 GeV, with a constant term of $\sim$ 200 ps [54].

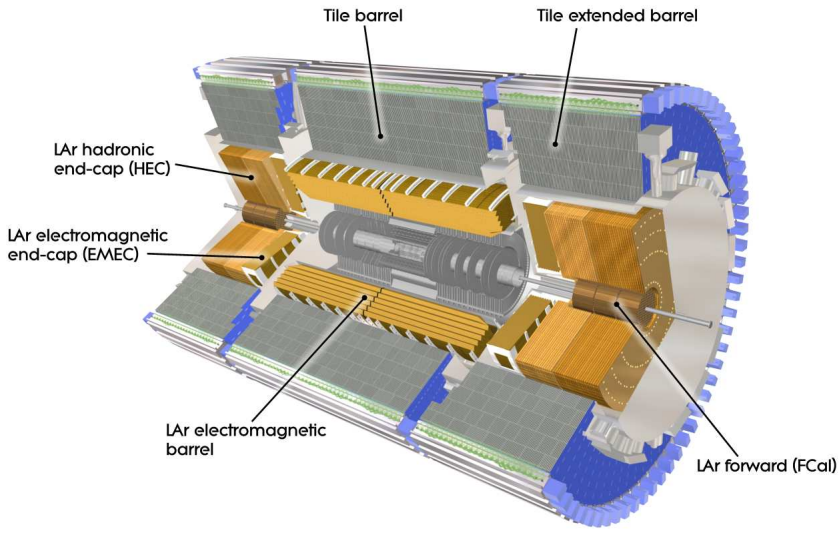


Figure 3.5: A cut-away view of the ATLAS calorimeter system [55].

The hadronic Tile Calorimeter. The ATLAS Tile calorimeter measures the four-momenta of the jets, hadrons, and hadronically decaying tau-leptons through their strong and electromagnetic interaction with the detector material. It also assists in muon identification and Level 1 triggering.

The central section of the ATLAS hadronic detector is represented by a segmented sampling Tile calorimeter. Layers of the active medium (the tiles of a plastic scintillator) are sandwiched with layers of steel absorber. The TileCal uses a plastic scintillator due to its good scintillating properties. The energy of passing

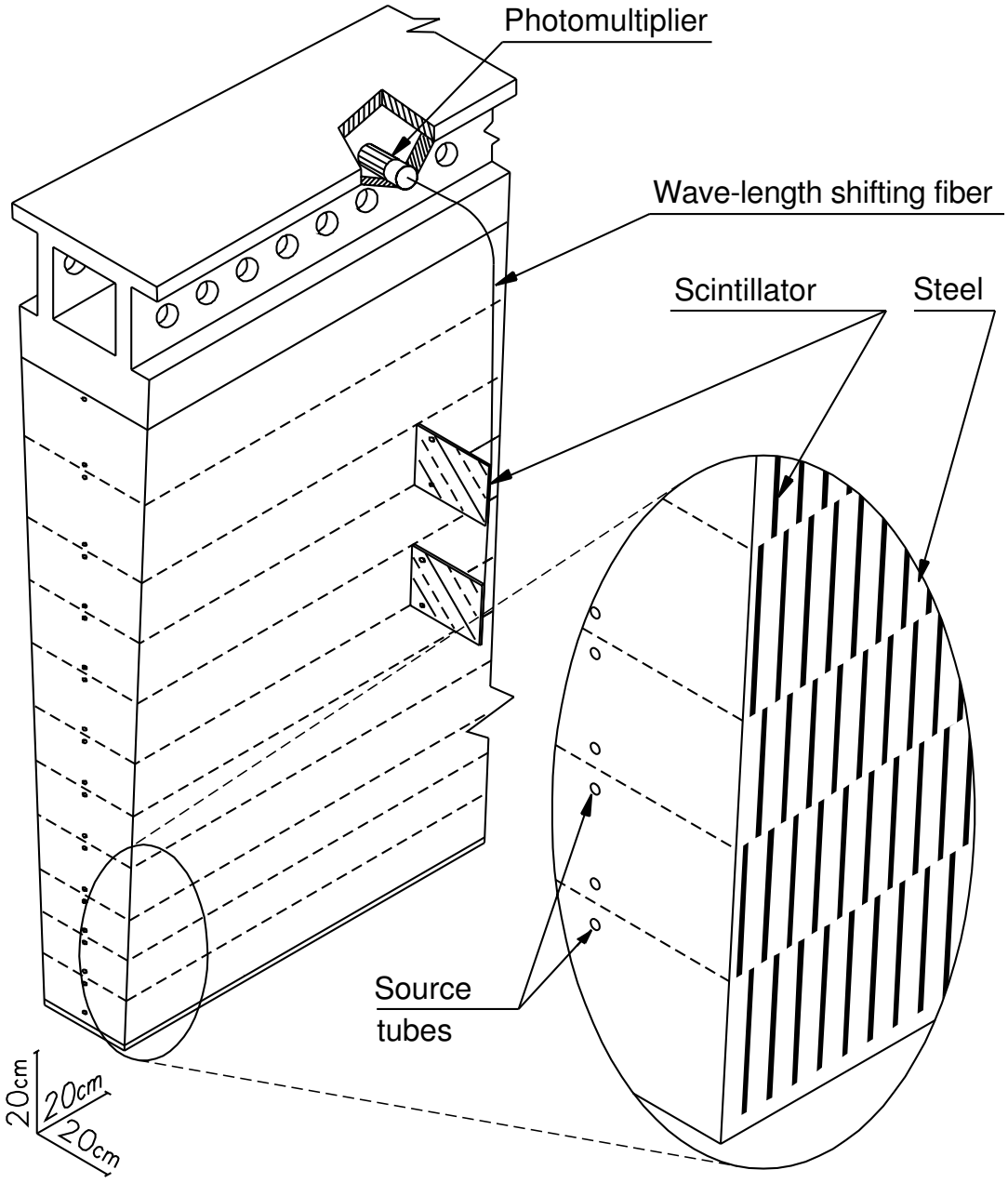


Figure 3.6: The mechanical assembly and the optical readout of the Tile Calorimeter module [55].

through particles is converted into light which is transmitted to photomultiplier tubes (PMT). In PMTs, the collected light is transformed into an electric current.

Overall, the TileCal comprises 5182 cells, and a cell is typically read out by two PMTs (channels). The cell granularity is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for two innermost radial layers and $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ for the outermost radial layer. The TileCal is azimuthally segmented in 64 modules (Figure 3.6).

With an effective nuclear interaction length of 20.7 cm [56] and the instrumented region of 1.64 m radial thickness [57], the TileCal provides detection throughout 7.4 λ_{nuclear} for particles emitted at 90 degrees to the beam line [57].

The electronic noise is measured to be at the level of $20 - 40$ MeV, in dedicated pedestal runs without signal exposure. The pile-up noise is measured with

zero-bias triggered events and compared to minimum bias Monte Carlo simulations [58]. The pile-up affects the most the TileCal cells in the innermost radial layer due to the highest particle flux close to the main interaction point. The pile-up noise is lower in two outer layers.

The calibration scale was derived from the Tile Cal responses to electrons during test beam campaigns. The response to collision muons and high-momentum isolated cosmic muons verifies the measured energy at the EM scale. Isolated hadrons are used as a probe of the hadronic response. The detector response azimuthal uniformity and linearity are observed [59], [60].

In Run 2, the hadronic Tile Calorimeter performed within the designed jet resolution $\sigma(E)/E = 50\%/\sqrt{E} \otimes 3\%$. The time resolution studied with multi-jet events is within 1 ns for the cell with energy input above 30 GeV [61]. The constant term of the time resolution approaches 0.3 ns.

The Data Quality (DQ) efficiency 99.65% was achieved by the TileCal in Run 2.

The LAr Hadronic End Cap Calorimeter (HEC). The LAr hadronic end-cap calorimeter employs copper-tungsten plate absorbers. Its geometry is beneficial for registering jets and hadronic showers in the forward region $1.5 < |\eta| < 3.2$. Drift regions of the 1.8 mm size are created with a gap structure of the electrodes [62]. It is read out by an electrostatic transformer with GaAs cryogenic preamplifiers [63] which improves the signal-to-noise ratio and reduce the rate of failures. Signal is promptly processed based on the concept of "active pads" [64].

The detector is capable to measure muons and jets escaping the LAr EM calorimeter in the higher-pseudorapidity region.

The Forward Calorimeter (FCal). The Forward Calorimeter (FCal) covers a very forward $3.1 < |\eta| < 4.9$ pseudorapidity region extending the ATLAS detector sensitivity for more than 10 λ_{nuclear}. The FCal uses LAr as an active medium and copper and tungsten as a passive medium in the two inner EM hadronic modules and an outer hadronic module, respectively. The LAr gaps are at least four times finer than in the LAr EM calorimeter. Such topology provides a fast readout in a busy environment.

The FCal functions as an electromagnetic as well as hadronic calorimeter in the regime of the extremely dense flux of energetic particles. It assists in the identification of forward jets and works as shielding for the outer muon system. It is also employed to register punch-through jets (the remnants of beams that are not registered by preceding detectors) [65]. The forward detector provides a solid-angle encapsulation of the whole detector, which reduces energy loses and improves the reconstruction of the missing transverse energy.

3.2.3 Muon spectrometer

An outermost muon spectrometer (MS) plays a key role in the detection of muons [66]. Together with the Inner Detector, the Muon Spectrometer identifies muons and measures their momenta. The MS provides efficient muon triggering.

The MS comprises superconducting air-core toroidal magnets. The average strength of the barrel and end-cap toroids is 0.5 T and ~ 1 T, respectively [39].

The muon detectors are represented by high-precision drift and multi-wire proportional chambers.

Muon momentum is defined via the muon trajectory curvature in the magnetic field. The measurement is performed at six-eight η points along the track in the chamber. The muon path is detected by **M**onitored **D**rift **T**ubes (MDTs) which are arranged in three radial layers along the beam axis. In the innermost End Cap $2 < |\eta| < 2.7$ range, the **C**athode **S**trip **C**hambers (CSCs) are deployed due to their ability to withstand high particle flux. **R**esistive **P**late **C**hambers (RPCs) ($|\eta| < 1.05$) and **T**hin **G**ap **C**hambers (TGCs) ($1.05 < |\eta| < 2.4$) play a role of muon triggers and measure muons (η, ϕ) coordinates.

The single-hit spatial resolution is about 80 μm for each MDT chamber [67]. The TGCs spatial resolution is better than 100 μm. The intrinsic time resolution of the RPC chambers is ~ 1.44 (1.64) ns in the η (ϕ) plane [68]. The MS relative transverse momentum resolution is below 3% for $p_{\mathrm{T}} > 22$ GeV and is about 10% at $p_{\mathrm{T}} \approx 1$ TeV as studied in $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ events.

3.2.4 Forward Detectors

There are forward detectors LUCID (LUminosity Cherenkov Integrating Detector), ZDC (Zero Degree Calorimeter), ALFA (Absolute Luminosity For ATLAS), and AFP (Atlas Forward Proton) in the ATLAS experiment. Placed at a distance from the interaction point, they measure outcomes of particle collisions at the very forward rapidity range. The data collected by the forward detectors in low pile-up conditions allows for studying diffractive events and soft QCD processes. The LUCID detector plays an important role in measuring luminosity at ATLAS.

3.2.5 Trigger system

Due to large number of events and the limited capacity of data storage, only events displaying features interesting for scientists are selected. The ATLAS trigger system is dedicated to identifying events of interest.

The hardware-based Level 1 trigger exploits data from the calorimeters and the muon spectrometers. Candidate electrons, photons, jets and muons are promptly reconstructed. The Level 1 trigger decides whether the event will be saved for physics analysis. The Level 1 trigger acceptance rate is 200 kHz.

If accepted by the Level 1 trigger, the event is passed to a two-stage software-based High-Level trigger (HLT). The HLT applies selection algorithms using the full-granularity information from the whole ATLAS detector, including tracking by the inner detector. After acceptance by HLT triggers, the event rate is reduced from the initial 40 MHz rate to 200 kHz.

The trigger system uses its own algorithms to promptly reconstruct objects, typically in a simplistic way. The trigger-level objects are often called online ones in this thesis due to the fact they are reconstructed at the very moment the event occurred. The offline objects are reconstructed afterwards based on the full recorded information. The offline estimation of time and energy may account for various corrections aiming to improve measurements.

4. Monte Carlo simulation

4.1 Monte Carlo methods in particle physics

The Monte Carlo (MC) simulation is an essential part of measurements and searches in high-energy experiments. It is widely used for modelling particle production and decay as well as their interaction with matter. The MC technique relies on the estimation of multi-factor probabilities for processes. The Monte Carlo method allows for a random outcome of the process and due to that it obtained the name of the gambling destination in Monaco.

A simplified process of the data sample generation is outlined below.

Firstly, Monte Carlo generators describe hard-scatter processes happening in pp collisions at a parton level. These processes are typically light-flavour quark-antiquark scattering, gluon-gluon scattering, or (anti)quark-gluon interaction with a large momentum transfer. In the hard scattering, SM particles are produced in known interactions: other quark pairs or gluons (via strong interaction), intermediate bosons (via electroweak interaction), and Higgs bosons (via Higgs interaction). A particular BSM interaction can be simulated within a chosen model as well, given a set of constraints, parameters, and conditions for particle decay. Modern MC generators support the modelling of hard scattering processes at a tree level with leading order (LO) accuracy and some of them achieve next-to-leading order (NLO) accuracy.

Then, the outcomes of the hard-scatter reaction and the final-state kinematic properties are calculated for a given decay mode(s). At this step of MC generations, the kinematic properties of product particles can be studied. The so-called generator-level (particle-level, truth-level) observables give an estimate of how particle would be measured by an ideal, perfectly calibrated detector with an infinite resolution [69].[1] The modelling at a generator level does not include detector simulation, and the particle-level observables are used to validate the MC simulation.

Once the hard process is generated, the particle decay cascade, electromagnetic and hadron shower development and underlying events are simulated.

In practice, the measurement of objects by a real detector is distorted due to reconstruction and detector imperfections. Thus, the Monte Carlo generated samples further undergo simulation of passing through the detector and the trigger system. The passage of particles through the matter of the detector is modelled by the `Geant 4` package [70]. This is a common tool for the simulation of interactions between particles and detector material. The ATLAS experiment supports two approaches: full and fast ATLAS simulation [71].

The simulated data are then digitized in the same way as it is done for real data (Figure 7.1 shows the data processing flow). As mentioned in Section 2.2, hard scatterings overlay with multiple underlying events happening at the same bunch crossing during pp collisions at LHC. In MC, the effect of pile-up is emulated by overlaying the hard-scatter events with inelastic pp events generated

[1] As these observables refer to particles and interactions properties themselves and are not biased with detector effects, these generator-level variables are often called "truth" in the jargon of the ATLAS experiment.

Process	Generator		PDF set		UE tune	ME order	XS order
	ME	PS	ME	PS			
V+jets	Sherpa v2.2.1		NNPDF30NNLO		Sherpa	NLO/LO	NNLO
$t\bar{t}$	PowhegBox v2	Pythia v8.230	NNPDF30NLO	NNPDF23LO	A14	NLO	NNLO NNLL
single top s	Powheg-Box v2	Pythia v8.230	NNPDF30NLO	NNPDF23LO	A14	NLO	NLO
single top t	Powheg-Box v2	Pythia v8.230	NNPDF30NLO	NNPDF23LO	A14	NLO	NLO
Wt	Powheg-Box v2	Pythia v8.230	NNPDF30NLO	NNPDF23LO	A14	NLO	NLO NNLL
Diboson	Sherpa v2.1.1		NNPDF30NNLO		Sherpa	NLO/LO	NNLO

Table 4.1: The list of MC generators, their settings, and accuracy of cross section calculation for SM background processes.

by `Pythia` v8.186 [72] employing the `NNPDF2.3LO` set of PDF [73] and the `A3` tune [74]. Consistency between $\langle\mu\rangle$ profiles between MC and real data is provided by the procedure of pile-up reweighting [75].

Each event simulated with Monte Carlo is assigned a weight. The event weight corrects for imperfections in the modelling of detector response. It includes factors of object reconstruction and trigger efficiencies.

4.2 MC modelling for SM and BSM processes

The research presented in this thesis analyses events with signatures of the Higgs boson decay into two tau-leptons. Another analysis studies events with a pair of tau-leptons and two jets. Numerous Standard Model processes mimic characteristics of events involving tau-leptons, jets and missing energy. Therefore, a precise estimation of the SM background is a must-have in these analyses (and a majority of HEP analyses in general).

SM background processes, except a fraction of fakes, are simulated with MC.[2] Majority of background processes is shared between two analyses presented ($H \to \tau\tau$ measurement and τ^* search).

The MC generators used for modelling background, versions of PDF sets and tunes, the accuracy of generation in QCD and cross section calculation are listed in Table 4.1. Production of the MC samples is undertaken centrally by the ATLAS **P**article **M**anagement **G**roup (PMG).

Excited tau-lepton and Higgs boson signal samples were also generated with MC. Their generation is described in the dedicated chapters.

For MC simulated events, the data processing chain includes additional steps in comparison to real ATLAS data. The MC event generator `EVNT` outputs (hard-scattering processes, hadronic showering and underlying events) undergo emulation of passage through the ATLAS detector. As result, the Geant 4 sets of energy depositions sampled in time and segmented spatially in the active detector material are collected in the `HITS` format. The hits are digitized as they would be read out by the real detector (`RDO` format).

[2]The fake background contribution is represented by events where a certain type of particle is misidentified.

5. ATLAS event reconstruction

5.1 Event reconstruction

The signals read out from the ATLAS subdetectors are reconstructed. The L1 triggers provide a fast estimation of the objects and based on their decision the event is kept or discarded for further reconstruction steps and subsequent storage.

The picture of the pp bunch collisions is restored based on the reconstructed tracks and calorimeter clusters. The signal in the Inner Detector can be then assigned to a vertex of a particle production or decay. Electrons, muons, jets, τ_{had}, and photons are reconstructed based on the combined information from the Inner Detector, the calorimeters, and the Muon Spectrometer. If several objects share the same detector signal, a procedure of so-called overlap removal (OLR) resolves ambiguities in object identification. It will be described in detail in Section 5.6. In a brief, the order of preference for objects is as follows: muons, electrons, hadronically decaying tau-leptons, and jets.

The energy disbalance in the plane perpendicular to the beam axis is understood as the energy taken by neutrinos.[1] The pile-up and underlying events can be estimated alongside.

5.2 Electrons

Electrons are identified as energetic clusters deposited in the EM calorimeter and matched with the charged-particle tracks identified by the Inner Detector [76].

The ATLAS software `Athena` performs calibration of electrons as well as photons with the corresponding tool.[2] At the next step, calibration adjusts the energy scales between different layers of the EM calorimeter as well as between MC and data [77]. The energy resolution is optimized with the multivariate regression algorithm and application of the supercluster method [78].

In Run 2, the ATLAS distinguished three working points of electron identification: `Loose`, `Medium`, and `Tight` [53]. They provide the efficiency of 93%, 88% and 80%, respectively.

A set of variables are calculated to discriminate prompt electrons against EM energy deposits driven by jets, photons, and non-prompt electrons. The latter can originate from background decays of heavy-flavour hadrons. A projective likelihood estimator is employed to further reject the events where electrons are mimicked by semi-leptonic decays of heavy quarks.

The electron isolation is defined as a ratio of transverse energy in a cone of the $\Delta R = 0.2$ angular distance to the total energy deposited by the corresponding cluster.[3] Several electron isolation categories are defined, including `Gradient` and `PLVTight` working points (WPs) [76].

[1] It is called the missing transverse energy as will be introduced later on.

[2] The electron and photon calibration tool `EgammaCalibrationAndSmearingTool` is employed.

[3] Here, the angular distance is defined by the pseudorapidity and azimuthal coordinates: $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

Electrons should have good Object Quality, i.e. their composite clusters pass good-quality criteria and the electromagnetic shower has the shape as expected.[4] Several selections are intended to remove the signal from cells with non-functional calorimeter readout or cells with disrupted High Voltage supply.[5]

Electrons are required to have transverse momentum above 15 GeV in both presented analyses. The pseudorapidity of an electron is defined as a barycentre of the electron's primary cluster in the second layer of the EM LAr calorimeter. The required $|\eta_{\mathrm{BE2}}| < 2.47$ range corresponds to the fiducial space of the Inner Detector and the high-granularity section of the EM calorimeter. The transition region $1.37 < |\eta_{\mathrm{BE2}}| < 1.52$ between the barrel and the end-cap modules is excluded.

Additional selections on the vertex configuration are devoted to rejecting non-prompt electrons originating from tau-lepton decay. The selections on the significance $|d_0|/\sigma_{d_0} < 5$ and $|z_0 \sin\theta| < 5$ mm are recommended by the ATLAS `EGamma` working group. Here z_0 (d_0) is the longitudinal (transverse) impact parameter with respect to the primary vertex and σ_{d_0} is the uncertainty on the transverse impact parameter. This selection is included in the search for excited tau-leptons. However, it was not applied in the $H \to \tau\tau$ analysis: as electron from τ_{lep} is generally displaced from the primary vertex in $H \to \tau\tau$ events, the requirements on d_0 and z_0 would decrease the yield of the signal signatures.[6]

Baseline electrons are required to satisfy the `Loose` ID criteria before entering the overlap removal. This is related to both analyses considered.[7]

The difference in electron reconstruction, identification, and isolation between data and MC is eliminated by applying appropriate scale factors for event weights in Monte Carlo.

5.3 Muons

Muons are reconstructed independently in the inner detector and in the muon spectrometer, and data from individual detectors are combined to form the final muon tracks [79].

The muon momentum is calibrated to make p_{T} of the simulated muons aligned with muon p_{T} properties in real data. The momentum resolution and scale corrections acknowledge non-uniformities of the magnetic field and the detector material, fluctuating energy loss in multiple scattering, and the impact intrinsic resolution [80].

It is important to distinguish between prompt muons and delayed muons. Non-prompt muons can emerge in decays of, for instance, tau-leptons, pions, or kaons. A proper identification consequently guarantees precision in p_{T} measurement. There are six identification categories supported by the ATLAS: `VeryLoose`, `Loose`, `Medium`, `Tight`, `LowPt`, and `HighPt` [81]. The `Tight` category ensures lower fake acceptance and a high identification efficiency (85% and

[4] The variable called `goodElectron` is used for this selection.

[5] The variable `passedDeadHVCellRemoval` is estimated.

[6] The impact parameter is defined as the shortest distance between the electron candidate track and the proton beam line.

[7] The $H \to \tau\tau$ analysis requires electron candidate to pass the `LOOSE_AND_BLAYER_LLH` identification WP. The `Loose` ID is required in the search for excited tau-leptons.

99% for muons with $p_{\mathrm{T}} > 10$ GeV and $p_{\mathrm{T}} > 50$ GeV, respectively). The `HighPt` WP provides better resolution for energetic tracks with $p_{\mathrm{T}} > 100$ GeV.

Muon isolation is estimated with the transverse energy of tracks and topological clusters reconstructed in a cone around the muon.

Muon candidates with $|\eta| < 2.47$ are considered. Baseline muons are required to satisfy the following p_{T} requirements before the OLR: $p_{\mathrm{T}} > 10$ GeV in the $H \to \tau\tau$ analysis and $p_{\mathrm{T}} > 7$ GeV in the search for excited tau-leptons.

In $H \to \tau\tau$ analysis, baseline muons have to pass the *Loose* identification selection. In the search for excited tau-leptons, muon should pass the `HighPt` identification criteria as this analysis explores high-energy range [81]. Similar to the electron selection, the criteria on the muon's impact parameter are applied in the ETL analysis but not in the $H \to \tau\tau$ measurement.

The scale factors (SFs) for the muon identification (ID) and isolation are introduced.

5.4 Jets

Jets are reconstructed from topological clusters [82, 83] or PFlow [84] objects by means of the anti-k_{T} algorithm [85] with a distance parameter $R = 0.4$.

In general, jets are calibrated in the following steps: jet calibration area, residual correction, eta jet energy scale, global scale correction, and smearing (in situ) for MC (data) [86]. This implies corrections for pile-up and UE activity, interaction vertex displacement, and partonic origin. The calibration reduces the difference between the detector-level energy deposition and the energy carried by the generator-level decay products. Since 2019, it is recommended to employ PFlow-reconstructed jets [84] in the ATLAS.

Jets are labelled with the Jet Vertex Tagger (JVT) [87] to distinguish those jets produced in hard scattering from pile-up induced jets. In the forward pseudorapidity region, a dedicated forward Jet Vertex Tagger (fJVT) [88] gives an advantage of further suppression of pile-up jets. There are multiple options for pile-up jet reduction based on the (f)JVT score, including `Medium` and `Tight` Working Points with 92% pile-up jet suppression as assessed in MC simulation. The (f)JVT scale factor is applied for jets with $p_{\mathrm{T}} < 60$ GeV to suppress likely pile-up jets.

Other taggers are used to indicate whether a given jet originates from quark or gluon as well as to label originating quark flavour (b, c, or light quarks). B-tagging [89] by the `DL1r` algorithm is applied to centrally located ($|\eta| < 2.5$) jets with p_{T} above 60 GeV. The efficiencies in estimating the (forward) jet vertex and heavy flavour scores are introduced as extra scale factors of the MC event weight.

There are basic kinematic selections for jets before the OLR: $p_{\mathrm{T}} > 20$ GeV and rapidity $|y| < 4.5$.[8] These limits are motivated by the Inner Detector spatial coverage ($|\eta| < 2.5$) as well as a lower identification and reconstruction efficiency for low-energetic and high-y jets.

[8]The rapidity of the particle is defined by the formula: $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$. The pseudorapidity approaches the rapidity in a high-energy and massless limit. In the ATLAS experiment, pseudorapidity is widely used due to the complexity of interactions in pp collisions. The magnitude of rapidity is largely applicable for jets due to rapidity's Lorentz invariance and the power of demonstration of jet separation.

5.5 Tau-leptons

Tau-leptons decay to pions and a neutrino in hadronic mode.[9] The neutrino does not leave a signal in the detector and its momentum is estimated as missing transverse energy introduced below (Section 5.7). Pions make up a visible hadronically decaying tau-lepton (τ_{had}).

Hadronically decaying tau-leptons τ_{had} are reconstructed as anti-k_{T} jets [85] with a distance parameter $\Delta R = 0.4$ [90]. Hadronic tau-lepton decays are identified as jets matching with track vertex candidate(s), whereas one or three charged tracks are associated with the vertex depending on the "prongness" of tau-lepton. Neutral pion candidates are reconstructed with the Tau Particle Flow [91].

Most of the τ_{had} decay leads to the production of one or three charged pions, and for that τ_{had} are one-prong and three-prong distinguished. The τ_{had} decay can produce up to two neutral pions, along with the charged hadrons. A dedicated ATLAS tool *PanTau* classifies the τ_{had} based on the multiplicity of pions in decay products.

A specific τ_{had} energy calibration is applied to the tau-lepton candidates to account for the difference between the detector energy deposition and the energy carried by the generator-level decay products. The tau energy scale is derived with multivariate analysis (MVA) techniques based on the data on the related tracks, clusters and neutral pions [90].

As the detector responds in a similar way to passing through jets and τ_{had} are reconstructed with the same principle as jets, a fraction of detector-level tau-leptons might be "fakes". This occurs when a QCD jet is misidentified as τ_{had}. To estimate the efficiency of the τ_{had} identification, specialized Boosted Decision Trees (BDT) [92] and Recurrent Neural Networks (RNN) [93] algorithms are trained to estimate the likelihood of a given τ_{had} candidate to be a genuine τ_{had} object. Several working points (Tight, Medium, Loose, and VeryLoose) correspond to different tau identification efficiency [94].

The isolation of reconstructed τ_{had} is evaluated based on the core tracks and topological clusters. Additional requirements on a track-jet distance and a fraction of track momentum associated with jets are set [95].

There is a $\sim 5 - 10\%$ probability that a jet will pass the ID criteria at the working point of 90% real τ_{had} efficiency [96]. Therefore, the background modelling should properly account for the fakes. The $j \to \tau_{\text{had}}$ fake contribution is estimated with data-driven methods as will be mentioned in Section 6.

Further, a hadronically decaying tau-lepton may appear to be a misreconstructed light lepton. BDTs are used to quantify the probability that a given reconstructed τ_{had} is an actual electron. Eventually, the muon veto can be applied to reject the events where a muon is mistakenly identified as τ_{had}.

The efficiency of objects identification and tagging efficiency are introduced with corresponding scale factors (SFs).

Tau-leptons candidates with $p_{\text{T}} > 20$ GeV are selected for further event processing. They require to be reconstructed in the $|\eta| < 2.5$ pseudorapidity range. The transition region between the end-cap and central barrels of the calorimeters ($1.37 < |\eta| < 1.52$) is excluded as reconstruction efficiency is low in this range.

[9] Candidates for leptonically decaying tau-leptons τ_{lep} are found in events with reconstructed light leptons and missing transverse energy.

The absolute electric charge of tau-lepton is required to be a unit. One- and three-prong tau-leptons are considered where the number of prongs is defined by the number of the associated charged tracks. τ_{had} candidates are required to pass at least `Loose` ID requirements [76] before the overlap removal.

Additionally, in MC samples, reconstructed tau-leptons may need to match certain generator-level particles in order to avoid ambiguities in background estimation.[10]

5.6 Overlap removal

The preceding paragraphs explain how the objects are identified, reconstructed, and calibrated. The reconstructed objects also undergo selection to make sure that they have physical meaning and that their identification is efficient. At this step of the event reconstruction, however, they can share tracks and clusters and, hence, overlap geometrically. Therefore, a procedure of overlap removal (OLR) is performed to remove ambiguities when several objects are reconstructed based on the same detector signal.

The OLR procedure checks geometrical overlaps between electrons, muons, jets, and tau-leptons.[11] In addition, the energetic and topological properties of overlapping objects and a whole-detector picture can be used for decision-making. Firstly, particles which do not overlap with others, are labelled as passing. Then, actual overlap removal is performed in the following steps:

- If there are an electron and a muon in very close proximity to each other ($\Delta R_y < 0.01$) and the muon is calorimeter-tagged by calorimeters, the electron is kept.
- If two electrons share tracks and second-layer clusters, an electron with larger momentum is kept.
- If a τ_{had} candidate overlap with an electron within the $\Delta R_y < 0.2$ cone, the electron is kept.
- If a τ_{had} candidate overlap with a muon within the $\Delta R < 0.2$ cone, the muon is kept.
- If a calorimeter-tagged muon shares an Inner Detector track with an electron, the muon is removed. Then, electrons sharing an Inner Detector track with the remaining muons are removed.
- If a jet is in the $\Delta R = 0.2$ cone of an electron and $p_{\text{T}}^{e}/p_{\text{T}}^{\text{jet}} < 0.8$, the electron is removed.

[10] There are two collections available in MC: the particles at the generator level and the reconstructed objects at the detector level. Objects from both collections are geometrically matched. The nature and source of a given reconstructed object can then be learnt from the labels on its matched generator-level particle.

[11] The angular distance in terms of rapidity and azimuthal coordinate is used in the OLR: $\Delta R_y = \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$.

- The overlap of a jet with a muon within $\Delta R = 0.2$ is resolved as follows. If the jet has less than three associated tracks and some of these tracks are matched to the muon's Inner Detector tracks or if $p_{\mathrm{T}}^{\mu} / \sum p_{\mathrm{T}}^{\text{jet track}} > 0.7$ and $p_{\mathrm{T}}^{\mu} / p_{\mathrm{T}}^{\text{jet}} > 0.5$, the muon has precedence.
- An electron within a non-pileup jet's $\Delta R = 0.4$ cone is removed.
- A muon within a non-pileup jet's $\Delta R = 0.4$ cone is removed.
- If there is a jet within the cone $\Delta R = 0.2$ around a hadronically decaying τ, the jet is removed.

In this way, a given detected signal is assigned to a single physical object after the overlap removal.

Electron candidates surviving the OLR are required to pass the Medium (Tight) isolation WP in the search for excited tau-leptons (measurement of the $H \to \tau\tau$ cross section).[12]

5.7 Missing transverse energy

In the ATLAS experiment, the missing transverse energy is calculated with two terms: the so-called hard one and the soft one.

The hard term of MET corresponds to hard-scatter signals from reconstructed and calibrated particles (electrons, muons, hadronically decaying tau-leptons, photons) and jets.

The soft term is represented by low-p_{T} charged-particle tracks and soft clusters which are associated with hard-scatter vertex but not linked to any reconstructed hard-scatter object.[13] The soft term is reconstructed using charged and neutral particle flow objects (PFOs).

Then, MET is reconstructed as a negative sum of the transverse momenta of reconstructed hard objects (jets, hadronically decaying tau-leptons, electrons, muons) and a soft term. The latter quantifies the underlying events' activity. The (Figure 5.1a). The corresponding soft term can be reconstructed as a **T**rack-based **S**oft **T**erm (TST) or a **C**alorimeter **S**oft **T**erm (CST) [97].

The calculation of the MET vector, $\vec{E}_{\mathrm{T}}^{\text{miss}}$, is based on the assumption of energy balance in the transverse plane (Figure 5.1b) and expressed by equation

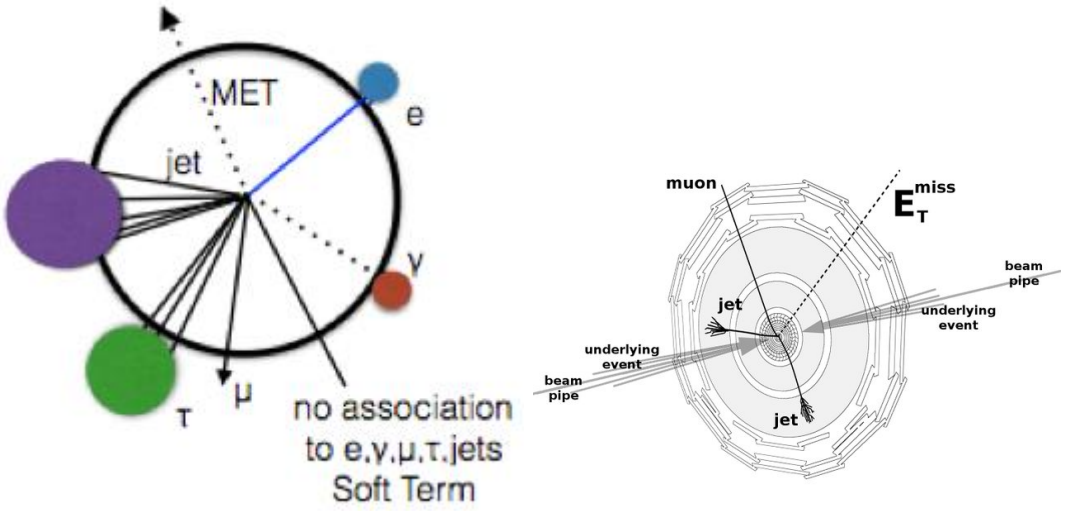
$$\vec{E}_{\mathrm{T}}^{\text{miss}} = -\sum_{\substack{\text{selected}\\\text{electrons}}} \vec{p}_{\mathrm{T}} - \sum_{\substack{\text{selected}\\\text{muons}}} \vec{p}_{\mathrm{T}} - \sum_{\substack{\text{selected}\\\tau_{\mathrm{h}}}} \vec{p}_{\mathrm{T}} - \sum_{\substack{\text{selected}\\\text{jets}}} \vec{p}_{\mathrm{T}} - \sum_{\substack{\text{soft}\\\text{tracks}}} \vec{p}_{\mathrm{T}} \tag{5.1}$$

The absolute value of $\vec{E}_{\mathrm{T}}^{\text{miss}}$ is commonly referred to as MET, $E_{\mathrm{T}}^{\text{miss}}$, in this thesis.

At early reconstruction steps, the MET is calculated by the `Athena` `METMaker` tool which takes original ("raw", i.e. non-calibrated) hard-scatter inputs and uses an association map between these objects. After overlap removal, MET is re-evaluated with the fully reconstructed, calibrated and filtered electrons, photons, hadronically decaying tau-leptons τ_{had}, muons, jets, and a soft term.

[12] The `Gradient` isolation WP [81] is applied to electron candidates in the $H \to \tau\tau$ analysis of the partial Run 2 data set based on the `Athena` release 20.7. This is discussed in Appendix C).

[13] The tracks and calorimeter clusters in the soft term have lower momentum $p_{\mathrm{T}} < 20$ GeV.



(a) (b)

Figure 5.1: The definition of the soft term is based on tracks and clusters not assigned to hard objects [98] (a). The missing transverse energy momentum is defined as an imbalance in the plane perpendicular to the beam axis [99] (b).

6. Fake background modelling

The definition of fake background covers the events where tau-leptons, electrons, and muons are mimicked by QCD jets or light leptons. QCD jets can mimic hadronically decaying tau-leptons (τ_{had}), and these events make up a sizeable fake background in final states with τ_{had}. The $j \to \tau_{\text{had}}$ fake background is typically estimated with data-driven techniques. Modelling of fake backgrounds with a small contribution (within $\approx 2\%$) relies on the Monte Carlo method.

6.1 Fakes of hadronically decaying tau-leptons

Multiple jets can be produced in pp collisions, and some jets can be misidentified as tau-leptons. Events with $j \to \tau_{\text{had}}$ become an important source of hadronically decaying tau-leptons fakes. Due to intricate MC modelling of multi-jet production, this fake $j \to \tau_{\text{had}}$ contribution is estimated with real data by the data-driven methods. The $H \to \tau\tau$ analysis and the search for excited tau-leptons with the complete Run 2 ATLAS data set employ the Fake Factor (FF) method [100] to predict $j \to \tau_{\text{had}}$ fakes.

The FF method presumes that fake rates are about the same in both phase spaces: the one prevalently containing studied types of events (a so-called *signal region*) and the one with a high contribution of fake tau-lepton (a so-called *control region*). The fake event rate is measured with real ATLAS data in the control region enriched with fake tau-lepton events. The FFs are parametrized as a function of pseudorapidity, transverse momenta, and number of tracks for both leading and subleading tau-leptons.[1] The obtained fake factor is then interpolated to the signal region. In this way, the contribution of events with fake tau-lepton is estimated. Sections 9.5 and 10.9 will discuss the FF method in application to the presented analyses.

Fake $j \to \tau_{\text{had}}$ background was modelled with the `notOS` ("not-opposite-sign") data-driven method in the $\tau_{\text{had}}\tau_{\text{had}}$ channel by the $H \to \tau\tau$ analysis with the partial Run 2 data set (collected in 2015-16) [4, 101] and in the first $H \to \tau\tau$ measurement by the ATLAS [3, 102]. The QCD fakes are derived in events where at least one τ_{had} candidate is not isolated (i.e. fail to pass the nominal isolation criteria). Furthermore, the two τ_{had} candidates have a combined electric charge not equal to -1 (for example, +1 or -3).

The fake $l \to \tau_{\text{had}}$ component is simulated by Monte Carlo, given its relatively small contribution. The reconstruction and identification algorithms effectively reduce the rate of $l \to \tau_{\text{had}}$ misidentification. Events with light leptons mimicking taus are estimated with Monte Carlo due to their contribution of typically within a few per cent. For instance, the BDT-based electron veto suppresses fakes due to misidentified electrons in regions where they are important. The cases of the overlap between a tau-lepton and a muon can be treated in the specified way at the stage of the final-state tau-lepton selection.

[1] The leading object is the one with the highest reconstructed transverse momentum p_{T} among all the reconstructed objects of this type. Thereupon, a subleading object has the second-largest transverse momentum.

6.2 Fakes of leptonically decaying tau-leptons

Misidentification of tau-leptons also occurs in their leptonic decay mode. A jet can be identified as an electron or muon. Also, non-prompt electrons and muons can be misidentified as products of tau-lepton decay. Tracks used for muon reconstruction can have other than muon decay origin. Similarly, calorimeter clusters can originate from sources other than electrons. The ambiguities are resolved by the overlap removal to some extent; however, residual misidentification can occur. Likewise, it is done for the fake τ_{had}, fake leptonically decaying tau-leptons are estimated with data-driven methods: the ABCD method [103, 104] and/or the matrix method (MM) [105, 106]. τ_{lep} fakes were estimated with the MM (the ABCD method) in $\tau_{\text{lep}}\tau_{\text{lep}}$ channel in the $H \to \tau\tau$ coupling analyses with the full Run 2 [107] and partial Run 2 [101] data sets.

7. Data analysis

7.1 Data management

Accepted by triggers, signals from the entire ATLAS detector are collected by the data acquisition (DAQ) system. The quality of data is controlled in real-time at the ATLAS Point 1.

A complete event is reconstructed based on the ATLAS detector's RAW inputs or the HITS simulated by Monte Carlo. After events recontruction, analysis object data (AOD) files are produced. Figure 7.1 depicts the ATLAS data processing chain.

The recorded ATLAS data set is highly extensive and reduced to lower-size derivation `DAOD`. The search for excited tau-leptons exploits `DAOD_PHYS` derivations, effectively becoming a unified format for ATLAS analyses since 2021. The Higgs boson to tau-leptons coupling measurement uses the dedicated `DAOD_HIGG4D2` format.

The actual analysis is performed on yet more succinct input files called NTuples. This approach ensures the computation is lightweight and modestly memory-consuming. NTuples for the excited tau-lepton search were produced with Prague's `A`thena **a**nalysis **c**ommon (`AAC`) framework [108].

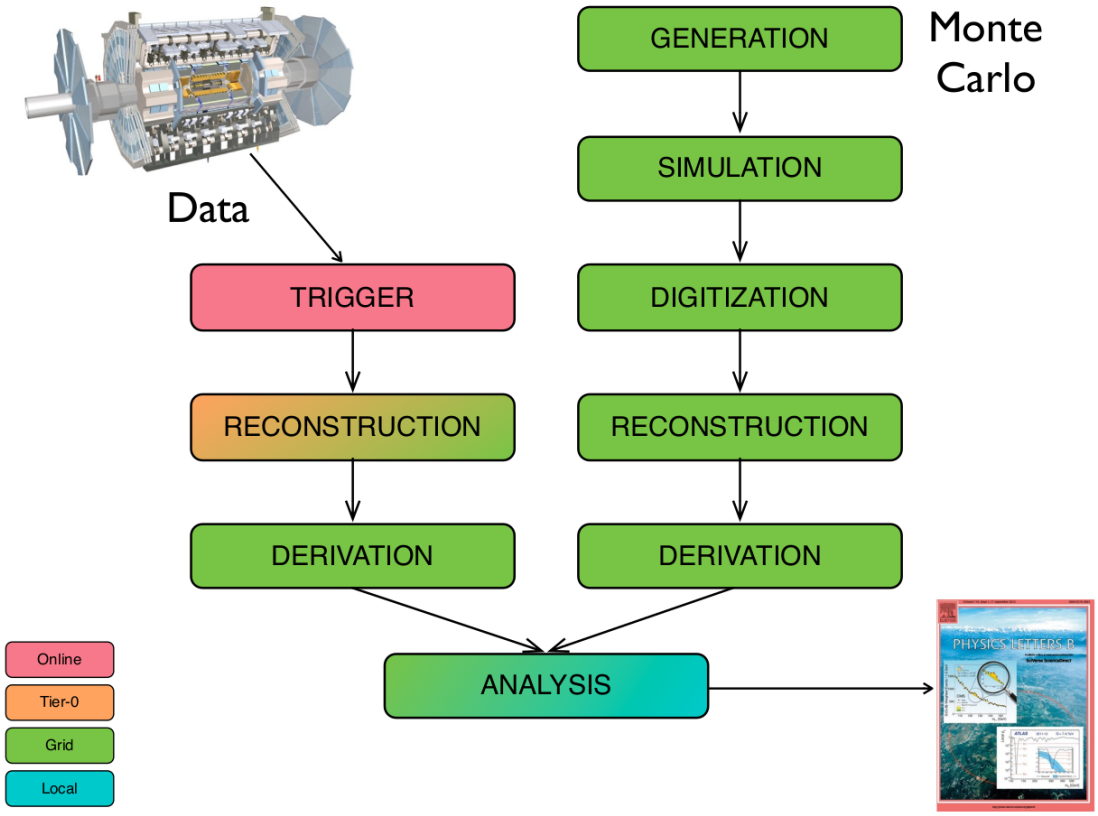


Figure 7.1: Preparation of real and MC simulated data in the ATLAS framework [109].

7.2 Requirements to ATLAS data

Depending on the status of particle beams and ATLAS sub-detectors, the recorded data can be used or not for physics analyses. Physics analyses are performed based on the event recorded during stable beam collisions. All the ATLAS detector sub-systems need to be ready for data-taking. A part of the detector can suffer from malfunctioning for a certain period of data-taking, and the data collected in the corresponding LBs or runs can be of lower quality. Therefore, only runs and luminosity blocks from the Good Run Lists (GRLs) are analysed for physics. The GRLs exclude the data collected when the ATLAS detector or its part was not fully operational, or proton beams were not stable.

The GRL data sets from 2015, 2016, 2017, and 2018 data-taking corresponds to the integrated luminosity of ~ 3.2, ~ 32.9, ~ 44.3, and ~ 58.5 fb^{-1}, respectively.

7.3 Baseline event selection

On top of that, a real-data event should satisfy the following criteria to ensure high data quality. When a part of the detector fails during the data-taking or the data collected are corrupted, a veto to the event should be applied. In particular, physics signatures are not looked up in events with noise bursts or data corruption in the LAr, data corruption in the TileCal, or data collected during the SCT recovery after single event upsets. The events with an incomplete record are not analysed for physics. As one is interested in physics signatures, an event (either from real data or MC simulation) should contain at least a single primary vertex with at least two associated tracks. The primary vertex is the most energetic one in terms of the sum $\sum p_{\mathrm{T}}^2$ of all the associated tracks with transverse momentum above 500 MeV. These event selection criteria are summarized in Table 7.1.

Hadronic jets is a dominant product of proton-proton collisions. As multiple jets can be reconstructed at a time, and non-collision background and detector noise can be falsely identified as jets. Therefore, an event-level jet cleaning is performed [110]. Rejection of fake-like jets is based on the evaluation of the sum of transverse momenta of tracks ($p_{\mathrm{T}} > 500$ GeV), jet's transverse momenta and maximal sampling fraction.

Further, the event is required to be accepted by an appropriate trigger. Events with two hadronically decaying tau-lepton candidates are required to be accepted by the lowest unprescaled di-tau trigger. Events with one or two τ_{lep} need to pass a corresponding single- or di-lepton trigger. The detailed conditions of acceptance by triggers and farther event selection will be described in the related sections on the $H \to \tau\tau$ and excited tau-lepton analyses.

criteria	all regions
ATLAS GRL	true
core flag	false
inner detector	false
LAr flag	false
TileCal flag	false
N_{PV}	$N_{\mathrm{PV}} >= 1$

Table 7.1: The list of event criteria applied to events in real data collected by the ATLAS detector.

7.4 Statistical analysis

Statistical analysis at the LHC is the main instrument to conclude with a certain degree of confidence in what is observed in collected data. There are two main strategies for data analysis at the ATLAS experiment: measurements of the Standard Model parameters and searches for the physics Beyond the Standard Model.

The analyses presented in this thesis are performed with the `HistFitter` [111], a software framework for statistical data analysis commonly used in high-energy physics. Both analyses presented in this thesis are designed as counting experiments. They rely on the measurement of the observable(s) which are sensitive to the studied signal in the phase space which is expected to be enriched in the studied events.

Discriminative variables (test statistics) are chosen based on the expected signal significance and/or the power of signal-to-background separation.[1] Those are S_{T}, a scalar sum of di-tau and dijet transverse momenta, and di-tau invariant mass $m_{\tau\tau}^{\mathrm{MMC}}$ in the excited tau-lepton search and the $H \to \tau\tau$ coupling measurement, respectively.

The measurement of the parameters of interest is performed in one or several signal regions (SRs), i.e. the phase space where the signal is expected. Figure 7.2 depicts a typical flow of an analysis made with the `HistFitter`. Background processes need to be understood as much as possible in SR(s) in order to achieve high accuracy.

A control region is formed by a set of topologies where a given SM background is prevalent. A perfect control region includes a minimal fraction of events with studied signatures. To validate the SM background prediction in the SR, CRs should collect events with topologies similar to the signal signatures.

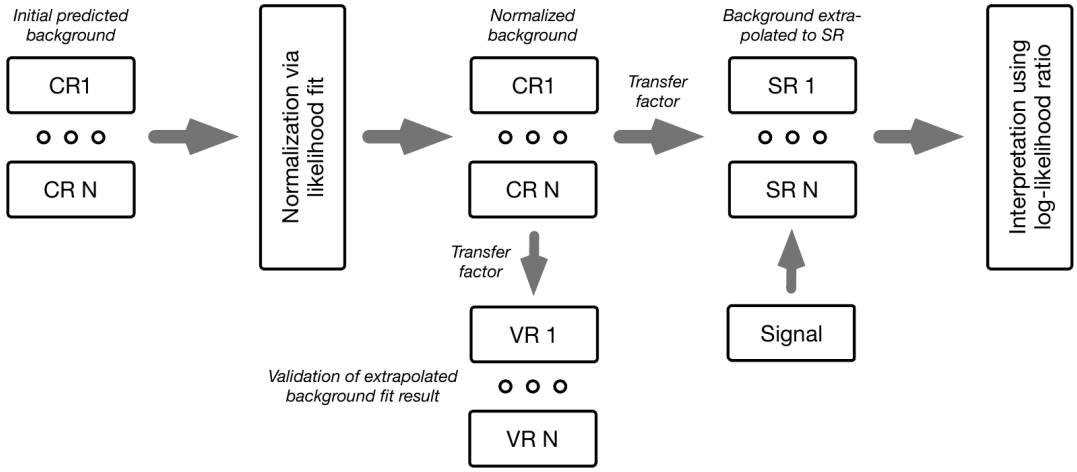


Figure 7.2: A strategy of an analysis with `HistFitter` [112]. Background modelling is constrained in control regions (CRs), and its prediction is extrapolated to the signal regions (SRs).

A set of cut-based selections allows distinguishing signal events from background ones (in the high-energy physics jargon, the signal and control regions

[1] These methods are described below.

are orthogonal). Background modelling for a given process is compared with real data. Based on the observations in a given control region, a template of that SM background process is built and then extrapolated to the signal region.

In addition, validation regions can be introduced to check the consistency between a background model and data. Validation regions typically contain events whose kinematics are more similar to those in signal regions than those in control regions, while still maintaining orthogonality between the control regions and signal regions.

Hypothesis testing targets to estimate the parameters of interest (POIs). Those are usually cross section of particle production and decay or signal strength. The latter is defined as the ratio between observed and theory-predicted cross section $\mu = \sigma^{\text{obs}}/\sigma^{\text{theory}}$. In both analyses presented, such a scheme with a single parameter of interest (μ) is used.

7.5 Optimization with respect to the studied signal

7.5.1 The method of ROC curve

The Receiver Operational Curve [113] shows the efficiency of acceptance of one kind of event (here called A) and simultaneous rejection of another (here called B). In the non-realistic case of a perfect separation between signals A and B, the ROC has a rectangular shape: signal A is always accepted or signal B is always rejected. With the ideal signal-to-background separation, the area under the curve (AUC) has the maximal possible value (1). In real experiments, the shape is concave or convex, and the AUC is below 1. The larger the AUC value is, the better separation of a signal from a background is achieved.

The method of the ROC curve was widely used in this thesis. For instance, the performance of the di-tau mass estimator was estimated with AUC in the $H \to \tau\tau$ analysis. The choice of the final fit variable in the search for excited tau-leptons was made based on the ROC comparison. The discussed method optimized the event selection in the search for excited tau-leptons.

7.5.2 Signal significance

In a large enough data set, the signal significance can be approximated by the following relation [114]:

$$Z = S/\sqrt{B} \tag{7.1}$$

The integral signal significance can be estimated as a sum in quadrature of per-bin $S/\sqrt{B}$ ratios over all the bins in a given analysed histogram in the following way:

$$Z = \sqrt{\sum_{i \in \text{histobins}} S_i^2/B_i} \tag{7.2}$$

For example, the integral signal Z was used as an auxiliary figure of merit in studies of signal separation in the ETL search.

7.5.3 Uncertainties

The results of the statistical analysis in general include statistical and systematic uncertainties. The statistical uncertainty is a consequence of the limited size of the experimentally collected or MC-simulated data set. The systematic uncertainties arise as experimental and theoretical uncertainties of the measurement. Experimental uncertainties are imminent due to the finite resolution of object reconstruction and uncertainties on the scale factors. Thus, resolution and scale uncertainties result in smeared objects four-momenta. In the ATLAS experiment, this source of uncertainty is often called kinematic and provided by the `Athena` smearing tools. As mentioned in Section 4, a weight is assigned to the MC-generated event. This weight might include a scale factor which corrects for identification, reconstruction, and trigger (in)efficiencies in order to make MC prediction compatible with real data. There are "weight" uncertainties imposed on these SFs.

The predicted cross section on the SM and BSM processes has uncertainties due to approximations and assumptions made in the theoretical model and MC simulation. This source of uncertainty is often called a theory one. The uncertainty on the size of the collected (or simulated) data sets impacts the statistical test as well.

7.6 Model PDF

The histograms with the discriminative variable distributions (so called workspace input) are provided to the analysis framework. The input data are parametrised by a probability density function (PDF) deduced by the `HistFactory` [115] tool.

The expected ν_{cb} number of events in a given bin can be parametrized with a set of parameters, such as, for instance, normalization factors, systematic uncertainties, and statistical uncertainties.

Taking into account multiple control and signal regions, MC statistical uncertainties $\vec{\gamma}$ and systematic uncertainties $\vec{\alpha}$, the likelihood to obtain certain data in n independent measurements is given by the formula [114]:

$$L(\vec{\mu},\vec{\gamma},\vec{\alpha}) = \prod_i Pois\{n_i;[\sum_p \mu_p(\nu_i^p + \sum_k \Delta_{ik}^p(\alpha_k))]\gamma_i\} Pois(m_i;m_i\gamma_i) \prod_k N(\alpha_k;0,1). \tag{7.3}$$

Here the i index runs over all bins of the input histograms, the p index loops over contributing processes, the k index denotes sources of systematic variations, n_i is the ith bin content, ν_i^p is the expected number of events in the ith bin for the pth process. m_i expresses the effective number of events in the ith bin in the histogram for the process p as estimated using the MC simulation.[2]. The Poissonian and Gaussian terms constrain the systematic and statistical uncertainties as discussed before.

The model parameters are defined by maximizing the likelihood function, i.e.

$$\frac{\partial L(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}} = 0 \tag{7.4}$$

[2]The effective number of events is defined with respect to weights assigned to each from N contributing events: $N_{\text{eff}} = (\sum_{i=1}^N \omega_i)^2 / \sum_{n=1}^N \omega_i^2$

Here all NPs are denoted through a set θ. This brings the maximum likelihood estimators for the POI $\hat{\mu}$ and the nuisance parameters $\hat{\theta}$. The best-fit estimates for the POIs and the NPs are usually found through a negative natural logarithm of the likelihood (an NLL function). For given values of the PoI(s), other NPs can be chosen in such a way that the resulting likelihood is maximized $L(\mu, \hat{\hat{\theta}}(\mu))$. This procedure is called profiling and the estimated likelihood is named a profile likelihood.

According to Neyman-Pearson lemma [116, 117], an optimal test statistic is a likelihood ratio. Accordingly, a hypothesis is tested at a given μ value in the following manner:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} \tag{7.5}$$

Here $L(\hat{\mu}, \hat{\theta})$ is an unconditional likelihood estimated with the best-fit values and $L(\mu, \hat{\hat{\theta}}(\mu))$ is a conditional likelihood with the NPs estimated for a given μ.

7.6.1 Discovery fit

In searches for BSM physics and SM precision measurements, a null hypothesis is usually a pure SM prediction (i.e. the background-only hypothesis). The alternative hypothesis is a combination of the SM background and a signal. In this setup, physicists compare the real data with the $\mu = 0$ assumption.

The likelihood function in the precision SM measurement and analysis for the discovery of new particles takes the following form:

$$q_0 = \begin{cases} -2 \ln \lambda(0), & \hat{\mu} > 0 \\ 0, & \hat{\mu} \leq 0 \end{cases} \tag{7.6}$$

The NLLs are calculated for $\mu = 0$ and $\hat{\mu}$ values, and $\ln(\max L(\mu = 0) / \ln \max L(\hat{\mu}))$ is used as a test statistic.

The p_0 value quantifies how much the null hypothesis is compatible with real data. The following integral defines it:

$$p_0 = \int_{q_0 \text{ obs}}^{+\inf} f(q_0|0) \mathrm{d}q_0 \tag{7.7}$$

The observation is statistically significant if the p value for the background-only hypothesis is as low as 2.9×10^{-7}.

Another value – significance – can express the p_0-valu. The significance is a number of standard deviations $Z_0 = \Phi^{-1}(1 - p_0)$ where Φ is a quantile of a cumulative of the standard Gaussian distribution with the zero mean and the unit variance $N(\alpha_k; 0, 1)$.

Based on Wilk's theorem [118], the χ^2 Pearson distribution asymptotically approximates the test statistics in a large data set. The p_0 value can then be extracted as follows: $Z_0 \approx \sqrt{q_0}$. The Z_0 significance quantifies the degree of confidence that scientists can reject the null hypothesis and claim the signal existence.

The significance Z states the level of their confidence in the observation of a new particle. For the first time, in 2012, the ATLAS experiment observed the

Higgs boson in di-photon events with a significance of 4.5σ (Figure 7.3b [119]). In this case, the obtained p_0 is low enough for the 126 GeV Higgs boson signal tested. It indicates that the null hypothesis can not reliably describe the data, and the alternative hypothesis is preferred.

7.6.2 Exclusion fit

The BSM physics exclusion is done by testing a null BSM+SM hypothesis against an alternative SM-only hypothesis. Of course, BSM searches aim to discover BSM effects, and this is achieved in the statistical analysis of a null SM hypothesis and an alternative extraSM+SM hypothesis. Afterwards, a given null hypothesis is rejected or accepted with a certain significance level.

Searches for new particles often reported limits on BSM parameters with a certain confidence level.

Upper limits on parameters are set in statistical tests with the following likelihood functions:

$$q_\mu = \begin{cases} -2\ln\lambda(\mu), & \hat{\mu} \leq \mu, \\ 0, & \hat{\mu} > 0 \end{cases} \tag{7.8}$$

For the exclusion purposes, likelihood is maximized for $\mu = 1$ and $\hat{\mu}$ POI values and the test statistics takes a form of $\ln\max L(\mu = 1) - \ln\max L(\hat{\mu})$. Similarly to the discovery significance Z_0, the significance with which the $b + \mu s$ hypothesis is rejected can be approximated based on Wilk's theorem as follows: $Z = \sqrt{q_\mu}$.

The p-value obtained for the hypothesized signal with the strength μ is estimated by equation $p_\mu = 1 - \Phi(\sqrt{q_\mu})$ [114]. The signal hypothesis is excluded if the p_μ value is lower than 0.05 with a 95% confidence level.

The following examples illustrate limit setting in the HEP analyses. The search by the LEP from 2002 [120] rejected the existence of the Higgs boson in the mass range below 114.1 GeV (Figure 7.3a). The search performed by the CMS collaboration [121] excluded the existence of excited electrons with mass below ~ 5.6 GeV with 95% CL.[3]

7.6.3 Confidence level limits

. The confidence level (CL) [122, 123] allows for estimating a variable of interest within a defined interval compliant with observation. The method evaluates the probability that the observed data conform with a given hypothesis i, namely:

$$\mathrm{CL}_b \equiv P(L \leq L_o bs) \quad \text{for background-only hypothesis} \tag{7.9}$$

$$\mathrm{CL}_{s+b} \equiv P(L \leq L_o bs) \text{ for background+signal hypothesis} \tag{7.10}$$

$$\tag{7.11}$$

The signal confidence level The $\mathrm{CL}_s = \mathrm{CL}_{s+b}/\mathrm{CL}_b$ quantifies to which extent the observed data is unlikely to conform to s but are rather compatible with b. In the HEP community, the S hypothesis is excluded with a 95% confidence level

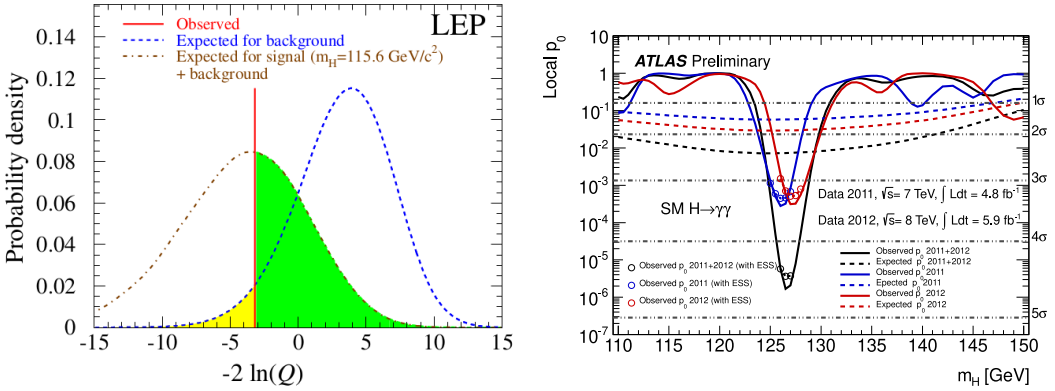
[3]This analysis considers that excited tau-leptons are produced and decay in the contact interaction at the compositeness scale $\Lambda = 10$ TeV. These parameters will be introduced in Chapter 10 dedicated to the search for excited tau-leptons.

if the following condition holds: $\mathrm{CL}_s \leq 0.05$. The following expression for CL_s is valid in the case of low sensitivity to the signal [114]:

$$\mathrm{CL}_s = \frac{p_\mu}{1 - p_0} \tag{7.12}$$

where $p_\mu \equiv p_{s+b} = \int_{q_\mu^{\mathrm{obs}}}^{+\infty} f(q_\mu|\mu)\mathrm{d}q_\mu$ and $p_0 \equiv p_b = \int_{-\infty}^{q_\mu^{\mathrm{obs}}} f(q_\mu|0)\mathrm{d}q_\mu$.

Figure 7.3a demonstrates the distribution of the test statistic in the search for Higgs boson $m_H = 116.6$ GeV at the LEP [124]. The yellow shaded area describes the compatibility with the background hypothesis. The compatibility with the background+signal hypothesis is depicted with the sector shaded in green.



(a) (b)

Figure 7.3: The test statistic distribution in the search for SM Higgs boson $m_H = 115.6$ GeV by the LEP working group [124] (a). The p_0-value in dependence on the assumed Higgs boson mass in a search in a $\gamma\gamma$ decay channel with a partial Run 1 data set collected at $\sqrt{s} = 7$ and 8 TeV by the ATLAS collaboration [119] (b).

8. Di-tau mass reconstruction

8.1 Di-tau decay

A pair of tau-leptons is a common system in the final states for both analyses introduced in this thesis. A characteristic variable of di-tau is its invariant mass $m_{\tau\tau}$.

Each of the tau-leptons undergoes subsequent decay to leptons (35.2%) or hadrons (64.8%) with the mean tau-lepton life time of 2.9×10^{-13} s ($c\tau \sim 85$ μm) [125]. There are three di-tau final states: fully leptonic $\tau_{\text{lep}}\tau_{\text{lep}}$, semi-leptonic $\tau_{\text{lep}}\tau_{\text{had}}$, and fully hadronic $\tau_{\text{had}}\tau_{\text{had}}$. They are observed in 12%, 46% and 41% of decays, respectively.[1] There are two neutrinos in the fully hadronic di-tau decay. The semi-leptonic and leptonic decays bring three and four neutrinos, respectively. Due to the very low interaction cross section, neutrinos escape the detector material so that their four-momenta are not reconstructed directly. Only a part of the di-tau decay products is visible, i.e. measurable by the ATLAS detector. Thus, the reconstruction of di-tau mass is not a trivial task.

In hadron colliders, the missing transverse energy is a common approximation for neutrino energy. This approach is adequate if colliding particles are strictly aligned along the beam axis. Such a configuration implies a zero momentum in the plane transverse to the beam axis. In that case, a momentum imbalance in the direction perpendicular to the beam direction signifies missing transverse energy (Figure 5.1b).

Di-tau invariant mass estimators can use any approximation for neutrino energy. There are also methods which rely on the quantities directly measured by the detector.

8.2 Overview of existing methods

Partial or approximate estimators for $m_{\tau\tau}$ employ information on the visible tau-lepton decay products[2] and missing transverse energy.

The methods described below are widely used in event selection and its optimization. They also provide a good proxy of di-tau reconstructed mass in measurements and searches for BSM physics, such as the search for lepton flavour violation in $H \to l\tau$ decay [126] and the search for excited tau-leptons presented in this thesis. However, precision measurements, including Higgs boson cross section measurements, would demand higher accuracy. Di-tau mass is estimated with the Missing Mass Calculator (see Section 8.2.5) in the $H \to \tau\tau$ coupling measurement.

[1] The $\tau_{\text{lep}}\tau_{\text{lep}}$, $\tau_{\text{lep}}\tau_{\text{had}}$, and $\tau_{\text{had}}\tau_{\text{had}}$ di-tau decay modes are occasionally called lep-lep, lep-had and had-had channels, respectively, in this thesis.

[2] It is defined by the four-momentum of visible τ_{had} decay products (light-leptons) in hadronic (leptonic) tau-lepton decay channel.

8.2.1 Visible mass

A visible di-tau mass is calculated as a mass of the visible leading and subleading tau-lepton decay products four-momenta ($\mathcal{P}^{\text{vis}}_{\tau 0}$ and $\mathcal{P}^{\text{vis}}_{\tau 1}$, respectively). This is given by equation

$$m^{\text{vis}}_{\tau\tau} = \sqrt{(\mathcal{P}^{\text{vis}}_{\tau 0} + \mathcal{P}^{\text{vis}}_{\tau 1})^2} \tag{8.1}$$

8.2.2 Effective mass

An effective mass is an immediate sum of the four-momenta of all the objects in the final state of di-tau decay, both visible and invisible. It is defined by equation

$$m^{\text{eff}}_{\tau\tau} = \sqrt{(\mathcal{P}^{\text{vis}}_{\tau 0} + \mathcal{P}^{\text{vis}}_{\tau 1} + \mathcal{P}^{\text{miss}})^2} \tag{8.2}$$

in which $\mathcal{P}^{\text{vis}}_{\tau 0}$ is the four-momentum for the visible decay products of leading (subleading) tau-lepton. $\mathcal{P}^{\text{miss}} := (E_{\text{T}}, -E_{\text{T}x}, -E_{\text{T}y}, 0)$ approximates the contribution from invisible decay products through the MET four-momentum.

8.2.3 Transverse mass

The transverse energy for a system of two particles with three-momenta $\vec{a}$ and $\vec{b}$ can be generally defined as follows:

$$m_{\text{T}}(\vec{a}, \vec{b}) = \sqrt{2p_{\text{T}}(\vec{a})p_{\text{T}}(\vec{b})(1 - \cos\Delta\phi(\vec{a}, \vec{b}))} \tag{8.3}$$

where tau-lepton mass is neglected. The transverse mass of a di-tau system is reconstructed by pairing three final state objects (leading tau-lepton τ_0, subleading tau-lepton τ_1 and MET) with the formula

$$m^{\text{total}}_{\text{T}} = \sqrt{m^2_{\text{T}}(E^{\text{miss}}_{\text{T}}, \tau_0) + m^2_{\text{T}}(E^{\text{miss}}_{\text{T}}, \tau_1) + m^2_{\text{T}}(\tau_0, \tau_1)} \tag{8.4}$$

8.2.4 Collinear mass approximation

The method of collinear mass approximation (CMA)[127] relies on the assumption that visible and invisible tau-lepton decay products are collinear: $\phi^{\text{miss}} \approx \phi^{\text{vis}}$, $\theta^{\text{miss}} \approx \theta^{\text{vis}}$. Further, the CMA method assumes neutrinos are the only source of the missing transverse energy. The components of missing transverse energy $E^{\text{miss}}_{\text{T}x}$ and $E^{\text{miss}}_{\text{T}y}$ can be described by the following system [128]:

$$\begin{array}{rcl} E^{\text{miss}}_{\text{T}x} & = & p^{\text{miss}}_0 \sin\theta^{\text{vis}}_0 \cos\phi^{\text{vis}}_0 + p^{\text{miss}}_1 \sin\theta^{\text{vis}}_1 \cos\phi^{\text{vis}}_1 \\ E^{\text{miss}}_{\text{T}y} & = & p^{\text{miss}}_0 \sin\theta^{\text{vis}}_0 \sin\phi^{\text{vis}}_0 + p^{\text{miss}}_1 \sin\theta^{\text{vis}}_1 \sin\phi^{\text{vis}}_1 \end{array} \tag{8.5}$$

The fraction of the energy taken by neutrinos can be factorized out (Figure 8.1), and system (8.5) can be simplified as follows:

$$m_{\tau\tau} = \frac{m^{\text{vis}}_{\tau\tau}}{\sqrt{x^{\text{coll}}_0 x^{\text{coll}}_1}} \tag{8.6}$$

where collinear fractions x^{coll}_0 and x^{coll}_1 are the momentum fractions carried away by the visible decays of the leading (subleading) tau-lepton: $x^{\text{coll}}_{0(1)} = p^{\text{vis}}_{0(1)}/(p^{\text{vis}}_{0(1)} + p^{\text{miss}}_{0(1)})$ and the di-tau visible $m^{\text{vis}}_{\tau\tau}$ as defined in Eq. (8.1).

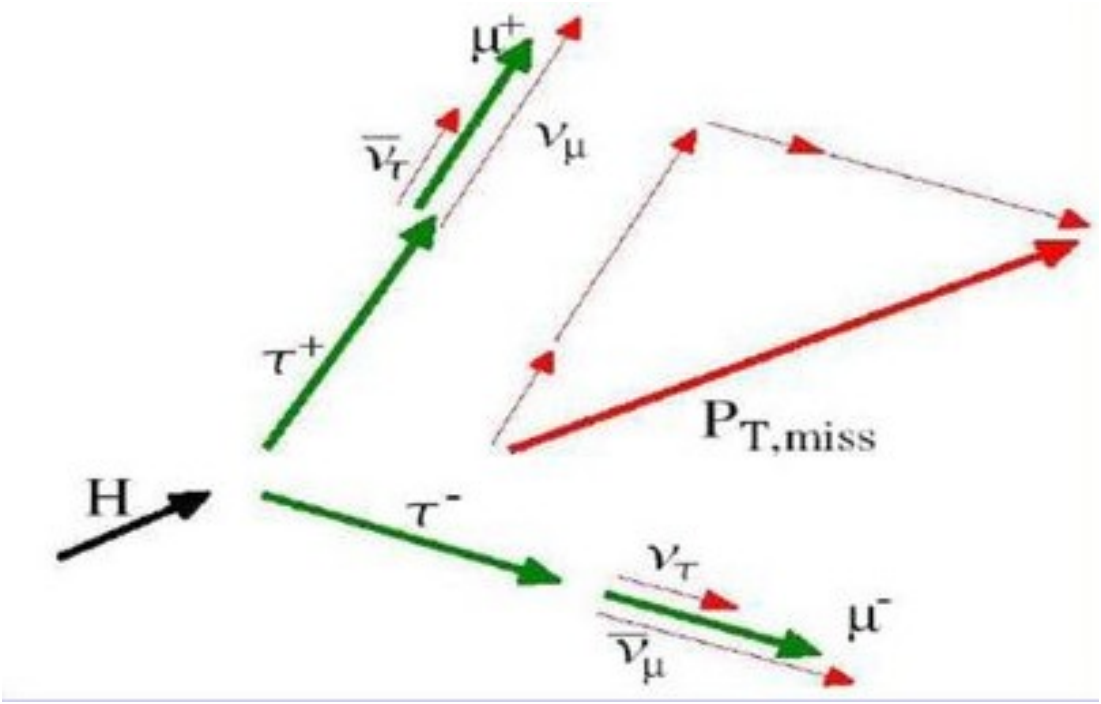


Figure 8.1: The collinear approximation in the Higgs boson decay to two tau-leptons [129]. The $\tau_\mu\tau_\mu$ decay is illustrated, with both neutrinos and muons assumed to be collinear. Accordingly, the missing transverse energy is assumed to be placed between the muons.

Solutions for the system of the equations (8.5) have physical meaning when the condition of collinearity between tau-leptons is satisfied. In general, the collinear mass approximation can be applied for di-tau mass calculation in the $H \to \tau\tau$ analysis. In the $H \to \tau\tau$ decay, the assumption on the collinearity between visible and invisible products is valid as $m_\tau \ll m_H$. Section 9.1.2 discussed limitations of the CMA for the $m_{\tau\tau}^{\text{coll}}$ estimation in the $H \to \tau\tau$ analysis.

The CMA also has limitations in the case of MET has much higher values than the τ_{had} transverse momentum. For example, the solutions where estimated collinear fractions x_0^{coll} and x_1^{coll} have a negative sign can be unphysical as this would imply negative neutrino energies. The collinear fraction values close to zero would lead to divergence in the form of infinite neutrino momentum (as stated $x_i^{\text{coll}} \sim 1/p_i^{\text{miss}}$ Eq.(8.6)). In this sense, the output values of x_0^{coll} and x_1^{coll} can help to distinguish events with large MET, e.g. overestimated in the result of jet misreconstruction and $j \to l$ misidentification. Introduction of thresholds on the collinear fractions x_0^{coll} and x_1^{coll} also removes non-physical solutions.

8.2.5 Missing Mass Calculator

The Missing Mass Calculator [128] is an advanced likelihood-based technique for invariant mass estimation in semi-visible final states. It was invented for di-tau mass calculation at the CDF in Fermi Lab. The method is used for mass reconstruction of Higgs boson in $H \to \tau\tau$ decay in the ATLAS Collaboration [4, 5]. The technique relies on knowledge of tau-lepton decay kinematics, which allows preferring one solution over another. The MMC takes into account that the

measured energy of visible decay products and the MET has a finite resolution. The MMC technique accounts for the kinematic constraints and likelihood while considering the energy variation and position of the particles in the decay cascades over the allowed phase space.

The kinematics of di-tau decay is described by the following under-constrained system of equations [128]:

$$\begin{aligned}
E_{\mathrm{T}x}^{\mathrm{miss}} &= p_0^{\mathrm{miss}} \sin\theta_0^{\mathrm{miss}} \cos\phi_0^{\mathrm{miss}} + p_1^{\mathrm{miss}} \sin\theta_1^{\mathrm{miss}} \cos\phi_1^{\mathrm{miss}} \\
E_{\mathrm{T}y}^{\mathrm{miss}} &= p_0^{\mathrm{miss}} \sin\theta_0^{\mathrm{miss}} \sin\phi_0^{\mathrm{miss}} + p_1^{\mathrm{miss}} \sin\theta_1^{\mathrm{miss}} \sin\phi_1^{\mathrm{miss}} \\
m_{\tau 0}^2 &= (m_0^{\mathrm{miss}})^2 + (m_0^{\mathrm{vis}})^2 + 2\sqrt{(p_0^{\mathrm{vis}})^2 + (m_0^{\mathrm{vis}})^2}\sqrt{(p_0^{\mathrm{miss}})^2 + (m_0^{\mathrm{miss}})^2} \\
&\quad -2p_0^{\mathrm{vis}} p_0^{\mathrm{miss}} \cos\Delta\theta_0^{\mathrm{vis,miss}} \\
m_{\tau 1}^2 &= (m_1^{\mathrm{miss}})^2 + (m_1^{\mathrm{vis}})^2 + 2\sqrt{(p_1^{\mathrm{vis}})^2 + (m_1^{\mathrm{vis}})^2}\sqrt{(p_1^{\mathrm{miss}})^2 + (m_1^{\mathrm{miss}})^2} \\
&\quad -2p_1^{\mathrm{vis}} p_1^{\mathrm{miss}} \cos\Delta\theta_1^{\mathrm{vis,miss}}
\end{aligned} \tag{8.7}$$

in which $E_{\mathrm{T}x}^{\mathrm{miss}}$ and $E_{\mathrm{T}y}^{\mathrm{miss}}$ are MET x- and y-components; $m_{0(1)}^{\mathrm{vis}}$ and $p_{0(1)}^{\mathrm{vis}}$ are the mass and the absolute momentum of the combined visible leading (subleading) tau-lepton decay products (i.e. hadrons, electrons, or muons); $m_{0(1)}^{\mathrm{miss}}$ and $p_{0(1)}^{\mathrm{miss}}$ is the mass and the magnitude of the combined invisible leading (subleading) tau-lepton decay products (one or two neutrinos). The angular variables ϕ and θ stand for the azimuthal and polar angle, respectively. $\Delta\theta_{0(1)}^{\mathrm{vis,miss}}$ is the polar separation between vectors of visible and invisible leading (subleading) tau-lepton decay products. Depending on tau-lepton decay mode, the system Eq. (8.7) includes from six ($\tau_{\mathrm{had}}\tau_{\mathrm{had}}$) to eight ($\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$) unknowns: the mass and the momenta of the final-state neutrinos, $m_{0(1)}^{\mathrm{miss}}$ and $\vec{p}_{0(1)}^{\mathrm{miss}}$.

There are more unknowns than constraints for di-tau decays, and to scan over possible values of unknowns is made. For each event, a scan over possible configurations of the visible and invisible tau-decay products is performed in a Markov chain.

If both tau-leptons decay hadronically, a single neutrino is produced and $m_{0(1)}^{\mathrm{miss}}$ can be taken as zero and scan over kinematically allowed states of the $(\phi_0^{\mathrm{miss}}, \phi_1^{\mathrm{miss}})$ grid is performed. At any point of the $(\phi_0^{\mathrm{miss}}, \phi_1^{\mathrm{miss}})$ grid, the vectors $\vec{p}_{0(1)}^{\mathrm{miss}}$ (Eq.(8.7)) can be then calculated precisely. A given solution is weighted with the probability to observe the current distance between visible and invisible products, $\mathcal{P}(\Delta R_{0(1)}^{\mathrm{vis,miss}})$. The likelihood of $\Delta R_{0(1)}^{\mathrm{vis,miss}}$ is estimated in well-known processes like $Z \to \tau\tau$ and parametrised as a function of the $p_{\mathrm{T}}^{\mathrm{vis}}$, and the tau-lepton decay mode or the number of associated tracks (named prongness).[3] Further, information on tau-leptons energy and MET resolution can be exploited in the likelihood estimation by including their PDFs, $\mathcal{P}(E_{0(1)}^{\mathrm{vis}})$ and $\mathcal{P}(E_{\mathrm{T}}^{\mathrm{miss}})$. Knowledge of the tau-lepton decay kinematics is used to account for the τ_{had} energy smearing. The data on the MET resolution allows for estimating the likelihood of the MET value in the current configuration.[4] In result, solution of Eq. (8.7) are found in a dense grid of $\phi_{0(1)}^{\mathrm{miss}}$, $E_{0(1)}^{\mathrm{vis}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$. In the case of leptonically decaying tau-leptons, a scan over the mass of final-state neutrinos, $m_{0(1)}^{\mathrm{miss}}$, provides additional constraints.

[3]The decay mode gives information on the number of charged and neutral particles reconstructed in the τ_{had}. The decay mode variable is provided by the `PanTau` classifier.

[4]A detailed definition of the MET probability term $\mathcal{P}(E_{\mathrm{T}}^{\mathrm{miss}})$ will be given below, in Section 9.1.4.

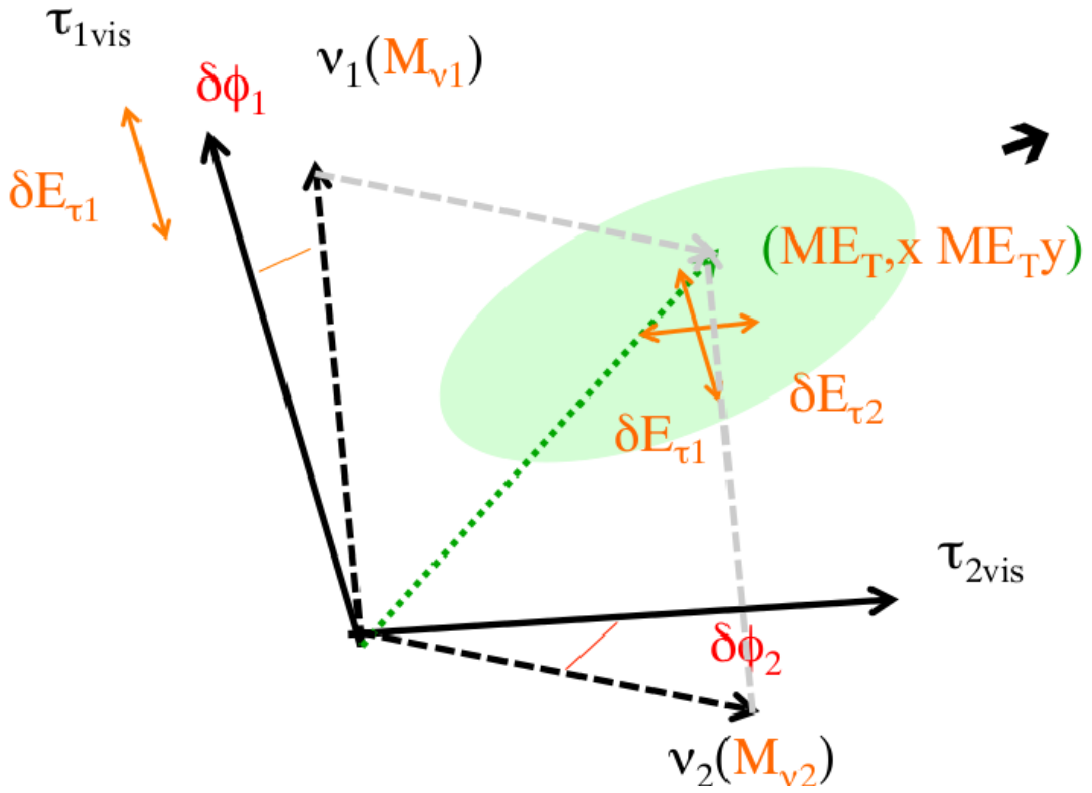


Figure 8.2: The principle of the multidimensional MMC scan of the kinematically allowed phase space in an event [98].

In each point, a solution for $m_{\tau\tau}$ is weighted with the likelihood of that given configuration of visible and invisible tau-leptons decay products (Figure 8.2). For each tried topology, the final weight is defined as a log-likelihood of its total probability, i.e. [5]

$$\mathcal{L} = -\lg(\mathcal{P}_{\text{total}}) = -\lg\Big(\mathcal{P}(\Delta R_0^{\text{vis,miss}}) \times \mathcal{P}(\Delta R_1^{\text{vis,miss}}) \times \mathcal{P}(E_{\text{T}}^{\text{miss}}) \times \mathcal{P}(E_0^{\text{vis}}) \times \mathcal{P}(E_1^{\text{vis}}) \times \mathcal{P}(m_0^{\text{miss}}) \times \mathcal{P}(m_1^{\text{miss}})\Big) \quad (8.8)$$

The final PDF is used to choose the most likely configuration of the di-tau decay products. In a particular event, the solution with the highest likelihood and minimal $-\lg(\mathcal{P}_{\text{total}})$ is taken as a final estimator of the Higgs boson mass.[6] In this way, the total di-tau invariant mass $m_{\tau\tau}^{\text{MMC}}$ is fully defined by the MMC algorithm.

The illustration: from knowledge to application

Figure 8.3 gives a glimpse of the MMC reconstruction method in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. Here, the Higgs boson production in vector boson fusion is illustrated.[7] The pair of two forward jets with a significant dijet invariant mass is characteristic

[5] The formula generalises the likelihood estimation: it describes the PDFs in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel, and in the case of hadronic decay mode, the corresponding m_i^{miss} term becomes the unit.

[6] The logarithm to the base ten is taken.

[7] Higgs boson production modes will be discussed later, in Section 9.1.

of the Higgs boson production via vector boson fusion. In this illustration, one tau-lepton decays to a muon and neutrinos, while the other decays to hadrons and a neutrino. As neutrinos are assumed to be the only source of the missing transverse energy, one expects the MET to be placed between the muon and τ_{had}. The $\tau_\mu\tau_{\text{had}}$ pair recoils against two jets so that it is likely to be placed between jets in terms of pseudorapidity and have a smaller η opening angle due to recoil. Furthermore, a suggestion on the visible energy of τ_{had} and muon can be applied. In VBF production mode, the produced Higgs boson is nearly at rest. As two neutrinos are produced in the leptonic muon decay, and a single ν_τ neutrino originates from the hadronic tau-lepton decay, the measured momentum of τ_{had} is preferably smaller than the muon's p_{T}. These facts enable physicists to estimate the likelihood of a given spatial and energetic final-state configuration.

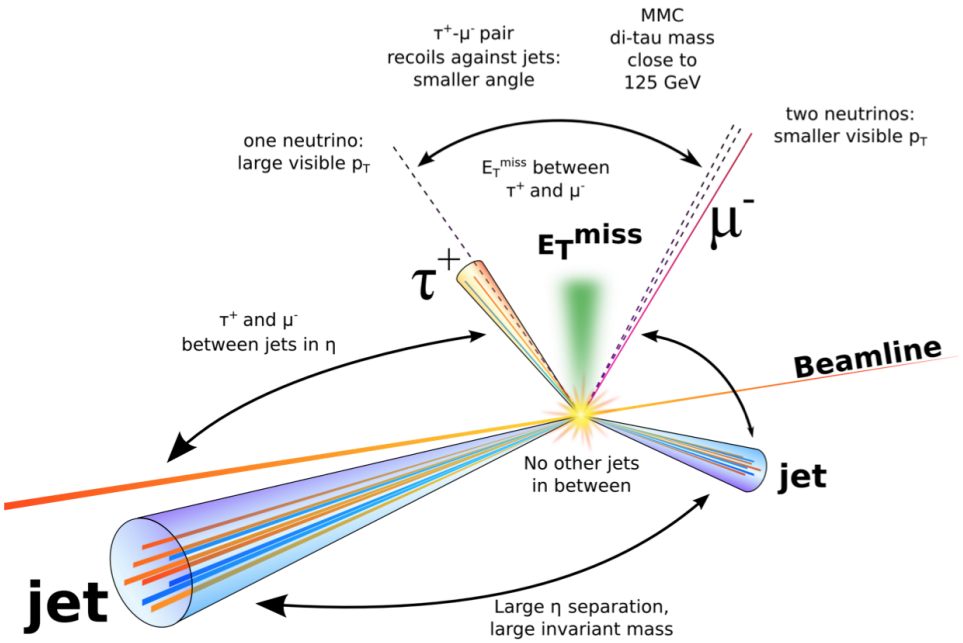


Figure 8.3: The illustration of the Higgs boson mass reconstruction with the MMC in the $H \to \tau_{\text{lep}}\tau_{\text{had}}$ decay [130].

9. Analysis of Higgs boson decay to a pair of tau-leptons

9.1 Introduction: Higgs boson production and decay and its studies at the ATLAS

After the Higgs boson discovery in 2012 [131, 132], the physics program of Large Hadron Collider (LHC) [133] has been focusing on precise measurements of the Higgs boson properties and searching for signs of the physics beyond SM in the Higgs sector.

There are four leading processes of the Higgs boson production (Figure 9.1): gluon-gluon fusion (ggH), vector boson fusion (VBF), associated production together with vector boson VH and associated production with a top-quark pair $t\bar{t}H$. In proton-proton collisions at the centre-of-mass energy of 13 TeV, gluon-gluon fusion is the leading mode of the Higgs boson production. It yields the estimated cross section of 48.6 pb (Figure 9.2a). Vector-boson fusion is the second largest contributor to the Higgs boson production at the LHC (3.78 pb).

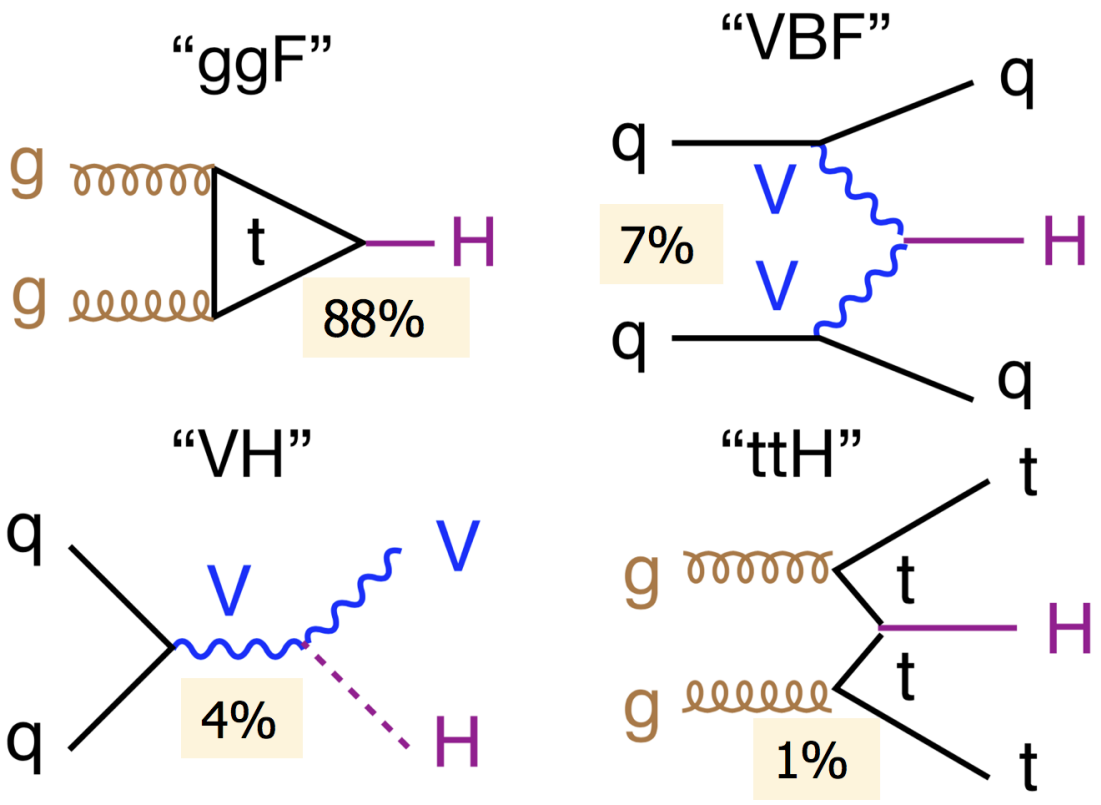
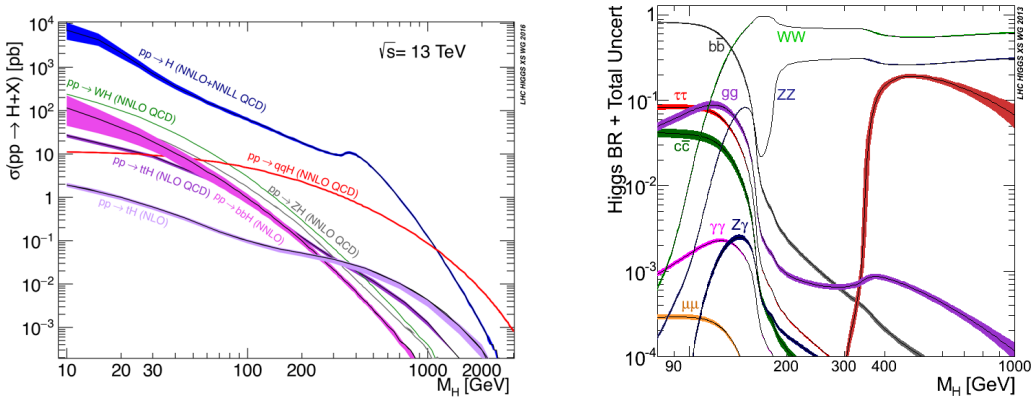


Figure 9.1: The Higgs boson production modes and their relative contribution in pp collisions at the LHC.

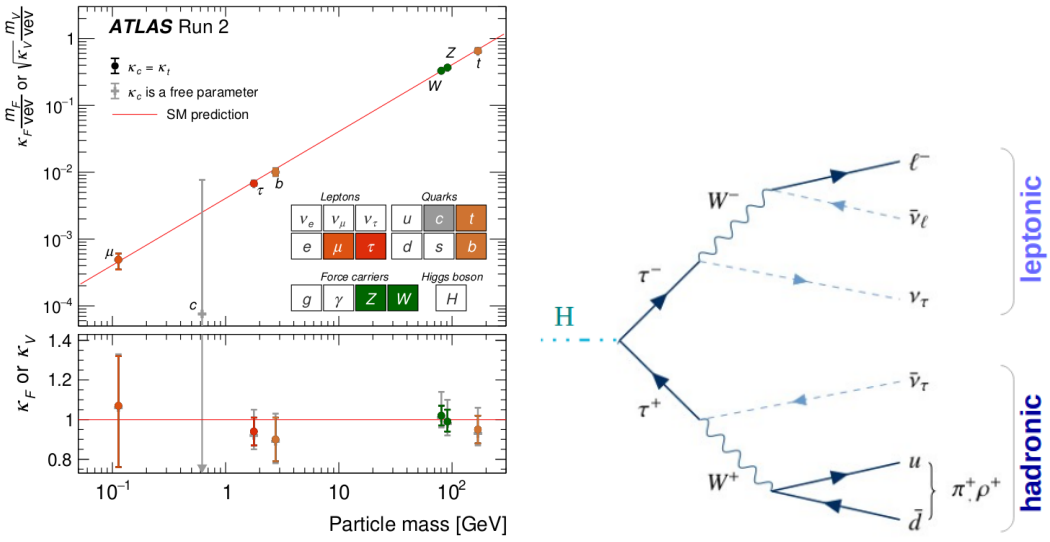
Studies of the Higgs boson properties and measurement of the $H \to \tau\tau$ coupling are necessary to explore both SM and BSM physics. Among the SM Higgs boson decays to fermions, $H \to \tau\tau$ is the second most frequent one with a branch-



(a) (b)

Figure 9.2: The Higgs boson production cross section at $\sqrt{s} = 13$ TeV (a) [134]. The Standard Model Higgs boson decay branching ratios (b) [135].

ing ratio of 6.32% (Figure 9.2b). The Yukawa coupling (Figure 9.3a) between tau-leptons and Higgs boson gives an opportunity to directly measure the Higgs boson properties such as its parity, spin, mass and decay width [22, 23, 136, 137, 138]. The presented research focuses on measuring the cross section of Higgs boson production in the four modes discussed. The actual $H \to \tau\tau$ coupling measurement aims to complement the result obtained with the ATLAS partial Run 2 [4] data set (the data collected in 2015-16).



(a) (b)

Figure 9.3: The coupling strength (a). The Feynman diagram for the $H \to \tau_{\text{lep}}\tau_{\text{had}}$ process (b).

The larger $H \to \tau\tau$ BR allows exploring coupling dependencies in different p_{T} regions. For example, BSM phenomena can manifest as patterns of deviation from the SM prediction in measured differential production cross section. Therefore, the $H \to \tau\tau$ analysis with full Run 2 data set targets measurement of the Higgs boson production fiducial and differential cross sections.

9.1.1 MC modelling and analysed data set

The presented $H \to \tau\tau$ cross section measurement uses the data collected by the ATLAS experiment [39, 40, 41] during LHC Run 2. The events are categorised into four groups following the Higgs boson production channel: namely, $ggH \to \tau\tau$, VBF, $t\bar{t}H$, and VH.[1] The measurement includes various di-tau decay modes: a fully hadronic channel $\tau_{\text{had}}\tau_{\text{had}}$, a semi-leptonic $\tau_{\text{lep}}\tau_{\text{had}}$ (Figure 9.3b), and a fully leptonic $\tau_{\text{lep}}\tau_{\text{lep}}$.[2]

The Higgs boson production and decay are modelled with Monte Carlo as described elsewhere [4, 5].

Numerous background processes mimic the Higgs boson decay to a pair of tau-leptons. The most considerable irreducible background arises from the Drell-Yan $Z \to \tau\tau$ process. Processes involving single or pair top-quark production also bring the final states with decaying tau-leptons. MC prediction for these processes is validated in corresponding control regions (CRs). Smaller background contributions originate from $Z \to ll$, di-boson VV and $H \to WW^*$ events; these are modelled with MC. Several properties of background simulation by Monte Carlo are described in Appendix B.

The contribution of events with fake tau-leptons is estimated with the data-driven techniques introduced briefly in Section 6. Section 9.5 will discuss some aspects of fake background estimation.

Several studies presented in this chapter were made with the partial Run 2 ATLAS data set. The data collected during 2015-16 data-taking comprise 36.1 fb^{-1} (or 32.9 fb^{-1}) of integrated luminosity.[3] This data set was prepared with the `Athena` release 20.7. If not stated otherwise, other results are based on the analysis of the 2015-18 data set of the integrated luminosity of 138.2 fb^{-1}. The corresponding NTuples were prepared with the `Athena` software release 21.

The plots shown in the chapter include statistical uncertainties.

9.1.2 Higgs boson signal discriminative variable: di-tau invariant mass

The di-tau invariant mass is an essential metric for the $H \to \tau\tau$ measurement. It approximates the mass of the decaying Higgs boson and becomes a final discriminant variable for the Higgs boson signal extraction from measured data. An accurate reconstruction of $m_{\tau\tau}$ ensures that the analysis is sensitive to the Higgs boson signal. An accurate $m_{\tau\tau}^{\text{coll}}$ estimation leads to the higher precision of the $H \to \tau\tau$ coupling measurement.

The main source of the irreducible background in $H \to \tau\tau$ coupling measurement is the Drell Yan process, $Z \to \tau\tau$. It makes from 50% to 90% of the total SM background depending on the analysed production category and di-tau decay mode. Masses of the Z boson and Higgs bosons are $\sim$91 GeV and $\sim$125 GeV, respectively. Due to the finite resolution, the measured mass spectra partially

[1] The discussed analysis of the $H \to \tau\tau$ coupling considered hadronic decays of top quarks and gauge bosons in the $t\bar{t}H$ and VH categories. One of the final-state tau-leptons is a product of the Higgs boson decay.

[2] These di-tau decay modes are also called lep-lep, lep-had and had-had channels, respectively.

[3] The size of the data set depends on the conditions of data processing. As will be indicated, several plots are based on the analysis of a larger data set.

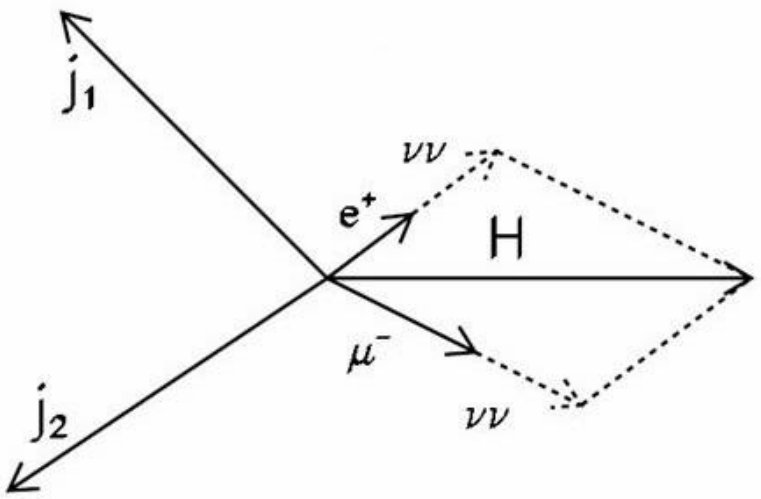


Figure 9.4: The collinear mass approximation applied to the Higgs boson decay to a pair of tau-leptons [129]. Jets balance the boosted di-tau in the transverse plane.

overlap. Hence, precise mass reconstruction is essential for signal discrimination from the background.

The reconstruction of di-tau invariant mass is challenging as the final-state neutrinos are elusive. Due to four neutrinos, the $m_{\tau\tau}$ estimation is particularly demanding in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel.

The mass estimation methods listed in Section 8 provide fast output as they exploit directly available information on the reconstructed objects. However, they suffer from a lack of accuracy or universality. Namely, the visible and transverse mass variables provide only a partial estimate and therefore have a limited $m_{\tau\tau}$ resolution.

The CMA method gives a reasonable proxy for $m_{\tau\tau}$ in boosted di-tau topologies. Given the presence of at least a single jet, it can estimate di-tau mass in the Higgs boson production through gluon fusion. In this case, the produced Higgs boson is boosted. Jets activity in an opposing hemisphere of the detector provides a boost of di-tau (Figure 9.4). However, if Higgs boson is nearly at rest, tau-leptons likely point in the opposite directions and $\Delta\phi \to \pi$. In the case that the collinear condition does not hold, Eq. 8.5 diverges as $p^{\text{miss}}_{0(1)} \sim 1/\sin(\phi^{\text{vis}}_0 - \phi^{\text{vis}}_1)$. Such back-to-back topology is likely for Higgs boson production in vector boson fusion as the produced Higgs boson is nearly at rest (Figure 9.5).

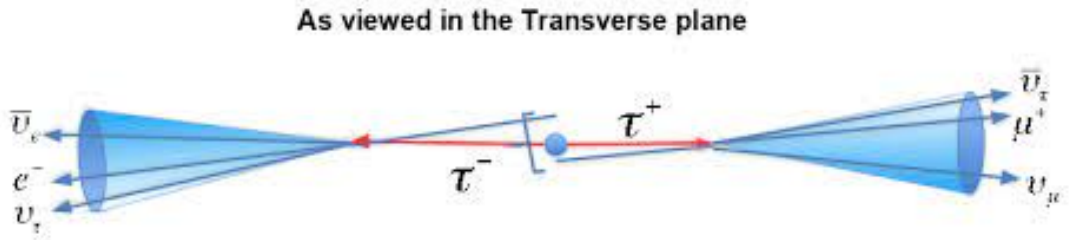


Figure 9.5: The neutral boson at rest decays to a pair of tau-leptons [139].

The collinear mass approximation is effectively applicable when tau-leptons

from $H \to \tau\tau$ decay are highly boosted ($m_\tau \ll m_H$). The CMA is inefficient for mass reconstruction in vector boson fusion when two tau-leptons are back-to-back. Due to the limitations of the visible, transverse, collinear mass estimators, the $H \to \tau\tau$ measurement requires another method.

The ATLAS experiment uses the Missing Mass Calculator (MMC) technique for di-tau mass estimation in the $H \to \tau\tau$ measurement in the ggH and VBF categories. The MMC provides an efficient reconstruction of Higgs boson mass. It showed good performance in Run 1 ATLAS $H \to \tau\tau$ analysis [3]. The MMC-based di-tau invariant mass was used as a fit variable in the $H \to \tau\tau$ analysis with the partial Run 2 data set (2015-16) [4].

The next sections describe the efforts made towards updating and improving the MMC method for the $H \to \tau\tau$ coupling measurement with the complete Run 2 ATLAS data set (2015-18). The MMC run-time was optimised, and the MET resolution estimation was updated. Employment of the object-level MET significance was scrutinised in pursuit of finding the best estimator for the MET resolution. Performed studies also answered whether the evaluation of the probability of a given MET configuration could be optimised. Finally, the ways of further MMC developments are outlined.

9.1.3 Missing Mass Calculator: probabilistic method and its status

The core part of the MMC is the evaluation of the probability of a given orientation and momenta of visible di-tau decay products to invisible ones. As said in Section 8.2.5, the global likelihood (Eq. (8.8)) describes how much the current configuration of MET and visible tau-lepton four-momenta is consistent with the kinematics of the tau-lepton decay products and the vent topology. Multiple configurations are tried for each event while scanning over kinematically allowed values of MET, tau-lepton p_{T}, and angular separation between them. Each configuration is assigned a weight reflecting its probability.

The decay topology and tau-leptons four-momenta are studied in well-known processes. Section 8.2.5 briefly mentioned this. Specifically, $H \to \tau\tau$ coupling measurement employs the MMC parametrisation with $Z \to \tau\tau$ events. The $Z \to \tau\tau$-based parametrisation is applicable in the $H \to \tau\tau$ analysis as the properties of decay products do not depend on tau-leptons' origin. In the MMC for the $H \to \tau\tau$ Run 2 analysis, the $\mathcal{P}(\Delta R^{\mathrm{vis,miss}})$ and $\mathcal{P}(E^{\mathrm{vis}})$ PDFs are parametrised as functions of the visible tau-lepton momenta and properties. Further, the MET probability term is parametrised as a function of the MET resolution, as explained in the next section.

9.1.4 MMC: MET probability term

In the MMC, the missing transverse momentum approximates the neutrinos' energy. The MET is defined as the deficiency of momentum in the transverse plane. The missing transverse energy is supposed to be driven by neutrinos only. The estimated MET values rely on a precise reconstruction, identification, and calibration of all the hard-scatter and soft-term objects found in a given event (Eq. (5.7)). However, mismeasurement of jets, miscalibration of hard-scatter

objects, or restrictions in the soft-term estimation can affect the resulting MET value [140]. Cosmic muons, beam halo particles, and detector noise can drive the MET overestimation. Furthermore, missing energy can signify the production of elusive particles, for example, Dark Matter candidates [141]). Hence, one should account for possible misinterpretation of MET as the neutrinos' energy proxy. The detector effects and mismeasurement lead to the Gaussian distribution of MET [128]. The MET resolution quantifies the accuracy of the MET estimation.

The MMC performs a scan over likely MET values, and the MET resolution plays a role of a factor quantifying the effect of non-precise measurement of MET. The MMC algorithm includes a MET likelihood term $\mathcal{P}(E_{\mathrm{T}}^{\mathrm{miss}})$ defined as follows:

$$\mathcal{P}(E_{\mathrm{T}}^{\mathrm{miss}}) = \mathcal{P}(E_{\mathrm{T}x,y}^{\mathrm{miss,sugg}}; E_{\mathrm{T}x,y}^{\mathrm{miss,meas}}, \sigma_{\mathrm{MET}}) = \exp\left\{ -(E_{\mathrm{T}x}^{\mathrm{miss,sugg}} - E^{\mathrm{miss,meas}})^2 + (E_{\mathrm{T}y}^{\mathrm{miss,sugg}} - E_{\mathrm{T}y}^{\mathrm{miss,meas}})^2)/2\sigma_{\mathrm{MET}}^2 \right\} \quad (9.1)$$

where $E_{\mathrm{T}x,y}^{\mathrm{miss,meas}}$ and $E_{\mathrm{T}x,y}^{\mathrm{miss,sugg}}$ are the values of the MET x- and y-components as measured in data and as suggested at the current point in the MMC scan over $E_{\mathrm{T}}^{\mathrm{miss}}$ values, respectively. The size of the $E_{\mathrm{T}}^{\mathrm{miss}}$ region scanned is dictated by the estimated MET resolution: $E_{\mathrm{T}}^{\mathrm{miss}} \pm N\sigma_{\mathrm{MET}}$. In this way, the MMC method relies on the MET resolution estimate and requires an adequate σ_{MET} description.

The MMC employed the MET resolution estimated with event-level information: tau-leptons p_{T} and decay modes, jet multiplicity, MET, and others. The initial parametrisation was derived with Run 1 $Z \to \tau\tau$ MC samples. This parametrisation needs a revision for the full Run 2 data set analysis. The following sections revisit the MET resolution estimation for the $H \to \tau\tau$ analysis with the full ATLAS Run 2 data set. Further, this thesis tests other approaches for the MET resolution evaluation. In particular, the applicability of the object-level MET significance (as introduced in Section 9.3.3 later) was extensively studied. Besides, the extension of the MET probability (Eq. (9.1)) was probed.

The scan over kinematically allowed MET values finds the configuration with the largest probability to be an actual $\tau\tau$ decay process. The likelihood MMC scan checks thousands of points in multi-dimensional phase space. The MMC's computational time is relatively large, making the $m_{\tau\tau}^{\mathrm{MMC}}$ calculation the longest stage in event processing. Moreover, ATLAS's object reconstruction and identification algorithms have constantly been improving throughout data-taking and post-Run 2. This progress can result in the sufficiency of a shorter MMC scan. The net sections present the efforts to minimise the MMC run time.

9.2 Event selection

Several event selection criteria help to select $H \to \tau\tau$ signatures from large amounts of data. Table 9.1 summarises the corresponding event selection, which is effective for data processed with `Athena` r. 21. The event selection applied in the analysis of the partial Run 2 data set largely resembles the event selection for the complete Run 2 data analysis (see Table C.1 in Appendix C and the supporting documentation [101]).

The corresponding trigger should accept the events: single-lepton and di-lepton triggers in the $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ channel, single-lepton trigger in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel,

and the di-tau trigger in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The di-electron, di-muon, and electron-muon triggers are employed in the $\tau_e\tau_e$, $\tau_\mu\tau_\mu$, and $\tau_e\tau_\mu$ channels, respectively. Single-electron (single-muon) triggers accept $\tau_e\tau_{\text{had}}$ ($\tau_\mu\tau_{\text{had}}$) events. Lepton candidates are required to be geometrically matched to corresponding trigger-level objects. One can find the complete list of the triggers used in the $H \to \tau\tau$ analyses with the partial and full ATLAS Run 2 data sets in [4] and [5], respectively. In this thesis, Appendix A illustrates the process of object reconstruction during the triggering of $\tau_{\text{had}}\tau_{\text{had}}$ events.

There are the requirements on electrons, muons, τ_{had} and jets applied before the overlap removal. Namely, the lepton candidate should pass proper identification requirements as said in Section 5. In addition, overlaps between τ_{had} and electrons are checked in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.[4] In the next step, the OLR procedure is standard (Section 5.6). Afterwards, jets are required to pass `Tight` WP selection on JVT/fJVT score.

The final event and object selection criteria list is given in Table 9.1 and described in detail in the parer [5].

criterion	$\tau_e\tau_\mu$	$\tau_{\text{lep}}\tau_{\text{had}}$		$\tau_{\text{had}}\tau_{\text{had}}$
		$\tau_e\tau_{\text{had}}$	$\tau_\mu\tau_{\text{had}}$	
N_e	1	1	0	0
N_μ	1	0	1	0
$N_{\tau_{\text{had}}}$	0	1	1	2
$N_{b\text{-jet}}$	0 (85% WP)	0 (85% WP)	0 (85% WP)	0 (70% WP)
p_{T}^e [GeV]	>15 to 27	>27		
p_{T}^μ [GeV]	>10 to 27.3		>27.3	
$p_{\text{T}}^{\tau_{\text{had}}}$ [GeV]		>30		>40, 30
Identification	e/μ: Medium	$e/\mu/\tau_{\text{had}}$: Medium		τ_{had}: Medium
Isolation	e: Loose, μ: Tight	e: Loose	μ: Tight	
Charge	Opposite charge			
$E_{\text{T}}^{\text{miss}}$ [GeV]	>20			
Kinematics	$m_{\tau\tau}^{\text{coll}} > (m_Z - 25)$ GeV $30\,\text{GeV} < m_{e\mu} < 100\,\text{GeV}$	$m_{\text{T}} < 70\,\text{GeV}$		
Leading jet	$p_{\text{T}} > 40\,\text{GeV}$			$p_{\text{T}} > 70$ GeV, $\lvert\eta\rvert < 3.2$
Angular	$\Delta R_{e\mu} < 2.0$ $\lvert\Delta\eta_{e\mu}\rvert < 1.5$	$\Delta R_{l\tau_{\text{had}}} < 2.5$ $\lvert\Delta\eta_{l\tau_{\text{had}}}\rvert < 1.5$		$0.6 < \Delta R_{\tau_{\text{had}}\tau_{\text{had}}} < 2.5$ $\lvert\Delta\eta_{\tau_{\text{had}}\tau_{\text{had}}}\rvert < 1.5$
Coll. fractions x_0^{coll}, x_1^{coll}	$0.1 < x_0^{\text{coll}} < 1.0$ $0.1 < x_1^{\text{coll}} < 1.0$	$0.1 < x_0^{\text{coll}} < 1.4$ $0.1 < x_1^{\text{coll}} < 1.2$		$0.1 < x_0^{\text{coll}} < 1.4$ $0.1 < x_1^{\text{coll}} < 1.4$

Table 9.1: The event and objects selection criteria in the $H \to \tau\tau$ analysis [107].

The Good Run List and event-level data quality requirements are applied. The baseline pre-selection implies compliance with data quality criteria and rejection of non-collision events ($N_{\text{vert}} \geq 1$ and $N_{\text{track}} \geq 2$). The corresponding trigger

[4]The `VeryLooseLlhEleOLR` variable is used for that.

should accept the events: electron-muon in the $\tau_e\tau_\mu$ channel, single-electron or single-muon in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, and di-tau in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

All electrons and muons must comply with the corresponding `Medium` identification WPs. In the lep-had and $\tau_e\tau_\mu$ channels, the electron (muon) isolation WP is set to `Loose` (`Tight`). The requirements on electron and muon p_{T} are tightened with respect to the p_{T} criteria before OLR due to the p_{T} thresholds of single-lepton and electron-muon triggers. The lowest p_{T} values applied in the final event selection are cited in Table 9.1.

Jet Vertex Tagging and forward JVT are applied to reconstructed jets. If not JVT/fJVT tagged, a jet is removed from the analysis. The `DL1r` b-tagger is operating at the `FixedCutBEff_85` WP in the lep-lep and lep-had channels. Due to lower $t\bar{t}$ contribution, the `FixedCutBEff_70` criterion is used in the had-had channel.

τ_{had} are required to pass the `Medium` RNN ID criteria in the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels. In the $\tau_e\tau_{\text{had}}$ channel, an additional selection of the `Medium` eBDT WP is applied to reduce a fraction of events where electron mimics τ_{had}. The muon veto rejects muons misidentified as τ_{had}. In $\tau_{\text{had}}\tau_{\text{had}}$ decays, each reconstructed τ_{had} needs to match trigger-level tau candidates.

The p_{T} selection for light leptons and τ_{had} is tightened with respect to the trigger thresholds. This helps to reduce background contribution and ensures stable trigger identification efficiency. The minimal values of transverse momentum for electrons and muons are listed in the paper [5].

Further event-wide criteria are imposed to select $H \to \tau\tau$ event candidates. The Higgs boson candidate is reconstructed from two τ_{had} and MET. Its mass is estimated with the MMC ($m_{\tau\tau}^{\text{MMC}}$). Lepton candidates are required to be of the opposite charge. The decay channel is defined by the numbers of selected electrons, muons and τ_{had}. Events with additional leptons satisfying only Loose identification criteria are rejected. To ensure MET compatibility with di-τ decay, only events with $E_{\text{T}}^{\text{miss}} > 20$ GeV are considered, and the requirements on the fractions of visible τ decay products x_0 and x_1 are introduced.

Resolved reconstruction of visible four momenta for both tau-leptons is complicated for largely boosted $\tau\tau$ pairs. Therefore, the analysis considered those final states where energetic jets balance the di-tau system (Figure 9.4). In particular, the presence of at least one high-energy jet is required.

The requirements on invariant mass estimates and absence of b-tagged jets are applied to reject V + jets events and top processes. The criterion on the distance between visible tau-decay products suppresses the contribution of non-resonant decays. $\tau_{\text{lep}}\tau_{\text{lep}}$ decays undergo additional selection on $m_{\tau\tau}^{\text{coll}}$ [4], which ensures orthogonality to the $H \to WW^*$ measurement. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, the leptonic W boson decays are suppressed by the transverse mass selection. The b-jet veto requirement suppresses the top-quark background.

The conditions on the di-tau angular opening $0.6 < \Delta R < 2.5$ are imposed following the di-tau trigger setups mentioned in Appendix A. Selections on the angular separation (ΔR, $\Delta\eta$) between two leptons enhance signal presence in all the considered channels.

9.3 MMC: estimating the MET resolution and the MET probability

The following paragraphs outline the estimation of the MET probability term in the MMC. Several approaches for the evaluation of the MET resolution are introduced. Performance of the proposed σ_{MET} estimation methods is primarily studied in the leptonic di-tau $\tau_{\text{lep}}\tau_{\text{lep}}$ decay channel. The MMC studies were performed mainly in $H \to \tau\tau$ events as the gluon-gluon fusion is the leading Higgs boson production mode. The ATLAS data set from 2015-16 (2015-18) is analysed as outlined previously.

9.3.1 MET resolution parametrisation

As discussed in Section 5.7, the MET values depend on the momenta of all the objects reconstructed in a given event. Consequently, the MET resolution can be presented as a function of the variable indicating the energy of all the reconstructed particles, jets, and the soft term. In particular, the MET resolution is commonly parametrised as a function of the total transverse energy $\sum E_{\text{T}}$. The total transverse energy is defined as the scalar sum of the transverse energy of the reconstructed objects which were used to estimate the MET value:[5]

$$\sum E_{\text{T}} = \sum p_{\text{T}}^{e} + \sum p_{\text{T}}^{\mu} + \sum p_{\text{T}}^{\tau} + \sum p_{\text{T}}^{\text{jet}} + \sum p_{\text{T}}^{\gamma} + \sum p_{\text{T}}^{\text{soft}} \tag{9.2}$$

The MET resolution can be then proxied with an uncertainty on the total event activity described by $\sum E_{\text{T}}$, i.e.

$$\sigma = f(\sqrt{\sum E_{\text{T}}}) \tag{9.3}$$

This functional representation was validated in numerous MET resolution and scale studies (for instance, [142], [97]). The MET resolution is usually proportional to $\sqrt{\sum E_{\text{T}}}$. An explicit form of parametrisation is process-dependent. The $\sum E_{\text{T}}$ dependence takes a different form for different di-tau decay modes. It depends on the data set, which reflects differences in reconstruction techniques, data-taking conditions or MC simulations.

In the $H \to \tau\tau$ analysis, the MET resolution was estimated by comparing the measured MET as reconstructed in a given event, and the particle-level MET as would be detected by an ideal detector. The ATLAS jargon names it “reco-truth" comparison as introduced earlier 4. The difference between the detector-level and particle-level MET is measured for each event. The width (σ) of the Gaussian fit for the obtained $E_{\text{T}x,y}^{\text{miss,reco}} - E_{\text{T}x,y}^{\text{miss,truth}}$ distributions is taken as the MET resolution σ_{MET}.[6] After this procedure is repeated for several bins of $\sum E_{\text{T}}$, the σ_{MET} is plotted as a function of $\sum E_{\text{T}}$ as presented in Figure 9.6.

[5]The total transverse mass is reconstructed from all stable ($\tau > 3 \times 10^{-11}$ s) charged and neutral particles with absolute momentum $|p|$ above 500 and 200 MeV, respectively.

[6]The label “true" denotes the information obtained at the level of generated particles. The label “reco" denotes the values reconstructed after detector effects simulation. The MC generation steps and detector effects simulation were introduced in the Section on Monte Carlo simulation 4.

The MET resolution in the lep-lep channel also depends on $\Delta\phi_{\mathrm{ll}}$, which is the azimuthal separation between the visible tau-lepton decay products. Figure 9.7 shows this relation. This dependence is a consequence of extra degradation of the MET resolution due to multiple MET sources (four neutrinos in the $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ final state).

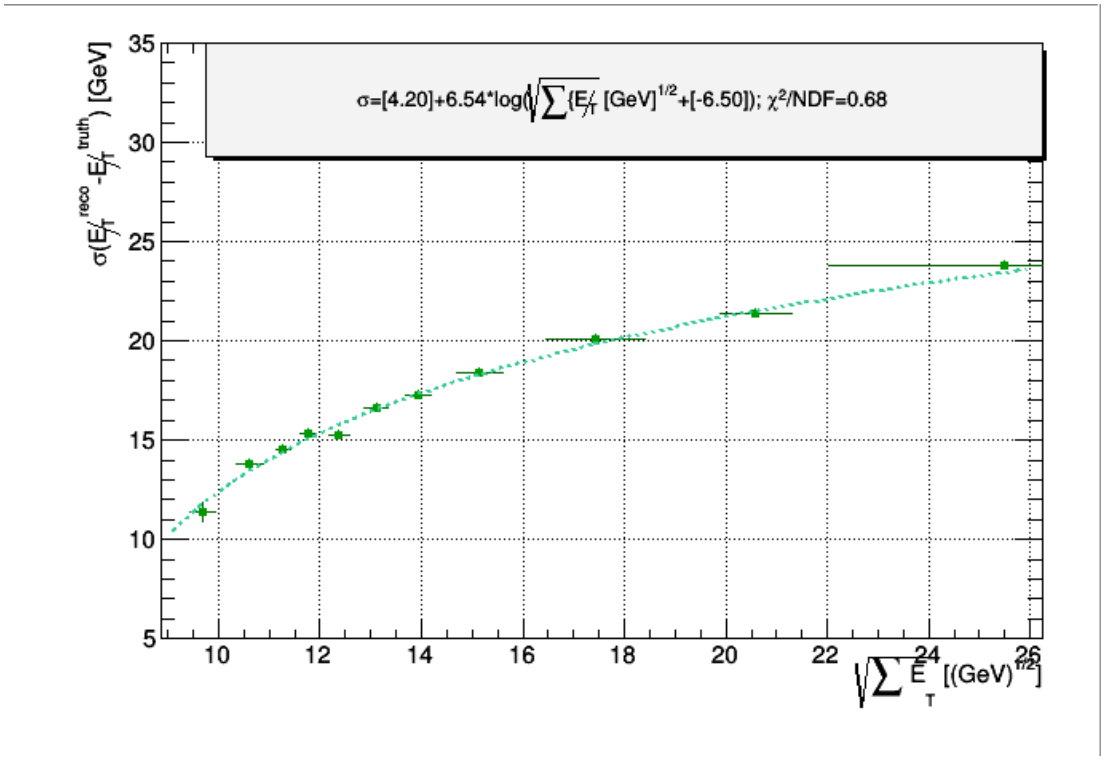


Figure 9.6: The MET resolution dependence on the per-event total transverse energy $\sum E_{\mathrm{T}}$ in $Z \to \tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ events. The cited function shows the parametrisation of σ_{MET} in the following function: $\sigma_{\mathrm{MET}}(\sum E_{\mathrm{T}}) = x_0 + x_1 \log(\sqrt{\sum E_{\mathrm{T}}} + x_2)$ where x_0, x_1, and x_2 are free parameters of the fit.

The parametrisation in Figures 9.6 and 9.7 was derived based on $Z \to \tau\tau$ modelling with Sherpa v.2.2.1 which is a widely used simulation for the Run 2 ATLAS analyses as for 2019. The difference between the initial and recent parametrisations is attributed to changes in MC modelling and to progress in object reconstruction achieved by the ATLAS Combined Performance groups.

9.3.2 MET resolution dependence on pile-up

Additional interactions occurring at the same bunch crossing can deteriorate MET reconstruction. The pile-up can lead to further degradation of the MET resolution. The pile-up ultimately affects the reconstruction of di-tau invariant mass $m_{\tau\tau}^{\mathrm{MMC}}$, as observed in the $H \to \tau\tau$ analysis [143]. Thus, the $m_{\tau\tau}^{\mathrm{MMC}}$ depends on the average number of interactions per bunch crossing $\langle\mu\rangle$.

The MET resolution dependency on the pile-up $\langle\mu\rangle$ was factorised out. Figure 9.8 shows how the residual MET resolution deteriorates with higher pile-up. Each additional interaction worsens the MET resolution by ~ 200 MeV. Following these observations, an additional parametrisation of the $\langle\mu\rangle$ was proposed for the MMC in the full Run 2 $H \to \tau\tau$ analysis.

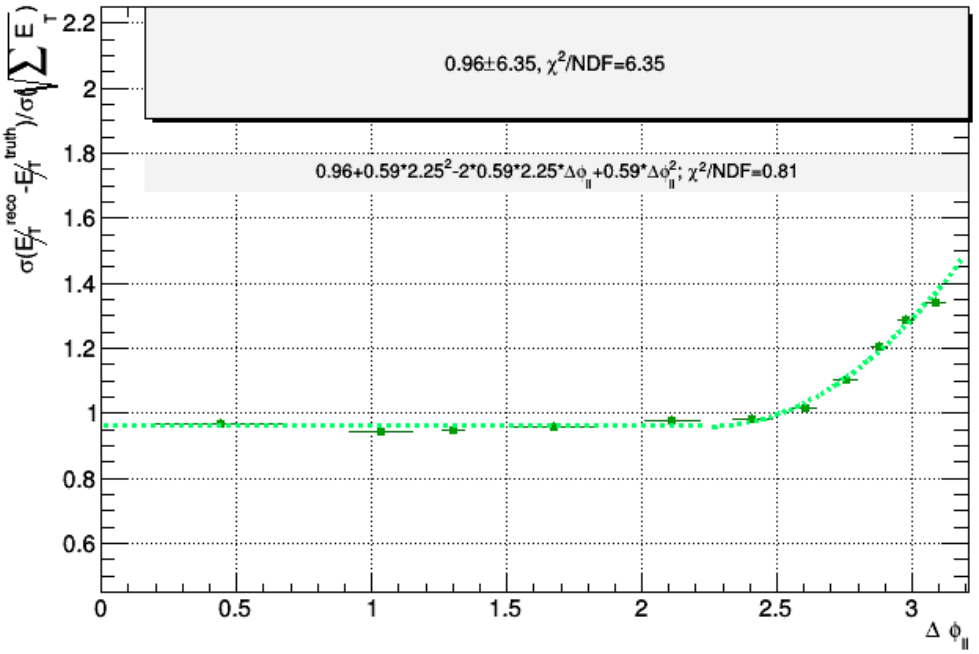


Figure 9.7: The residual MET resolution dependence on the azimuthal separation between two light leptons $\Delta\phi_{\mathrm{ll}}$ in $Z \to \tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ events. The residual dependence was parametrised with two different functions in the ranges of $\Delta\phi_{\mathrm{ll}}$. A linear function was used to parametrise in the $\Delta\phi_{\mathrm{ll}} < 2.25$ as written in the first line in the legend. A parabolic parametrisation $\sigma_{\mathrm{MET}}(\Delta\phi_{\mathrm{ll}})/\sigma_{\mathrm{MET}}(\sum E_{\mathrm{T}}) = x_0 + x_1(2.25 - \Delta\phi_{\mathrm{ll}})^2$ was applied in the $\Delta\phi_{\mathrm{ll}} > 2.25$ range as shown in the second line in the legend.

As will be seen later on (Section 9.3.3), the application of the $\langle\mu\rangle$ factor improves the MET resolution description. The extended parametrisation of the MET resolution is in better consistency with another σ_{MET} estimate.

The parametrisations on $\sqrt{\sum E_{\mathrm{T}}}$, $\Delta\phi_{\mathrm{ll}}$, and $\langle\mu\rangle$ allow estimating MET resolution on an event-by-event basis. This σ_{MET} estimation relies on the general information about the event and can suffer from disregarding the actual nature of objects. For example, a reconstructed jet originates in hard scattering or pile-up. While the former are full contributors in MET, the latter ought to partially contribute to the jet term. Moreover, this approach does not account for the directional anisotropy of the MET resolution.[7]

Using the object-based MET resolution estimation $\mathcal{S}$ can help tp overcome this drawback. The $\mathcal{S}$ can potentially help to overcome the outlined features of the σ_{MET} parametrisation.

9.3.3 MET resolution through the MET significance

The object-based MET significance $\mathcal{S}$ is designed in the ATLAS to validate the missing transverse energy [144]. The `METSignificanceTool` tool was developed

[7]Note the ellipsoidal shape of the MET resolution depicted by shaded green in Figure 8.2.

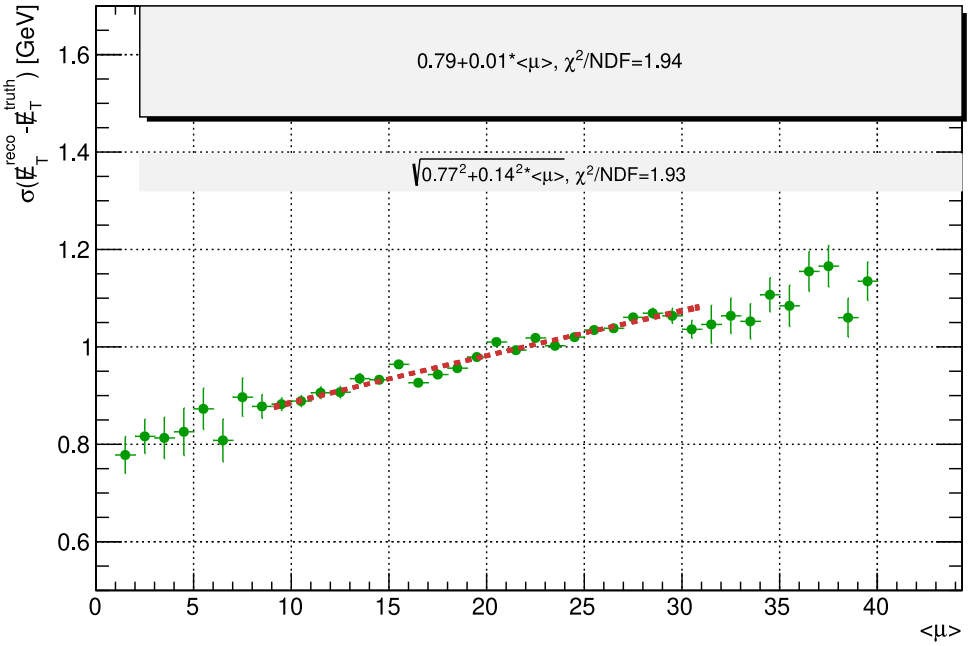


Figure 9.8: The underlying MET resolution dependence on the average number of interactions per bunch crossing $\langle\mu\rangle$. The σ_{MET} normalised to the already parametrised $\sigma_{\text{MET}}(\sum E_{\text{T}}, \Delta\phi_{\text{ll}})$ is implied (the y-axis). Two parametrisations – with a linear $\sigma_{\text{MET}}(\langle\mu\rangle)/\sigma_{\text{MET}}(\sum E_{\text{T}}, \Delta\phi_{\text{ll}}) = x_0 + x_1\langle\mu\rangle$ and square-root $\sigma_{\text{MET}}(\langle\mu\rangle)/\sigma_{\text{MET}}(\sum E_{\text{T}}, \Delta\phi_{\text{ll}}) = \sqrt{x_0^2 + x_1^2\langle\mu\rangle}$ functions – reveal similar performance (ξ^2/n.d.f.), as shown in the legend in the first and second lines, respectively.

by the ATLAS JetETmiss group. The MET significance became available for analyses since `Athena` release 21.

The MET significance quantifies the result of comparing two hypotheses on MET: that MET is non-zero in a given event and that MET is zero [144]. This definition of S_{MET} is given by equation

$$S_{\text{MET}} = 2\ln\Big(\frac{\max_{p_{\text{T}}^{\text{inv}}\neq 0}\mathcal{L}(E_{\text{T}}^{\text{miss}}|p_{\text{T}}^{\text{inv}})}{\max_{p_{\text{T}}^{\text{inv}}=0}\mathcal{L}(E_{\text{T}}^{\text{miss}}|p_{\text{T}}^{\text{inv}})}\Big) = (E_{\text{T}}^{\text{miss}})^{\text{T}}(\sum_i \mathbf{V}^i)^{-1}(E_{\text{T}}^{\text{miss}}) \tag{9.4}$$

in which the likelihood function implicates the total covariance matrix V^i for the resolution of the hard objects and the soft term. The covariance matrix $\mathbf{V}_i$ propagates the effects of resolution from reconstructed hard-scatter and soft-term objects.[8] This way, the resolution of all the hard objects and the soft term contribute to the MET variation.

In general, the low S_{MET} values (in the range of $S_{\text{MET}} \in (0, 2)$) usually indicate the absence of the MET source in a given event. Conversely, larger MET significance values (above 5) can signify that the source of the missing transverse energy is not well understood. Namely, the MET of a smaller value or in another

[8]The paper [144] explains the approach of the $\mathcal{S}$ calculation in details. Namely, it defines the covariance matrix through spatial components of the resolution of a given object.

direction is expected in this event based on the information on the hard-scatter objects as shown in Figure 9.9.

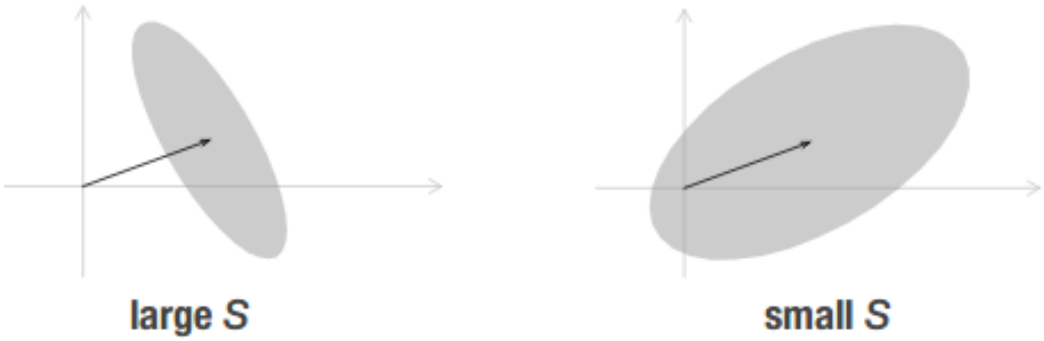


Figure 9.9: Generic illustrations of possible configurations of the measured and expected MET. The arrow indicates the MET vector (as defined by Eq. 5.1), while the shaded area depicts the MET resolution. The left sketch shows a case of the large estimated MET significance. The right one illustrates the event with an understood source of the MET. In this case, one can estimate the MET resolution based on the properties of measured hard-scatter and soft-term objects.

The MET significance values are characteristic of the process type. In the decay of boson to a pair of tau-leptons, $\mathcal{S}$ preferably falls between ≈ 2 and ≈ 5. The $\mathcal{S}$ range is typical for events with a genuine MET source and well-reconstructed objects.

The object-based MET significance can be used for straightforward calculation of the MET resolution: $\sigma_{\mathrm{MET}} = \sum E_{\mathrm{T}}/\mathcal{S}$. This σ_{MET} definition was tested against the parametrised one as shown in Figure 9.10. The two methods bring similar results in most events.

The MET significance, however, plays a role in MET resolution smearing in events where the S_{MET} is low despite the higher measured $E_{\mathrm{T}}^{\mathrm{miss}}$ value. Furthermore, when the MET significance is low while the estimated MET is high, the method results in large σ_{MET} values (see the high-σ_{signif} tail in Figure 9.10). Such topologies can occur in the case of object misreconstruction or misidentification. In this way, the MET significance has the potential to spot the events which can not be qualified as Higgs or Z boson decays. Appendix E tells on this in detail.

9.3.4 Anisotropy of MET resolution

The default version of the MMC assumes that the MET resolution is smeared isotropically around its direction. However, the resolution depends on the mutual configuration of reconstructed objects and the MET. For example, the resolution is estimated more precisely if it is oriented along a highly energetic hard-scatter object. Vice versa, the MET resolution projection is high if no hard-scatter objects point in that direction.

The actual MET resolution is generally anisotropic, and one can decompose it into two components (Figure 9.11). The longitudinal component σ_{L} is the resolution in the direction of the MET, whereas the transverse component σ_{T} is the resolution perpendicular to the MET. The ATLAS `METSignificance` tool

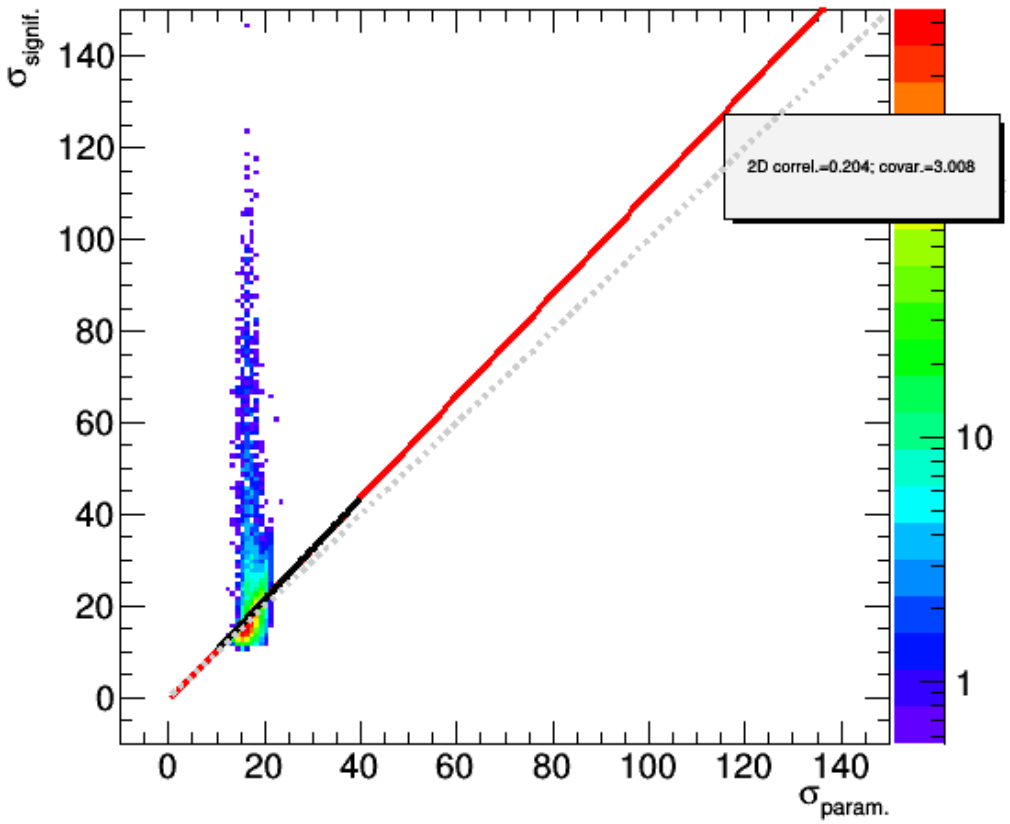


Figure 9.10: The MET resolutions in units of GeV): $\sigma_{\text{MET}}^{\text{param}}$ defined as a function of event-level parameters $\sum E_{\text{T}}$, $\Delta\phi_{\text{ll}}$ and $\langle\mu\rangle$ (on the x-axis) and $\sigma_{\text{MET}}^{\text{signif}}$ based on object-level MET significance $E_{\text{T}}^{\text{miss}}/S_{\text{MET}}$ (on the y-axis). The red line depicts the observed correlation trend between $\sigma_{\text{MET}}^{\text{param}}$ and $\sigma_{\text{MET}}^{\text{signif}}$. The grey dotted line corresponds to a perfect correlation between $\sigma_{\text{MET}}^{\text{param}}$ and $\sigma_{\text{MET}}^{\text{signif}}$ (therefore, the x-axis range spans up to 150 GeV).

estimates the values of these individual MET components. It also provides a factor of correlation between them ρ. The two-component σ_{MET} technique applies dedicated coefficients for pile-up-like jets [144], which results in high MET resolution values. The MMC di-tau mass reconstruction can profit from knowing the detailed MET description.

One can expand the MET resolution PDF definition using the information on the longitudinal and transverse MET resolution components and the correlation between them. The formula Eq. (9.1) is then replaced by the 2D Gaussian

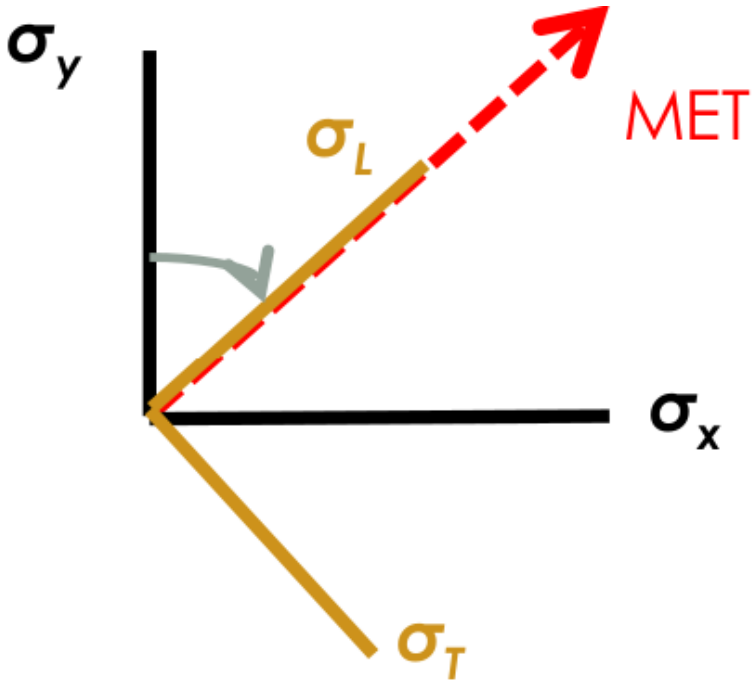


Figure 9.11: The scheme of the MET resolution components [98]. σ_{L} stands for the longitudinal component of the MET resolution, i.e. the one projected in the direction of the MET. σ_{T} is the transverse MET resolution obtained as the MET resolution projection to the plane perpendicular to the MET direction. Similarly to Figure 8.2, σ_{x} and σ_{y} are the MET resolution in the transverse plane where the MET is defined.

function

$$\mathcal{P}(E_{\mathrm{T}\parallel}^{\mathrm{miss,sugg}}, E_{\mathrm{T}\perp}^{\mathrm{miss,sugg}}; E_{\mathrm{T}\parallel}^{\mathrm{miss,meas}}, E_{\mathrm{T}\perp}^{\mathrm{miss,meas}}, \sigma_{\mathrm{MET}\parallel}, \sigma_{\mathrm{MET}\perp}, \rho) = \frac{1}{2\pi\sigma_{\mathrm{MET}\parallel}\sigma_{\mathrm{MET}\perp}\sqrt{1-\rho^2}} \times \exp\Big\{ -\frac{1}{2(1-\rho^2)}\Big[\Big(\frac{E_{\mathrm{T}\parallel}^{\mathrm{miss,sugg}} - E_{\mathrm{T}\parallel}^{\mathrm{miss,meas}}}{\sigma_{\mathrm{MET}\parallel}}\Big)^2 + \Big(\frac{E_{\mathrm{T}\perp}^{\mathrm{miss,sugg}} - E_{\mathrm{T}\perp}^{\mathrm{miss,meas}}}{\sigma_{\mathrm{MET}\perp}}\Big)^2 - 2\rho\Big(\frac{E_{\mathrm{T}\parallel}^{\mathrm{miss,sugg}} - E_{\mathrm{T}\parallel}^{\mathrm{miss,meas}}}{\sigma_{\mathrm{MET}\parallel}}\Big)\Big(\frac{E_{\mathrm{T}\perp}^{\mathrm{miss,sugg}} - E_{\mathrm{T}\perp}^{\mathrm{miss,meas}}}{\sigma_{\mathrm{MET}\perp}}\Big)\Big]\Big\} \quad (9.5)$$

In this way, the MET probability term can be estimated with the input data provided by the `METSignificanceTool`: the parallel and perpendicular MET resolution ($\sigma_{\mathrm{MET}\parallel}$ and $\sigma_{\mathrm{MET}\perp}$ MET resolution) as well as their correlation (ρ).

9.3.5 MMC performance: finding an optimal MET likelihood term

The separability between signal and background was studied in the three σ_{MET} approaches mentioned above. The first one used the parametrisation on $\sum E_{\mathrm{T}}$, $\langle\mu\rangle$, and $\Delta\phi_{\mathrm{ll}}$. The second method exploited the object-level MET significance: $\sigma_{\mathrm{MET}} = E_{\mathrm{T}}^{\mathrm{miss}}/S_{\mathrm{MET}}$. The third method employed the 2D MET PDF (Eq. (9.5)).

The MET resolution and the MET PDF term were appropriately varied in the Missing Mass Calculator.

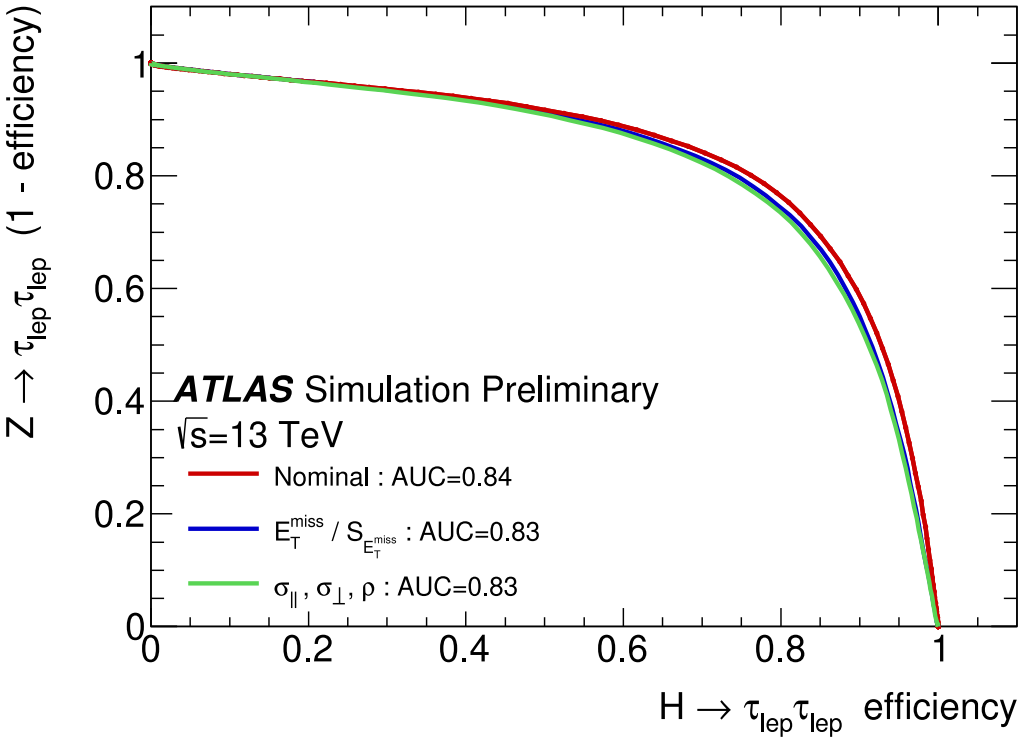


Figure 9.12: ROC curves for $H \to \tau\tau$ signal separation from the $Z \to \tau\tau$ background in the $\tau_{\text{lep}}\,\tau_{\text{lep}}$ channel. Three methods for the MET resolution estimation in the MMC $m_{\tau\tau}$ calculation are shown. The "Nominal" graph depicts the outcome from the σ_{MET} parametrisation as a function of $\sum E_{\text{T}}$, $\langle\mu\rangle$, and $\Delta\phi_{\text{ll}}$.

Figure 9.12 shows the ROC curves for selecting signal $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ events and rejecting background $Z \to \tau_{\text{lep}}\tau_{\text{lep}}$ processes when using the reconstructed di-tau invariant mass, $m_{\tau\tau}^{\text{MMC}}$. The methods perform with similar signal-to-background discriminator efficiency. The observed difference in the AUC is within 1%. The $m_{\tau\tau}^{\text{MMC}}$ distributions obtained with the three considered methods are given in Figure 9.13. Overall, all three approaches for σ_{MET} estimation perform at the same level. Application of the S_{MET} reveals prospects in rejecting $Z \to \tau\tau$ events (there is a drop by 10% in the 90 GeV $< m_{\tau\tau}^{\text{MMC}} <$ 110 GeV range).

9.3.6 Parametrisation on the suggested $E_{\text{T}x,y}^{\text{miss}}$

The event with maximal weight (Eq. (8.8)) becomes a solution for $m_{\tau\tau}^{\text{MMC}}$. The MET value corresponding to the solution generally differs from the input MET. The difference between measured MET and MMC output $E_{\text{T}}^{\text{miss,reco}} - E_{\text{T}}^{\text{miss,MMC}}$ has a distribution with a finite width (Figure 9.14).

The preference for one MET topology over another can play a role of an additional factor in the MET probability. The $E_{\text{T}}^{\text{miss,reco}} - E_{\text{T}}^{\text{miss,MMC}}$ has a Gaussian shape, and its factor enters the likelihood of a particular MET solution as follows:

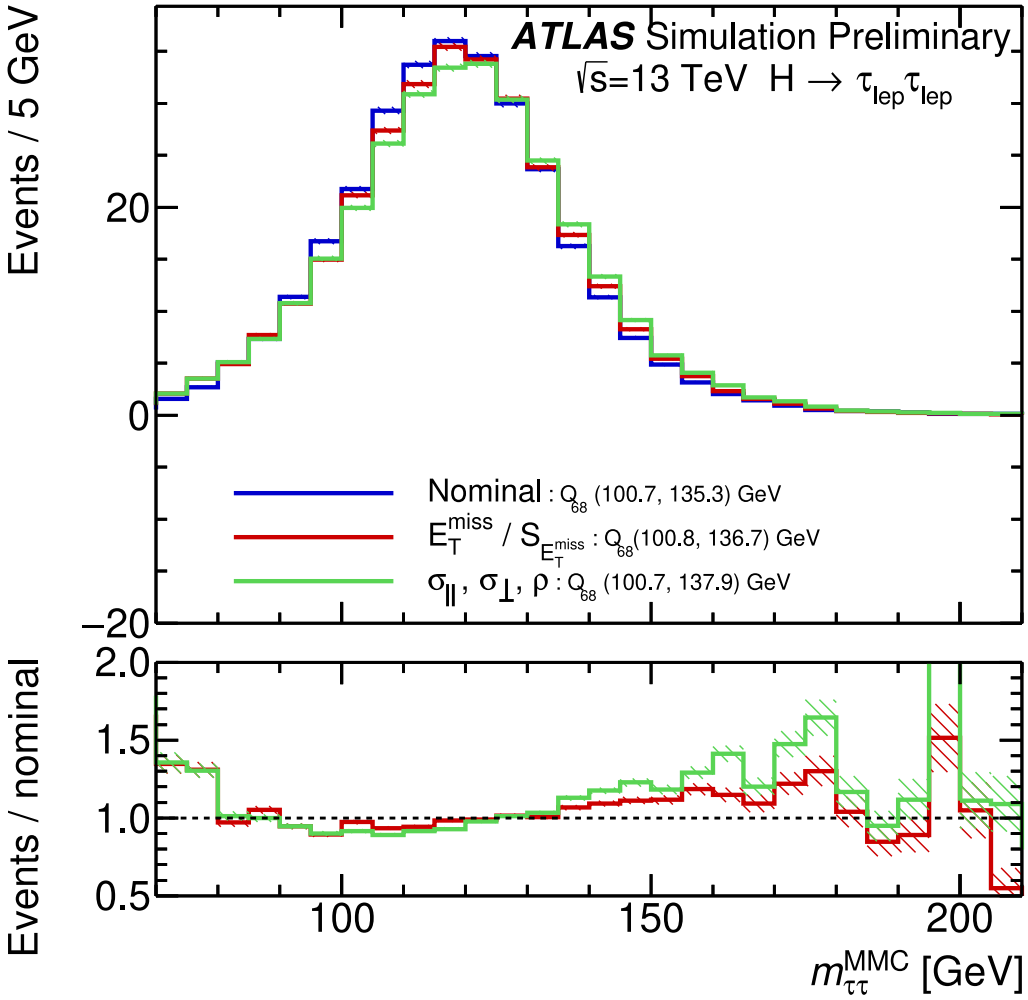


Figure 9.13: The $m_{\tau\tau}^{\text{MMC}}$ distributions obtained with three ways of the MET probability estimation: the parametrised MET resolution and the MET PDF as in Eq. (9.1); the MET resolution from the MET significance; and the MET PDF term as in Eq. (9.5) with $\sigma_{\text{MET}\parallel}$,$\sigma_{\text{MET}\perp}$, and ρ as provided by the `METSignificanceTool`.

$$\mathcal{P}(E_{\text{T}x,y}^{\text{miss}}) = \exp\Big(-\frac{(\Delta E_{\text{T}x,y}^{\text{miss}})^2}{2\sigma_{\text{MET}}^2}\Big)\exp\Big(-\frac{(E_{\text{T}x,y}^{\text{miss,reco}} - E_{\text{T}x,y}^{\text{miss,sugg}})^2}{2\sigma_{\text{sugg}}^2}\Big) \tag{9.6}$$

Here the $E_{\text{T}x,y}^{\text{miss,reco}}$ stands for a reconstructed value of the MET components as measured in data, whereas $E_{\text{T}x,y}^{\text{miss,sugg}}$ is the MET value corresponding to the MMC solution. The variance of the MET solutions σ_{sugg} was estimated as a function of $\sum E_{\text{T}}$ (Figure D.3 in Appendix D).

The results of testing the MMC calculator with the changed MET PDF are shown in Figures 9.15 and D.4 (Appendix D). No significant impact of the $m_{\tau\tau}^{\text{MMC}}$ shape nor MMC performance parameters was found.

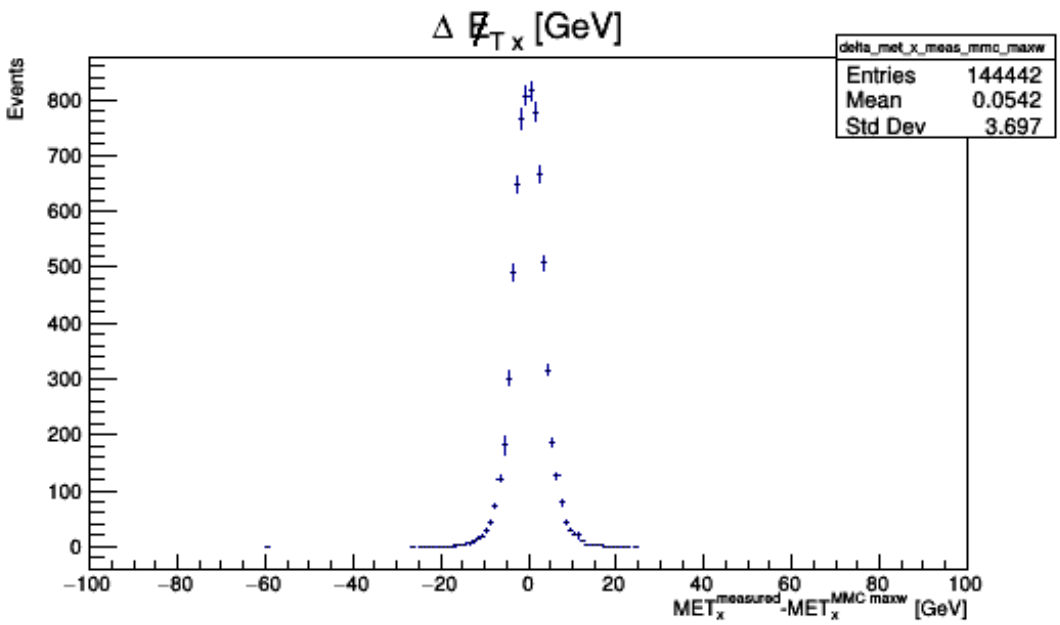


Figure 9.14: The distribution of the $E_{\mathrm{T}x}^{\mathrm{miss}}$ difference between the $E_{\mathrm{T}x}^{\mathrm{miss}}$ reconstructed value as measured by the ATLAS detector and the $E_{\mathrm{T}x}^{\mathrm{miss}}$ suggested value as solved by the MMC.

9.3.7 Exploring other definitions of $\mathcal{S}_{\mathrm{MET}}$

Besides the primary definition of the MET significance (Eq.(9.4)), the ATLAS `MET significance` tool also takes into account the configuration of the di-tau system. One can also estimate the MET significance with respect to the azimuthal location of the di-tau system ($S_{\mathrm{MET}}^{\tau\tau}$). For that, the covariance matrix is rotated by ϕ_{MET} angle. Another approach estimates the MET significance in the longitudinal direction: $S_{\mathrm{MET}}^{\mathrm{long}} = E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{\sigma_{\mathrm{MET}\parallel}}$. Respectively, the $S_{\mathrm{MET}}^{\mathrm{long}\,\tau\tau}$ is estimated with respect to the longitudinal MET resolution $\sigma_{\mathrm{MET}\parallel}$ and the azimuthal topology of the tau-lepton pair. Figures 9.16 and D.5 (Appendix D) present the $m_{\tau\tau}^{\mathrm{MMC}}$ performance with $S_{\mathrm{MET}}^{\mathrm{long}}$ and $S_{\mathrm{MET}}^{\tau\tau}$. These plots compare the alternative and standard S_{MET} definitions. There are indications that $S_{\mathrm{MET}}^{\tau\tau}$ performs better than the conventional S_{MET}. However, the level of improvement is very small (within 1%).

9.3.8 Studies of the MET scan range

As described earlier, the kinematically allowed phase space is scanned to find the most likely solution. The size of the scanned subspace is defined by the variation of the energy of τ_{had} and the MET. The scan over the MET values is performed in the $(E_{\mathrm{T}}^{\mathrm{miss}} \pm n\sigma_{\mathrm{MET}})$ range. The MET resolution σ_{MET} determines the size of the $(E_{\mathrm{T}}^{\mathrm{miss}} \pm n\sigma_{\mathrm{MET}})$ range. The presented studies aimed to find the most optimal size of the $E_{\mathrm{T}}^{\mathrm{miss}}$ range.

The following options of the MET scan range were tested: $(E_{\mathrm{T}}^{\mathrm{miss}} \pm 2\sigma_{\mathrm{MET}})$, $(E_{\mathrm{T}}^{\mathrm{miss}} \pm 3\sigma_{\mathrm{MET}})$, $(E_{\mathrm{T}}^{\mathrm{miss}} \pm 4\sigma_{\mathrm{MET}})$, and $(E_{\mathrm{T}}^{\mathrm{miss}} \pm 5\sigma_{\mathrm{MET}})$. The obtained $m_{\tau\tau}^{\mathrm{MMC}}$ distributions and ROC curves are given in Figures 9.17 and D.6 (Appendix D).

With the MET resolution defined via MET significance, the range of the

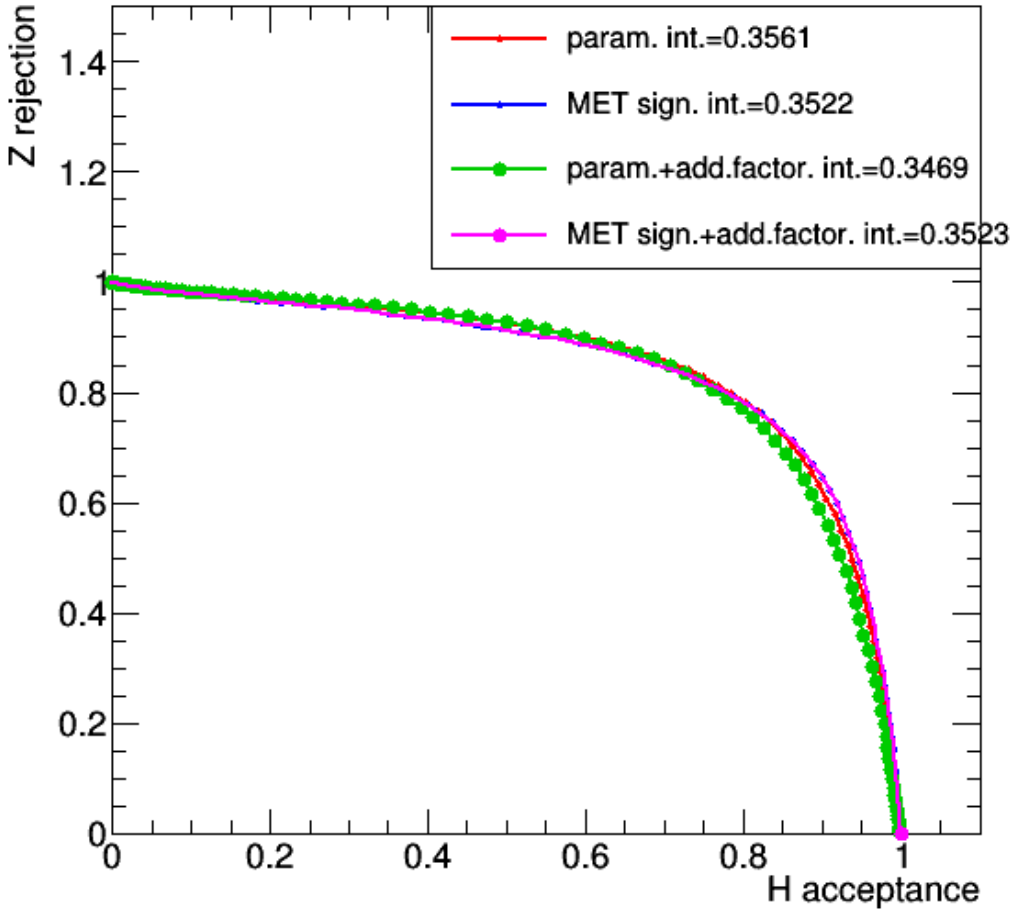


Figure 9.15: The ROC curves of the $H \to \tau\tau$ signal separation from the $Z \to \tau\tau$ background obtained with a number of alternative settings: where the MET resolution is parametrised (the red curve) and with the additional factorisation on $E_{\mathrm{T}}^{\mathrm{miss,reco}} - E_{\mathrm{T}}^{\mathrm{miss,MMC}}$ (the green line); where MET resolution is calculated through object-level MET significance (the blue line) and with the additional factorisation on $E_{\mathrm{T}}^{\mathrm{miss,reco}} - E_{\mathrm{T}}^{\mathrm{miss,MMC}}$ (the magenta line).

scan does not influence the $m_{\tau\tau}^{\mathrm{MMC}}$ shape and $H \to \tau\tau$ vs $Z \to \tau\tau$ separation power significantly. All the tried MMC layouts find solutions for di-tau mass very effectively: the inefficiency rate is at the level of 10^{-6}. No substantial differences in performance are observed in the case of the parametrised MET resolution, either. Although the differences are tiny, the 4σ option is the best choice based on the mean and/or RMS metrics (Figure D.6 in Appendix D). The area under signal-to-background separation curves has a slightly higher value for the 4σ option as demonstrated in Figure 9.17.

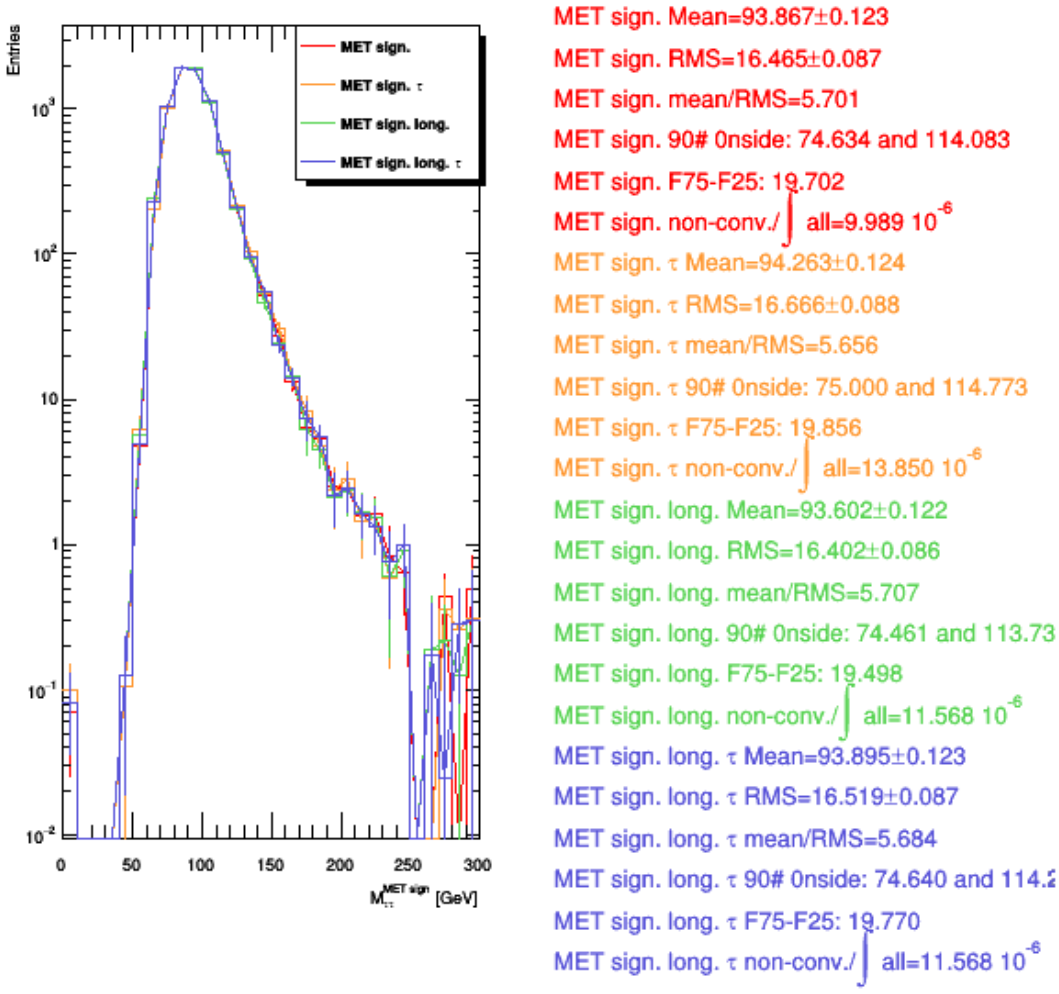


Figure 9.16: The $m_{\tau\tau}^{\mathrm{MMC}}$ spectra are compared in several options to define MET significance. The red, yellow, green, and blue graphs plot the $m_{\tau\tau}^{\mathrm{MMC}}$ outcomes obtained when the MET resolution was estimated as $E_{\mathrm{T}}^{\mathrm{miss}}/S_{\mathrm{MET}}$, $E_{\mathrm{T}}^{\mathrm{miss}}/S_{\mathrm{MET}}^{\tau\tau}$, $S_{\mathrm{MET}}^{\mathrm{long}}$, and $E_{\mathrm{T}}^{\mathrm{miss}}/S_{\mathrm{MET}}^{\mathrm{long}\ \tau\tau}$, respectively. The label includes the mean and RMS, the mean-to-RMS ratio, the inter-quantile range between the 5th and 95th, the 25th and the 75th percentiles, and the rate of the method non-convergence.

9.4 The Missing Mass Calculator run-time optimisation

The initial MMC version was adopted for the analysis of the ATLAS Run 1 data set. For Run 2, however, the ATLAS hardware conditions and triggering schemes were improved. Object reconstruction schemes, including MET description, were elaborated. As a result of these improvements, one can expect that the most likely solution would be in proximity to the measured τ_{had} and MET. This would mean faster finding of the solution and the scan (Eq. (eq:mmc:likelihood)) with a lower number of scan iterations N_{iter}.

The default MMC version prescribed 200 K iterations. The relation between N_{iter} and the MET PDF during the scan was studied. Figure 9.18 demonstrates that the N_{iter} in the MMC can eventually be optimised by reducing the number of

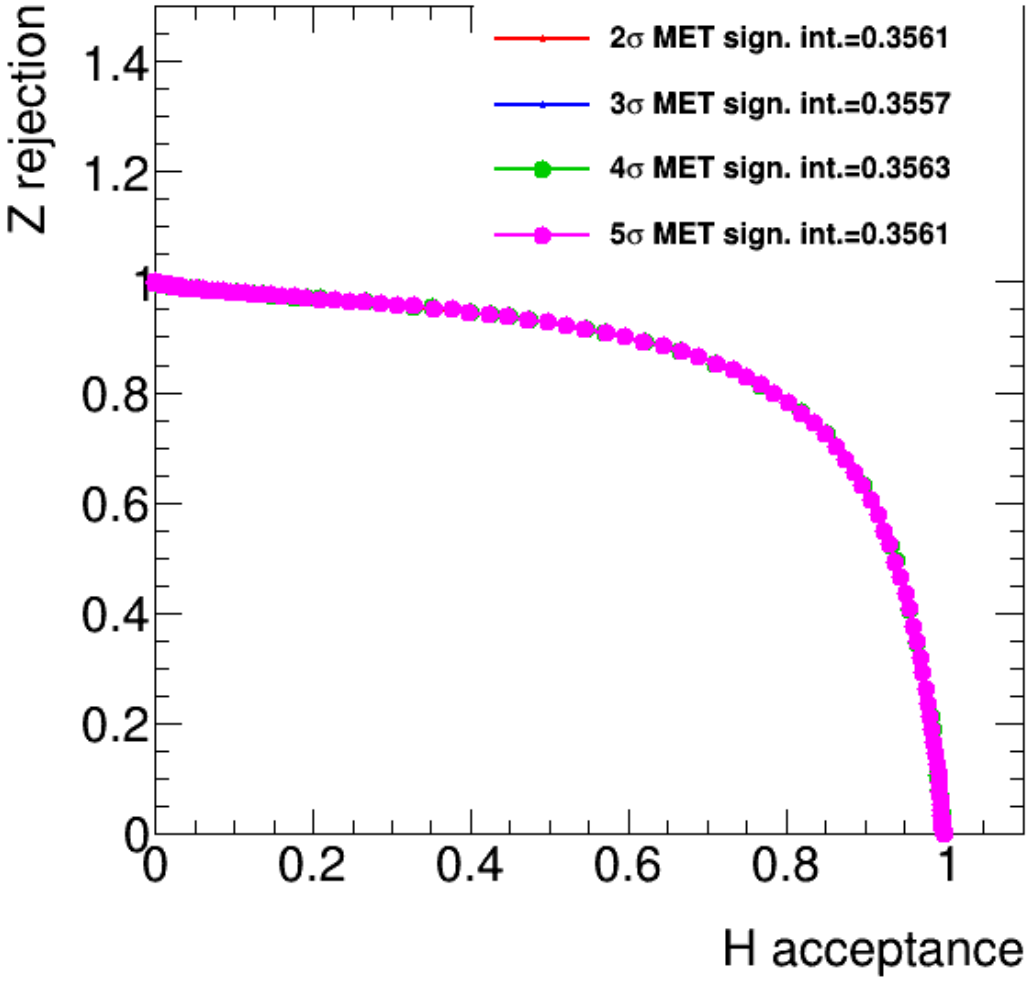


Figure 9.17: The ROC curves of the $ggH \to \tau\tau$ signal separation from the $Z \to \tau\tau$ background while observing $m_{\tau\tau}^{\text{MMC}}$ obtained with varying the $E_{\text{T}}^{\text{miss}}$ scan range size. The MET resolution is estimated with the object-level MET significance.

steps. Around the 50,000th step, solutions with lower probability become more frequent. This is ineffective as these solutions have a lower likelihood and are not chosen for a final di-tau mass estimate. Analysis of the total probability showed that an optimal solution could be found after about a thousand steps (Figure D.1 in Appendix D).[9] The applicability of the MMC with a shorter scan was tested. The MMC performance with the nominal $N_{\text{iter}} = 200$ K and the reduced $N_{\text{iter}} = 50$ K settings was compared.

The analysed data set included $Z \to \tau\tau$ and $H \to \tau\tau$ events with hadronic and leptonic di-tau decays ($\tau_{\text{had}}\,\tau_{\text{had}}$ and $\tau_{\text{lep}}\,\tau_{\text{lep}}$, respectively). The considered MC samples comprised an integrated luminosity of 36.1 fb^{-1}, which corresponds to the 2015-16 ATLAS data set.

The distribution of the di-tau invariant mass $m_{\tau\tau}^{\text{MMC}}$ for two different numbers of iterations used in the phase space scan by the Markov chain in the MMC is shown in Figure D.8 (Appendix D). The shapes of the $m_{\tau\tau}^{\text{MMC}}$ distribution in

[9]The deviation from low $-\log(\mathcal{P}_{\text{total}})$ in long iterations is even more pronounced if the $E_{\text{T}}^{\text{miss}} \pm 2\sigma_{\text{MET}}$ range is scanned (see Figure D.2 in Appendix D).

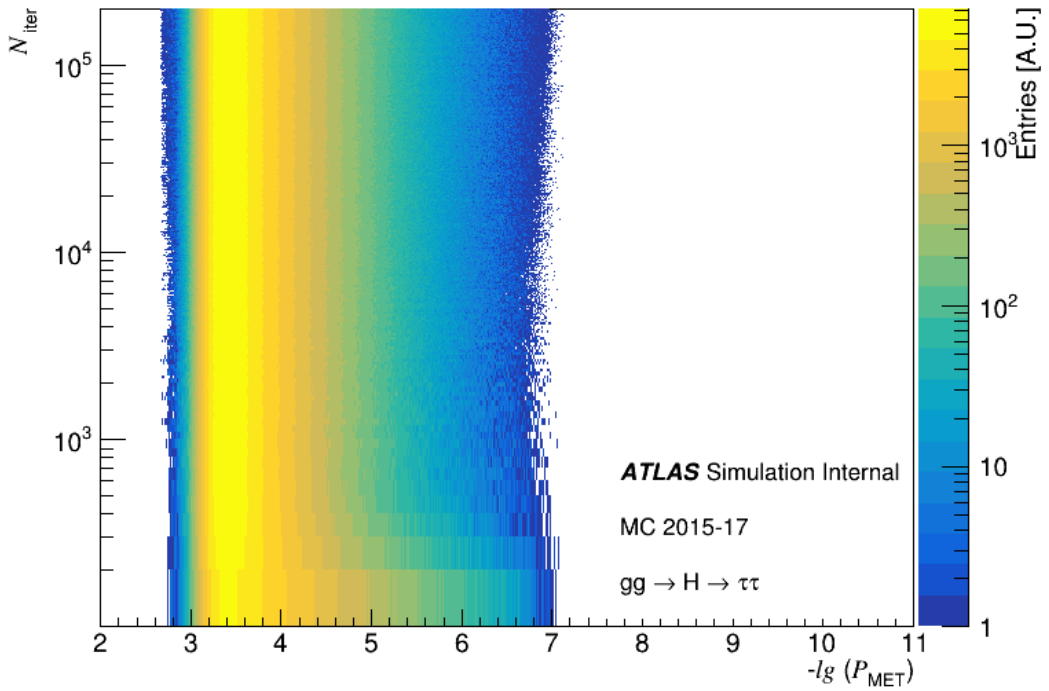


Figure 9.18: The MET likelihood versus the number of the current iteration in the MMC Markov Chain over possible configurations of di-tau products. The distribution is obtained in the analysis of $gg \to H \to \tau_{\text{lep}}\tau_{\text{lep}}$ events.

both versions are in agreement. The mean value, standard deviation, and inter-quantile range between the 32nd and the 68th percentiles are comparable between the two MMC settings. Figure 9.20 shows the $m_{\tau\tau}^{\text{MMC}}$ resolution dependence on the number of iterations.[10] Likewise, the ROC curves show consistent performance in the two options as seen in Figure 9.19, which shows the relative AUC of ROCs is depicted as a function of the number iterations in the Markov chain scan. The AUC values are normalised to the AUC value in the nominal MMC version with 200 K iterations. The AUC becomes nearly constant at $N_{\text{iter}} = 50$, and increasing N_{iter} to 200 K improves the AUC only by 0.2%. Figure D.7 in Appendix D shows the ROC curves for the separation between $ggH \to \tau\tau$ and $Z \to \tau\tau$.

An optimal configuration with the maximal PDF can be found with decreased number of iterations. The scan of 50 K iterations is applicable in the Run 2 MMC. On the other hand, the adopted version gains from the point of view of CPU time. The computation time was reduced by ~ 4 times (Figure D.9). Such MMC modification optimises memory consumption and fastens the $H \to \tau\tau$ analysis approximately twice.

[10]The resolution is defined as the Gaussian width determined in the range of ± 1.5RMS around the mean of the $m_{\tau\tau}^{\text{MMC}}$ distribution.

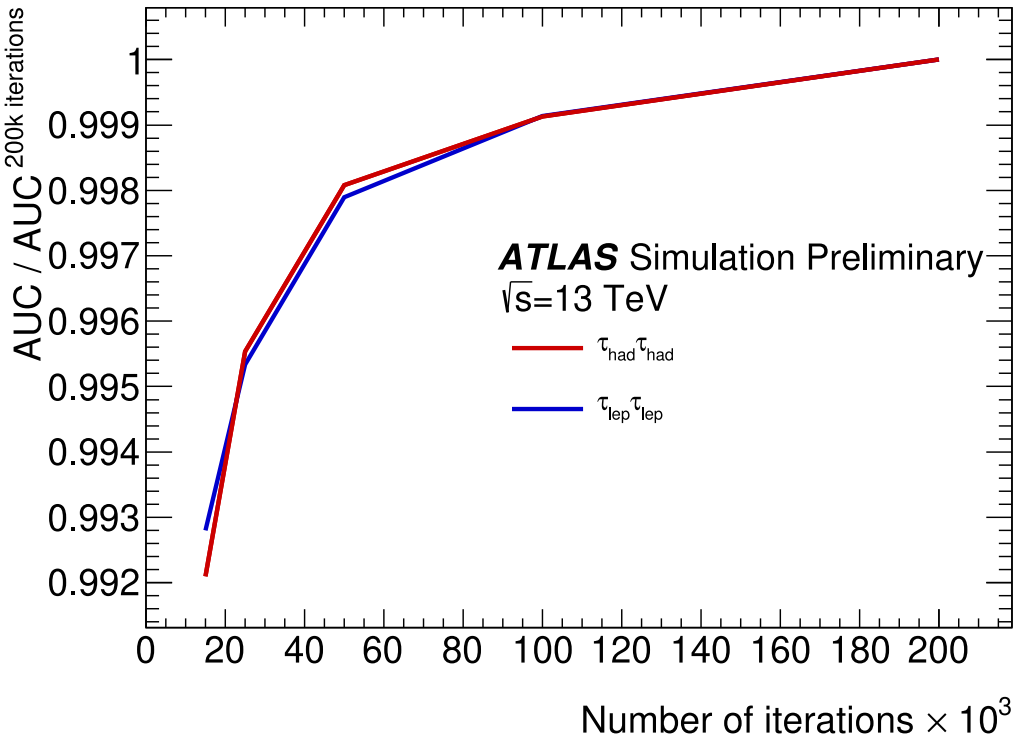


Figure 9.19: The AUC of the $H \to \tau\tau$ signal to the $Z \to \tau\tau$ background separation in the adopted MMC with 50 K iterations normalised to the AUCs obtained with the default 200 K scan. The AUC values for $\tau_{\text{had}}\ \tau_{\text{had}}$ and $\tau_{\text{lep}}\ \tau_{\text{lep}}$ events are presented.

9.5 Background estimation in the $H \to ll$ channel

As discussed in Section 9.1.1, background in the $H \to \tau\tau$ coupling measurement is predicted from a mixture of data-driven methods and Monte Carlo simulation. Typically, background processes involving real leptons are modelled with MC, while the contribution from fake leptons is defined with the data-driven approach.

As it was introduced in Section 6, hadronically decaying tau-leptons can be faked by light leptons, namely electrons and muons, and QCD jets. The $e \to \tau_{\text{had}}$ and $\mu \to \tau_{\text{had}}$ fakes are simulated with MC. These events include real leptons, but these real electrons and muons were misidentified as hadronically decaying τ_{had} during the event and object reconstruction. Jet can happen to be misidentified as lepton, i.e. $j \to e$, $j \to \mu$, $j \to \tau_{\text{had}}$ (later on mentioned as $j \to \tau$). This QCD-driven fake background is estimated with real data in the $H \to \tau\tau$ analysis. One needs to ensure that different methods address appropriate background and there are no ambiguities in defining a given background source. A common way to prevent fake double counting relies on subtracting events A when fakes are estimated in events B. For example, this is done in the Fake Factor method, which predicts the fake background in the $\tau_{\text{lep}}\tau_{\text{had}}$ decay channel.

In brief, the Fake Factors are measured with real data in a subset of events where tau-lepton fails to pass the nominal ID requirements (Figure 9.21a). The

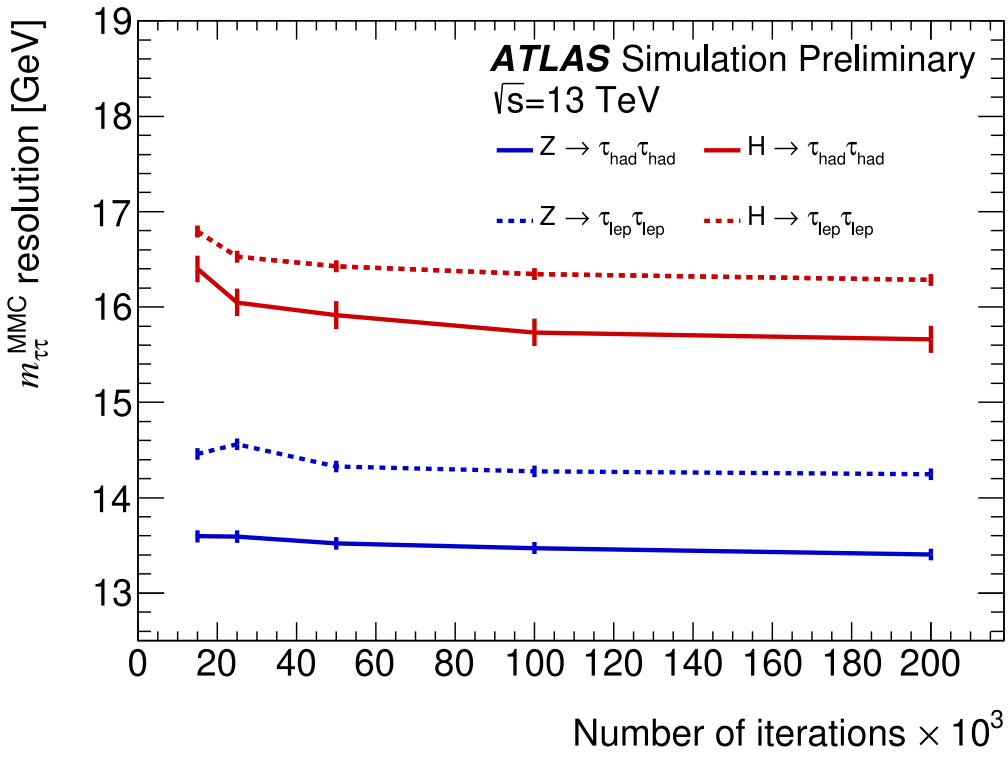


Figure 9.20: The $m_{\tau\tau}^{\mathrm{MMC}}$ resolution as a function of the number iterations used in the phase space scan by the Markov chain in the MMC.

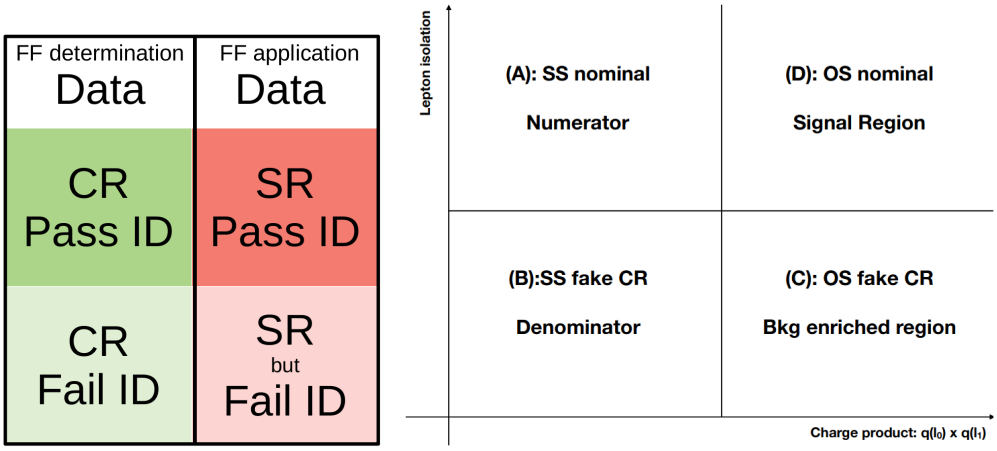
technique is designed to describe a fraction of the $j \to \tau_{\mathrm{had}}$ fakes. At the same time, the contribution from the $l \to \tau_{\mathrm{had}}$ fakes is modelled with MC. The contribution from real leptons $l \to \tau_{\mathrm{had}}$ are subtracted. This step ensures that the FF solely include the $j \to \tau_{\mathrm{had}}$ fakes.

This real-lepton subtraction is based on knowledge of genuine particle type. The information on the particle type is commonly labelled with Particle Data Group Identification number (PDG ID) [145] in MC-generated events. Final-state leptons are matched to generator-level particles, as stated in 5.5. The information on the matched truth particles signifies the source and nature of the reconstructed lepton.

In the lep-lep channel, one or two final-state light leptons can be mimicked by a jet. This fake contribution is assessed with either the ABCD or the Matrix method. In the ABCD method, fake-enriched phase space estimates a fraction of events where a QCD jet is misidentified as an electron or muon. The Transfer Factors (TFs) are derived in the fake-enriched control region (labelled B in Figure 9.21b). The fake CR is defined by changing several nominal, i.e. signal region, selections.[11] During TFs calculation, events with real leptons are estimated with MC and subtracted as follows:

$$f_{\mathrm{trans}} = \frac{A}{B} = \frac{N_{\mathrm{iso}}^{\mathrm{SS,Data}} - N_{\mathrm{iso}}^{\mathrm{SS,MC}}}{N_{\mathrm{non-iso}}^{\mathrm{SS,Data}} - N_{\mathrm{non-iso}}^{\mathrm{SS,MC}}} \tag{9.7}$$

[11] Explicitly, the fake CR (labelled B in Figure 9.21b) selections are: subleading lepton does not pass the `Gradient` isolation WP and $q_l \times q_l = +1$ (same sign (SS of leptons' electric charge). The requirement of Electron ID is relaxed to `Loose` whereas the SR includes `Medium` ID electrons.



(a) (b)

Figure 9.21: The Fake Factor derivation in the FF method. Courtesy of Vojtěch Pleskot (a) [146]. Sketch of the different regions considered in the fake background estimation for the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel (b) [101].

The obtained transfer factors are then applied to the nominal, i.e. opposite-sign (OS), region. This way, the data-driven template of $j \to l$ fakes is obtained as $D = A/B \times C$. At the same time, other backgrounds are modelled with Monte Carlo. In principle, the data-driven $j \to l$ background estimated can account for possible $j \to l$ fakes modelled with MC. For instance, the subtraction of MC $j \to l$ fakes can be done in the following way:

$$D = f_{\text{trans}} \times (N^{\text{OS,Data}}_{\text{non-iso}} - N^{\text{OS,MC}}_{\text{non-iso}}) \tag{9.8}$$

An analyser must ensure that MC prediction does not include MC-based $j \to l$ fakes.

In MC, jets misidentified as light leptons can be spotted using PDG ID labelling of the generator-level counterpart (as described at the beginning of this section). The presented studies answer the question of whether there is double counting of the $j \to l$ fakes. This research ought to find out if the criteria on matching "not-to-quark" would exclude excessive $j \to l$ fakes, if such, in the $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ analysis.

9.5.1 $j \to l$ background modelling

The studies were made to learn the ID, type and origin of the final-state particles in the MC-predicted background. The partial Run 2 data set was analysed, and the applied event selection is described in Table C.1 (Appendix C).

The contribution by quarks and gluons mimicking light leptons was seen from the PDG IDs of the matched generator-level particles. Figure 9.22 maps the relationship between leading and subleading tau-lepton PDG ID in the total MC-based background prediction. No $j \to l$ events with jets mimicking light leptons are observed. Consequently, one can conclude that there is no need to introduce a dedicated selection to remove the MC leptons matched to quarks or gluons.[12]

[12]Technically, this is done by rejecting final-state light leptons if their generator-level counterpart has a quark or gluon PDG ID.

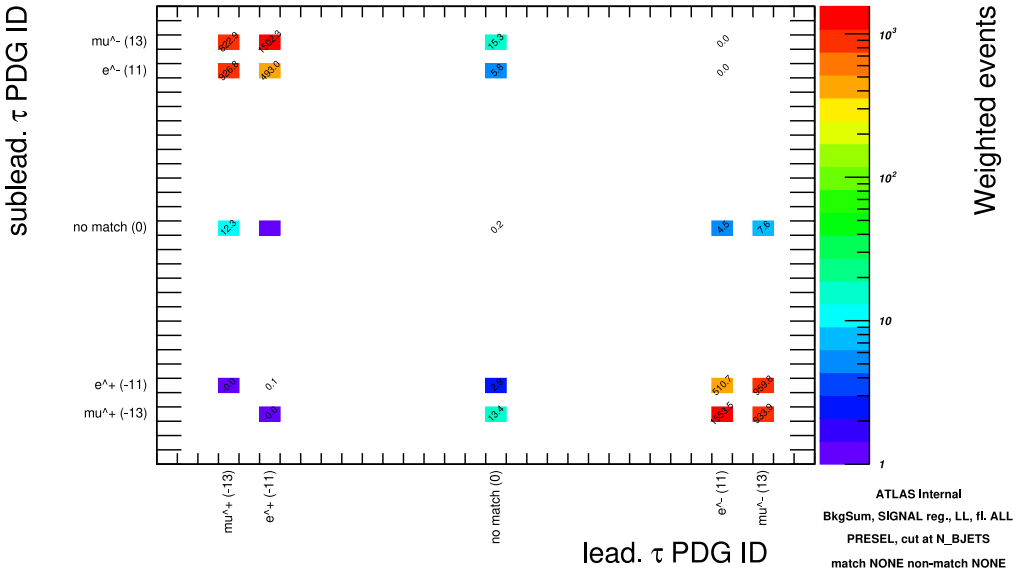
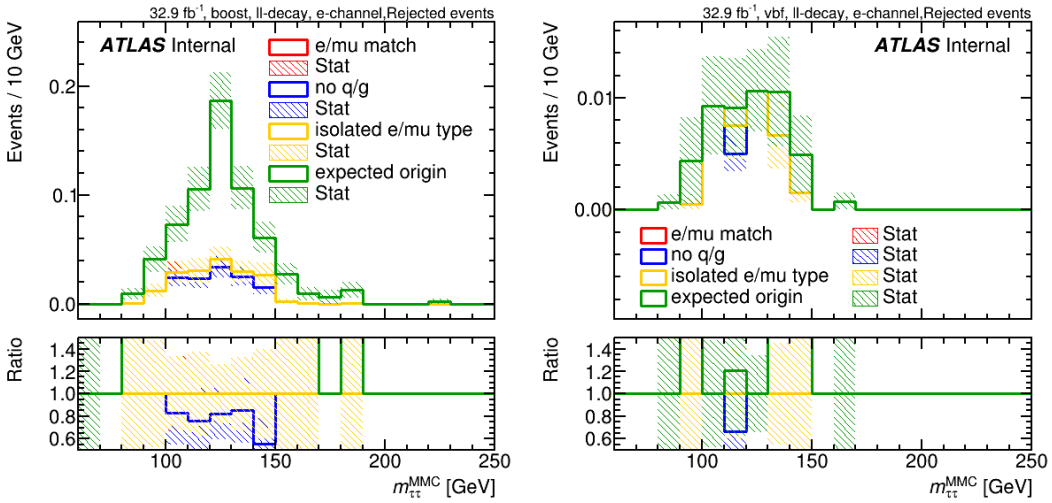


Figure 9.22: The leading tau-lepton PDG ID versus subleading tau-lepton PDG ID in preselected background events.

Figure 9.23 depicts the $m_{\tau\tau}^{\mathrm{MMC}}$ distributions in the events rejected by the requirement on the matched particle PDG ID not to be quark or gluon. In the boosted category, the fakes by the background electrons yield about 0.4% of the total background. Their fraction comes mainly from $Z \to ee$ QCD production (3.56%) and $Z \to \tau\tau$ EWK production (2.48%). The impact of the tested selections is negligible in the VBF category.



(a) (b)

Figure 9.23: The $m_{\tau\tau}^{\mathrm{MMC}}$ spectra in the events which do not pass the selection criteria in the boosted (a) and VBF (b) event categories. Several options for the event selection are applied.

As seen from Figures 9.22 and F.1, there are events where one of the tau-

leptons does not match any generator-level particle.[13] The genesis of these events with non-matched detector-level tau-leptons was investigated.

9.5.2 Unmatched light leptons

Studies of the unmatched tau-leptons nature were done with *Type* and *Origin* variables which are available for MC-modelled particles. The type and origin particle classification is provided by the `MCTruthClassifier` tool [147], which assigns two numbers to particles:[14]

- The *Type* which indicates its properties and behaviour, and
- The *Origin* points to the mother particle(s).

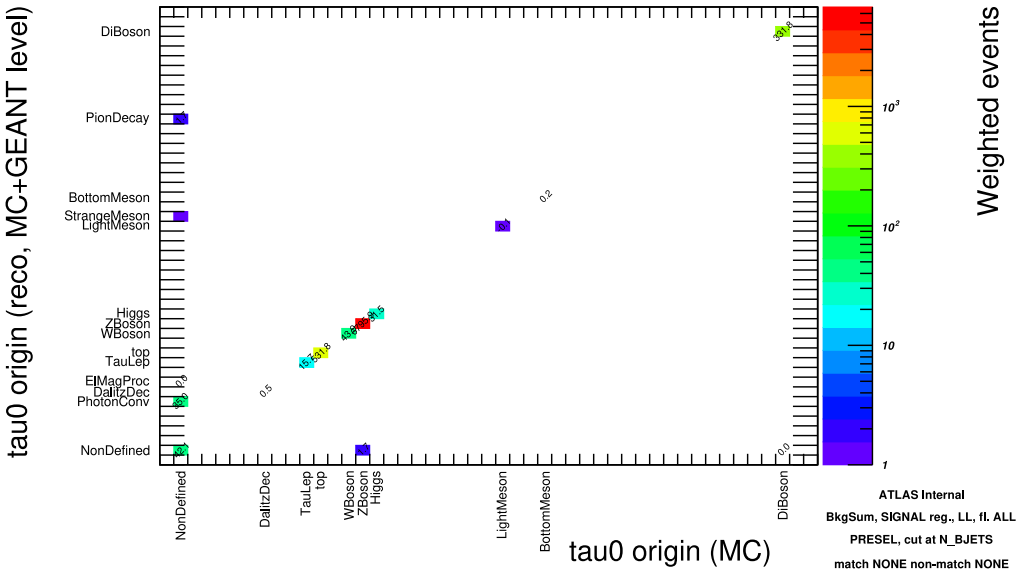


Figure 9.24: The relation between the “MC" and “MC + Geant 4" *Origin* variables for leading tau-lepton.

The analysed NTuples contain two variables specifying the *Type*, defined at different event reconstruction steps. The one, let’s name it “MC", is determined with a pure Monte Carlo particle generator record. Another one, let’s name it “MC + Geant 4", is determined after detector response is simulated (Figure 9.25). The second option includes `Geant 4` simulated particles and, consequently, provides a complete description of the particle nature and source. The *Origin* variables are defined at the two reconstruction steps as well, so that one can observe both the *Origin* MC and *Origin* MC + Geant 4.

Figure 9.24 shows the relation between “MC + Geant 4" and “MC" *Origin* in background events. The two definitions on the *Type* variable are compared in Figure F.2 in Appendix F. As a result of particles’ passage through the detector matter, secondary electrons and muons occur. This information becomes available

[13] If a reconstructed particle can not be matched to any generator-level particle, a zero PDG ID is returned.

[14] These classifications are available as particle decorations in the Truth DAODs.

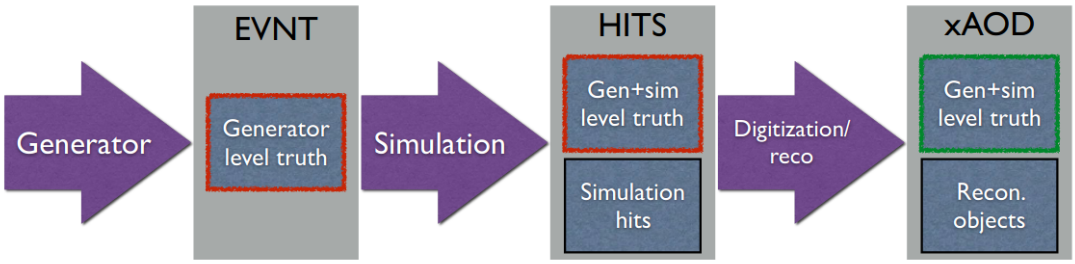


Figure 9.25: The flow of event reconstruction in Monte Carlo modelling and Geant 4 simulation [147]. Two collections of “truth" particles are available: particles at the MC generator level and particles at the generator+simulation level.

with “MC + Geant 4" variables. Further, the “MC + Geant 4" were analyses for the non-matched leptons.

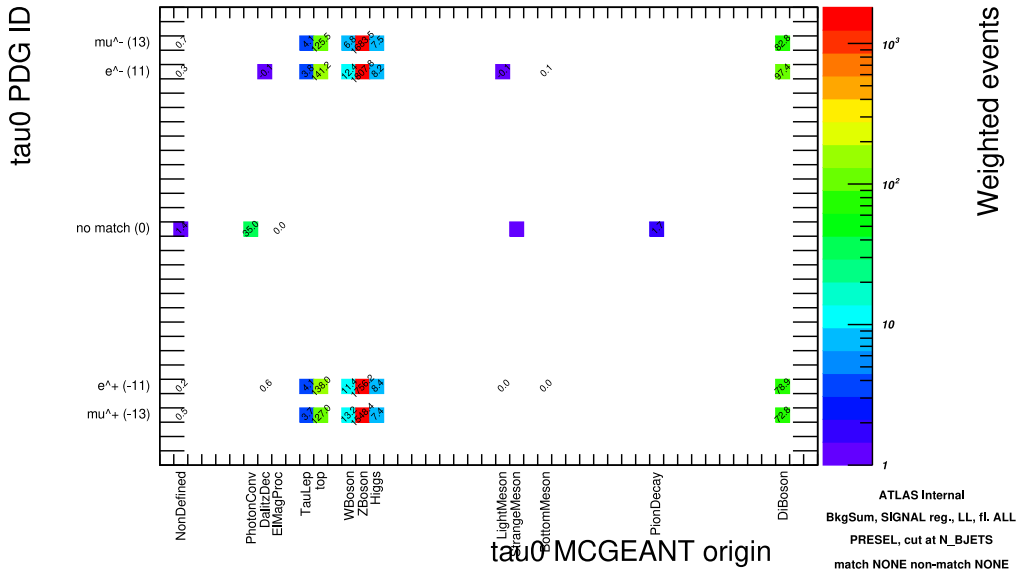


Figure 9.26: The origin of the matched particle-level counterpart of the leading tau-lepton in relation to its PDG ID.

Figure 9.26 compares the origin of the particle-level objects matched to the leading reconstructed tau-lepton. About $\sim 90\%$ of the non-matched leading leptons are electrons produced in photon conversion. A similar trend is observed for non-matched subleading tau-lepton: $\sim 54\%$ of them originate from photon conversion (Figure F.3 in Appendix F). About 92% (56%) of non-matched leading (subleading) tau-lepton are classified as background electrons as seen in Figure 9.27 (Figure F.4 in Appendix F). Figure 9.28 shows the origin of background leptons in three analysed $\tau_{\text{lep}}\tau_{\text{lep}}$ channels. The background light leptons occur after the hard-scattering phase due to interaction with the detector material. This subsequent electromagnetic and hadron activity can lead to the production of background light leptons. No direct parent candidate is found among generator-level particles for background leptons. This explains the source of the non-matched leptons.

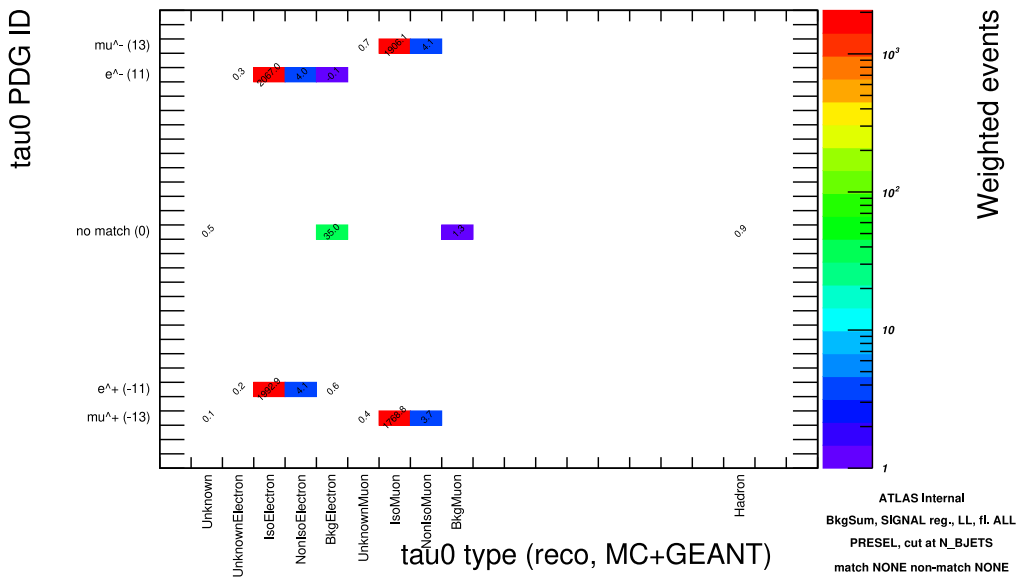


Figure 9.27: The type of the matched particle-level counterpart of the leading tau-lepton in relation to its PDG ID.

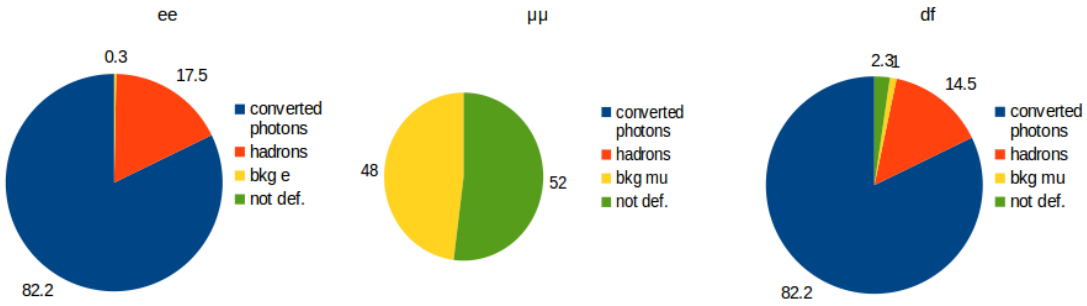


Figure 9.28: The origin of the leptons, which were not matched to any generator-level particle. The composition is shown for the $\tau_e\tau_e$ (left), $\tau_\mu\tau_\mu$ (middle), and $\tau_e\tau_\mu$ (right) final states.

9.5.3 Other options to estimate $j \to l$ fakes

Other potential options to exclude the events with no real light leptons in the $\tau_{\text{lep}}\tau_{\text{lep}}$ final state are the following:

- To accept only the events where a reconstructed light lepton is matched to a generator-level light lepton. This option implies an exclusive match to electrons and muons. Technically, the PDG ID of the generator-level particle needs to correspond to those of the reconstructed lepton.

- To accept only the events where reconstructed light lepton is of an isolated type. This can be done by requesting the *Type* of the matched generator-level particle to be “Isolated electron" (or “Isolated muon").

The *Type* and *Origin* of the leptons selected with the exact $e \to e$ and $\mu \to \mu$ criteria is shown in Figure 9.30 and Figure F.6 (Appendix F), respectively. They

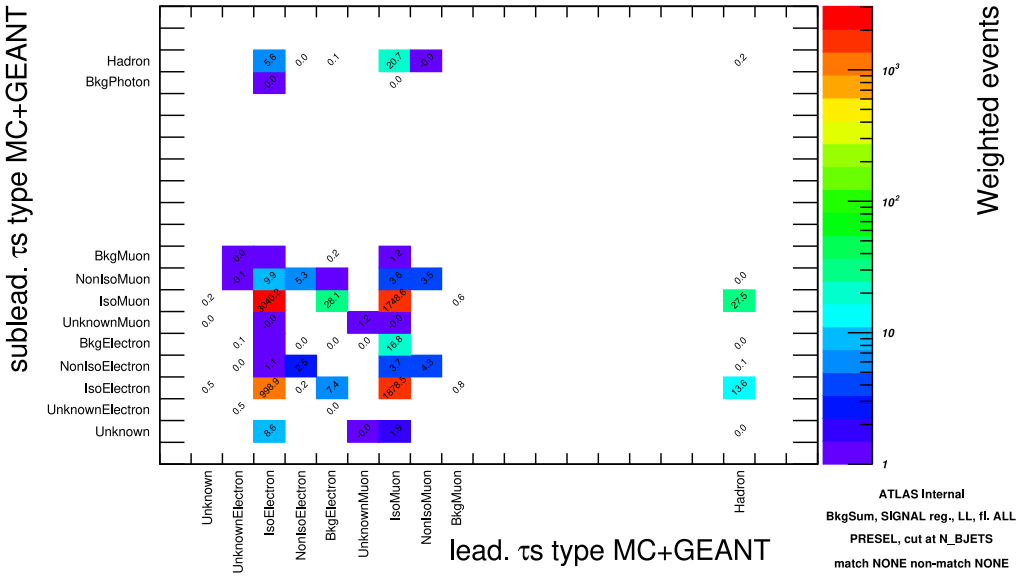


Figure 9.29: The type of matched particle-level counterpart of the leading tau-lepton and subleading tau-lepton. The preselected MC-modelled fake background events are analysed.

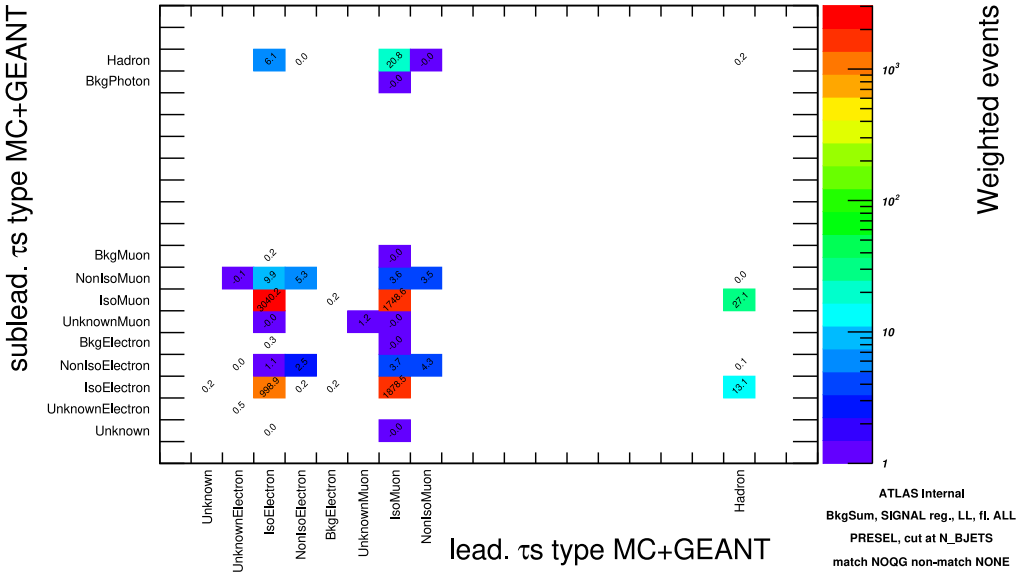


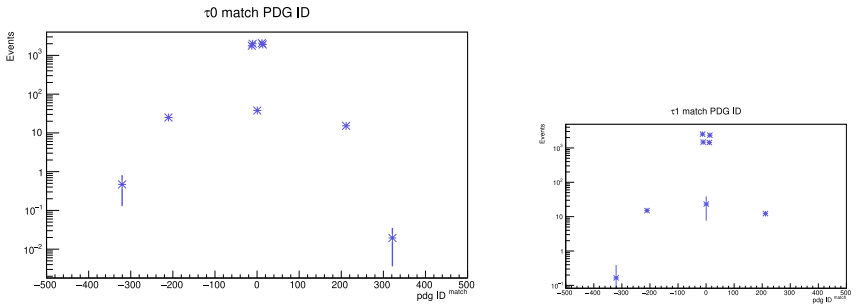
Figure 9.30: The type of matched particle-level counterpart of the leading tau-lepton and subleading tau-lepton. The preselected MC-modelled fake background events are analysed, with the selection removing reconstructed leptons matched to the quarks or gluons (based on the PDG ID of the generator-level particle) and removing leptons unmatched to any generator-level particle..

can be compared to the *Type* and *Origin* variables in Figures 9.29 and F.5 (Appendix F) which were obtained in inclusive events, i.e. with any selection intended to reduce fake leptons. Figure F.8 shows leptons' origin when only "Isolated" elec-

tron and muons are selected.[15] Figure 9.23 shows the $m_{\tau\tau}^{\mathrm{MMC}}$ distributions for the signal events which do not satisfy the tested selection criteria.

The option of selecting the isolated leptons only is, in a sense, somewhat artificial. It removes genuine light leptons although produced not in tau-lepton decays. These processes unavoidably occur during pp collisions, and the ATLAS data contain them. As the data-driven fake estimation does not describe them, the analysis should keep the MC description.

It can be seen that the requirement of the exact $e \to e$ and $\mu \to \mu$ matching removes events of combinatorial background: where one light lepton arises in decays of top-quarks, tau-leptons, W and Z bosons. At the same time, another one can not be matched to any particle, neither MC-only nor MC+Geant 4 simulated.[16] Indeed, these events include leptons mimicked by hadrons ($h \to l$). As one can see in Figure 9.31, pions (PDG ID 211) and kaons (PDG ID 321) mimic leptons. Light leptons faked by hadrons are assumed to be modelled with the ABCD method from real data. Therefore, this selection can be applicable in this analysis.



(a) (b)

Figure 9.31: The PDG ID of the generator-level particles matched to the leading (a) and subleading (b) reconstructed lepton.

Table 9.2 holds the yields by the events which do not pass the requirement on the PDG ID to be exactly light lepton. $h \to l$ fakes make up about 0.4% of the total background across the decay modes. This contribution is small and stays within the uncertainty as seen in the $m_{\tau\tau}^{\mathrm{MMC}}$ distribution (Figure 9.32). Therefore, the potential gain in introducing this selection is low.

9.6 Discussion and summary

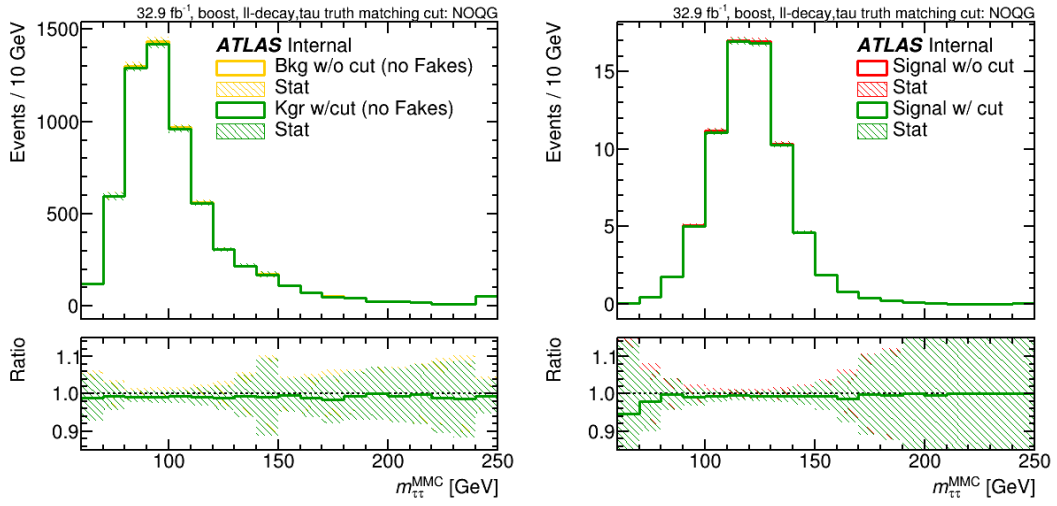
The sections above outlined the performance of the Missing Mass Calculator. Several approaches to improve this algorithm of the di-tau mass reconstruction were tested. Their outcomes were compared with several metrics, including the $m_{\tau\tau}^{\mathrm{MMC}}$ shape and the mean, width and inter-quantile of the $m_{\tau\tau}^{\mathrm{MMC}}$ distribution. The level of separation between the $H \to \tau\tau$ signal and $Z \to \tau\tau$ background events

[15] The *Type* is not shown, as it includes only isolated leptons.

[16] The leptons of the `Unknown` *Type* or `Non-Defined` *Origin* are interpreted as outcomes of mismeasurements, misreconstruction, or misidentification. It can point out the missing information in the event record.

cut	no match to isolated l type			no match to e/μ PDG ID		
channel	τ^0_{lep}	τ^1_{lep}	$\tau_{\text{lep}}\tau_{\text{lep}}$	τ^0_{lep}	τ^1_{lep}	$\tau_{\text{lep}}\tau_{\text{lep}}$
$\tau_e\tau_e$	0.109	0.111	0	0.109	0.111	0
$\tau_\mu\tau_\mu$	0.006	0.021	0	0.001	0.017	0
$\tau_e\tau_\mu$	0.261	0.132	0.004	0.266	0.131	0

Table 9.2: The yields of events in which the reconstructed leading (subleading) τ^0_{lep} (τ^1_{lep}) light lepton or both leptons $\tau_{\text{lep}}\tau_{\text{lep}}$ does (do) not pass the type requirements. The type and PDG ID are defined as the type and PDG ID of the matched generator-level particle.



(a) (b)

Figure 9.32: The $m^{\text{MMC}}_{\tau\tau}$ distributions in the total background prediction (a) and $ggH \to \tau\tau$ (b) events: before and after the requirement on the PDG ID on the matched generator-level particle.

was studied with the ROC curve method. Signal-to-background discriminating power was estimated with the AUC value.

The MMC mass resolution is $\sim 16\%$, $\sim 17\%$, $\sim 18\%$ for τ_{had} τ_{had}, τ_{had} τ_{lep},τ_{lep} τ_{lep} channels, respectively. In the case tau-lepton decays leptonically, the worse resolution is observed due to the presence of additional neutrino(s). This emphasised the importance of an accurate MET estimation and evaluation of its possible miscounting (i.e. MET resolution). Presented research evaluated how the precision of the missing transverse energy estimation impacts the $H \to \tau\tau$ analysis. The applicability of the object-based MET significance was scrutinised.

Various methods of MET resolution estimation were tested. Notably, the MET resolution estimation through the MET significance was advantageous due to its independency of the data set and decay channel and improved readability. Specifically, three methods were compared: the first one uses the MET resolution parametrisation on event parameters, the second one estimates the MET resolution through the object-level MET significance $\sigma_{\text{MET}} = E^{\text{miss}}_{\text{T}}/S_{\text{MET}}$, and the third one employs the values for longitudinal and transverse MET resolution components as calculated by the ATLAS `METSignificanceTool`. The three compared approaches led to consistent MMC performance. The MMC provides the

signal acceptance at the level of 85% with the background rejection of $\sim 70\%$.

The best MMC performance is achieved when the MET resolution σ_{MET} is parametrised as a function of $\sum E_{\text{T}}$, $\langle\mu\rangle$, and, for $\tau_{\text{lep}}\tau_{\text{lep}}$ events, $\Delta\phi_{\text{ll}}$. This new parametrisation treats the pile-up effect, which caused additional smearing of the MET resolution. This final MET resolution parametrisation in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel was updated for the data-taking and reconstruction conditions. Such σ_{MET} estimation through multiple parameters resulted in the steady performance of the MMC. This method was used in the MMC for di-tau mass reconstruction in the final version for publication of the $H \to \tau\tau$ analysis with the Run 2 data set.

In addition, the computational time of the MMC was optimised. The phase space scan with 50 k iterations in the Markov chain was sufficient for Run 2 data set analysis due to the improved data-taking conditions and object reconstruction schemes. The CPU time of mass calculation was reduced, which resulted in about twice the faster event processing. The final MMC setup improved mass resolution by $\sim 1\%$ [5] with respect to the previous measurement based on 2015-16 data [4].[17]

The author gives an outlook at perspective on ways of mass reconstruction improvement. Namely, the object-level MET significance usage in event selection and/or mass reconstruction is promising. There are indications that better mass resolution and background rejection can be reached.

Fake background modelling was studied in the $\tau_{\text{lep}}\tau_{\text{lep}}$ decay channel in the $H \to \tau\tau$ coupling measurement. QCD jets mimicking light leptons are modelled by the data-driven method. Other sources of fake τ_{lep} background are intentionally modelled with Monte Carlo simulation. This approach prevents potential ambiguities in the fake background prediction.

Several other options for validating the $j \to l$ fake estimation were tested. The requirement of non-matching to quarks or gluons was ineffective as no light leptons were mimicked by jets found in Monte Carlo simulated events. The studies described earlier confirm that light lepton fakes by jets are estimated purely with the data-driven method. As the data- and MC-based definitions do not interfere, there is no need to subtract the $j \to l$ contribution in MC (specifically, in formulae Eq. (9.8)).

At one, the presented analysis showed that about 0.4% of events include light leptons faked by hadrons. The selection on the *Type* or PDG ID of the matched counterpart removes $h \to l$ events. This can be technically achieved by requiring PDG ID to correspond to light leptons. Alternatively, potential double-counting of the $h \to l$ fakes can be prevented with the criterion on the *Type* variable. The requirement on the isolated *Type* of light lepton is placed. Compared to the tested "isolated-only" *Type* selection, this criterion is relaxed, but it preserves non-prompt light leptons from background processes. Simultaneously, it would remove $h \to l$ fakes and prevent a bias on the fake estimation.

Further, the author studied the nature and genesis of the reconstructed light leptons, which do not match any generator-level particle. About 82% of events with non-matched leptons are represented by background electrons originating in photon conversion. As these events include real light leptons but not genuine

[17] Besides the parametrisation of the MET resolution and the optimisation of the computational time, the final MMC version included the update of several factorisations, which were made by Michael Hübner [148].

τ_{lep}, they need to be included in the background prediction in the $H \to \tau\tau$ measurement. These fakes are predicted based on MC, as shown in this thesis. This analysis also revealed that more descriptive variables on the type and origin are defined after detector effects simulation.

10. Search for excited tau-leptons

10.1 Introduction

10.1.1 Motivation

This search aims to find excited tau-leptons. Several experimental results are incompatible with the Standard Model predictions, and some of the inconsistencies can be explained by the composite nature of the SM particles. According to the idea of compositeness, detected up-to-now particles are compounds of even smaller building blocks.[1]

Electron anomalous magnetic momentum deviates from the SM predictions by 2.4σ according to to [6]. The experimental results of the measurements of muon magnetic momentum $\mu - 2$ deviate from the SM expectation by 4.2σ [7]. While a new set of constraints on tau-lepton anomalous magnetic momentum [150] are compatible with the SM prediction with the full uncertainty, the best-fitted α_τ value is outside of the SM predicted 95% confidence-level interval.[2] These observations may imply the compositeness of leptons. Likewise, the large gyromagnetic factor of a proton is a consequence of its compound nature. The anomalies observed in heavy meson decays [152, 153, 154, 155, 156] may also impose ideas of lepton non-universality and compositeness [157]. Eventually, compositeness can be a clue to explain the existence of three particle generations in the Standard Model.

The experimentally referred evidence of excited leptons can be one of the signatures of lepton compositeness. At the atomic level, transitions from excited to ground atomic states are typically accompanied by the emission of photons. Orbitally excited resonances (for example, B_s mesons) decay to states with lower orbital momentum and γ-quantum. Similarly, excited D-mesons de-excite to D-mesons with the emission of pions. Observation of two octets of mesons - scalar and vector ones - with the same quark composition but different isospin configuration indicates the bounding of structural quarks inside them. In a like way, observation of leptons and their excited states can be a manifestation of the composite nature of both.

If excited states of tau-leptons exist, they can be observed in their transitions to Standard Model particles. The search is based on the ATLAS Run 2 data set. The following sections describe the search for excited tau-leptons, the analysis strategy, and sensitivity to excited tau-leptons (or, abbreviated, ETL).

10.1.2 Theory of excited tau-leptons

The benchmark model for this search is described in the paper by U. Baur, M. Spira, and P.M. Zerwas [158]. Within the model, leptons are composite par-

[1] They are often referred to as preons following the model and the terminology suggested in the paper [149].

[2] The recent results on α_τ measurement from the CMS Collaboration [151] differ from the findings by the ATLAS, but these results are defined with larger uncertainty for the moment. A future measurement from both experiments with an extended data set could clear it up.

ticles and can exist in standard and excited states. The former is the ordinary Standard Model lepton, and the latter is the hypothetical excited lepton.

Similarly to SM leptons, excited leptons are fermions. The spin and isospin of the excited fermions are set to 1/2 [159]. The hypothetical left- and right-handed excited leptons form $SU(2)$ isospin homodoublets:[3]

$$\begin{pmatrix}\nu_e^*\\ e^{*-}\end{pmatrix}_L, \begin{pmatrix}\nu_e^*\\ e^{*-}\end{pmatrix}_R, \begin{pmatrix}\nu_\mu^*\\ \mu^{*-}\end{pmatrix}_L, \begin{pmatrix}\nu_\mu^*\\ \mu^{*-}\end{pmatrix}_R, \begin{pmatrix}\nu_\tau^*\\ \tau^{*-}\end{pmatrix}_L, \begin{pmatrix}\nu_\tau^*\\ \tau^{*-}\end{pmatrix}_R$$

Excited leptons carry a unit lepton number, a negative unit weak hypercharge and a unit electric charge.

The model suggests that there is a contact interaction between leptons and quarks. The Feynman diagram of contact interaction with a four-fermion vertex is drawn in Figure 10.1a. Due to the contact interaction, excited tau-leptons directly could couple to SM quarks and leptons and can be produced in pp collisions at the LHC.

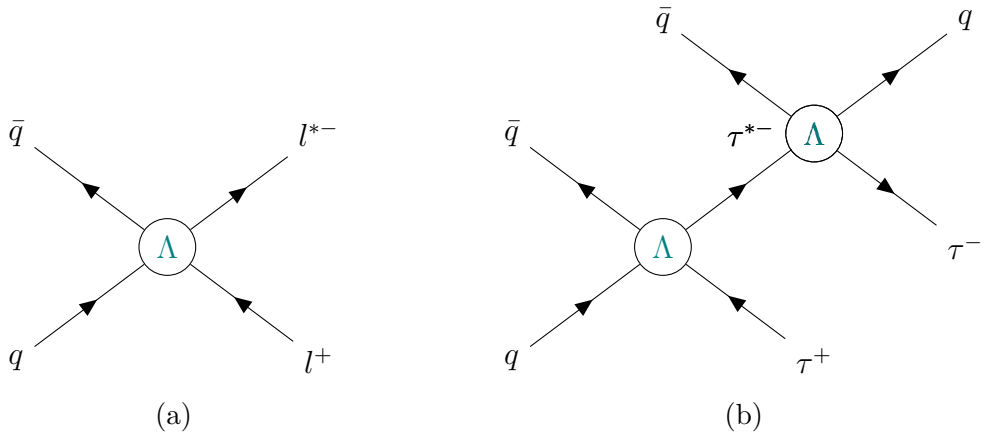


Figure 10.1: (a) Schematic Feynman diagram for scattering at a 4-vertex with the production of single excited lepton. (b) Feynman diagram for single excited tau-lepton production and decay through the four-fermion contact interaction.

The following effective Lagrangian describes contact interaction between four fermions:

$$\mathcal{L}_{\text{contact}} = \frac{g_*^2}{\Lambda^2}\frac{1}{2}j^\mu j_\mu \tag{10.1}$$

where g_* is the strength of contact interaction, j_μ is the contact-interaction current is defined:

$$j_\mu = \eta_L \bar{f}_L \gamma_\mu f_L + \eta'_L \bar{f}^*_L \gamma_\mu f^*_L + \eta''_L \bar{f}^*_L \gamma_\mu f_L + h.c + (L \to R) \tag{10.2}$$

where η_L, η'_L and η''_L are factors for currents of contact interaction between SM fermions, excited and SM fermions, and excited fermions, respectively. ETL could be produced in pairs or singly in association with an SM tau-lepton.

The considered contact interaction model introduces two free parameters: the mass of excited lepton $m_{\tau*}$ and the compositeness scale Λ. The mass defines the kinematics of the final state. The compositeness scale scales the probability of

[3]Excited leptons can also be assigned to isosinglet. Its treatment generally repeats the treatment of isodoublet and therefore is omitted from now on.

ETL production and decay. The compositeness of excited leptons does not reveal itself at the energies below the compositeness scale Λ.

Excited leptons can also participate in electroweak interactions. The gauge-mediated interaction between excited fermions f^* and gauge bosons is described by the effective Lagrangian:

$$\mathcal{L}_{\text{gauge}} = \bar{f^*}\gamma^\mu \Big[g_s \frac{\lambda^a}{2} G^a_\mu + g\frac{\tau}{2}\mathbf{W}_\mu + g^{'}\frac{Y}{2}B_\mu \Big] f^* \tag{10.3}$$

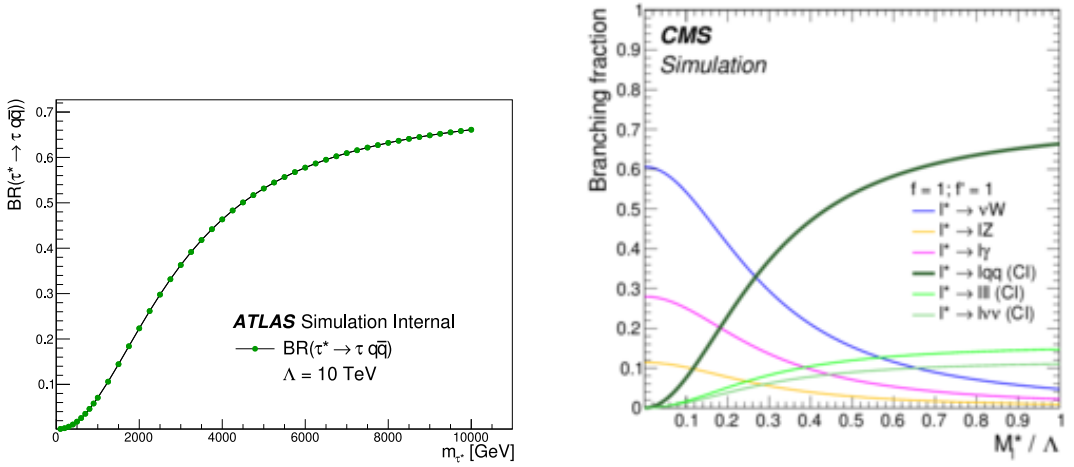
where G^a_μ, $\mathbf{W}_\mu$, and B_μ stand for the gluon, $SU(2)$, and $U(1)$ field, respectively. g_s, $g = e/\sin\theta_W$, and $g^{'} = e/\cos\theta_W$ are strong, $SU(2)$ electroweak, and $U(1)$ electroweak coupling constant, respectively.

The following effective Lagrangian characterizes the magnetic-transition interaction between ordinary and excited leptons:

$$\mathcal{L}_{\text{trans}} = \frac{1}{2\Lambda}\bar{\psi^*_R}\sigma^{\mu\nu}\Big(gf\frac{\tau}{2}\mathbf{W}^{\mu\nu} + g'f'\frac{Y}{2}\mathbf{B}_{\mu\nu}\Big)\psi_L + h.c. \tag{10.4}$$

In the formulae above, ψ denotes lepton spinor, g and g' are $SU(2)$ and $U(1)$ coupling scales, $\mathbf{W}^q_{\mu\nu}$ and $\mathbf{B}_{\mu\nu}$ are the $SU(2)$ and $U(1)$ field strengths.

10.1.3 The model: excited tau-lepton production and decay



(a) (b)

Figure 10.2: The branching ratio of the excited tau-lepton production by contact interaction as calculated by the Pythia 8 Monte Carlo generator upon signal sample simulation (a). The branching ratios of excited lepton decay channels as a function of the ratio of the excited lepton mass (noted in this plot specifically as M^*_l) and compositeness scale (Λ) for fixed values of the model parameters $f = f^{'} = 1$ as assumed [121].

In proton-proton collisions, excited tau-leptons are predominantly produced by the contact interaction in quark-antiquark scattering. Their generation in vector boson fusion yields only a negligible fraction of excited tau-lepton production

at the LHC [160]. This analysis focuses on the single ETL production since its double production of excited leptons has a lower cross section. ETLs can decay in electroweak and contact interactions. The former can occur in $l^* \to lZ$, $l^* \to l\gamma$, $l^* \to W\nu$ processes, and the latter occurs in four-fermion interactions such as $l^* \to lq\bar{q}$. They preferably decay to an ordinary tau-lepton and a pair of quarks if $m_{\tau*}/\Lambda > 0.3$ as noted in [161] and shown in Figure 10.2. Therefore, this search focuses on the contact interaction single production and decay of excited tau-leptons. The diagram of this process is illustrated in Figure 10.1b. This search's final state of interest includes two hadronically decaying tau-leptons and two jets.

In this search, contact-interaction currents between left-handed particles are considered (i.e. the η factors introduced in Eq. (10.2) are assumed to be equal to one).[4] The constant of the contact interaction strength g_* is taken equal to 4π for normalization purposes. The coupling strength parameters (g_*, η_L, $\eta_L^{'}$, $\eta_L^{'}$) are conservatively assumed to be unit.

To consider only contact-interaction mediated generation and decay of ETLs, ETL decays in other interactions are ommitted.[5]

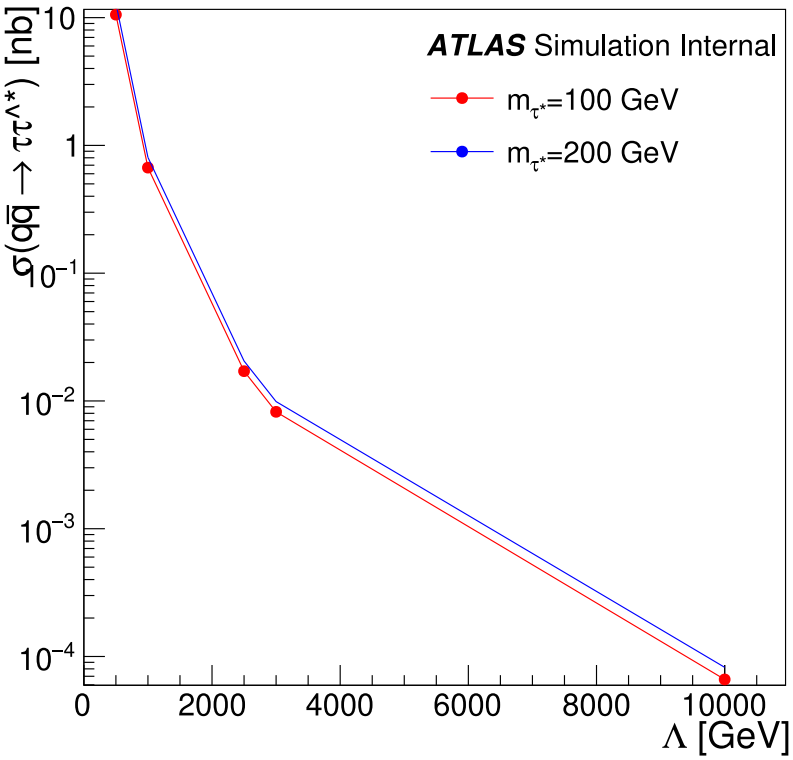


Figure 10.3: The expected cross section of the process as a function on the compositeness scale of the contact interaction Λ. The production cross sections for excited tau-leptons with masses of 100 GeV and 200 GeV are shown.

[4]For simplicity, right-handed currents are neglected as they interact in the same manner.

[5]Dynamics of contact interaction and kinematic properties of the final state do not depend on whether gauge-mediated interaction included in simulation or not. This is shown in Appendix G.

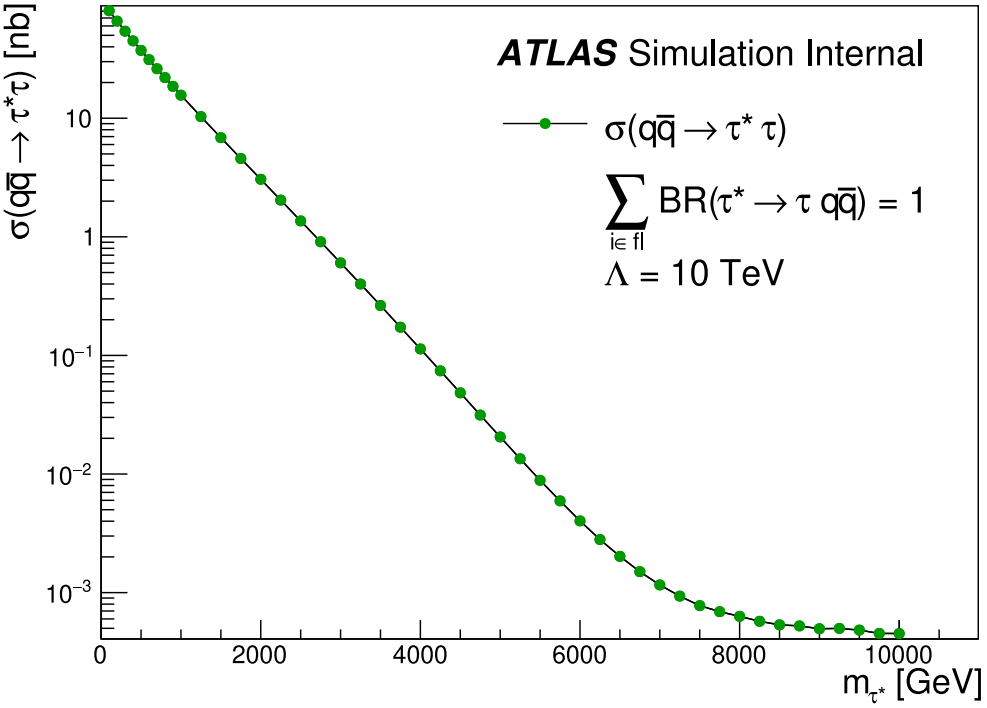


Figure 10.4: The excited tau-lepton production cross section as a function of the ETL mass.

As mentioned above, the produced excited leptons' mass fundamentally defines the final-state kinematics. The production cross section is proportional to $1/\Lambda^4$, and the shape of kinematic distribution follows $m_{\tau*}$. Figure 10.4 illustrates the expected cross section of excited tau-lepton production in quark-antiquark scattering. Figure 10.3 shows the dependence of the cross section of ETL production in contact interaction on the compositeness scale Λ of contact interaction.

10.1.4 Overview of existing results on excited leptons and contact interaction

High-energy physics experiments have been searching for excited leptons for the last decades. The Large Electron-Positron Collider (ALEPH [162], DELPHI [163], L3 [164], OPAL [165]), HERA (H1 [166] and ZEUS [167]), Tevatron (CDF [168] and D0 [169]), and LHC [170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 191, 192, 192, 193, 194] set limits on the excited lepton and contact interaction parameters.

The LHC experiments set the most stringent limits on the excited lepton masses. Figure 10.5 depicts the values. The excited electron and muon of masses below 5.6 and 5.7 TeV, respectively, were the most stringently, for the time being, excluded by the CMS collaboration when the compositeness scale is assumed to be equal to the excited tau-lepton mass [121]. In this search, the CMS Collaboration excludes the existence of For the case of $\Lambda = 10$ TeV, the upper excited electron (muon) of mass below 3 (3.5) TeV [121]. The only to date cited lower limit on

the excited tau-lepton mass of 2.5 TeV is obtained in the $m_{\tau*} = \Lambda$ scenario by the ATLAS collaboration in the model-independent search in multi-lepton final states [195].

The most up-to-date limits on the compositeness scale Λ_{llqq} are established in searches in high-mass di-lepton final states [196, 197]. Similarly, the LHC measurements of dijet final states [198, 199] set limits on the Λ_{qqqq} interaction scale. Figure 10.6 presents the ATLAS and CMS results, where the upper limits for Λ are higher than 9 TeV.

10.2 Goals

The presented search aims to discover excited tau-leptons in the Run 2 ATLAS data. In the case of no BSM signature, upper limits on the excited tau-lepton properties are to be determined.

The analysed data set includes 139 fb^{-1} of data collected by the ATLAS detector in pp collisions at the Large Hadron Collider during Run 2.[6] Signatures of excited tau-leptons are searched in the final states with two hadronically decaying tau-leptons and at least two jets. The expected signal significance is to be calculated. For the case of no excess of events, 95% confidence level limits on excited tau-lepton production cross section and excited tau-lepton mass will need to be set.

The next sections demonstrate the steps made, from the model implementation for the Monte Carlo generation of the excited tau-lepton signal to the estimation of the expected discovery significance and expected lower limits on the excited tau-lepton mass.

10.3 Signal and background modelling

10.3.1 Generation of excited tau-lepton signal

The production and decay of hypothesised excited tau-leptons were simulated in proton-proton collisions at 13 TeV centre-of-mass energy. The excited tau-lepton samples are generated with the Pythia 8 (version 8.243) Monte Carlo generator. The generation used the `NNPDF2.3LO` set of parton distribution functions (PDFs) and the A14 tune for modelling underlying events. An excited tau-lepton is produced in quark-antiquark scattering through contact interaction. An ordinary (SM) tau-lepton is produced as well. The excited tau-lepton further decays to a tau-lepton and a quark-antiquark pair via the contact interaction. In this way, this search purely focuses on contact interaction. All flavours of quarks in the final state are considered.[7] The coupling strength parameters (g_*, η_L, $\eta_L^{'}$, $\eta_L^{'}$) are set to a unit, and the contact interaction strength parameter g_* is set to 4π as discussed above (Section 10.1.2).

[6] On several plots shown in this chapter, an attentive reader can notice a different label "138.2 fb^{-1}". These results were obtained based on a slightly smaller data set before a part of the data from 2017 had been reprocessed.

[7] Decay $\tau^* \to \tau t\bar{t}$ is kinematically allowed for excited tau heavier than ~ 360 GeV.

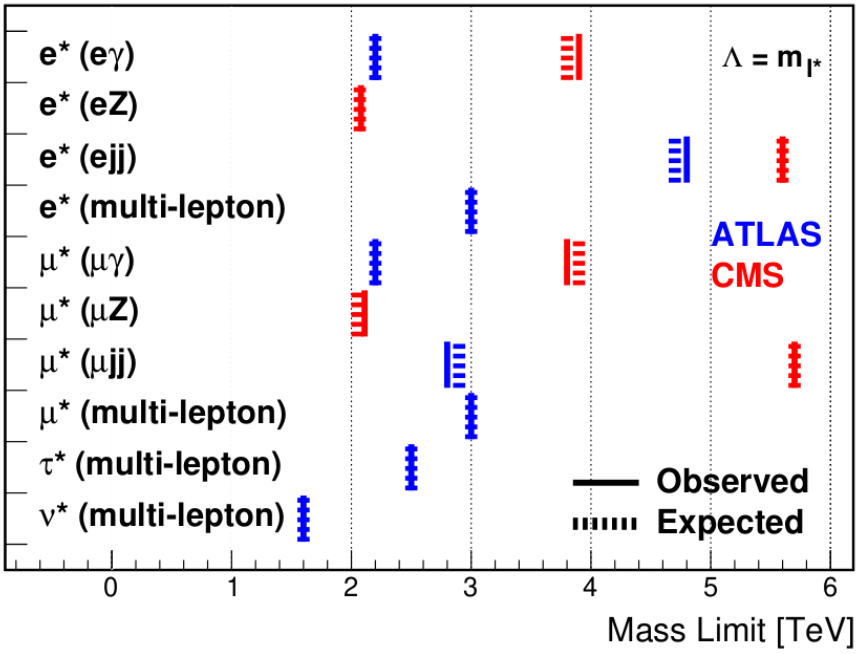


Figure 10.5: 95% confidence level lower limits on the mass of excited leptons set by the LHC experiments [200].

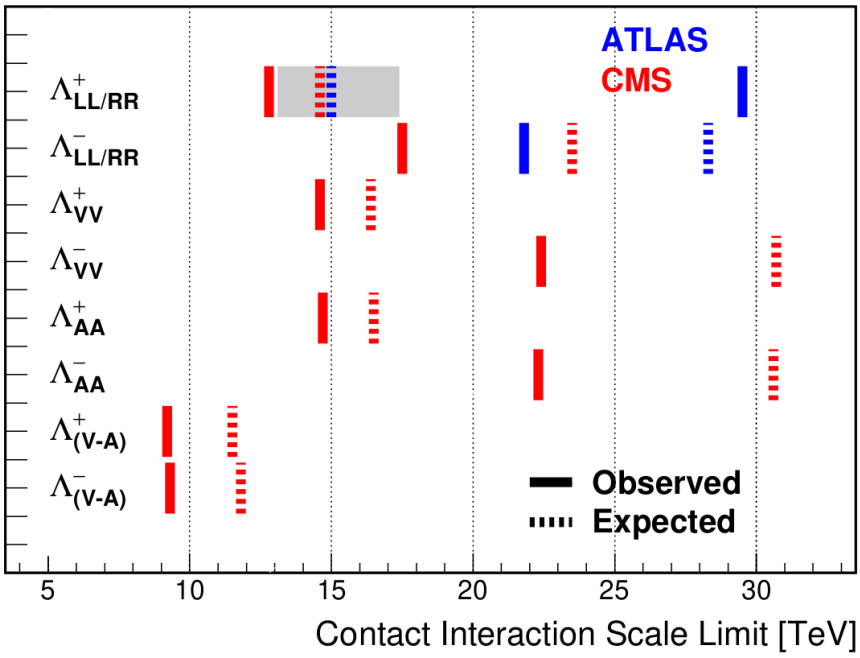


Figure 10.6: Observed (solid lines) and expected (dashed lines) 95% confidence level lower limits on the contact interaction scale Λ obtained by the LHC experiments (ATLAS [199] and CMS [198]) [200].

This search is performed with the compositeness scale Λ fixed at 10 TeV. This choice is mainly motivated by the existing limits on the compositeness scale Λ (Figure 10.6). The unitarity condition allows searching for ETL of any mass up to 10 TeV [201].[8]

[8]It is worth mentioning that rescaling to another Λ value is rather trivial given the fact that

The MC signal samples were generated for excited tau-leptons of a mass between 200 GeV and 10 TeV.[9] This allows for determining limits on the ETL mass and its production cross section. The scan over ETL mass between 1 TeV and 10 TeV is performed with a step of 250 GeV. For the mass points below 1 TeV, the search is performed in a finer grid with a step of 100 GeV.

Complete information on the simulation of ETL production in pp collisions is contained in the generated `EVNT` data set. In the ATLAS workflow, the MC production is made with the `AthGeneration`.

Modelling of excited tau-leptons was validated by studying kinematic properties of the final state at the generator level.[10] The obtained "truth" record includes information on hard-scatter production and consequent particles decays. The generator ("truth") level studies provide a good tool to quickly evaluate MC generation outcomes.[11] Figure 10.7 shows the generator-level kinematic distributions for the $\tau\tau jj$ final state.

The excited tau-lepton samples were then produced centrally by the ATLAS Particle Modelling Group with the `Athena` production framework.[12]. The hard scattering is generated by the `Pythia 8`. The produced HepMC record was then passed to `EvtGen` [202], which simulates spins and polarisation of heavy-flavour particles, particularly decays of B- and D-mesons. Table 10.1 lists the generated excited tau-lepton samples. The pile-up reweighting configuration with the $\langle\mu\rangle$ distributions was derived for each MC signal sample. The simulation of the detector response was made with the `Geant 4` toolkit. The particle propagation at the ATLAS calorimeter was modelled with the **A**TLAS **F**ast Simulation II (FastSim2, or AFII).

The generation of ETL samples was split into three campaigns: mc16a, mc16d, and mc16e. These campaigns correspond to ATLAS Run 2 data-taking periods: 2015+2016, 2017, and 2018. The number of generated events for a generation was adopted from the search for excited electron and muon ATLAS [160] and doubled due to multiple decay modes of $\tau_{\text{had}}\tau_{\text{had}}$. A proportion between the integrated luminosity collected in mc16a:mc16:d:mc16e periods was kept. In total, 100 k:140 k:180 k (30 k:40 k:50 k) events were generated for ETL with mass below (above) 1 TeV.

Further, the simulated signal samples undergo digitisation and reconstruction as outlined in Section 7.1.

10.3.2 Modelling of Standard Model background

Several Standard Model processes can produce the di-tau and dijet final state. The SM contribution has to be estimated thoroughly for effective signal discrimination from the SM contributions.

---

it changes only the cross section but not the kinematic properties of the final state. This is demonstrated in Appendix H.

[9]The 100 GeV mass point was initially included in this analysis. However, it was eventually decided to explore masses from 200 GeV as will be pointed out later.

[10]For that, the `EVNT` containers were converted to "truth"-level data: `DAOD_TRUTH1` or `DAOD_TRUTH3` depending on the required level of details.

[11]As said in Section 4, the generator-level ("truth"-level) information is also very useful to study resolution and reconstruction effects.

[12]MC generation was performed with `AthGeneration` release 21.6.42.

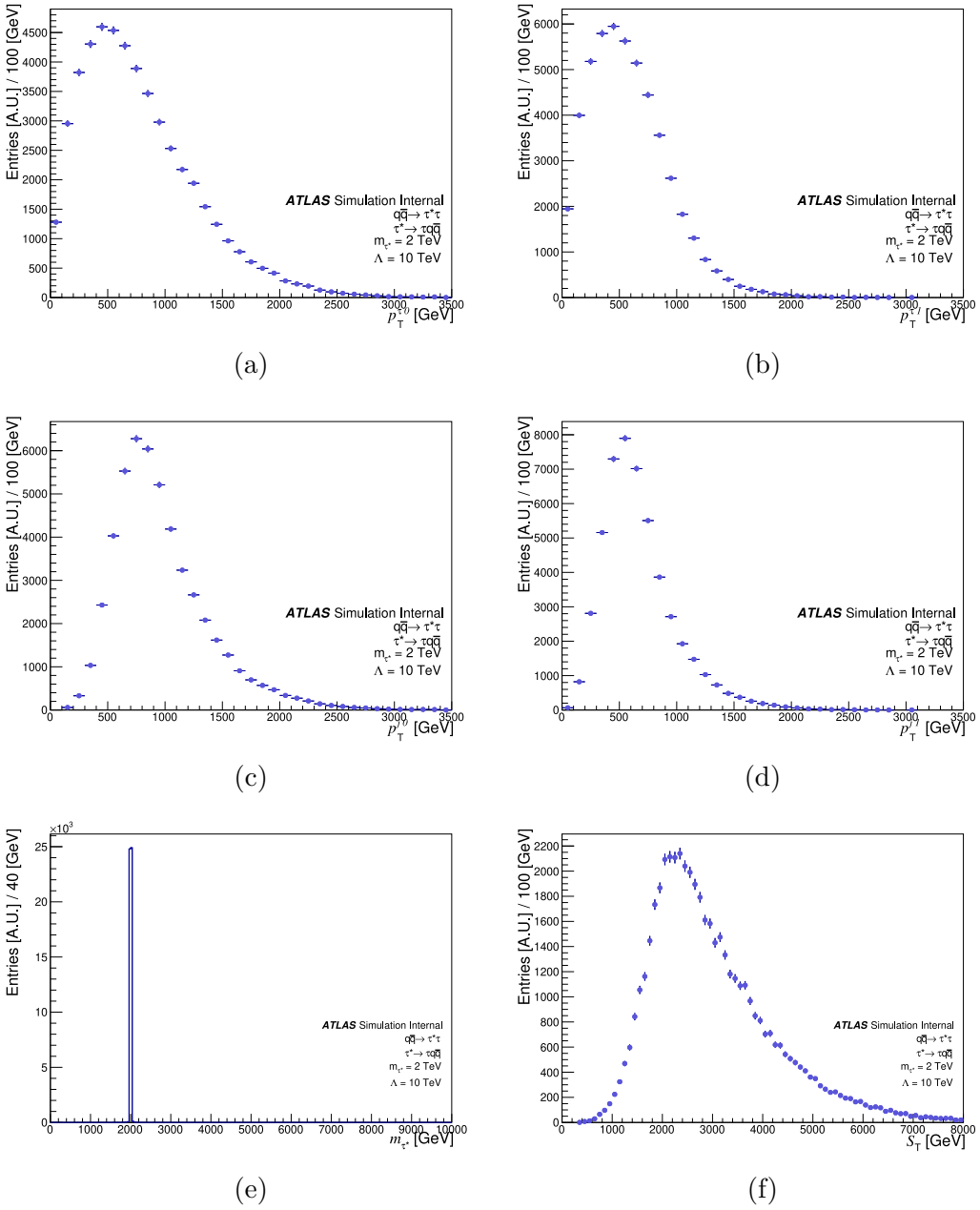


Figure 10.7: The kinematic distributions for the di-tau plus dijet final state in a process of production and decay of excited tau-lepton of 2 TeV mass at the compositeness scale $\Lambda = 10$ TeV: leading jet p_{T} (a), subleading jet p_{T} (b), leading tau-lepton p_{T} (c), subleading tau-lepton p_{T} (d), the invariant mass of the produced ETL (e), the S_{T} (scalar sum of p_{T} of two final-state tau-leptons and two leading in p_{T} jets) (f).

An irreducible background is represented by processes leading to $\tau\tau jj$ final states with real τ_{had} and jets. The following sources of the irreducible background are considered: di-tau production in Drell Yan process $Z \to \tau\tau$ and top-quark pair production in di-lepton decay mode (the decay chain $t\bar{t} \to WbWb \to \tau\nu_\tau b\tau\nu_\tau b$). Irreducible background rises from the top-quark production associated with W boson tW (the decay chain $Wt \to WWb \to \tau\nu_\tau\tau\nu_\tau b$) and in the case of leptonic W

DSID	$\tau*$ mass	Λ	$N_{\text{evt}}^{\text{gen}}$	DSID	$\tau*$ mass	Λ	$N_{\text{evt}}^{\text{gen}}$
800603	100 GeV	10 TeV	420k	800626	4.5 TeV	10 TeV	120k
800604	200 GeV	10 TeV	420k	800627	4.75 TeV	10 TeV	120k
800605	300 GeV	10 TeV	420k	800628	5 TeV	10 TeV	120k
800606	400 GeV	10 TeV	420k	800629	5.25 TeV	10 TeV	120k
800607	500 GeV	10 TeV	420k	800630	5.5 TeV	10 TeV	120k
800608	600 GeV	10 TeV	420k	800631	5.25 TeV	10 TeV	120k
800609	700 GeV	10 TeV	420k	800632	5.5 TeV	10 TeV	120k
800610	800 GeV	10 TeV	420k	800633	5.75 TeV	10 TeV	120k
800611	900 GeV	10 TeV	420k	800634	6 TeV	10 TeV	120k
800612	1 TeV	10 TeV	420k	800635	6.25 TeV	10 TeV	120k
800613	1.25 TeV	10 TeV	120k	800636	6.5 TeV	10 TeV	120k
800614	1.5 TeV	10 TeV	120k	800637	6.75 TeV	10 TeV	120k
800615	1.75 TeV	10 TeV	120k	800638	7 TeV	10 TeV	120k
800616	2 TeV	10 TeV	120k	800639	7.25 TeV	10 TeV	120k
800617	2.25 TeV	10 TeV	120k	800640	8 TeV	10 TeV	120k
800618	2.5 TeV	10 TeV	120k	800641	8.25 TeV	10 TeV	120k
800619	2.75 TeV	10 TeV	120k	800642	8.5 TeV	10 TeV	120k
800620	3 TeV	10 TeV	120k	800643	8.75 TeV	10 TeV	120k
800621	3.25 TeV	10 TeV	120k	800644	9 TeV	10 TeV	120k
800622	3.5 TeV	10 TeV	120k	800645	9.25 TeV	10 TeV	120k
800623	3.75 TeV	10 TeV	120k	800646	9.5 TeV	10 TeV	120k
800624	4 TeV	10 TeV	120k	800647	9.75 TeV	10 TeV	120k
800625	4.25 TeV	10 TeV	120k	800648	10 TeV	10 TeV	120k

Table 10.1: The list of data set identification (DSID) numbers for the Monte Carlo simulation of excited tau-lepton production and decay in contact interaction.

boson decay. Another contributor is the electroweak production of di-boson with their decay to final states with two leptons and two jets. The listed background processes are simulated with Monte Carlo.

A reducible background is represented with events where QCD jets or light leptons (electrons, muons) mimic one or two tau-leptons. A leading contributor to the reducible background is multi-jet and $W+$jets events. Among other contributors, there is Z boson production in di-lepton decay mode $Z \to ll$, semi-hadronic and fully hadronic $t\bar{t}$ decays, hadronic W boson decays in Wt production, and, less frequently, single top quark production in s- and t-channel. Background, where one or two hadronically decaying tau-lepton is mimicked by jet ($j \to \tau_{\text{had}}$), is estimated with a data-driven Fake Factor method, which was introduced in Section 6.1. Events where τ_{had} faked by light leptons ($l \to \tau_{\text{had}}$) make up a small fraction of the reducible background. These processes are modelled by MC. Multi-top and Higgs boson production are neglected in this analysis due to their small cross section.

Monte Carlo modelling of the background processes is discussed earlier in Section 9.1.1 and detailed in AppendixB.

10.3.3 Data preparation and analysis

This analysis exploits a unified format for AOD derivations – the `DAOD_PHYS` format. The derivation of `DAOD_PHYS` for data, ETL and background samples was undertaken by the responsible liaisons in the ATLAS Exotics group.

The final set of NTuples was prepared with Prague's Athena-based framework (aka AAC) [108]. The pre-fit studies were performed with Prague's framework (aka RAC) [203] based on the ROOT software [204]. The AAC and RAC software were purposefully set up for the needs of the ETL search. A majority of presented plots are made with plotting software [205].

10.4 The strategy of the search for excited tau-leptons

As discussed above, signatures of the excited tau-lepton production are searched in events with two ordinary (Standard Model) tau-leptons and two jets. A pair of jets and tau-leptons are occasionally referred to as the final-state quadruplet in this text. The events undergo selection to reduce the contribution of Standard Model background processes and, in this way, to increase the analysis sensitivity to the excited tau-lepton signal. The event selection is cut-based: final-state events are selected by applying criteria on several observables. The selection is applied to transverse momenta of leading and subleading tau-leptons and jets, the multiplicity of jets and tau-leptons. Further requirements for the di-tau system provide an optimized separation of the excited tau-lepton signal from the SM background.

The signal region was defined as the phase space where the production of the excited tau-leptons is preferable. The contribution from Drell-Yan events ($Z \to \tau\tau$, $Z \to ll$) is reduced with Z boson mass veto. Background processes are further suppressed by the requirement on the L_{T} variable, which is the scalar sum of transverse momenta of leading and subleading tau-leptons.

The most important observable in this search is the scalar sum of transverse momenta of each object making the final-state quadruplet: $S_{\mathrm{T}} = p_{\mathrm{T}}^{\tau 0} + p_{\mathrm{T}}^{\tau 1} + p_{\mathrm{T}}^{j0} + p_{\mathrm{T}}^{j1}$. The S_{T} discriminant was found to be most sensitive to the presence of ETL. It also allows for discrimination between signals from excited tau-leptons of different masses.

The leading sources of the Standard Model background are the $Z \to \tau\tau$, Top (pair $t\bar{t}$ and single top-quark) production, and Fakes, i.e. events with one or both τ_{had} fakes.[13] The $Z \to \tau\tau$ and Top contribution is estimated with the Monte Carlo modelling. The corresponding control regions – $Z \to \tau\tau$ CR and Top ones – are introduced in the fit (as it will be described in Section 10.11.1). This helps to constrain and normalise MC prediction with real data. The Fake background is estimated with the data-driven Fake Factor method. This fake background prediction is validated in the phase space maximally close to the signal region.[14] Modelling of other SM background processes is based on the MC,

[13] The combined contribution from the top-quark pair production $t\bar{t}$ and single top-quark production (Wt, s- and t-channels) are combined in the presented analysis.

[14] The Fakes are validated in the same-sign control region, i.e. in the region analogous to the SR beside the inverted requirement on the same sign of electric charge for both tau-leptons.

and their contribution is normalised to the theory-predicted cross section.

The statistical analysis is based on the collection of fit variable S_{T} histogram inputs. The fit employs a single parameter of interest: the ETL signal strength. The signal region and the $Z \to \tau\tau$ control region are built with $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ signatures, while the top-quark background CR is defined in the $e\mu$ final state. The Z and Top control regions are used to obtain freely floating NF_Z and NF_{Top} normalisation parameters. The fit includes nuisance parameters standing for per-bin statistical uncertainty (γ parameters) and the uncertainty on luminosity.

The expected cross section of the excited tau-lepton production is estimated with the CL_s method. The expected significance of the excited tau-lepton signal is defined.

The reported results are based on the fit of the model to the Asimov data set. In the Asimov data, real data are replaced by the predicted SM background (and model-predicted BSM signal if relevant). Another result is obtained in the simultaneous fit to the Asimov data in the SR and to real data in $Z \to \tau\tau$ CR and Top CR. The $Z \to \tau\tau$ ($t\bar{t}$ and Wt) SM prediction is fitted to real data in the $Z \to \tau\tau$ (Top) control region, respectively. This allows analysers to obtain the nuisance parameters: normalisation factors (NF_Z and NF_{Top}) and γ parameters for the Drell Yan and Top backgrounds. In this way, the $Z \to \tau\tau$ and Top SM contributions are constrained by data in the SR. Finally, fit in the SR is made to the defined Asimov data set (the Monte Carlo signal prediction and MC background estimation, with Top and $Z \to \tau\tau$, constrained and normalised in the respective CRs).

10.5 Event selection

There are general requirements applied in physics analysis regardless of its specific subject (Section 7.2). Each ATLAS subsystem should operate smoothly. The recorded data must be of high quality and include physical signatures (at least one primary vertex in the event). Table 7.1 lists these criteria.

Further, events must be accepted by the ATLAS unprescaled di-tau trigger, and the final reconstructed tau-object should match the related trigger-level objects. The choice of the trigger requirements follows the composition of the final state: two τ_{had} and at least two jets. Appendix A gives the specification of the used triggers.

Object selections before the overlap removal are listed in Table 10.2. The variables were described earlier in Section 5.[15] The overlap removal proceeds in the standard setup (as illustrated in Section 5.6 earlier). These steps are implemented in the AAC framework during the NTuple production.

Once the event is fully reconstructed and the overlaps are removed, requirements on the jet and τ_{had} multiplicity are imposed. Each event needs to include exactly two τ_{had} and at least two jets. Jet cleaning is performed on an event-by-event basis. The absence of reconstructed electrons and muons is required.

Further selections (Table 10.3) are designed to enhance the presence of the excited tau-lepton signal. Several cuts are introduced to reduce the SM back-

[15] Perhaps, it is worth repeating here explicitly that the requirement on the vertex variables z_0 and d_0 reject events where light leptons originate from leptonic tau-lepton decay.

	jet	τ	e	μ
p_{T} [GeV]	20	20	15	7
polar angle	$\lvert y\rvert < 4.5$	$\lvert\eta\rvert \in (0, 1.37)$ $\bigcup$ $\lvert\eta\rvert \in (1.52, 2.5)$	$\lvert\eta_{\mathrm{BE2}}\rvert \in (0, 1.37)$ $\bigcup$ $\lvert\eta_{\mathrm{BE2}}\rvert \in (1.52, 2.47)$	$\lvert\eta\rvert < 2.5$
ID	-	RNN > 0.01	LHLoose	HighPt
N_{trk}	-	1 or 3	-	-
$\lvert q\rvert$	-	1	-	-
σ_{d0}	-	-	< 5	< 5
$\lvert z_0 \sin\theta\rvert$	-	-	< 0.5	< 0.5
isolation	-	-	PLVTight	PLVTight

Table 10.2: The requirements on reconstructed jets, visible signatures of hadronically decaying tau-leptons, electrons, and muons in the search for excited tau-leptons.

ground.

10.5.1 Selection of tau-leptons

Both final state hadronically decaying tau-leptons are required to be geometrically matched to τ_{had} candidates as defined by the di-tau trigger (see Appendix A for details).

For tau-leptons, the minimal offline p_{T} should be 5 GeV above the trigger threshold: 40 GeV and 30 GeV for leading and subleading tau-lepton, respectively.

The jet rejection BDT (estimated with a **b**oosted **d**ecision **t**ree technique) score should correspond to `Medium` and `Loose` working points (WPs), respectively.

The `Medium` ID WP for leading tau-lepton is a compromise between signal efficiency and fake background rejection. This choice provides a substantial background rejection by a factor of $\sim 10 - 90$ and provides $\sim 75\%$ ($\sim 60\%$) efficiency in real 1-prong (3-prong) τ_{had} identification [94].[16] The `Loose` WP provides 85% (75%) efficiency in the identification of 1-prong (3-prong) tau-leptons. The choice of the ID WP for subleading tau-lepton is described in the following paragraph.

Two tau-leptons candidates should be of an opposite charge as is expected in events with excited tau-lepton production.

In Monte Carlo simulation, the reconstructed τ_{had} needs to be geometrically matched to generator-level τ_{had}, e, or μ.[17] Matching within the cone of $\Delta R < 0.4$ size is required. This allows for the inclusion of events with real tau-leptons and for estimation of the $l \to \tau$ fakes modelled with MC. To remind the reader, the $j \to \tau$ fakes are estimated with the fake factor method.

[16] In general, 3-prong fakes are rejected more efficiently: background rejection is about 4-5 times higher in comparison to the identification of 1-prong τ_{had}. The RNN-based techniques significantly outperform BDT-based ones from the point of view of the rejection of misidentified τ_{had} in dijet events (see more in [94]).

[17] The "truth" (generator-level) tau-leptons, electrons, and muons are stored in the corresponding `TRUTH` containers: TruthTaus, AntiKt4TruthDressedWZJets, TruthElectrons, TruthMuons.

selection	requirement
N_τ	2
N_j	≥ 2
$N_{b-\text{jet}}$	-
N_e	0
N_μ	0
RNN ID τ_0	Medium
RNN ID τ_1	Loose
$p_{\text{T}}^{\tau 0}$ [GeV]	> 40
p_{T}^{j1} [GeV]	> 30
p_{T}^{j0} [GeV]	> 70
p_{T}^{j1} [GeV]	> 60
$\lvert\eta^{j0}\rvert$	< 2.4
$\lvert\eta^{j1}\rvert$	< 2.4
$q_{\tau 0} \times q_{\tau 1}$	-1
$\Delta R(\tau_0, \tau_1)$	> 0.8
x_0^{coll}	> 0.1
x_1^{coll}	> 0.05
τ_{had} truth matching	yes

Table 10.3: The definition of the signal region and $Z \to \tau\tau$ control region in the search for excited tau-leptons.

Identification working point for subleading tau-lepton. In the ATLAS analyses performed with the `Athena` r. 21, several working points for tau-lepton identification are defined based on the RNN ID score. The supported WPs are quantified with the ID scores estimated with BDT or RNN as mentioned in Section 5.5.

In this analysis, two options for the identification working point for subleading tau-lepton were studied `Loose` ID WP and `Medium` ID WP. The *Loose* ID provides higher identification efficiency for real τ_{had}. On the other hand, the `Medium` ID was designed to provide higher rejection of non-τ_{had} background.[18] Studies of the S_{T} spectra and signal and background yields were done to choose an optimal WP for the subleading tau ID. Figure 10.8 demonstrates how switching from Loose to Medium subleading tau-lepton ID WP affects the signal and background contributions.

Figure 10.8a depicts S_{T} spectra for the total Standard Model background obtained with two considered τ ID WPs. Tightening the selection from `Loose` to `Medium` ID decreases background contribution by ∼20% on average. The Fake background was found to be the most sensitive to the choice of the WP: its contribution is suppressed by about 60% if the `Medium` ID is applied instead of the `Loose` one (Figure 10.8b).

The tau-lepton ID WP also impacts the signal selection. Figure 10.8c shows the effect on the 1 TeV excited tau-lepton signal. The signal yield decreases by ∼10% throughout the phase space. Hence, the tighter selection diminishes both

[18] A yet more stringent, `Tight`, tau ID WP was not considered as it provides lower identification efficiency of 60% (45%) for 1-prong (3-prong) tau-leptons [206].

the background and the signal yields.

After all, the $(S+B)/B$ ratio can quantify sensitivity to the signal. As shown in Figure 10.8d, tightening the criterion to `Medium` does not bring significant benefit in terms of $(S+B)/B$. The compared WP options show compatible signal sensitivity. Given the higher τ_{had} acceptance, the *s*election.

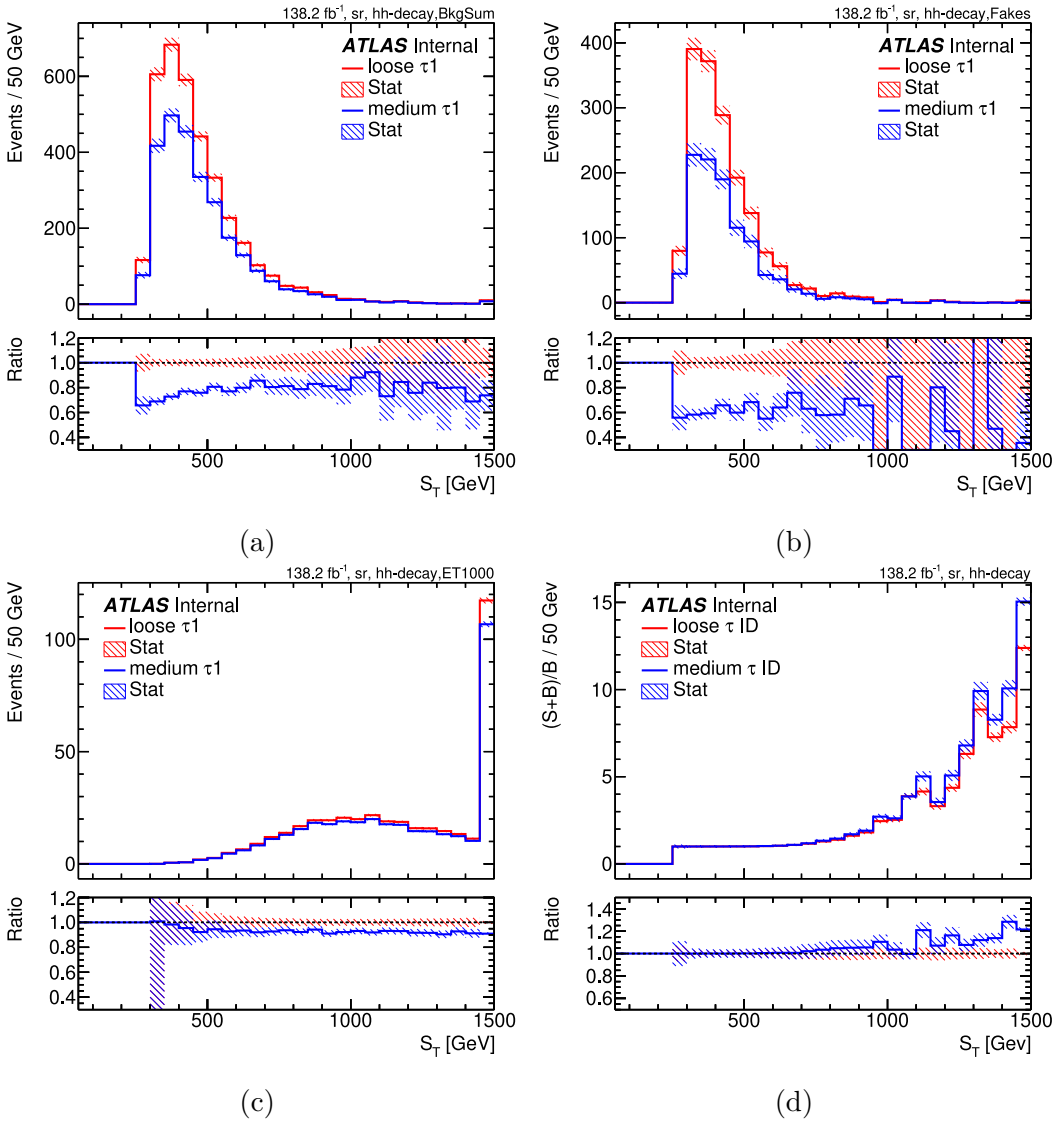


Figure 10.8: The comparison of the S_{T} distributions obtained with the `Loose` and `Medium` tau-lepton ID working points: the total SM background (a), fake background (b), excited tau-lepton of 1 TeV mass (c) and the $(S+B)/B$ ratio for $m_{\tau*} = 1$ TeV excited tau-lepton signal (d).

10.5.2 Selection of jets

The leading in p_{T} jet is required to have $p_{\text{T}} > 70$ GeV. The p_{T} requirement harmonised the selection procedure between 2015 and 2016-18 data-taking conditions when the di-tau and di-tau+jet triggers were employed, respectively. Subleading jet needs to pass the requirement $p_{\text{T}} > 60$ GeV. This selection removes a significant portion of pile-up-induced jets, especially in the forward regions of the

ATLAS detector. Two leading jets are required to lay in the rapidity range of $|y| < 2.4$ which corresponds to the Inner Detector coverage.

10.5.3 Selection on di-tau opening

Reconstructed tau-leptons are matched to the trigger-level objects as specified in Appendix A. The di-tau trigger accepts events with the requirement of the angular opening between two tau-leptons: $\Delta R_{\tau\tau} > 0.8$.[19] Following this trigger-level requirement, the analysis introduces the $\Delta R_{\tau\tau} > 0.8$ selection.

The impact of the $\Delta R_{\tau\tau} > 0.8$ criteria on the signal yields was studied (Figure 10.9). The selection influence the low-mass excited tau-lepton yields within 5% as shown for ETL of 200 GeV and 700 GeV in Figure 10.9a and 10.9b, respectively. The contribution from boosted di-tau increases with higher mass excited tau. However, the selection of the di-tau opening angle leads to a moderate ETL signal loss within $\sim 5\%$ for masses of 1 TeV and 2.5 TeV (Figure 10.9c and 10.9d, respectively).

10.5.4 Selection on collinear fractions

Section 8.2.4) described the collinear approximation for the di-tau mass reconstruction. In this search, collinear di-tau mass $m_{\tau\tau}^{\text{coll}}$ is used to define background- and signal-enriched phase space. In particular, it provides a way to distinguish the phase space highly populated by Z boson production.

Figure 10.10 shows the distributions for collinear fractions in so far selected events. In the x_0^{coll} and x_1^{coll} range between 0.4 and 1.2 signal dominated background contribution (normalized plots shown in Appendix I). To increase sensitivity to the excited tau-lepton signal and provide a robust estimation of the di-tau mass, events with $x_0^{\text{coll}} > 0.1$ and $x_1^{\text{coll}} > 0.05$ are selected.

For low-mass excited tau-leptons (300 GeV and below), the selection results in a signal yield drop by about 10% (Figure 10.11a). However, $\sim$95% of the signal in sensitive region ($S_T > 500$ GeV) pass this selection for the tested mass of 700 GeV and 1000 GeV (Figures 10.11b, 10.11c).

Moreover, this selection is helpful for background suppression, as seen in Figure 10.12b. Specifically, this selection vetoes a fraction of background events with no real transverse mass where, for example, light leptons mimic τ_{had} while miscalibration of jets results in sizeable MET. Thus, the $x_0^{\text{coll}} > 0.1$ and $x_1^{\text{coll}} > 0.05$ selections notably prevent the high-S_{T} spectra from population by $Z \to ll$ events (Figure 10.12a).

10.6 Signal region and control region design

Event selection with a set of variables separates a given process from others. In searches, a signal region is designed in such a way that it is predominantly populated with signal events. The SR is the space where an actual search is performed. In the presented search, the SR cut flow defines a subset of events

[19] In general, reconstruction of highly boosted di-tau requires more elaborated methodology [207] which is out of scope of this thesis.

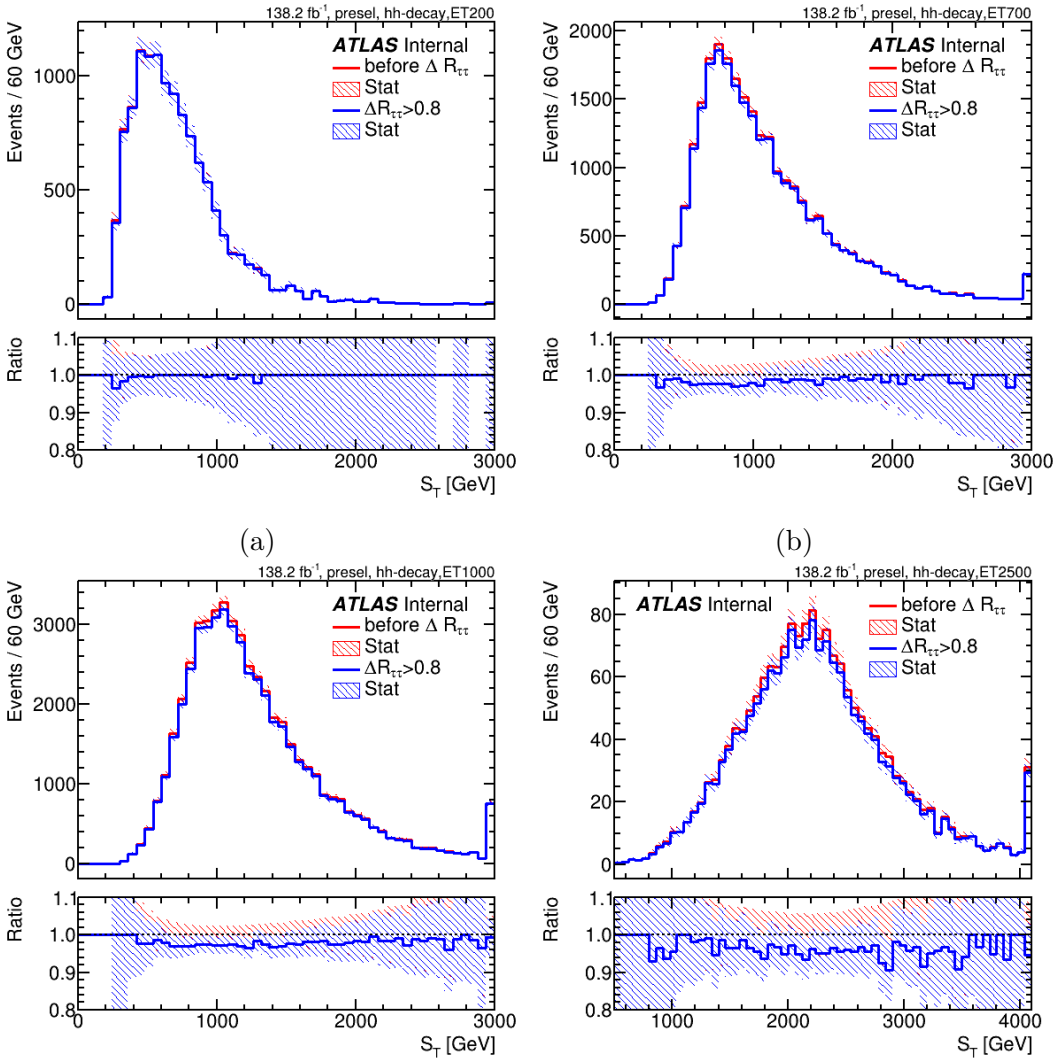


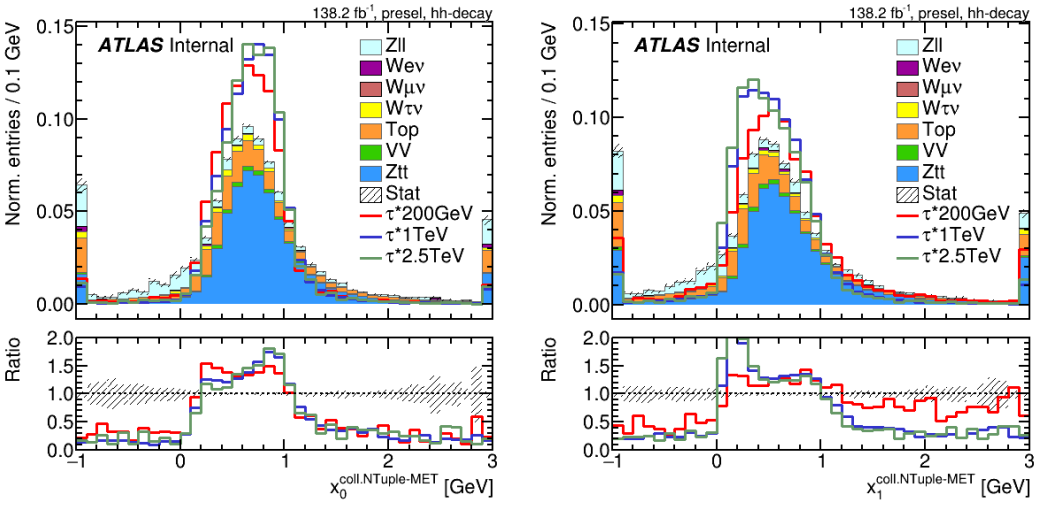
Figure 10.9: The impact of selection of the di-tau opening $\Delta R_{\tau\tau} > 0.8$ on the S_T distribution for excited tau-lepton signal of 200 GeV (a), 700 GeV (b), 1 TeV (c), and 2.5 TeV (d).

enriched with excited tau-lepton production. The final selection includes cuts on the di-tau collinear mass and the L_{T}, which is a scalar sum of both tau-lepton p_{T}.

To ensure that the Standard Model background is well understood, one needs to determine the phase space enriched with a particular background process. Typically several SM processes contribute to the SR, and their prediction is estimated and validated in corresponding control and validation regions.

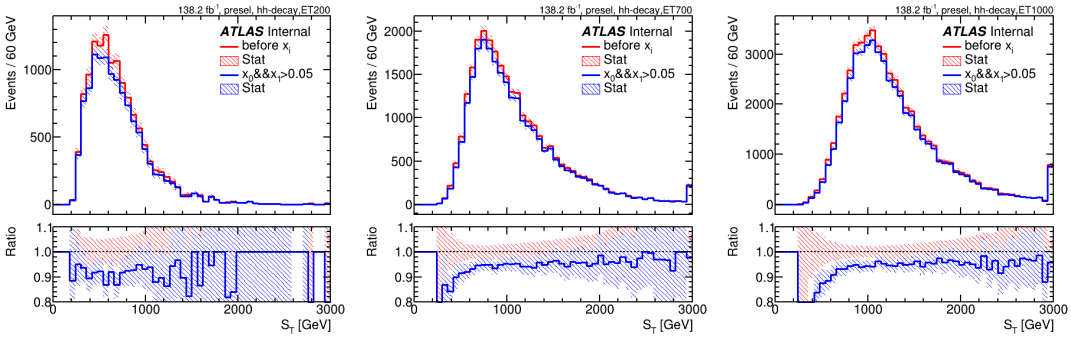
In this search, the respective signal and control regions have a cut-based definition with the observables characterising the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}jj$ system. As seen from the composition of the processes falling into the SR, the main SM background sources are $Z \to \tau\tau$ processes, top-quark production, and events with fake τ_{had}.

The selection on the di-tau collinear mass $m_{\tau\tau}^{\mathrm{coll}}$ separates the SR and the $Z \to \tau\tau$ CR. The following section discusses this choice. The constrained by data $Z \to \tau\tau$ prediction is then employed in the final scan for the excited tau-



(a) (b)

Figure 10.10: The distribution for collinear fractions for leading and subleading tau-leptons x_0 (a) and x_1 (b) in background and signal events. The result is obtained with interim NTuples.



(a) (b) (c)

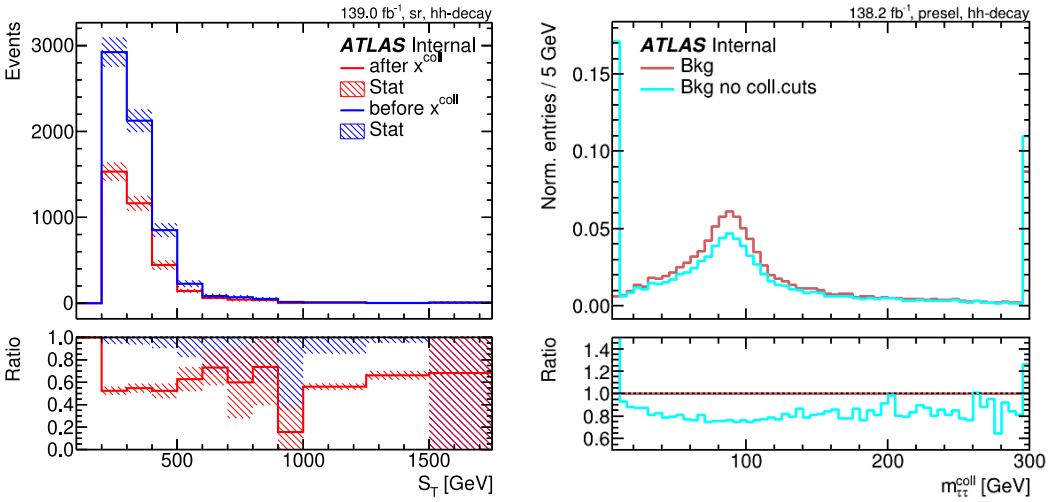
Figure 10.11: The comparison of S_{T} spectra in excited tau-lepton events before and after selection on the collinear fractions are applied. The ETL of mass 200 GeV (a), 700 GeV (b), and 1 TeV (c) is demonstrated.

leptons in the SR.

In this search, the contributions from $t\bar{t}$ and single-top production are constrained with real data in the Top CR. The Top CR is defined in a statistically independent final state with light leptons. Fully leptonic decay mode – $e\mu$ – is used to constrain and normalise the top-quark contribution.

The $j \to \tau_{\mathrm{had}}$ fake contribution was estimated with the data-driven Fake Factor method. The Fake Factors were measured in the phase space orthogonal to the SR. The fake-enriched phase space was populated with same-sign ($q_\tau \times q_\tau = 1$) events failing to pass at least one from definitive SR criteria on the di-tau observable ($m_{\tau\tau}^{\mathrm{coll}} > 110$ GeV, $L_{\mathrm{T}} > 140$ GeV).

The division of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ phase space into the SR and $Z \to \tau\tau$ CR is outlined in Figure 10.13.



(a) (b)

Figure 10.12: The comparison of S_{T} spectra in $Z \to ll$ (a) and all SM background (b) events before and after the selections on the collinear fractions are applied.

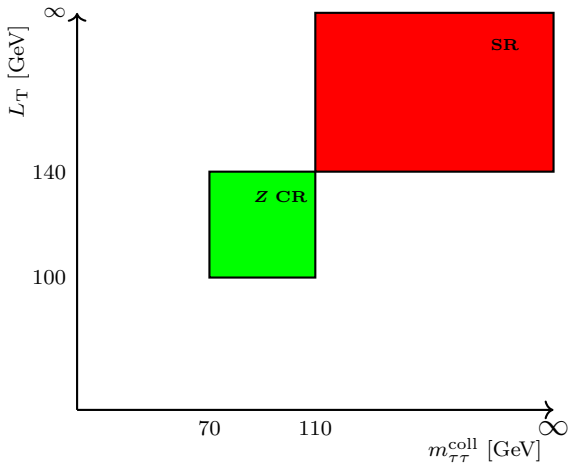


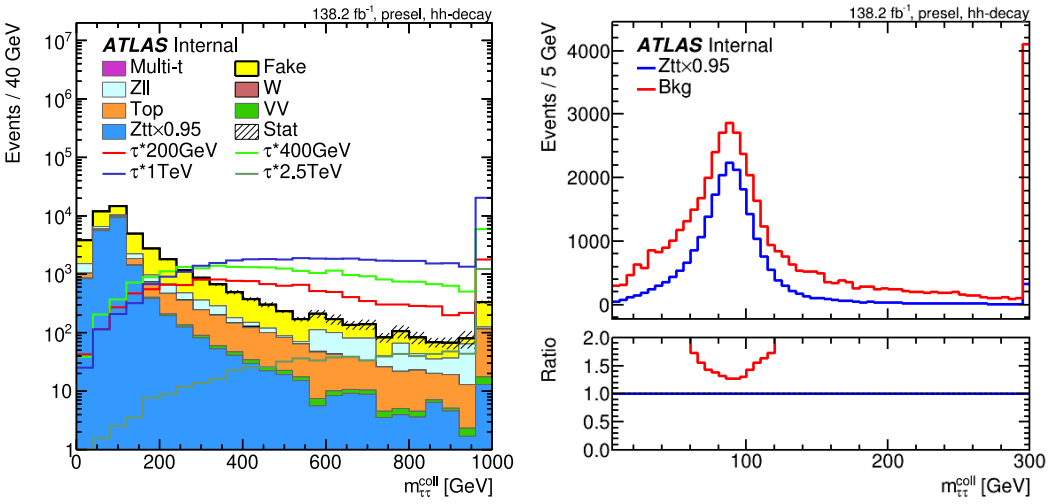
Figure 10.13: The orthogonality of the SR and the $Z \to \tau\tau$ CR selection on $m_{\tau\tau}^{\mathrm{coll}}$ and L_{T}.

10.6.1 Selection on di-tau collinear mass

Di-tau collinear mass is a good candidate for distinguishing regions with ETL signal prevalence and $Z \to \tau\tau$ background contributor. Figure 10.14 shows the di-tau collinear mass distribution in preselected SM and ETL events before and after the $m_{\tau\tau}^{\mathrm{coll}}$ selection is applied.

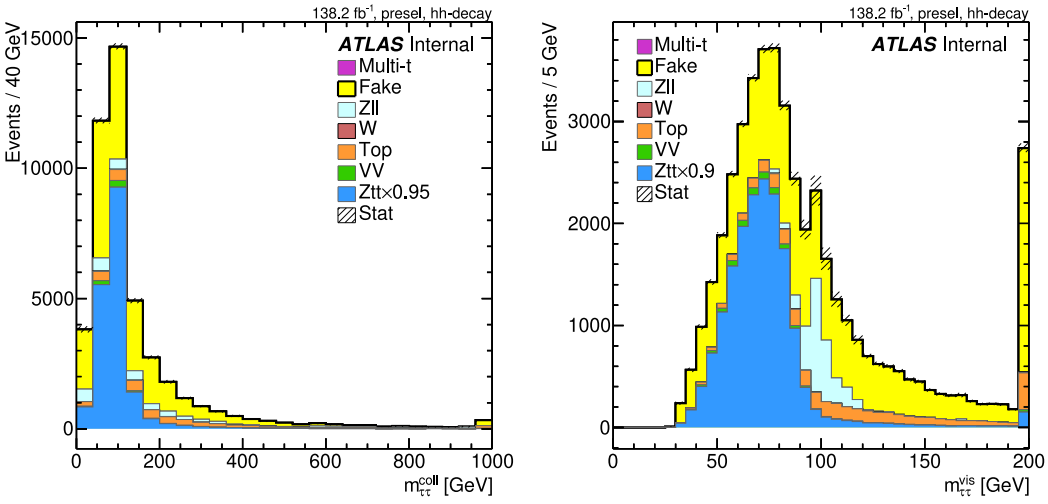
An alternative definition with visible di-tau mass 50 GeV $< m_{\tau\tau}^{\mathrm{vis}} <$ 100 GeV was considered. However, the visible $m_{\tau\tau}^{\mathrm{vis}}$ mass has a lower resolution (as seen in Figure 10.15) and, consequently, lower power of process-to-process separation.

The Z boson mass reconstruction resolution is about 20 GeV. The range of di-tau collinear mass between 70 and 110 GeV is enriched in Drell Yan process. The $Z \to \tau\tau$ contribution achieves $\sim 90\%$ from the total predicted background in this region (Figure 10.14b). Therefore, the phase space where $m_{\tau\tau}^{\mathrm{vis}}$ falls in the $(70\ \mathrm{GeV}, 110\ \mathrm{GeV})$ range is used as the $Z \to \tau\tau$ control region. To exclude



(a) (b)

Figure 10.14: The distribution of collinear mass (a) in preselected SM and ETL events (a). The spectrum and the yield of $Z \to \tau\tau$ process in relation to the total SM background (b).



(a) (b)

Figure 10.15: The distribution of collinear (a) and visible (b) mass in preselected events. The result obtained with interim NTuples is shown.

this large fraction of the Drell Yan background, the orthogonal selection $m_{\tau\tau}^{\text{coll}} > 110$ GeV defines the signal region.

10.6.2 Variable for fit

The choice of the variable for the final fit is usually made to maximise sensitivity to a BSM signal in searches for New Physics. It is important that the data are distributed in the observable, reasonably separating BSM signatures and the expected SM background. In the search for excited tau-leptons, this optimal observable ought to provide good discrimination between ETL signals driven by different hypothesized masses. The fit variable was chosen with the ROC curve

method. The ROC curve method gives a cross-section independent measure of discriminating power which is not relying on any assumption on the BSM particle production cross section.[20]

The list of the checked observables includes effective di-tau mass, transverse masses of di-tau and di-tau+MET systems, and invariant mass of the whole quadruplet (two tau-leptons and two jets). The scalar sums: of the leading and subleading jets transverse momenta $H_{\mathrm{T}} = p_{\mathrm{T}}^{j0} + p_{\mathrm{T}}^{j0}$, of the leading and subleading tau-leptons transverse momenta $L_{\mathrm{T}} = p_{\mathrm{T}}^{\tau 0} + p_{\mathrm{T}}^{\tau 1}$, and the leading and subleading jets and tau-leptons transverse momenta $S_{\mathrm{T}} = p_{\mathrm{T}}^{j0} + p_{\mathrm{T}}^{j1} + p_{\mathrm{T}}^{\tau 0} + p_{\mathrm{T}}^{\tau 1}$ were probed as well.

Figure 10.16 demonstrates the performance of tested discriminators. The S_{T}, $m_{\mathrm{T}}^{\mathrm{total}}$, m^{eff} values variables demonstrated good potential in separation the low-mass (200 GeV) ETL signal from the SM background. Further, the S_{T} variable outperformed others in classification of the excited tau-lepton signal of different masses (Figures 10.16c, 10.16d, 10.16f). A comparison of the extensive sets of the tested variables is given in Appendix J (Figures J.1-J.2). Therefore, data for statistical analysis are binned in the S_{T} variable.

10.6.3 Selection on L_{T}, a scalar sum of tau-leptons p_{T}

One can further reduce the SM background in the SR by applying another event selection. Applicability of selections on several energetic and angular variables, as well as multi-body invariant masses, were studied (Figure 10.17).

The selection on the L_{T} variable, a scalar sum of the transverse momenta of leading and subleading tau-leptons, was found to suppress the SM contribution while preserving the ETL events. In this sense, the selection on L_{T} increases sensitivity for excited tau-lepton signal. Figure 10.17d shows the L_{T} spectrum for the SM background and signal. The region $L_{\mathrm{T}} < 160$ GeV is populated by background processes ($Z \to ll$ and $j \to \tau$ fakes in particular).

The choice of a specific cut on L_{T} in the range {50,160} GeV was based on a comparison of the ROC curves in scan over the final binning variable S_{T}: for signal-to-background and signal-to-signal classification. The performance was estimated with several figures of merit: the AUC and the integral signal significance defined by Eq. (7.2). There is also cited the level of the background rejection at the point where the 95% of the signal is accepted. Figure 10.18 shows the performance of the selections on L_{T}. For comparison, performance with the selections on alternative variables (effective mass defined with Eq. (8.2) and total transverse mass defined with Eq. (8.4)) are given in Figures K.1 and K.2 in Appendix K. The criteria on L_{T} outperform the other testes options.

All the tested choices of the L_{T} cut were found equally effective in signal-from-signal separation as illustrated in Figures 10.20a and 10.20b. Therefore, the definition is based on the power of background rejection. It is worth mentioning that signal-to-signal discrimination is also robust when criteria on m_{eff} or $m_{\mathrm{T}}^{\mathrm{total}}$ are applied (see Figure K.2 in Appendix K. This emphasizes that the L_{T} (m_{eff}, or $m_{\mathrm{T}}^{\mathrm{total}}$) selection enhances the signal from excited tau-lepton regardless of its

[20] Such a model-independent approach is especially relevant for this search as the results can be interpreted in terms of other BSM models (like leptoquarks) and/or can be interpreted to other compositeness scale values.

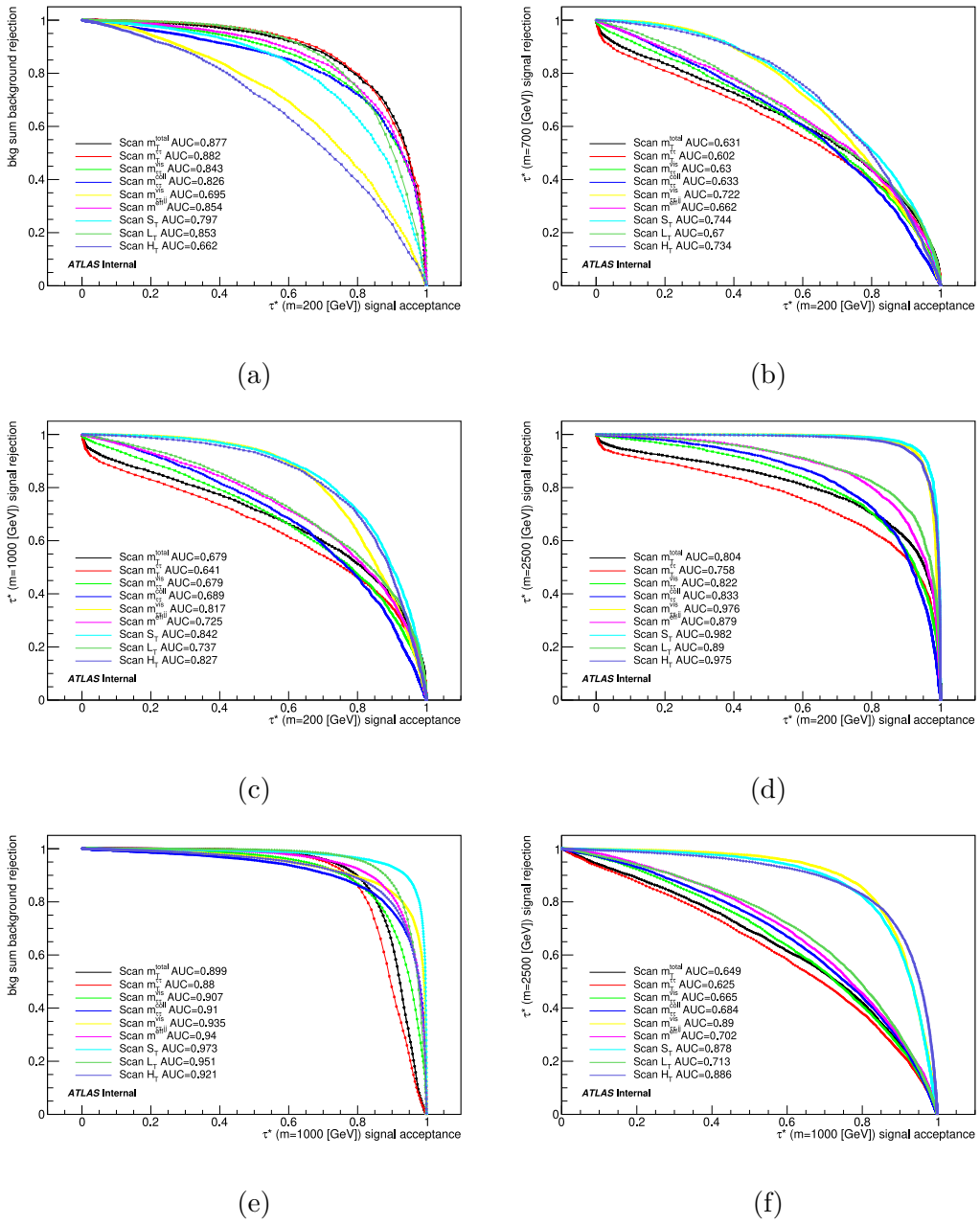


Figure 10.16: The ROC curves showing acceptance of excited tau-lepton m = 200 GeV signal versus rejection of SM background (a), ETL m = 700 GeV (b), ETL m = 1 TeV (c), and ETL m = 2.5 TeV (d). The ETL $m_{\tau*}$ = 1 TeV signal separation from the total SM background prediction (d). The ROC curve for separation between two excited tau-leptons signals with different hypothesised ETL masses, 1 TeV and 2.5 TeV (e).

actual mass. Consequently, this approach can be applied in the search for ETL in a wide mass range.

While the background rejection is the most effective for the tightest cut, signal significance decreases with a more stringent selection. As a trade-off between keeping signal events and removing background ones, the range of the search for the eventual cut was narrowed down to 110 GeV $< L_{\mathrm{T}} <$ 150 GeV.

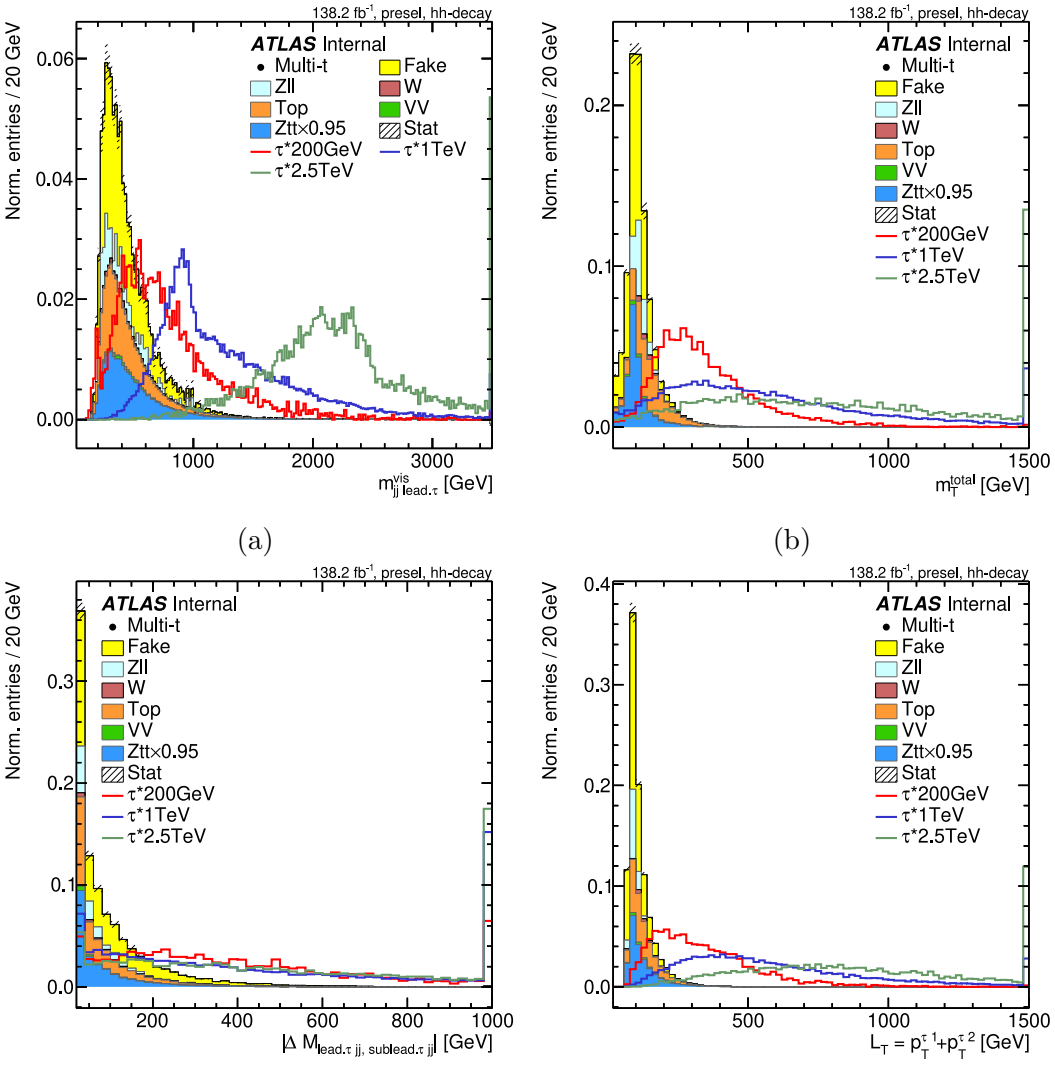


Figure 10.17: The $m_{(\text{lead}\tau)jj}$ (a), m_T^{total} (b), $\Delta p_T(\text{lead}\tau)jj, (\text{sublead}\tau)jj)$ (c), and L_T (d) distributions in SM background and excited -tau-lepton processes.

Furthermore, the separation between individual SM contributors and the ETL signal was studied. Figures 10.19 show the values of AUC, the integral ETL signal significance Z, and the level of $Z \to ll$, $Z \to \tau\tau$, and Fakes. The L_{T} selection around 140 GeV results in a higher signal significance estimate, which makes it a candidate for the final L_{T} criteria. This selection reduces $Z \to ll$ background while keeping the low-mass signal (Figure 10.19a). It also provides a good balance between preserving higher-mass signal and reducing the background processes contributing in the sensitive S_{T} bin, such as $Z \to \tau\tau$, Fakes (Figure 10.19d, 10.19e, 10.19f).

Therefore, the final event selection for the SR and CRs includes the $L_{\text{T}} >$ 140 GeV requirement. Figure K.3 in Appendix K illustrates the ROC curves for the fully defined SR. SM background rejection of 25% and signal acceptance of 95% is observed for the very low-mass ETL (200 GeV) as outlined in Figure K.3 (Appendix K). For heavier ETL, background rejection is more efficient. For instance, as seen in the same Figure K.3 (Appendix K), 88% of SM events are

rejected at the level of 95% ETL signal acceptance. The constructed event selection guarantees good signal-to-signal discrimination with the AUCs above 80% (see Figure K.3 (Appendix K)).

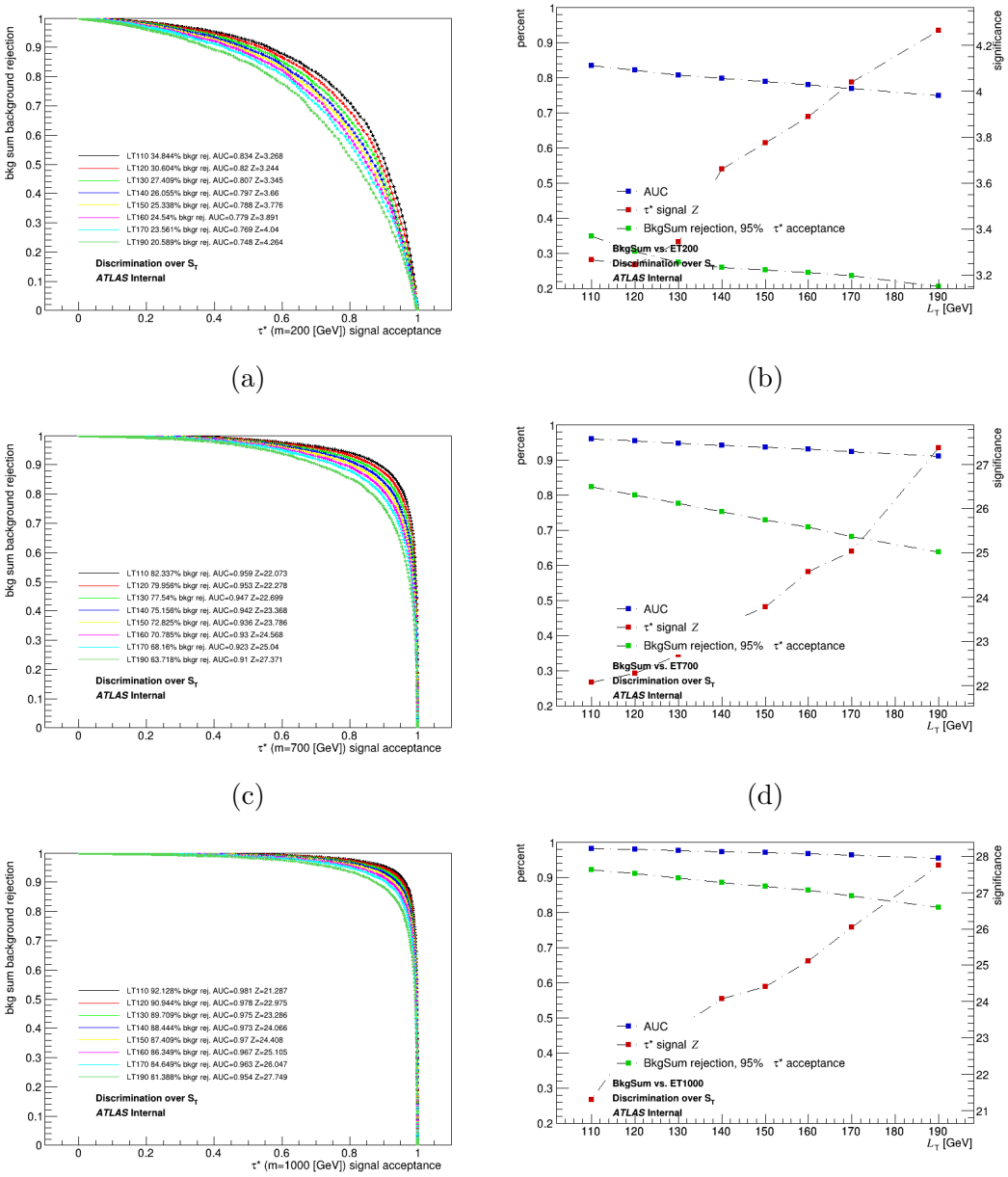


Figure 10.18: The discrimination between excited tau-lepton signal and the total SM background obtained in the scan over the S_{T} variable in preselected events with additional criteria imposed on the L_{T} variable. The values of the L_{T} selection are varied, and the corresponding ROC curves are depicted with different colours: excited tau-lepton signal ($m_{\tau*} = 200$ GeV) and the SM background (a), ETL $m = 200$ GeV and ETL $m = 1$ TeV (c), ETL $m = 1$ TeV and ETL $m = 2.5$ TeV (e). Plots (b), (d) and (f) show the integral significance, AUC, and the level of the acceptance of events X versus the rejection of events Y). Scanning over the S_{T} spectrum is performed.

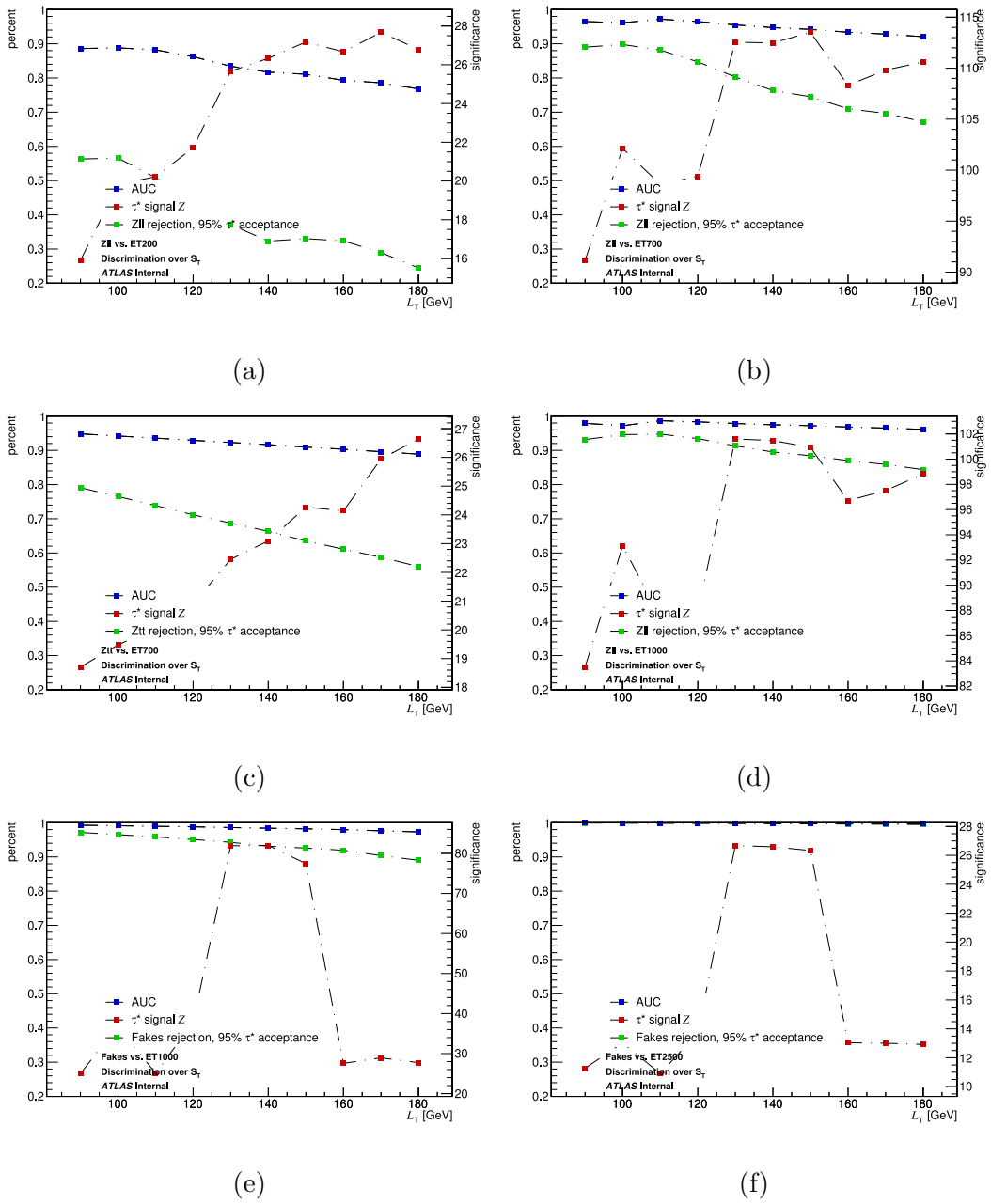
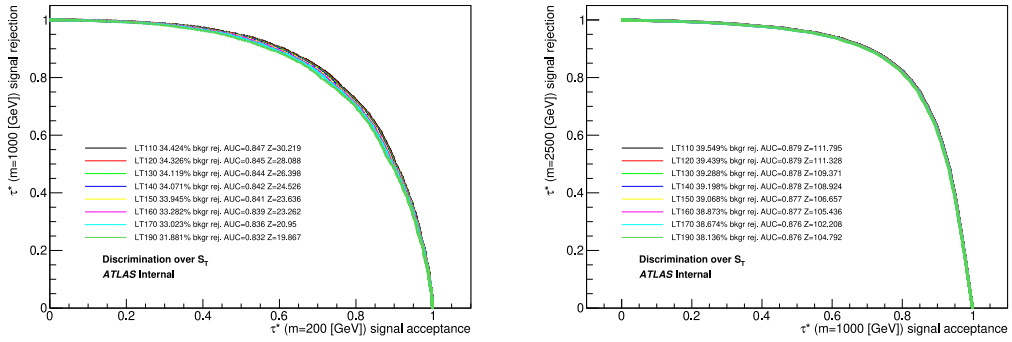


Figure 10.19: The discrimination between excited tau-lepton signal and individual SM background processes obtained in the scan over the S_{T} variable in preselected events with additional criteria imposed on the L_{T} variable. The values of the L_{T} selection are varied, and the corresponding ROC curves are depicted with different colours: excited tau-lepton signal ($m_{\tau*} = 200$ GeV) and $Z \to ll$ (a), ETL $m = 700$ GeV and $Z \to ll$ (b), ETL $m = 700$ GeV and $Z \to \tau\tau$ (c), ETL $m = 1$ TeV and $Z \to ll$ (d), ETL $m = 1$ TeV and Fakes (e), and ETL $m = 2$ TeV and Fakes (f). Plots show the integral significance, the AUC, and the level of the rejection of SM events at the 95% ETL signal acceptance.

10.7 Signal region

The selection described in the Sections above is intended to define an optimal signal region. The eventual signal regions represent the phase space with the



(a) (b)

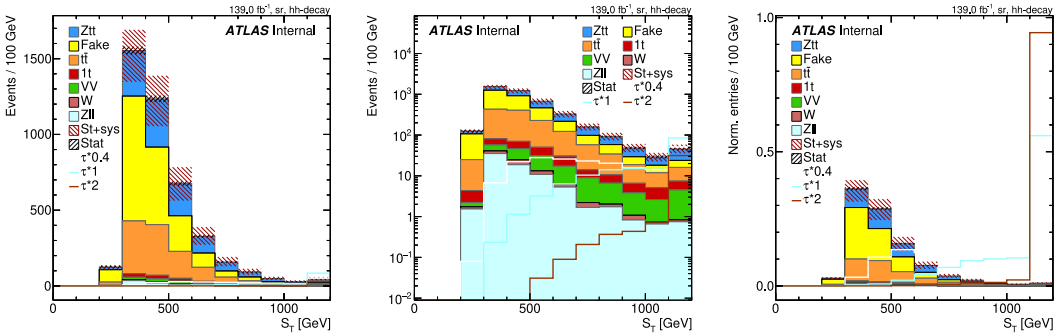
Figure 10.20: The discrimination between excited tau-lepton signals of 200 GeV and 1 TeV (a) as well as ETL signals of 1 TeV and 2.5 TeV (b) mass. The scan over the S_{T} variable is performed in preselected events with additional criteria imposed on the L_{T} variable.

enhanced presence of the excited tau-lepton signal.

The excited tau-lepton search employs a single signal region binned in the S_{T} variable. The signal region is defined by applying a series of cut-based selections discussed above and summarized in Table 10.3.

The SR uses equidistant binning with the last bin, including events with $S_{\mathrm{T}} > 1$ TeV. Such a binning scheme ensures that the SM estimate includes enough predicted events in the $S_{\mathrm{T}} > 1$ TeV bin sensitive to the ETL signal. This allows validating the $Z \to \tau\tau$, top-quark, and fake background contribution with real data.

Figure 10.21 illustrates the distribution of the S_{T} variable in the constructed signal region.



(a) (b) (c)

Figure 10.21: The S_{T} distribution in the signal region: in a linear scale (a), in a logarithmic scale (b), and normalized to a unit (c).

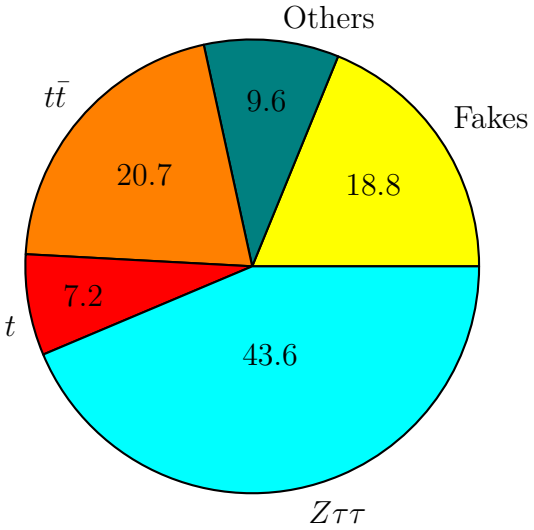
10.7.1 Background contribution

Figure 10.22a and Table 10.22b show the composition of event yields in the $S_{\mathrm{T}} > 1$ TeV phase space which is the most sensitive for ETL signal. The ratio $s/\sqrt{b}$ is

used to quantify the level of sensitivity to the ETL signal (here, s and b denoted the total background and signal event yields, respectively).

Table 10.4 shows a breakdown of the yields in the SR for three ETL mass points and SM background prediction. The absolute ratios of the yields of signal to background events are quoted.

The lowest mass point in this search is 200 GeV, as BSM effects are excluded in the low-energy range by numerous studies in the high-energy physics community. In addition, the S_{T} range for the very low-mass ETL is largely populated by the SM processes, which leave no room to define appropriate control and validation regions for the SM background.



(a)

SM Process	Abs. yield	Rel. yield
$Z\tau\tau$	32.5	43.6%
$t\bar{t}$	15.4	20.7%
Fakes	14	18.8%
VV	5.6	7.5%
single top	5.4	7.2%
Zll	1.4	1.9%
W + jets	0.16	0.2%
Total	74.4	
BSM process	Abs. yield	$s/\sqrt{b}$
τ^* 400 GeV	65.2	7.6
τ^* 1 TeV	100.2	11.6
τ^* 2 GeV	32.1	3.73

(b)

Figure 10.22: The contributions of each SM background process related to the total background yield in the signal region; the "Others" category combines W + jets, $Z \to ll$ and di-boson background (a). The breakdown of the absolute and relative contribution of SM background in high-S_{T} range ($S_{\mathrm{T}} > 1$ TeV); yields of excited tau-lepton processes and the ratios $s\sqrt{b}$ are quoted (b).

10.7.2 $(N-1)$ plots

After the event selection cut flow for the signal region was fully defined, further studies estimated the impact of individual selections on p_{T} of tau-leptons and jets.

So-called $(N-1)$ distributions check how a given selection affects the signal events yields and kinematics. The name of these distributions comes from the fact that they are obtained when applying all other (N) cuts beside one (1) of the current interest, i.e. $N-1$ selections are applied overall.

Figure 10.23 shows the p_{T} distributions for each object of the quadruplet (four final-state objects: leading and subleading tau-lepton, leading and subleading jet) when all selections are applied except for the selection on the given objects p_{T}. As seen in Figure 10.23a, the $p_{\mathrm{T}}^{j0} > 70$ GeV criterion keeps yields non-affected for heavy excited tau-leptons as it is demonstrated for 2.5 TeV, 5 TeV and 8 TeV mass points. The p_{T} criteria for leading jet removes about

process	$Z \to \tau\tau$ CR	Top CR	SS CR	SR
Fakes	501.9	-	1,262	1,834
$Z \to \tau\tau$	2,709	36	17.7	1,108
$t\bar{t}$	106.4	89,385	48.7	1,067
single top	7.5	2,713	5.1	98.9
$Z \to ll$	7.5	-	11	77.1
W + jets	2.5	-	10	16.5
VV	88.7	74	18	86.9
SM total	3,579	92,208	1,372	4,288
purity	76%	99%	92%	-
Data	3,741	88,293	1,368	-
Data/Bkgr	1.05	0.96	0.99	-
τ*400 GeV	1.9	0.033	6.4	207.6
τ*400 GeV S/B	0.053%	$< 0‰$	0.47%	4.8%
τ*1 TeV	0.47	1.778	7.5	150.8
τ*1 TeV S/B	0.013%	0.002%	0.55%	3.5%
τ*2 TeV	0.04	0.45	2.5	33.3
τ*2 TeV S/B	$< 0‰$	$< 0‰$	0.18%	0.77%

Table 10.4: The yields of SM and ETL events in the $Z \to \tau\tau$ CR, the Top CR, the same-sign CR, and the SR. The purity of the CRs, yield ratio between data and MC, and signal-to-background ratio are quoted.

$\sim 6.5\%$ of events for the lowest tested ETL mass of 100 GeV. Other p_{T} selection showed no significant impact on the selection efficiency for high ETL mass points. Figures 10.23b, 10.23c, 10.23d illustrate the $(N-1)$ S_{T} distributions for all but p_{T} selection on the subleading jet, leading tau-lepton, and subleading tau-lepton, respectively. The impact of other kinematic selections, the requirements on $m_{\tau\tau}^{\mathrm{coll}}$ and L_{T}, is demonstrated in Figure 10.24. The contribution of events with $m_{\tau\tau}^{\mathrm{coll}} < 110$ GeV ($L_{\mathrm{T}} < 140$ GeV) is within 5% (10%) for the smallest tested ETL mass of 200 GeV (Figure 10.24a and 10.24b). The $m_{\tau\tau}^{\mathrm{coll}}$ and L_{T} requirements do not effectively impact the signal selection for heavy excited tau-leptons (as seen for masses from ~ 2.5 GeV onwards in Table 10.6).

The effect of individual selections can also be seen from the event yields summarized in Table 10.6.

10.7.3 Signal selection efficiency

The efficiency of the signal selection in the signal region is shown in Table 10.5. It is estimated as the $N^{\mathrm{SR\,sel}}/N^{\mathrm{exp}}$ ratio of the number of the events passing all the SR criteria, $N^{\mathrm{SR\,sel}}$, to the total number of generated signal events.

The acceptance times efficiency of the ETL signal selection with the established cut flow is estimated and cited in the rightmost column in Table 10.5.

As seen from Table 10.5, only a part of generated ETL events passes the criteria of the SR. The main reason for such event yield suppression is the explicit choice of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ final state, which makes up about 42% of all possible di-tau decay modes. The further drop is due to the tau-lepton trigger and identification efficiency. Trigger selection efficiency is about 0.8 for τ_{had} with 30 GeV $< p_{\mathrm{T}} <$ 40 GeV [208]. It reaches a 90% efficiency for p_{T} above 50 GeV. The 1-prong

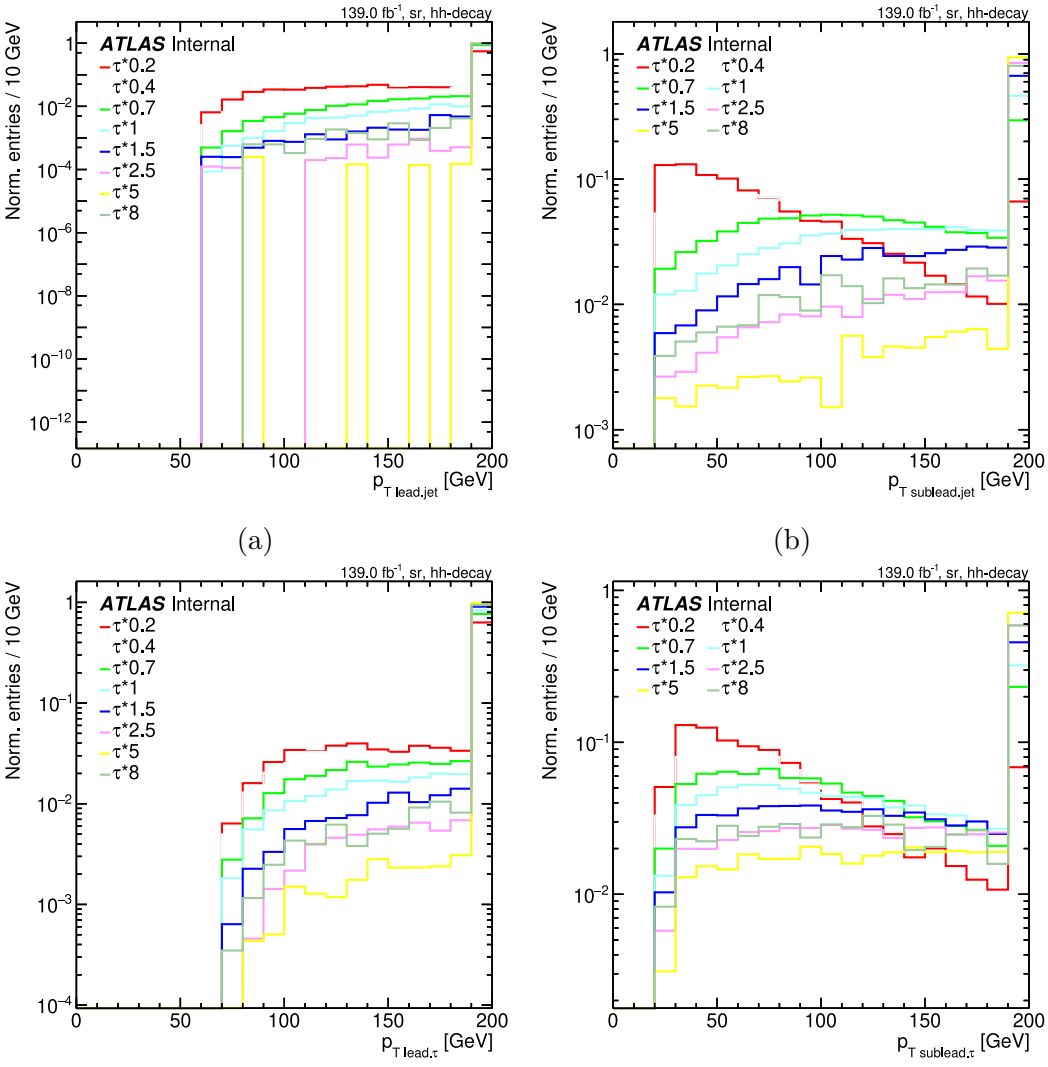
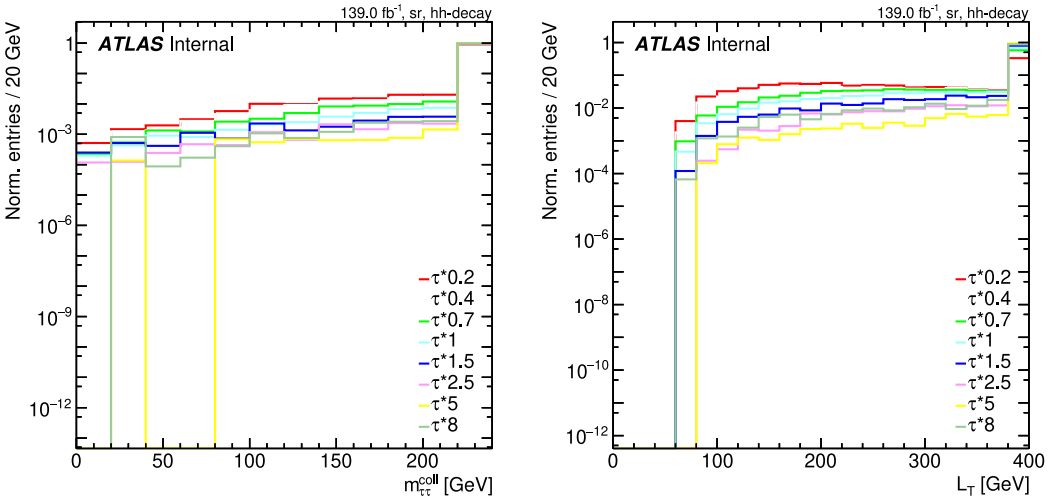


Figure 10.23: The p_{T} distributions for each object from the quadruplet: leading jet (a), subleading jet (b), leading tau-lepton (c), and subleading tau-lepton (d) with all the SR selections applied except the exact p_{T} criteria on this particular object.

(3-prong) tau-lepton identification efficiencies are 85% (85%) and 75% (60%) for the `Loose` and `Medium` ID working points [206] which are used in this analysis.

The effect of the trigger efficiency is notable for low-mass ETL points as shown in Table 10.5 with examples of 200 GeV and 400 GeV mass poles. Specifically, one can notice lower trigger selection efficiency (about 1/7 as can be seen from $\varepsilon^{\mathrm{NTuple}}$ for 200 GeV ETL) due to smaller tau p_{T}. Also, for the studied low-mass points of 200 GeV and 400 GeV, the signal selection efficiency drops further due to the applied kinematic selections on tau-lepton and jets p_{T} as well as $m_{\tau\tau}^{\mathrm{coll}}$ and L_{T}. As shown in Table 10.6, the kinematic selection influence the signal selection efficiency by 20% in total: $\sim$ 10% for the p_{T} (p_{T} and η) criteria on tau-leptons (jets) and then other $\sim$ 10% suppression due to the selections on $\Delta R_{\tau\tau}$, x^{coll}, $m_{\tau\tau}^{\mathrm{coll}}$, and L_{T}, combined.

The signal efficiency of about 7% is observed for mass points above $\sim$ 700 GeV



(a) (b)

Figure 10.24: The $N-1$ plots for the $m_{\tau\tau}^{\text{coll}}$ (a) and L_{T} (b) variables.

sample	N_{gen}	$N_{\text{sel}}^{\text{NTuple}}$	$\varepsilon^{\text{NTuple}}$	$N_{\text{sel}}^{\text{SR}}$	ε^{SR}
200 GeV	420 K	22620	12.6%	4047	0.96%
400 GeV	420 K	41442	23%	13494	3.2%%
700 GeV	420 K	53714	29.8%	23347	5.6%
1 TeV	420 K	71262	29.6%	34247	8.2%
1.5 TeV	120 K	18332	36.7%	8816	7.4%
2.5 TeV	120 K	19123	38.2%	9059	7.5%
5 TeV	120 K	18255	36.5%	7763	6.5%
8 TeV	120 K	17735	35.5%	7710	6.4%

Table 10.5: The numbers of reconstructed and selected excited tau-lepton events throughout data preparation. The efficiency of the signal selection in the constructed signal region is defined as $\varepsilon = N^{\text{SR sel}}/N^{\text{gen}}$. This estimate is based on the full signal MC data set in NTuples as of November 2022.

as shown in Table 10.5. This value was expected given suppression due to the choice of fully hadronic decay mode, acceptance, and identification efficiency. The effect of the kinematic selections on objects p_{T}, $m_{\tau\tau}^{\text{coll}}$ and L_{T} becomes smaller for high-mass points (as can be seen from data on 5 TeV and 8 TeV mass poles in Table 10.6).

With the signal selection efficiency in the considered SR estimated, one can predict the number of events which can be observed in the Run 2 ATLAS data set if the considered ETL model is valid. The expected number of events was estimated as $N_{\text{exp}}^{\text{SR}} = \varepsilon^{\text{SR}}\mathcal{L}\sigma_{q\bar{q}\to\tau\tau*}$. Here, $\sigma_{q\bar{q}\to\tau\tau*}$ is the production cross section calculated during MC generation, $\mathcal{L}$ is the integral luminosity of the analysed data set, and ε^{SR} stands for the signal acceptance as defined above. Figure 10.25 illustrates the expected number of selected ETL events as a function of hypothesised mass.

cut\$m_{\tau*}$ [TeV]	0.2	0.4	0.7	1	1.5	2.5	5	8
$N_\tau = 2, N_j >= 2$	497.8	620.2	468.5	308.6	146.1	29.9	0.43	0.013
$p_{\mathrm{T}}^{j0} > 70$ GeV	465.6	604.6	465.6	307.8	145.9	29.9	0.43	0.013
$\eta^{j0} < 2.4$	456.9	585.9	456.9	304.0	144.7	29.9	0.43	0.013
$p_{\mathrm{T}}^{j1} > 60$ GeV	396.3	423.8	396.4	279.7	138.2	29.1	0.43	0.012
$\eta^{j1} < 2.4$	208.0	398.9	378.4	268.8	133.9	28.5	0.42	0.012
$p_{\mathrm{T}}^{\tau 0} > 40$ GeV	206.5	397.4	377.6	268.5	133.7	28.5	0.42	0.012
$p_{\mathrm{T}}^{\tau 1} > 30$ GeV	187.1	374.6	364.5	261.0	130.8	28.0	0.41	0.012
ID τ_0	146.2	312.3	293.9	214.7	106.1	22.3	0.30	0.009
ID τ_1	123.7	257.4	251.4	180.2	87.5	18.2	0.26	0.008
$q_{\tau 0} \times q_{\tau 1} = -1$	118.7	248.1	239.9	170.5	81.7	16.6	0.23	0.007
truth match τ_0	118.5	247.9	239.8	170.4	81.7	16.6	0.22	0.007
truth match τ_1	116.8	245.9	238.5	169.6	81.4	16.6	0.22	0.007
$\Delta R_{\tau\tau} > 0.8$	116.3	243.0	233.2	165.1	78.6	15.9	0.21	0.007
x_0^{coll}	111.4	235.8	226.9	160.5	76.4	15.3	0.20	0.006
x_1^{coll}	105.3	226.7	218.7	154.7	73.6	14.6	0.19	0.006
$m_{\tau\tau}^{\mathrm{coll}} > 110$ GeV	102.1	222.7	216.6	153.8	73.2	14.6	0.19	0.006
$L_{\mathrm{T}} > 140$ GeV	92.0	207.6	209.5	150.7	72.44	14.6	0.19	0.006

Table 10.6: The yields of events with excited tau-leptons in the signal region. The breakdown of the yields observed at each selection step (aka cut) is shown. Several ETL mass points from the lowest considered one (200 GeV) through intermediate and up to 8 TeV are presented. These results were obtained with NTuples (version as for November 2022).

10.8 Control regions

The leading sources of SM background require their estimation to be checked in real data. The dedicated control regions are determined as the phase spaces where $Z \to \tau\tau$ and Top events are dominant in the phase space of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ and $e\mu$ final states, respectively. The shape, normalisation, and uncertainties on the $Z \to \tau\tau$ and Top background are estimated by fitting their MC prediction to real data in the CRs. Their prediction is then propagated to the SR in the fit. The $j \to \tau_{\mathrm{had}}$ is fully data-driven. The fake background is validated in the region defined in the same way as the SR except for an inversion of the requirement of tau-lepton electric charge: $q_\tau \times q_\tau = 1$ in the same-sign CR.

10.8.1 $Z \to \tau\tau$ control region

The event selection for the $Z \to \tau\tau$ control region is described in Table 10.3. The mass window 70 GeV $< m_{\tau\tau}^{\mathrm{coll}} <$ 110 GeV is chosen to constrain MC prediction for the $Z \to \tau\tau$ process as described in Section 10.6.1. The exact values for the range lower and upper cut values were chosen based on the fraction of $Z \to \tau\tau$ events in the total background input as described in Section (Figure 10.14b).

Alternatively, the $Z \to \tau\tau$ CR definition with visible di-tau mass was considered. However, the purity of such a CR was lower than one based on di-tau collinear mass (Table 10.7).

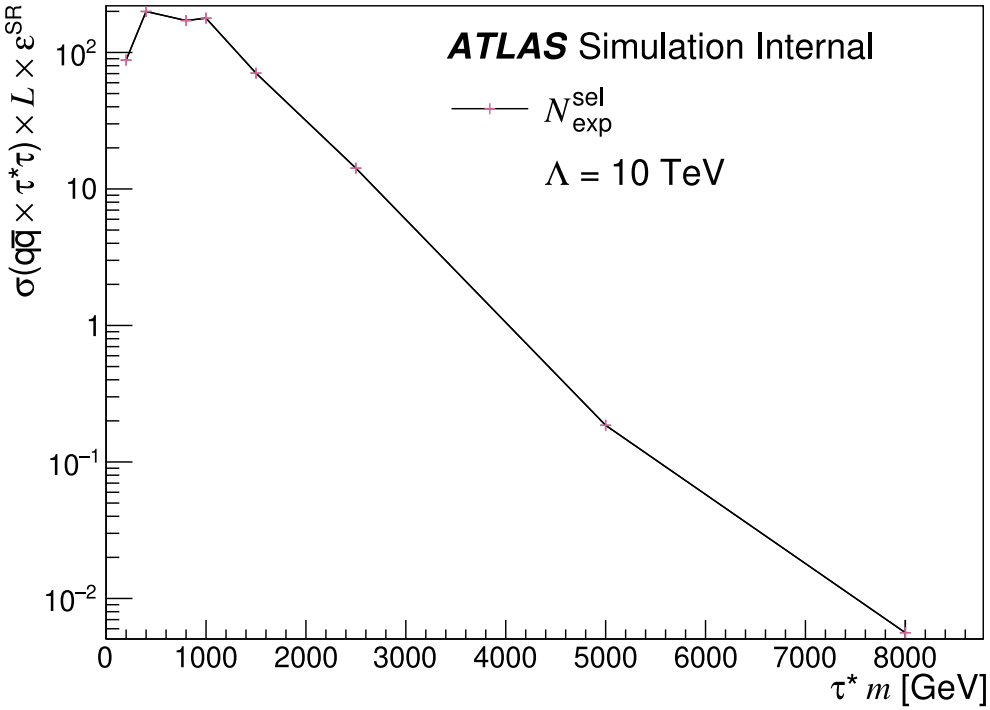


Figure 10.25: The expected number of excited tau-lepton events selected in the SR as a function of the ETL mass.

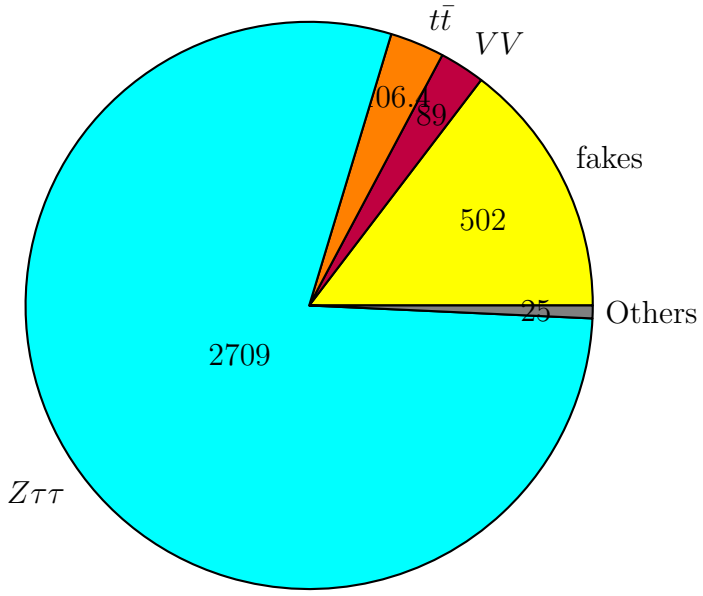


Figure 10.26: The composition of the absolute expected yields of SM processes in the constructed $Z \to \tau\tau$ CR. The $Z \to ll$, di-boson, and single-top background are combined in the "Others" category.

The consistency between the MC prediction and real data was checked in a set of observables. Figure 10.28 shows the kinematic distribution obtained in the Ztt

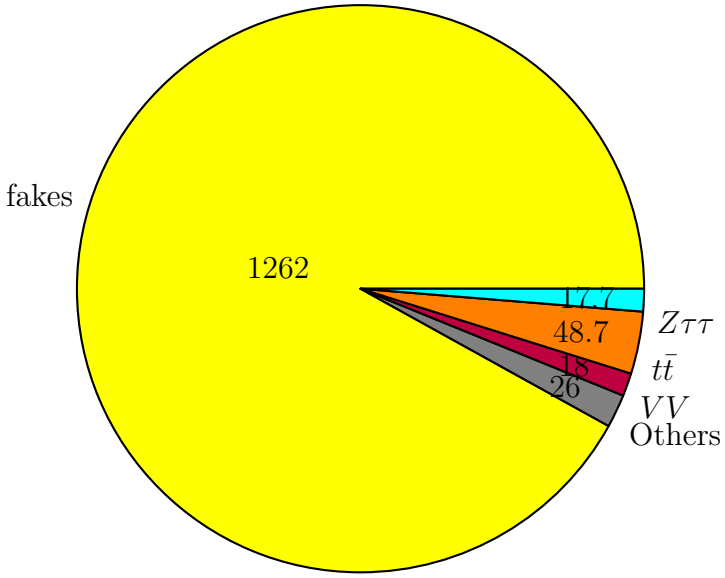


Figure 10.27: The composition of the absolute expected yields of SM processes in the constructed $Z \to \tau\tau$ CR. The $Z \to ll$, W + jets, and single-top background are combined in the "Others" category.

sample / sel.	50 GeV $< m_{\tau\tau}^{\text{vis}} <$ 100 GeV	70 GeV $< m_{\tau\tau}^{\text{coll}} <$ 110 GeV
total bkgr	32298.45	9459.15
$Z \to \tau\tau$	23130.87	8847.98
$Z \to ll$	3724.82	81.74
data	37284	9608
data/bkgr	1.15	1.02

Table 10.7: The event yield for SM background prediction and real data in two compared $Z \to \tau\tau$ CR defined with the selections: 50 GeV $< m_{\tau\tau}^{\text{vis}} <$ 100 GeV and 70 GeV $< m_{\tau\tau}^{\text{coll}} <$ 110 GeV. The results shown are obtained with interim NTuples.

CR. The MC modelling is in agreement with real data within the uncertainty.[21] The composition of the predicted SM background in the $Z \to \tau\tau$ CR and the SS CR is diagrammed in Figure 10.26 and 10.27, respectively.

10.8.2 Top control region

In this search, top-quark production is a considerable source of SM background. It makes up about 25% of the predicted background in the high-S_{T} range above 1 TeV (diagram 10.22a). Both top-quark pair and single-top production contribute to the di-tau and dijet final state.

An intuitive option is to construct the Top CR with a few selections orthogonal to the SR definition. Several attempts were eager to define a $t\bar{t}$ control region

[21] The pre-fit plots are given here. It means that they are prepared before the MC model was fitted to data, without profiling of individual uncertainties. The total uncertainty shown in the error band is obtained as a root of a sum of the individual systematic uncertainties, each squared, and the statistical uncertainty squared.

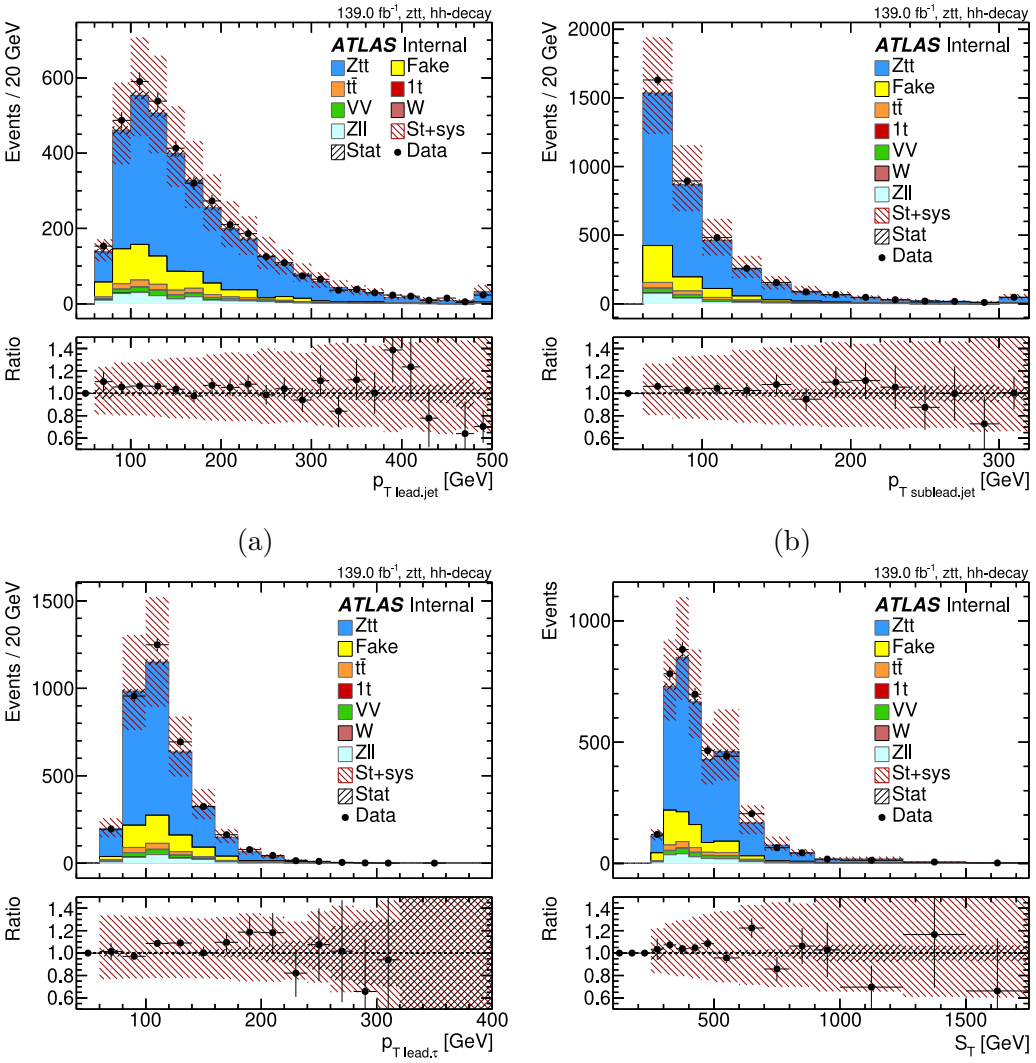


Figure 10.28: The distributions of p_{T}^{j0} (a), p_{T}^{j1} (b), $p_{\mathrm{T}}^{\tau 0}$ (c), and the scalar sum of p_{T} of the final-state quadruplet S_{T} (d) in the $Z \to \tau\tau$ control region.

with $\tau_{\mathrm{had}}\tau_{\mathrm{had}}jj$ final states. Figures 10.29 and 10.30 outline the resulting Top CR prototypes.

Specifically, tests were done to see how orthogonal selections could increase the fraction of top-quark production events (Figures 10.29 and 10.30). The list of the tested selections includes requirements on di-tau transverse mass, total transverse mass, MET and MET significance. Although the selection on m_{T} brought a relatively pure Top CR, it did not yield enough events in the S_{T} spectrum tail (Figures 10.29c and 10.29d). The advantage of the MET significance to discriminate events with significant MET was tested by applying the $S_{\mathrm{MET}} > 3(4)$ selection. The selection L_{T} variable inverted with respect to the SR definition, i.e. $L_{\mathrm{T}} < 140$ GeV, was tried in conjunction with the criterion on larger MET significance ($S_{\mathrm{MET}} > 4$) as shown in Figure 10.30b. Another attempted Top CR preserved the $m_{\tau\tau}^{\mathrm{coll}}$ range, which would be intermediate between the $Z \to \tau\tau$ CR and the SR: 110 GeV $< m_{\tau\tau}^{\mathrm{coll}} <$ 150 GeV. When applied together with the $S_{\mathrm{MET}} > 4$ criterion, this $m_{\tau\tau}^{\mathrm{coll}}$ selection revealed a low representation of the

Top processes as can be found in Figure 10.30c. Moreover, a significant signal contribution is not a good practice for CR. A similar effect is observed when introducing the criterion on effective mass along with selection on MET significance (Figure 10.30a). One more Top CR prototype was defined with selections on invariant mass of di-tau and dijet (i.e. $M_{\tau\tau jj} = \sqrt{(\mathcal{P}_{\tau 0} + \mathcal{P}_{\tau 1} + \mathcal{P}_{j0} + \mathcal{P}_{j1})^2}$) and S_{MET} and its phase space occurred to be limited in the S_{T} range. It also contained a large fraction of the signal events, as Figure 10.30d indicates.

Yet another prospective approach proxied the mass of the decaying top quark with the invariant mass of a tau-lepton and a jet paired in a way that maximizes the resulting invariant mass value. Although this method provides top-quark mass reconstruction, the obtained Top CR prototype was underpopulated in the region where the ETL signal is expected (see Figures 10.29a and 10.29b.).

These attempted Top CR prototypes were useless for control regions since they either yielded insufficient statistics, included significant fractions of other SM backgrounds, were sparsely populated high-S_{T} tail, or contained a considerable BSM signal contribution. Therefore, it was concluded that the Top CR could not be found in the $\tau_{\text{had}}\tau_{\text{had}}$ final state. This analysis purposefully does not apply b-jet tagging to distinguish top-quark decays. Such a flavour-inclusive approach allows for keeping the analysis flavour-inclusive which allows for considering final-state jets originating from quarks from all flavours inclusively.

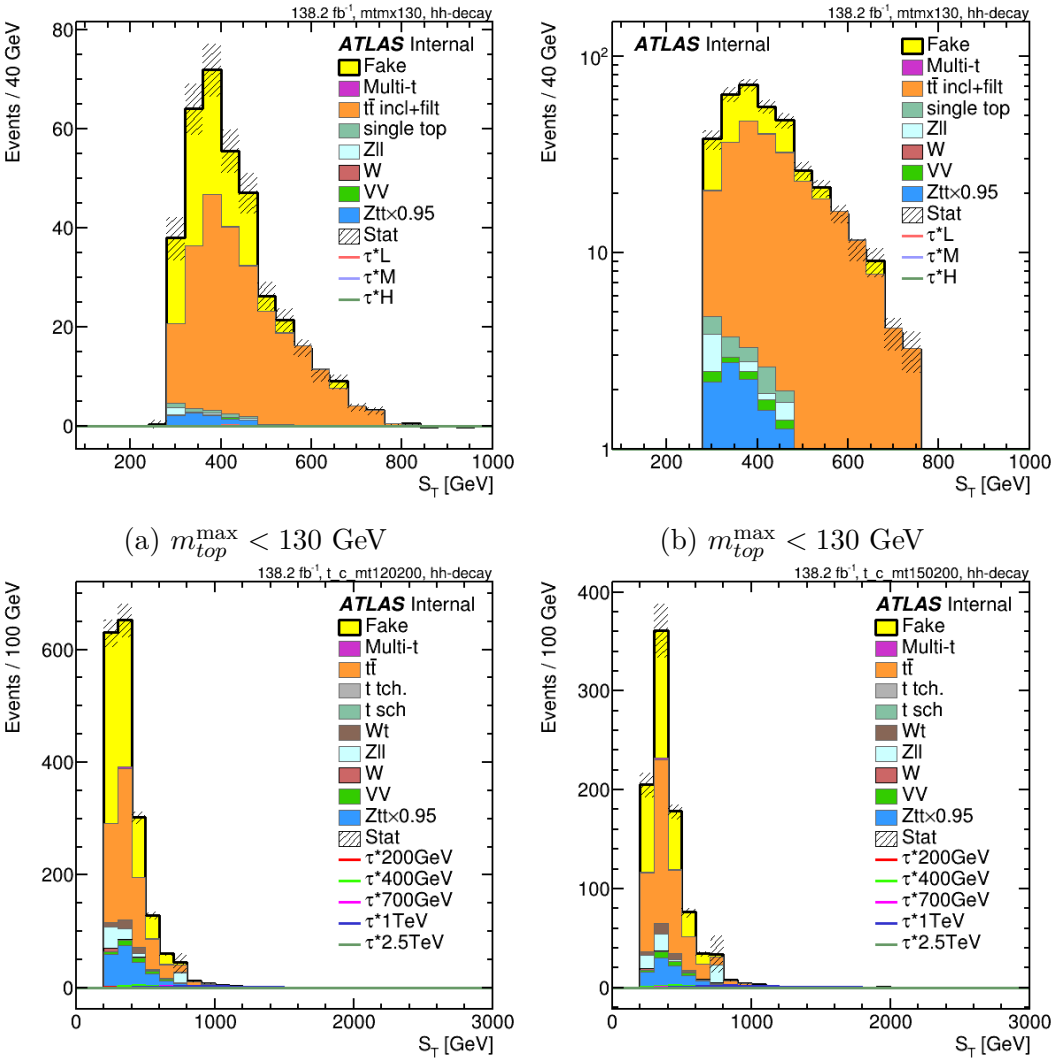
The problem of defining the Top CR can be viewed from a different angle. The Top control region can be built with a completely different final state which provides the advantage of being statistically independent of the SR. Namely, a $e\mu$ decay mode has higher branching ratio than the $\tau\tau$ mode of $t\bar{t}$. As a result, one can expect ~ 35 times more background events in the $e\mu$ final state than the di-tau one. Such a Top CR with light-lepton events provides a statistically independent phase space to estimate $t\bar{t}$ and Wt background.

In this Top CR, events should be accepted by the unprescaled single-muon trigger. Events are required to include exactly one electron and exactly one muon. At least two jets should be reconstructed in a given event. To enhance the contribution of the $t\bar{t}$ production, the presence of exactly two b-tagged jets is required.[22] The electron and muon should be of the opposite sign of electric charge and comply with the reconstruction, identification, isolation, and other requirements listed in Table 10.2. The final lower bound on kinematic variables are the following: $p_{\text{T}} > 60$ GeV and $|y| < 2.4$ for jets, $p_{\text{T}} > 30$ GeV for light leptons. The plots comparing the MC prediction to data are given in Figure 10.31.

10.9 Fake background

As mentioned in Section 6.1, the ETL search relies on fakes estimation with the fake factor (FF) method. Relative FFs are measured in the fake-enriched phase space in a subset of events with same-sign tau-leptons. The same-sign tau-leptons requirement enhances the share of events with jets misidentified as τ_{had}. The fake enriched region is defined orthogonally to the SR by requiring at least one of the following SR selections to fail, h.e. $m_{\tau\tau}^{\text{coll}} < 110$ GeV or/and $L_{\text{T}} < 140$ GeV. The FFs are obtained in function on the p_{T}, η, and the decay mode of τ_{had}. Two sets

[22] The `DL1r` b-quark tagger [89] is operating at the `FixedCutBEff_85` working point.



(a) $m_{top}^{\max} < 130$ GeV

(b) $m_{top}^{\max} < 130$ GeV

(c) 120 GeV $< m_{\mathrm{T}}^{\mathrm{total}} <$ 200 GeV

(d) 150 GeV $< m_{\mathrm{T}}^{\mathrm{total}} <$ 200 GeV

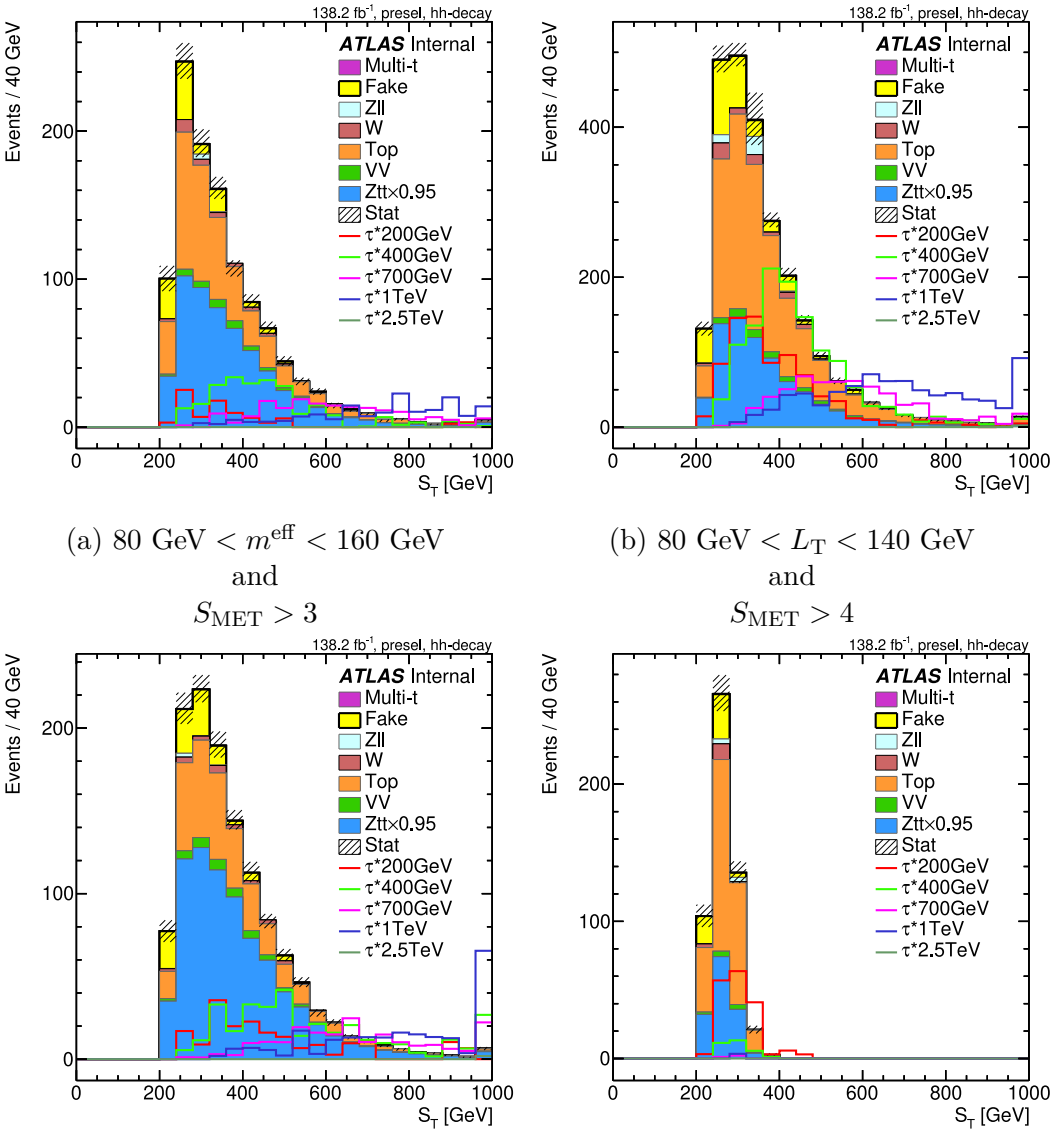
Figure 10.29: The S_{T} distributions in the tested Top CR prototypes (beginning). The selections applied on top of the preselections are mentioned in the captions.

of FFs are calculated in the following events: where leading tau-lepton passes the nominal (`Medium`) ID and where subleading tau-lepton passes the nominal (Loose) ID criterion. The former set results in "Loose" FFs measurement, and the latter one is used to measure "Medium" FFs.

In this way, the Fake background contribution can be estimated in the SR as explained in [209].

10.9.1 Same-sign control region

The FF method for $j \to \tau_{\mathrm{had}}$ estimation was validated in the same-sign control region (SS CR). Since studied processes (Drell Yan, ETL production and decay, $t\bar{t}$ production) do not gather same-sign tau-leptons in the final state, the inversion of the $q_\tau \times q_\tau$ selection can be applied to estimate the description of fake background component. The rest of the selections for SS CR are identical to the signal region definition (Table 10.3).



(a) 80 GeV $< m^{\text{eff}} <$ 160 GeV and $S_{\text{MET}} > 3$

(b) 80 GeV $< L_{\text{T}} <$ 140 GeV and $S_{\text{MET}} > 4$

(c) 110 GeV $< m_{\tau\tau}^{\text{coll}} <$ 160 GeV and $S_{\text{MET}} > 4$

(d) 200 GeV $< m_{\tau\tau jj} <$ 320 GeV and $S_{\text{MET}} > 4$

Figure 10.30: The S_{T} distributions in the tested Top CR prototypes (continuation). The selections applied on top of the preselections are mentioned in the captions.

The obtained same-sign control region is highly pure in τ_{had} fakes: the Fake background yields about 94% of the total predicted background (Figure 10.27). Kinematic distributions in the same-sign control region are shown in Figure 10.32. Figure 10.33 shows that the discriminative variable S_{T} shapes in the SR and the SS CR are in alignment.

10.9.2 Same-sign data-driven fake template

Alternatively, the fake background was estimated with a naïve same-sign data-driven template. This straightforward estimation of fakes contribution appeals

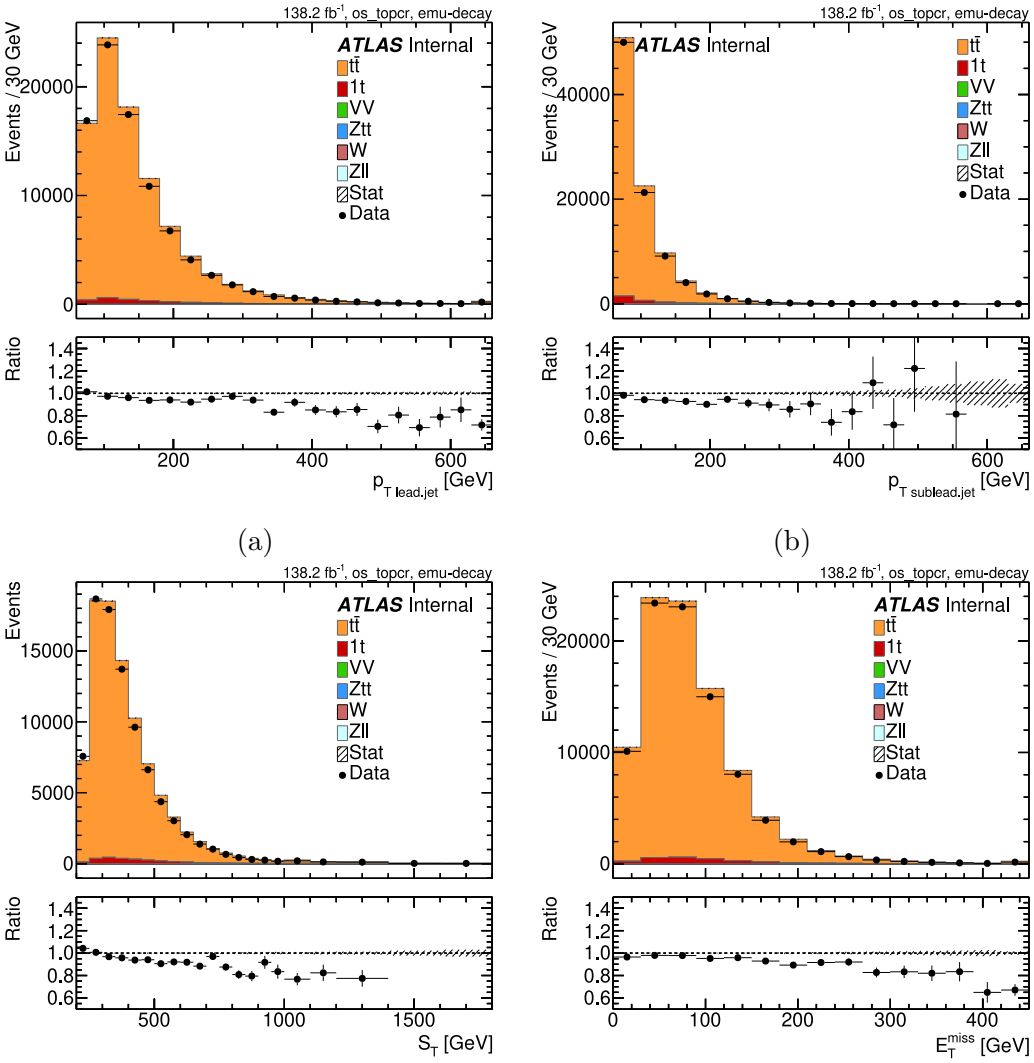


Figure 10.31: Kinematic distribution of p_{T}^{j0} (a), p_{T}^{j1} (b), S_{T} defined as $S_{\mathrm{T}} = p_{\mathrm{T}}^{e} + p_{\mathrm{T}}^{\mu} + p_{\mathrm{T}}^{j0} + p_{\mathrm{T}}^{j1}$ (c), and $E_{\mathrm{T}}^{\mathrm{miss}}$ (d) in the Top control region.

to the usage of events with same-sign tau-leptons in events passing all the SR selection criteria 10.3.

Figure 10.34 plots the distribution of S_{T} and di-tau collinear mass.

10.10 Systematic uncertainties

The shown distributions in the SR, the $Z \to \tau\tau$ and same-sign CRs include a set of systematic uncertainties. A part of the related systematic uncertainties cover differences in the kinematics of reconstructed objects in data and MC. In Monte Carlo, one needs to account for the detector acceptance, objects' reconstruction efficiency, and accuracy in the estimation of physical variables. For that, MC objects are transferred to detector-level objects by applying related resolution

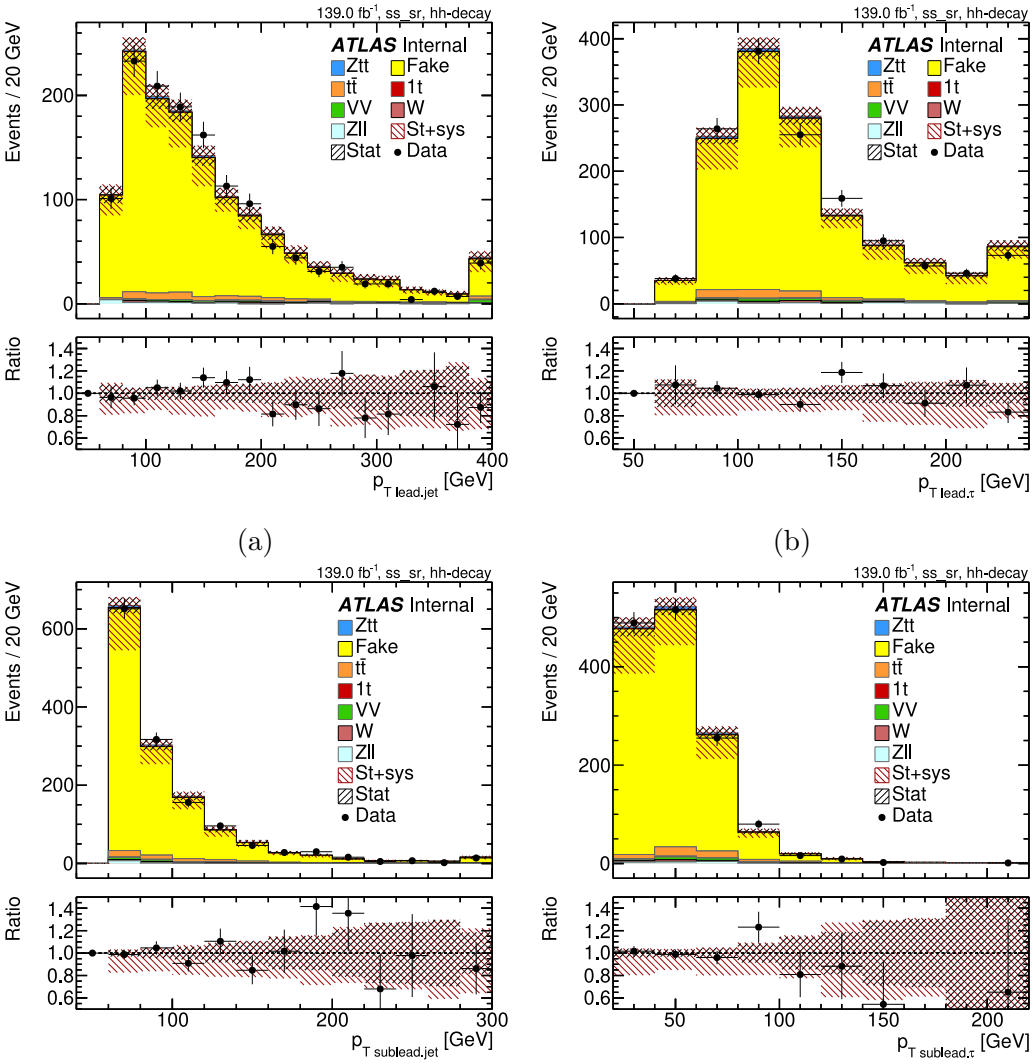


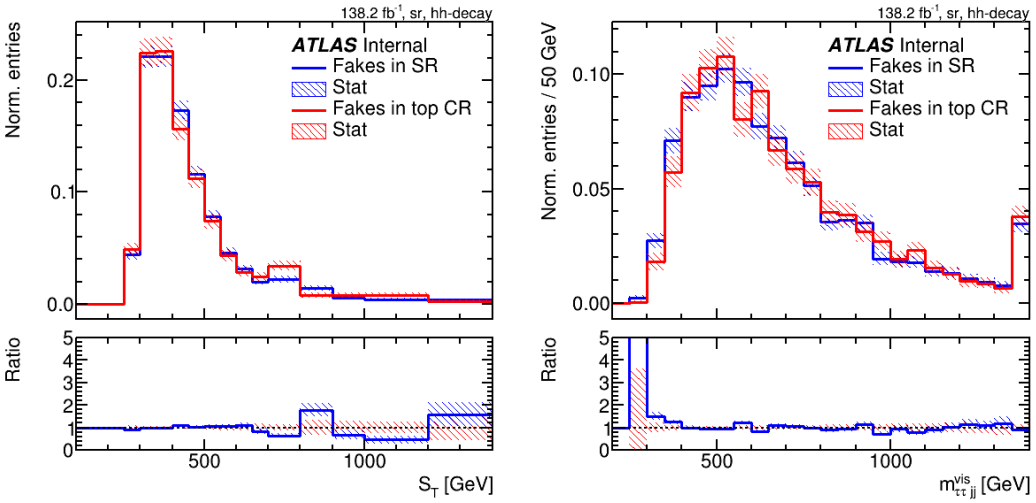
Figure 10.32: Kinematic distribution of p_{T}^{j0} (a), $p_{\mathrm{T}}^{\tau 0}$ (b), p_{T}^{j1} (c), and $p_{\mathrm{T}}^{\tau 1}$ (d) in the same-sign control region (SS CR).

response function and reconstruction efficiencies.[23] In this way, the energy resolution and scale in MC are adjusted to their expectation in real data, and the uncertainty on the energy resolution and scale are then propagated throughout the analysis. In this search, tau-lepton and jet energy scale and resolution uncertainties are considered.[24] Other experimental uncertainties cover MC weight variations. Tau-leptons and light leptons are assigned with the scale factors for efficiencies of reconstruction, identification, isolation, and trigger. As these scale factors are known with uncertainty, scale factors and, therefore, MC weights are varied respectively. The uncertainty on the jet tagging scale factor is included as well. In addition, the MC weight uncertainty on pile-up reweighting is added.

For the $Z \to \tau\tau$ prediction, all mentioned sources of systematic uncertainty

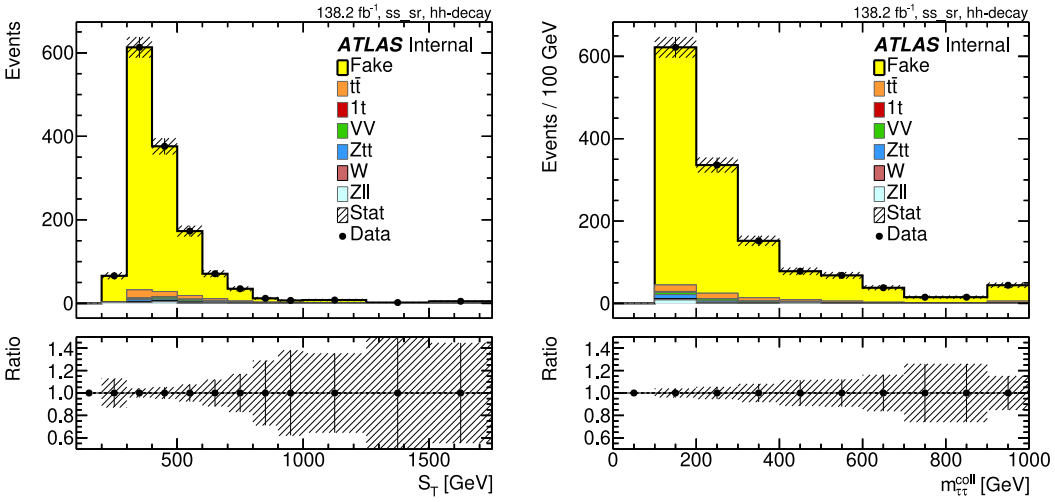
[23]This analysis step is typically called "simulation" as schematized in Figure 7.1 and briefly described in Section 4.

[24]This is done according to the usual ATLAS practice based on `TauSmearingTool`, `JetCalibrationTool`, `JetUncertaintiesTool`, `JetSmearingTool`.



(a) (b)

Figure 10.33: Comparison of kinematic distributions between SS CR and SR: S_{T} (a) and $m_{\tau\tau}^{\mathrm{coll}}$ (b).



(a) (b)

Figure 10.34: The S_{T} (a) and di-tau collinear mass $m_{\tau\tau}^{\mathrm{coll}}$ (b) in predicted SM background events where fake background contribution is estimated with the same-sign events.

are summed in quadrature in the pre-fit plots (for example, in Figure 10.8.1). Thereby, a large error band is seen. Respectively, the total pre-fit event yield has high uncertainty (as can be seen in Table 10.4).[25]

The excited tau-lepton signal modelling account for uncertainty on parton

[25]Before the fit, the uncertainties undergo normalisation, symmetrisation, smoothing, and pruning. In this search, the normalisation factor for $Z \to \tau\tau$ is correlated with some of $Z \to \tau\tau$ theory uncertainties. Once the normalisation effect is removed to decorrelate its effect from other NPs, the systematic uncertainty shrinks. Furthermore, the estimated systematic uncertainty magnitude is then verified with real data by fitting in the $Z \to \tau\tau$ control region. Once data are statistically analysed, each source of uncertainty becomes an individual nuisance parameter, and the likelihood function is maximised for a given value of a particular NP. In this way, the uncertainty becomes profiled. The resulting post-fit uncertainties will be significantly

distribution functions as well as the systematic variation due to a choice of a particular PDF set. The uncertainty on the running scale α_S is also included. These uncertainties were estimated with the `LHAPDF` tool [210, 211].

10.11 Statistical analysis

Once the signal region is defined, and the background contribution is estimated, it is time to answer the question of whether there is a hint at the existence of excited tau-leptons. The statistical analysis aims to find signatures of excited tau-lepton production and decay which would manifest as an excess of events in the measured S_T distribution.

10.11.1 The fit setup

The statistical analysis is performed with the YAW (Yet Another Wrapper) framework [212], which is a Prague's wrapper [212] around the `HistFitter` [111] machinery. The workspace input data to the fit machinery are prepared with the adaptation of Prague's **R**OOT **a**nalysis **c**ommon (RAC) framework [203]. The excited tau-lepton signal strength μ is a parameter of interest (POI) in this analysis.

The signal region definition is the same for all analysed τ^* mass points (from 200 GeV to 9.75 TeV). Two control regions are used for SM background constraining. Drell-Yan $Z \to \tau\tau$ events are constrained and normalised in the $Z \to \tau\tau$ CR in the $\tau_{\text{had}}\tau_{\text{had}}$ final state (Section 10.8.1). Top pair production is constrained in the CR in $t\bar{t} \to e\mu$ events with the $e\mu$ final state (Section 10.8.2). The Drell Yan ($Z \to \tau\tau$ and $Z \to ll$) background is normalised with a freely floating normalisation factor NF_Z. Likewise, the top-quark background (from $t\bar{t}$ and Wt production) is scaled with the NF_{Top}. The normalisation factors (NFs) enter the model as nuisance parameters.

Other nuisance parameters included in the model are the statistical uncertainty in Monte Carlo modelling and the uncertainty on the size of the collected data set (the integrated luminosity). The former enters the likelihood as the γ parameters defined as the statistical uncertainty in each bin of the input S_T distributions for each considered region. The latter describes the variation of the event yield in the data set and is taken as large as 1.7%, as measured by the LUCID detector conditions [213, 214].

In this way, the presented fit is made with the model simplified with respect to Eq. (7.3) and including statistical but systematic uncertainties. Then the likelihood of observing events as they are measured in real data becomes the following:

$$L(\mu, \gamma) = \prod_i Pois\{n_i; \sum_p \mu_p \nu_i^p \gamma_i\} Pois(m_i; m_i \gamma_i) \tag{10.5}$$

Here the likelihood is parametrised with the μ NPs (the excited tau-lepton signal strength μ, the freely floating normalisation factors NF_Z and NF_{Top}, and fixed NF for other SM backgrounds) and the γ NPs for statistical uncertainty in the binned template. The i and p indices stand for the ith bin in the input S_T distributions

smaller.

and the considered SM and ETL processes. As introduced in Section 7.4, n_i is the ith bin content, ν_i^p is the expected number of events in the ith bin for the pth process, and m_i is the effective number of events in the ith bin of the input S_{T} histogram of the process p.

The negative logarithmic likelihood is calculated regarding the set of nuisance parameters (γ's, $NF_{Z\to\tau\tau}$, NF_{Top}, the uncertainty on luminosity).

The presented results are based on the maximum likelihood fit to real data in control regions. This allows estimating a set of NPs for the $Z \to \tau\tau$ and Top background as precisely as possible. Once these background sources are constrained and normalised in the corresponding CRs, their estimated NPs are propagated to the SR. In the SR, the model is fitted to the Asimov data set, i.e. real data are replaced by the total background estimate.[26]

This analysis aims to discover the excited tau-lepton. The expected significance is to be calculated. In the case of no BSM signatures, the limits on the excited tau-lepton production cross section are set. Given the considered production and decay model, one can also set limits on the ETL mass.

10.11.2 Fit to the Asimov data set

The fit with the background-only ($\mu = 0$) alternative hypothesis is made to the Asimov data set in all the regions simultaneously. The background-only fit allowed analysers to verify that the statistical machinery is set up correctly. Although the Asimov data set does not have the power to constrain background processes, its results can highlight possible biases in NP estimation.

Figure 10.35 demonstrates the values of the μ and γ nuisance parameters. Their values centre around the unit and larger constraints on γ parameters are observed in less populated bins. No correlation above 0.3 is seen between individual NPs (Figure L.1 in Appendix L). The fitted yields are as expected and shown in Table L.1 (Appendix L).

10.11.3 Fit to real data in control regions

Figure 10.38 shows pre-fit and post-fit S_{T} distributions in the $Z \to \tau\tau$ CR, the Top CR, and the SR. As seen from the pre-fit and post-fit plots in the CRs, a better agreement between MC and data is achieved after the $Z \to \tau\tau$ and Top ($t\bar{t}$ and Wt combined) background are constrained with real data. The Top background prediction will require further reweighting to NNLO calculations. The number of background events predicted in the SR post-fit coincides with the SM expectation (Table 10.8). This is expected as the fit to the Asimov data is done in the SR. The expected Z and Top background shape and yield are estimated by extrapolation from the corresponding CRs to the SR.

The constraints and pulls of the NPs are shown in Figure 10.37. Likewise, it was seen in the fit to the Asimov data set (Figure 10.35) larger constraints are obtained to the statistical uncertainties in less populated bins. The fit to real data helped to define proper normalisation for the Z and Top background

[26] In ATLAS jargon, this type of fit is called "blinded": one does not look at real data in the region sensitive to the expected signal. This is a common approach in the ATLAS community to make sure that the strategy of a given analysis is solid.

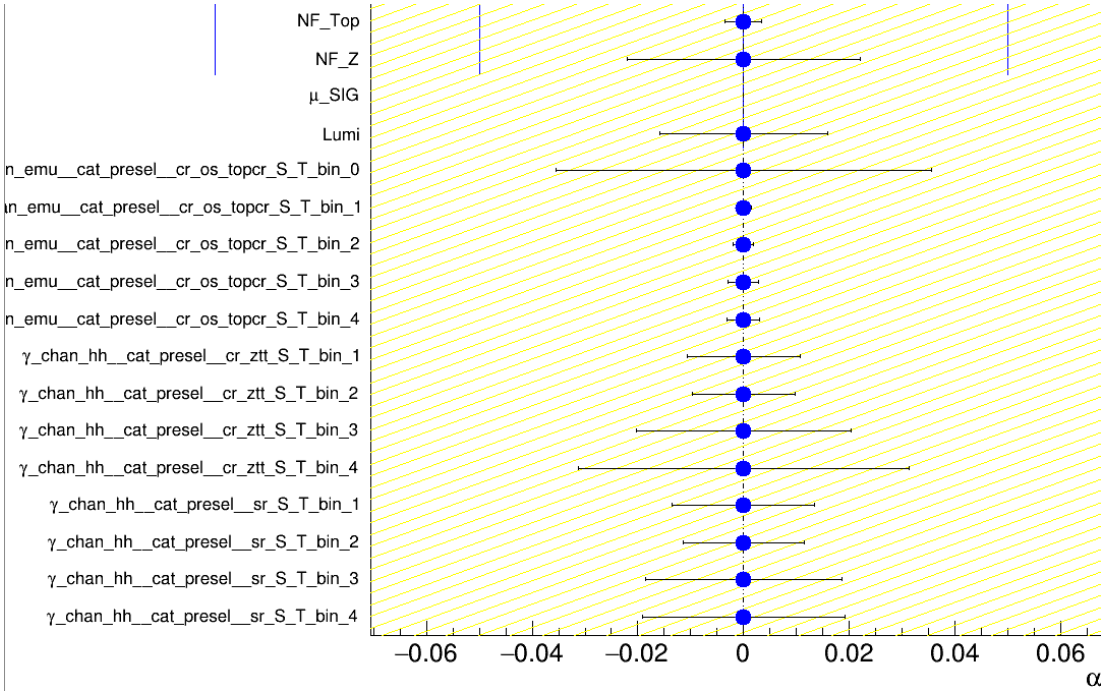


Figure 10.35: The pulls and constraints on the nuisance parameters in the fit to Asimov data set in the considered CRs and the SR. This result is obtained in the fit for excited tau-lepton of 1 TeV mass.

processes. They are pulled to the post-fit values $NF_Z = 1.08$ and $NF_{\text{Top}} = 0.96$. The correlation between individual NPs is presented in the matrix (Figure 10.36).

In the basic setup of the statistic-only fit, no significant correlation between NPs was found. Correlation up to 0.25 is observed between the normalisation factor for $Z \to \tau\tau$ and the statistic uncertainty in $S_{\text{T}} < 600$ GeV bins, which are the most populated in the $Z \to \tau\tau$ spectrum and, hence, provide a source of NPs constraints. Similarly, the correlation about 0.25 is seen between NF_{Top} and the statistical uncertainty in the 200 GeV $< S_{\text{T}} <$ 400 GeV range.

Another fit to real data in the CRs was made by including sets of systematic uncertainties in the $Z \to \tau\tau$ CR and SR. The systematic uncertainties on tau-lepton and jets scale factors were included. The SM and ETL modelling accounted for the tau-lepton and jet energy scale and resolution uncertainties. This fit also incorporated the uncertainty on the pile-up weight. The resulting post-fit S_{T} distributions are shown in Figure 10.39. Figure 10.40 demonstrates the correlation between individual NPs. A majority of the introduced uncertainties are not significantly correlated. The only high correlation is between NF_{Top} and the uncertainty of the tau-lepton energy scale (TES). Its value results from significant normalisation components of systematic uncertainties in top-quark production processes. Figures M.1 and M.2 in Appendix M show the profile scans for the Z and Top normalisation factors and the TES uncertainty, respectively. Figure 10.41 illustrates the post-fit values and uncertainties on the systematic uncertainties. No significant pulls or constraints are observed.

It is worth noting that the systematic uncertainties did not cover the Top CR. The SM prediction needs to include systematic variations, and the Top background modelling requires reweighting, as mentioned above (Section 10.11.3).

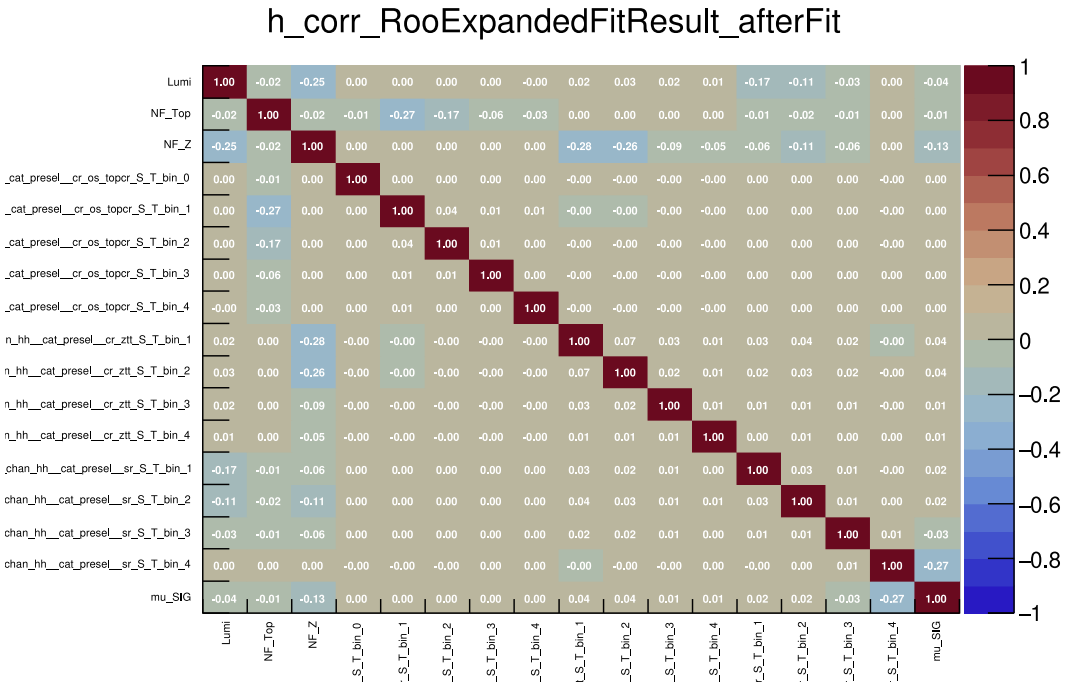


Figure 10.36: The correlation between NPs in the fit to real (Asimov) data in the CRs (SR). The result for the fit in the ETL with the mass of 1 TeV is shown.

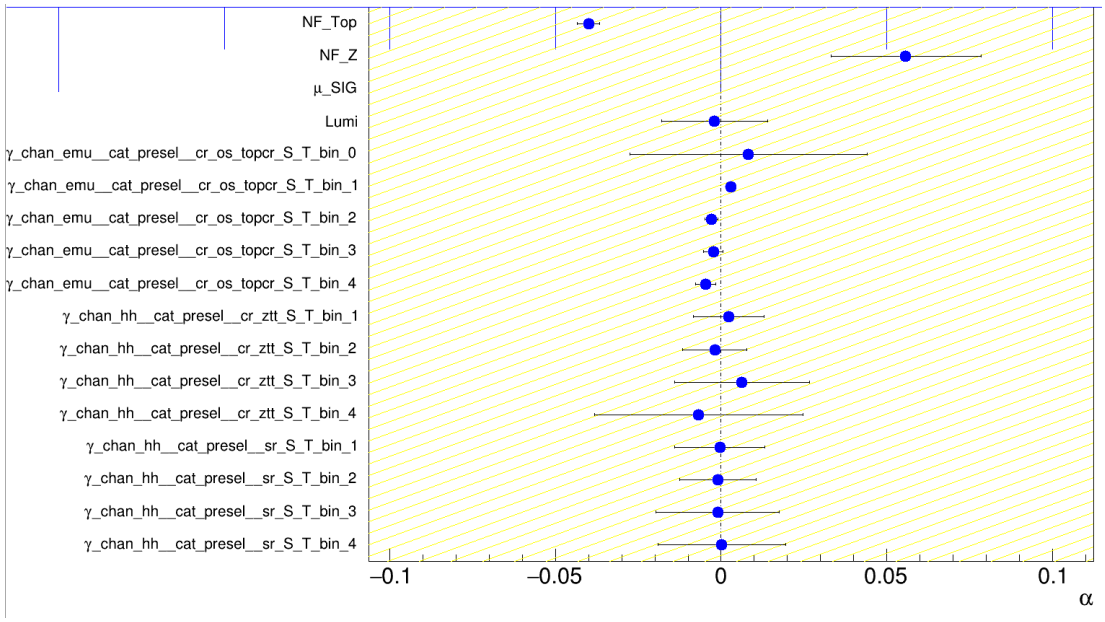


Figure 10.37: The pulls and constraints on the nuisance parameters in the fit to real (Asimov) data in the CRs (SR). This result is obtained in the fit for excited tau-lepton of 1 TeV mass.

After these steps, the post-fit (Figure 10.38d) agreement between the prediction and data improves within the uncertainty.[27]

[27]Furthermore, the systematic uncertainty between the single and pair top-quark production defines the proportion of their yields.

Process, N_{events}	$Z \to \tau\tau$ CR	Top CR	SR
Observed events	88777	3741	4287
Fitted bkgr events	88778.62 ± 297.70	3735.04 ± 58.53	4297.16 ± 43.83
Fitted Z events	39.57 ± 0.85	3033.48 ± 60.22	1250.30 ± 27.23
Fitted Top events	88658.51 ± 297.74	109.36 ± 0.87	1118.07 ± 9.45
Fitted Fakes events	0.00 ± 0.00	501.25 ± 8.93	1828.98 ± 30.17
Fitted Other events	80.58 ± 1.29	90.97 ± 1.60	103.16 ± 1.71
Fitted ET1000 events	$-0.04^{+0.22}_{--0.04}$	$-0.01^{+0.06}_{--0.01}$	$-3.35^{+17.84}_{--3.35}$
MC exp. SM events	92397.02	3579.53	4438.73
MC exp. Z events	37.51	2872.23	1185.17
MC exp. Top events	92276.80	113.86	1165.56
MC exp. Fakes events	0.00	501.86	1833.81
MC exp. Other events	80.87	91.11	103.43
MC exp. ET1000 events	1.84	0.47	150.76

Table 10.8: The event yields in the fit to measured data in two control regions ($Z\tau\tau$ CR and $e\mu$ Top CR) and fit to the Asimov data set in the signal region. The signal of 1 TeV excited tau-lepton is considered. The Z background combines $Z \to \tau\tau$ and $Z \to ll$ processes. The Top background includes $t\bar{t}$ and Wt. The "Other" background category includes W + jets and VV events.

10.11.4 Expected discovery significance

The expected discovery significance is set in the model-dependent exclusion fit mode. Tests of the null B-only hypothesis against the alternative $S + B$ one are made. Results of hypothesis tests show the level of compatibility between data and SM-only hypothesis and quantified by a p_0-value. The p_0 value is the probability of seeing an upward fluctuation of the event yield as high or higher than the observed data.

The Asimov data set is constructed as the $S + B$ model. The post-fit NPs are illustrated in Figure N.1 in Appendix N.

Tests to obtain p_0 significances were repeated for a set of S models with ETL of varying mass. The result is presented in Figure 10.42. This analysis has the potential to claim the discovery of ETL with mass in the range between 300 GeV and 1.4 TeV with 5σ significance.

10.11.5 Upper limits on excited tau-lepton mass and production cross section

The upper limits of the excited tau-lepton signal strength are estimated in the model-dependent exclusion fit setup in the `HistFitter`. A null $S + B$ hypothesis is tested against the alternative B hypothesis. The analysis established the level of compatibility between data and the $B + S$ hypothesis (SM background and BSM excited tau-lepton combined). The acquired p_0-value is the probability of a downward fluctuation which would be lower than the observed test statistic under the SM+signal hypothesis.

In the fit, the signal strength is a parameter of interest, and several hypotheses $\mu S + B$ are tested in order to find the ETL signal strength, which data can

exclude. For each tested μ value, the test statistic defined according to Eq. (7.8) is estimated.[28] For limit settings and the p_0 values is calculated as the integral Eq .(7.7). Figure 10.43 shows p_0-values as a function of the hypothesized value of μ in the fit to real (Asimov) data in CRs (SR). The signal models $\mu S + B$ are excluded at 95% confidence level if the $p \leq 0.05$ (as said in Section 7.6.3). The upper limit on the BSM signal strength $\mu_{\mathrm{UL}}^{\mathrm{exp}}$ is then defined by inverting the hypothesis test (by solving Eq. 7.12).

This procedure was repeated for the excited tau-leptons hypotheses in a mass range between 200 GeV and 5 TeV. The ELT production cross section as a function of the ETL mass is demonstrated in Figure 10.44. The expected cross section is estimated in the fit to the Asimov data in all the considered regions.[29] The observed cross section is obtained in the fit to real (Asimov) data in CRs (SR). Two results – with and without the systematic uncertainties included – are demonstrated. Statistical uncertainty becomes dominant in searches for high-mass (> 2 TeV) ETL.

Ultimately, an upper limit on the cross section can be calculated as: $\sigma_{\mathrm{UL}}^{\mathrm{exp}} = \sigma^{\mathrm{theory}} \mu_{UL}^{\mathrm{exp}}$ where σ^{theory} is the ETL production cross section predicted by the considered signal model. The lower limit on the ETL mass can be set at 2 TeV (Figure 10.44). The existence of the excited tau-leptons with mass below ~ 2 TeV is excluded with 95% CL.

10.12 Results and conclusions

This chapter described the analysis aiming to find the excited tau-leptons in the Run 2 ATLAS data set. The considered BSM model assumes that excited tau-leptons are produced and decay through contact interaction. The ETL production's signature is looked for in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}jj$ final states.

The production and decay of excited tau-leptons in the contact interaction were simulated with the Pythia Monte Carlo generator. The generated events were then passed through the simulation and reconstruction chain at the ATLAS computing Grid. Further, necessary NTulpes were derived from large DAOD data files.

The Monte Carlo simulation and data-driven methods predict the SM background. Leading background sources are constrained and/or validated with real data in the corresponding control regions.

Sets of event selections are applied to distinguish between the phase spaces enriched in the SM and BSM processes. The final event selection was optimized in terms of signal-to-background separation and signal selection efficiency. The final selection provides $\sim 26\%$ ($\sim 88\%$) SM background rejection at the 95% level of 200 GeV (1 TeV) ETL signal acceptance.

Statistical analysis is made with the S_{T} observable, which was found to be sensitive to the presence of ETL events. This variable was found to provide steady discrimination of ETL events against background events.

The model was fitted to real data in the $Z \to \tau\tau$ and Top CRs, and the fit to the Asimov data set was made in the signal region. The presented analysis

[28] The signal strength μ varies in the unconditional likelihood.

[29] An example of the p-value estimation in the tests of multiple ETL signal strength hypotheses is given in Figure L.2 (Appendix L).

can claim the discovery of the excited tau-lepton with mass between 300 GeV and 1.4 TeV. The expected upper limit on the excited tau-leptons with mass is ~ 2 TeV.

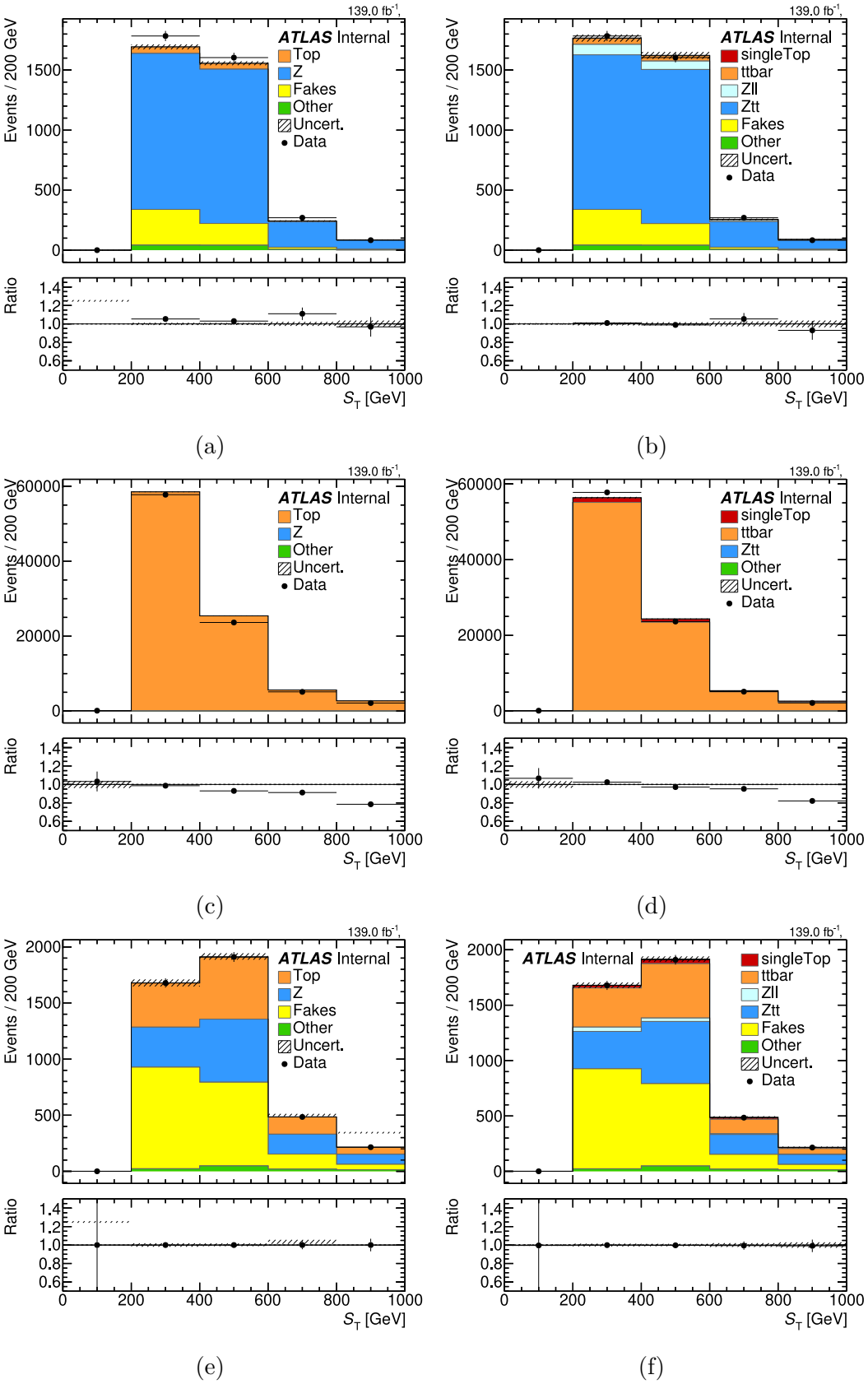


Figure 10.38: The pre-fit (on the left side) and post-fit (on the right side) S_T distributions in the $Z \to \tau\tau$ CR (a, b), the Top CR (c, d), and the SR (e, f). The results of the statistical-only fit to the real (Asimov) data in CRs (SR) are shown. The excited tau-lepton of 1 TeV mass is included as a BSM signal.

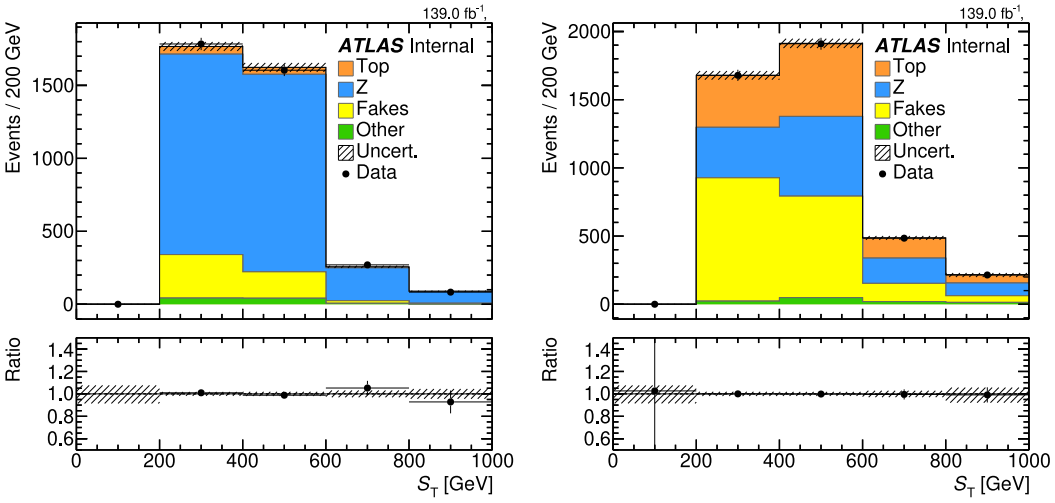


Figure 10.39: The post-fit S_T distributions in the $Z \to \tau\tau$ CR (a) and the Top CR (b). The results of the fit to the real (Asimov) data in CRs (SR) are shown. The excited tau-lepton of 1 TeV mass is included as a BSM signal. The statistical uncertainties, and the jet and tau-lepton uncertainties are accounted for in the fit.

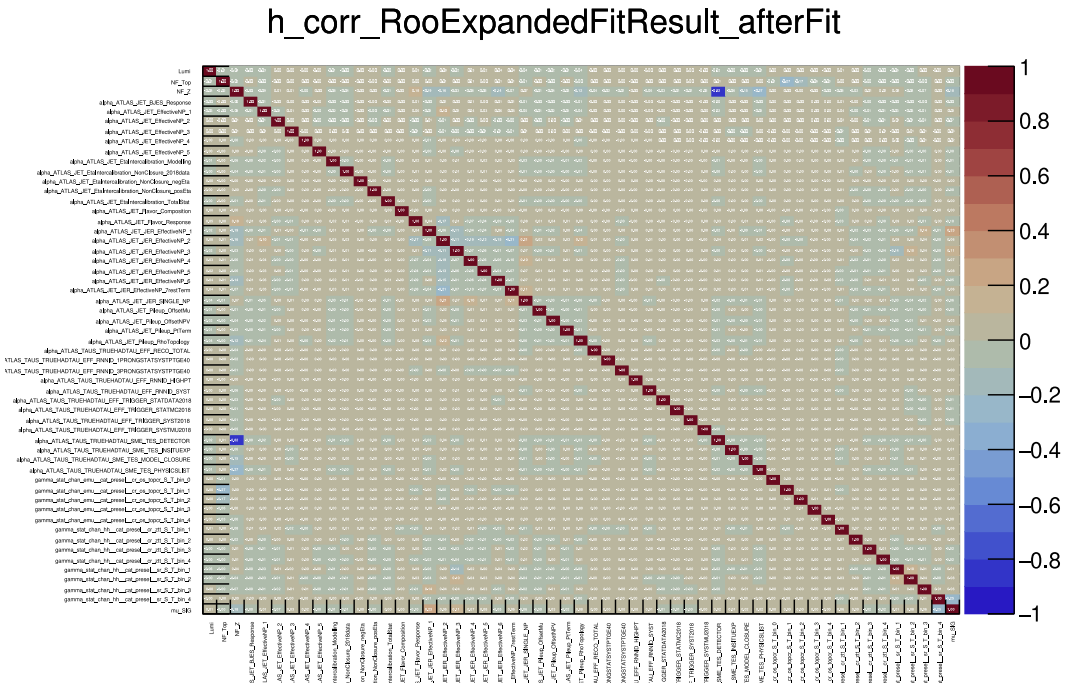


Figure 10.40: The correlation between NPs in the fit to real (Asimov) data in the CRs (SR) with statistical and tau-lepton and jet systematic uncertainties included. The result for the fit in the ETL with the mass of 1 TeV is shown.

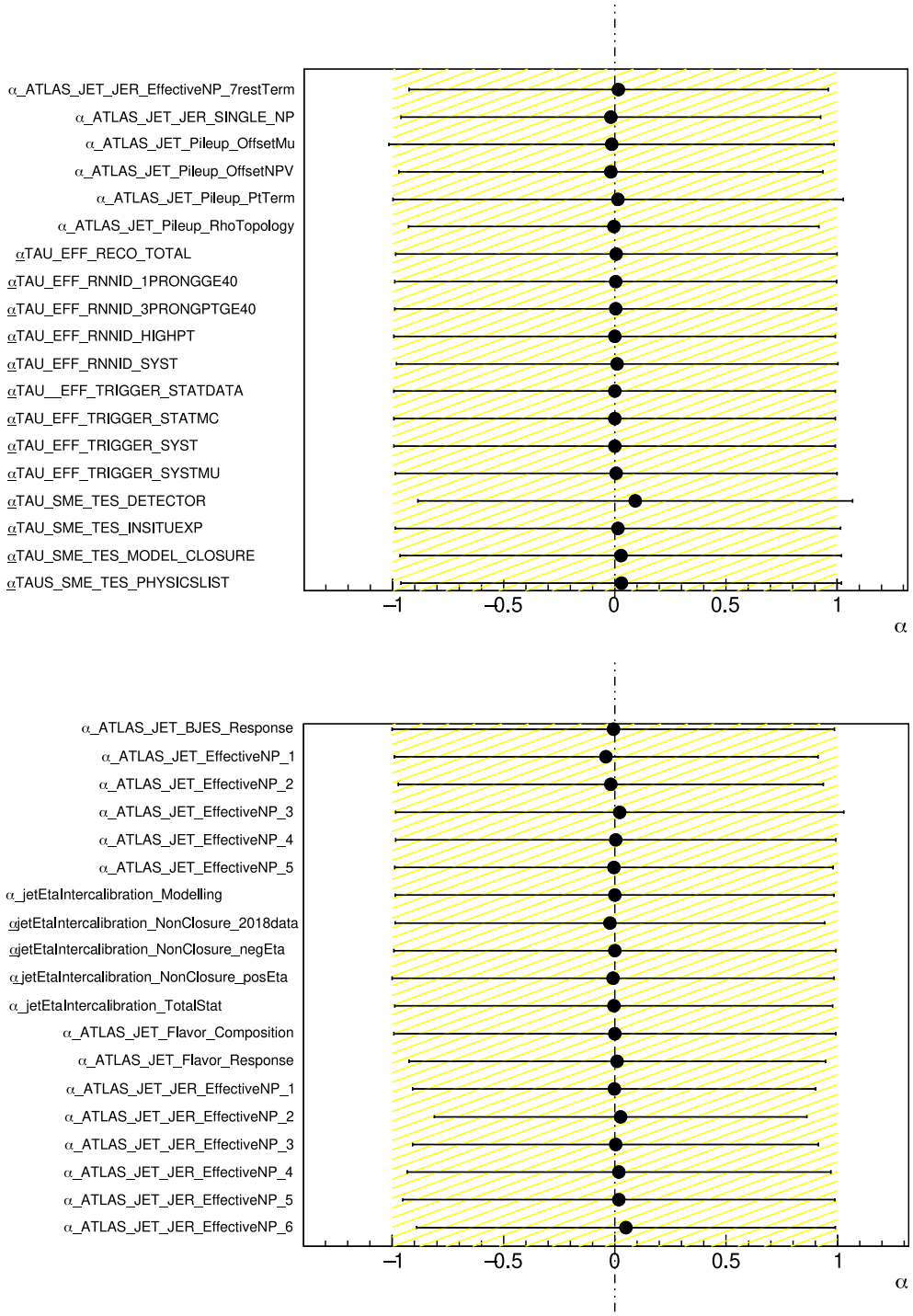


Figure 10.41: The post-fit values and uncertainties on the tau-lepton (a) and jet (b) systematic uncertainties in the fit to real (Asimov) data in the CRs (SR).

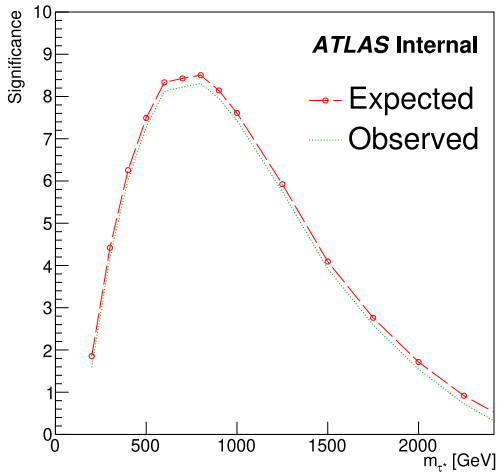


Figure 10.42: The expected discovery significance in dependence on the excited tau-lepton mass.

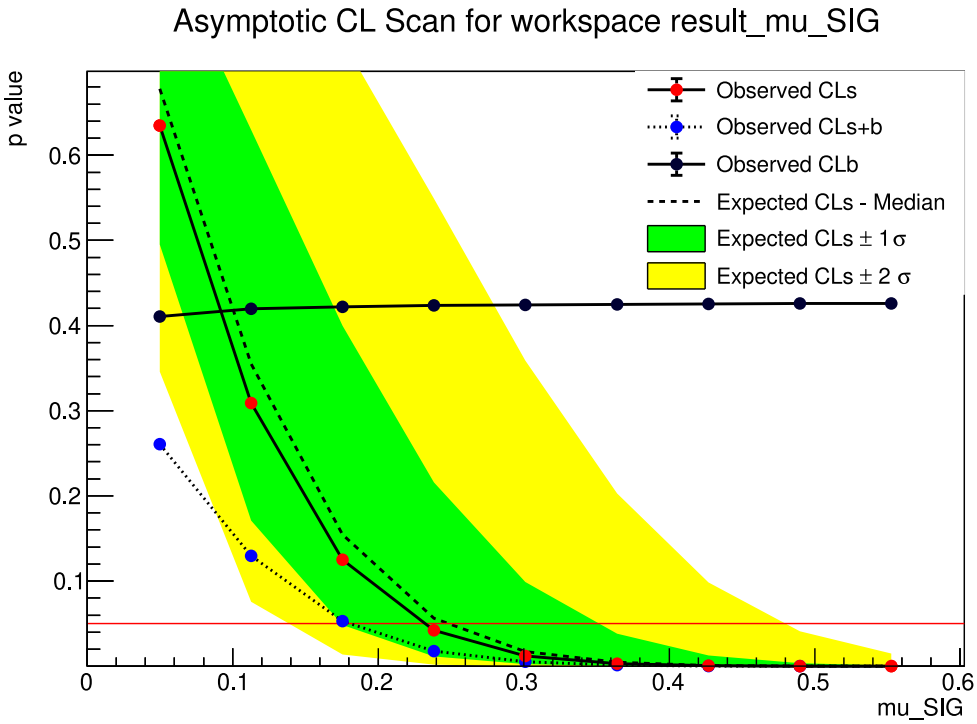


Figure 10.43: The dependence of the p_0-value on the excited tau-lepton (1 TeV) signal strength μ. Here the CL_b (CL_{s+b}) depicts the level of confidence in the background (signal+background) model.

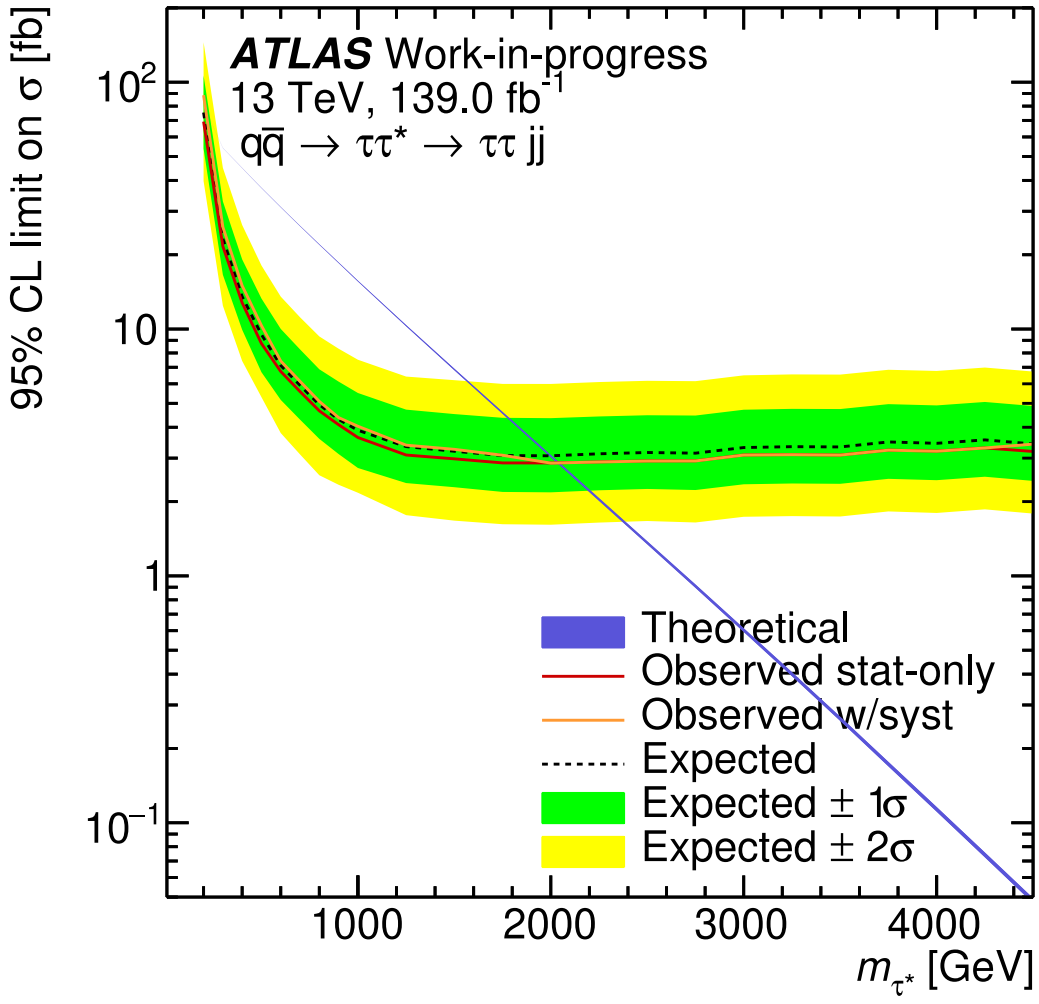


Figure 10.44: The 95% CL limit on the excited tau-lepton production cross section as a function of ETL mass obtained with the Asimov data set.

Conclusion

This thesis describes research conducted at the world's largest high-energy laboratory, the Large Hadron Collider. The presented studies contribute to expanding our current understanding of elementary particle physics. The experimental data collected by the ATLAS detector during Run 2 at the LHC are analysed.

One of the two presented analyses measured the cross section of Higgs boson production in events where Higgs boson decays to a pair of tau-leptons. The research mainly concentrated on the $\tau_{\text{lep}}\tau_{\text{lep}}$ final state. Reconstruction of the signal-sensitive variable - the invariant mass of a di-tau system - was performed with the Missing Mass Calculator (MMC). This PhD project multilaterally studied the MMC algorithm. It established the optimal parameters of the scan over possible MET values: the MET resolution σ_{MET} and the MET scan range. Specifically, the MET resolution was parametrised to correspond to the Run 2 data-taking conditions. Smearing of MET under higher pile-up was detected. The additional parametrisation on the pile-up level provided a more comprehensive description of the MET resolution. The scan range of $4\sigma_{\text{MET}}$ was confirmed to perform efficiently. This research verified that the object-based MET significance S_{MET} is applicable for the σ_{MET} estimation. Moreover, the studies suggest that the $H \to \tau\tau$ analysis can benefit from applying S_{MET} in effective event selection. Overall, the MMC tool revealed its robustness to modifications, reaffirming that this algorithm is highly suitable for the Higgs boson mass calculation. Further, the presented analysis finds that more than a less rigorous scan over possible solutions is required due to ever-improving data-taking conditions and event reconstruction algorithms at the ATLAS experiment. Consequently, the computation time of di-tau mass calculation was accelerated by a factor of four. The ultimate MMC setup calculates Higgs boson mass with a resolution of $\sim$ 16.5 (14.5) GeV in the $\tau_{\text{had}}\tau_{\text{had}}$ ($\tau_{\text{lep}}\tau_{\text{lep}}$) channel. Besides, the strategy of fake background modelling in the $H \to \tau\tau$ coupling measurement was validated. The studies show that the data-driven method and Monte Carlo simulation address different sources of lepton fakes unambiguously. The MC predicts that about 0.4% and 0.4% of the total background are leptons from photon conversion and non-prompt leptons originating in hadron decays.

The improvement of the Higgs boson mass reconstruction algorithm resulted in a 1% better resolution. The Higgs boson production cross section in $H \to \tau\tau$ decays by the ATLAS was measured in the Run 2 ATLAS data set. The measurement was published in JHEP [5]. The author described her findings in two ATLAS internal notes [215], [107]. The author presented her studies of the Missing Mass Calculator in two talks at international workshops on Higgs boson physics, Higgs 2019 and Higgs 2021 [216], [217].

Another introduced analysis of the ATLAS data aimed at discovering excited tau-leptons. Excited fermions could hypothetically exist if Standard Model fermions turned out to be composite. The presented analysis searched for excited tau-leptons produced and decaying through contact interaction at the compositeness scale $\Lambda = 10$ TeV. The analysis studies the events with quark-antiquark scattering, producing an ordinary tau-lepton and an excited tau-lepton. The excited tau-lepton decay to a Standard Model tau-lepton and a quark-antiquark

pair is considered. The hadronic decay mode of both tau-leptons is analysed, and the final state of interest includes two τ_{had} and at least two jets. A set of event selections define the event phase space enriched in the excited tau-lepton production (the ETL signal region). The presented studies determined the variable sensitive to the excited tau-lepton signal. The S_{T} variable, the scalar sum of transverse momenta of tau-leptons and two leading in p_{T} jets, effectively discriminates between the ETL signal and the SM background. It also provides strong separation of BSM signals from ETLs of different hypothesised masses. MC simulation models the expected SM background, except for events with $j \to \tau_{\text{had}}$ fakes which are predicted with the data-driven method.

Further, statistical analysis estimates the expected ETL signal strength and the expected discovery significance. Maximum-likelihood fit to real data is performed in the control regions for $Z \to \tau\tau$ and Top processes. This allows for constraining $Z \to \tau\tau$ and top-quark background with real data. The fit to Asimov data is made in the signal region. The analysis has statistically significant sensitivity for excited tau-leptons of mass between 300 GeV and 1.4 TeV.

No excess of events is observed, and no sign of excited tau-leptons was found in the analysed data set. The analysed data set does not contain signs of excited tau-leptons production and decay in contact interaction. This analysis sets the upper limit on excited tau-lepton mass. Presence of excited tau-leptons lighter than 2 TeV in the analysed data set is excluded at 95% confidence level.

The author presented the search for excited tau-leptons at working groups meetings and described the analysis steps in the internal documentation [146]. Currently, the related paper is targeting the final approval in the ATLAS Collaboration.

Bibliography

[1] ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, Sep 2012.

[2] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, Sep 2012.

[3] ATLAS Collaboration. Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector. *JHEP*, 04:117, 2015.

[4] ATLAS Collaboration. Cross-section measurements of the higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the atlas detector. *Phys. Rev. D*, 99:072001, Apr 2019.

[5] ATLAS Collaboration. Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 08:175, 2022.

[6] Peter J. Mohr, David B. Newell, and Barry N. Taylor. Codata recommended values of the fundamental physical constants: 2014. *Rev. Mod. Phys.*, 88:035009, Sep 2016.

[7] "Muon g−2 Collaboration". Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr 2021.

[8] LHCb Collaboration. Measurement of *CP*-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay. *Phys. Rev. Lett.*, 125:011802. 13 p, Mar 2020.

[9] LHCb Collaboration. Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays. *Physical Review Letters*, 122(19), May 2019.

[10] LHCb Collaboration. Test of lepton universality with $b^0 \to k^0\ell^+\ell^-$ decays. *Journal of High Energy Physics*, 2017(8), Aug 2017.

[11] F. Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. 1984.

[12] J. Horejsi. *Fundamentals of Electroweak Theory*. Karolinum Press, 2002.

[13] T. D. Lee and C. N. Yang. Question of parity conservation in weak interactions. *Phys. Rev.*, 104:254–258, Oct 1956.

[14] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957.

[15] M. Goldhaber, L. Grodzins, and A. W. Sunyar. Helicity of neutrinos. *Phys. Rev.*, 109:1015–1017, Feb 1958.

[16] UA1 Collaboration. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV. *Phys. Lett. B*, 122:103–116, 1983.

[17] UA2 Collaboration. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett. B*, 122:476–485, 1983.

[18] UA1 Collaboration. Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider. *Phys. Lett. B*, 126:398–410, 1983.

[19] UA2 Collaboration. Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider. *Phys. Lett. B*, 129:130–140, 1983.

[20] CDF Collaboratio. High-precision measurement of the W boson mass with the CDF II detector. *Science*, 376(6589):170–176, 2022.

[21] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, the SLD Collaboration, the LEP Electroweak Working Group, and the SLD electroweak, heavy flavour groups. Precision electroweak measurements on the Z resonance. *Phys. Rep.*, 427:257. 302 p, 2006.

[22] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[23] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[24] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[25] J. Horejsi. *Introduction to Electroweak Unification: Standard Model from Tree Unitarity*. World Scientific, 1994.

[26] Mathematical formulation of the standard model. `https://en.wikipedia.org/wiki/Mathematical_formulation_of_the_Standard_Model`.

[27] Antonio Pich. The Standard Model of Electroweak Interactions. In *2010 European School of High Energy Physics*, pages 1–50, Jan 2012.

[28] Muon $g-2$ Collaboration. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Physical Review Letters*, 126:141801, Apr 2021.

[29] Jorg Wenninger. The LHC collider. *Comptes Rendus Physique*, 16(4):347–355, 2015.

[30] J. Reuter. Monte Carlo event generators for the LHC. `https://www.desy.de/~reuter/downloads/vancouver_09.pdf`, note = LHC Theory 2009, TRIUMF, Vancouver.

[31] P. Weber. Jet algorithms. http://www.kip.uni-heidelberg.de/atlas/seminars/SS2009_JC/jet_algorithms.pdf. LHC Physics: Experimental Potential and Detector Performance, seminars.

[32] M. Mangano. QCD and the physics of Hadronic Collisions. *CERN Yellow Rep. School Proc.*, 4:27–62. 36 p, 2018.

[33] L. Evans and P. Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001–S08001, Aug 2008.

[34] H. Wiedemann. *Particle accelerator physics*. Springer, Berlin, 2007.

[35] TOTEM Collaboration. First measurement of elastic, inelastic and total cross-section at *sqrts* = 13 TeV by TOTEM and overview of cross-section data at LHC energies, 2017.

[36] Public ATLAS Luminosity Results for Run-2 of the LHC. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. Accessed: 2022-07-29.

[37] ATLAS Collaboration. Operation of the ATLAS trigger system in Run 2. *JINST*, 15(10):P10004, 2020.

[38] P. Newman. Lecture on electromagnetic and hadronic showers and calorimeter design. http://epweb2.ph.bham.ac.uk/user/newman/appt08/lecture10.pdf.

[39] ATLAS Collaboration. The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation*, 3(08):S08003–S08003, Aug 2008.

[40] ATLAS Collaboration. ATLAS Insertable B-Layer Technical Design Report Addendum. CERN-LHCC-2012-009, May 2012.

[41] ATLAS Collaboration. Production and integration of the ATLAS insertable b-layer. *Journal of Instrumentation*, 13(05):T05008–T05008, May 2018.

[42] How atlas detects particles: diagram of particle paths in the detector. https://cds.cern.ch/images/CERN-EX-1301009-01.

[43] ATLAS Collaboration. Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2. *Eur. Phys. J. C*, 77:673, 2017.

[44] ATLAS Collaboration. The evaporative cooling system for the ATLAS inner detector. *Journal of Instrumentation*, 3(07):P07003–P07003, Jul 2008.

[45] Y. Takubo. *Operational Experience and Performance with the ATLAS Inner Detector at the Large Hadron Collider at CERN*.

[46] ATLAS Collaboration. Operation and performance of the ATLAS semiconductor tracker in LHC Run 2. *Journal of Instrumentation*, 17(01):P01013, Jan 2022.

[47] ATLAS Collaboration. ATLAS HLT Tracking: Public Results. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/HLTTrackingPublicResults.

[48] ATLAS Collaboration. The ATLAS Inner Detector Trigger performance in *pp* collisions at 13 TeV during LHC Run 2. *Eur. Phys. J. C*, 82, Jul 2021.

[49] Y. Kulchitsky. On the e/h ratio of the electromagnetic calorimeter. 2000.

[50] Y. Kulchitskii, S. Tokar, and V. Vinogradov. Study of the ATLAS Hadronic Tile Calorimeter non-compensation sensitivity. Technical report, CERN, Geneva, Mar 2006.

[51] ATLAS Collaboration. Readiness of the ATLAS liquid argon calorimeter for LHC collisions. *Eur. Phys. J. C*, 70:723, 2010.

[52] A. Milic. ATLAS LAr Calorimeter Performance in LHC Run-2. https://indico.cern.ch/event/577856/contributions/3420168/. European Physical Society Conference on High Energy Physics.

[53] ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *Journal of Instrumentation*, 14(12):P12006–P12006, dec 2019.

[54] LAr Calo Public Results 2015. https://twiki.cern.ch/twiki/pub/AtlasPublic/LArCaloPublicResults2015/h_eResolution_EMB.pdf. Accessed: 2022-08-01.

[55] Published Tile Calorimeter figures. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/PublishedTilecalFigures.

[56] Ana Maria Henriques Correia. The ATLAS Tile Calorimeter. Technical report, CERN, Geneva, Mar 2015.

[57] "J. Abdallah J, P. Adragna, and all". Mechanical construction and installation of the ATLAS tile calorimeter. *Journal of Instrumentation*, 8(11):T11001–T11001, Nov 2013.

[58] ATLAS Approved Plots: Tile PileUp Noise Performance. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTilePileUpNoiSeperformance. Pile-up Noise (2016). Accessed: 2022-08-02.

[59] ATLAS Approved Plots: Tile Energy Calibration. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTileEnergyCalibration. Cosmic muon energy deposition profile as a function of track ϕ impact point.

[60] ATLAS Approved Plots: Tile Single Particle Response. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTileSingleParticleResponse. Calorimeter response to single isolated charged hadrons, characterised by energy over momentum. Accessed: 2022-08-02.

[61] Tile Calorimeter Public Results: Timing. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TileCaloPublicResultsTiming. Accessed: 2022-08-02.

[62] H. Wilkens. The ATLAS Liquid Argon calorimeter: An overview. *Journal of Physics: Conference Series*, 160(1):012043, Apr 2009.

[63] Y. Christoforou and O. Rossetto. GaAs preamplifier and LED driver for use in cryogenic and highly irradiated environments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 425(1):347–356, 1999.

[64] P. Schacht. ATLAS Liquid Argon Endcap Calorimeter R and D for sLHC. Technical report, CERN, Geneva, Oct 2009.

[65] D. Gillberg. Performance of the ATLAS forward calorimeters in first LHC data. *Journal of Physics: Conference Series*, 293:012041, Apr 2011.

[66] ATLAS Collaboration. Commissioning of the ATLAS Muon Spectrometer with cosmic rays. *Eur. Phys. J. C*, 70:875, 2010.

[67] ATLAS Collaboration. Resolution of the ATLAS muon spectrometer monitored drift tubes in LHC Run 2. *Journal of Instrumentation*, 14(09):P09011–P09011, Sep 2019.

[68] ATLAS Collaboration. Performance of the ATLAS RPC detector and Level-1 muon barrel trigger at $\sqrt{s} = 13$ TeV. *JINST*, 16(07):P07029, 2021.

[69] A. Buckley, D. Kar, and K. Nordström. Fast simulation of detector effects in Rivet. *SciPost Physics*, 8(2), Feb 2020.

[70] A. Ribon et al. Status of Geant4 hadronic physics for the simulation of LHC experiments at the start of the LHC physics program. 2010.

[71] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *European Physical Journal C. Particles and Fields*, 70(3):823–874, 2010.

[72] T. Sjostrand, S. Mrenna, and P.Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[73] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244, 2013.

[74] 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.

[75] W. Buttinger. Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data. Technical report, CERN, Geneva, May 2015.

[76] 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(8):639, 2019.

[77] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data. *Journal of Instrumentation*, 14(03):P03017–P03017, Mar 2019.

[78] ATLAS Collaboration. High Precision Electron and Muon Reconstruction Performance with ATLAS at LHC Run-2. Jul 2022.

[79] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 76(5):292, May 2016.

[80] *Muon reconstruction performance in ATLAS at Run 2*, Trieste, 2015. SISSA. Proceedings EPS-HEP 2015.

[81] ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13$ TeV . *Eur. Phys. J. C*, 81(7):578, 2021.

[82] W. Lampl, S. Laplace, and all. Calorimeter Clustering Algorithms: Description and Performance. ATL-LARG-PUB-2008-002, Geneva, Apr 2008.

[83] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *The European Physical Journal C*, 77(7):490, Jul 2017.

[84] ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS detector. *The European Physical Journal C*, 77(7), Jul 2017.

[85] M. Cacciari, G. P. Salam, and Gr. Soyez. The anti-k_{T} jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008.

[86] ATLAS Collaboration. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev.*, D96(7):072002, 2017.

[87] ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in *pp* collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *The European Physical Journal C*, 76(11), Oct 2016.

[88] ATLAS Collaboration. Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector. *The European Physical Journal C*, 77(9), Sep 2017.

[89] ATLAS Collaboration. ATLAS *b*-jet identification performance and efficiency measurement with $t\bar{t}$ events in *pp* collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(11), Nov 2019.

[90] 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.

[91] ATLAS Collaboration. Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment. *The European Physical Journal C*, 76(5), May 2016.

[92] L. Breiman, J. Friedman, and all. *Classification and Regression Trees*. Chapman and Hall/CRC, 1984.

[93] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning Representations by Back-propagating Errors. *Nature*, 323(6088):533–536, 1986.

[94] ATLAS Collaboration. Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment. ATL-PHYS-PUB-2019-033, 2019.

[95] ATLAS Collaboration. Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using *pp* collisions at $\sqrt{s} = 13$ TeV. 2017.

[96] ATLAS Collaboration. Tau Identification and Energy Scale performance: Preliminary Public Plots. `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/TAU-2018-001/`.

[97] ATLAS Collaboration. Expected performance of missing transverse momentum reconstruction for the ATLAS detector at $\sqrt{s} = 13$ TeV. ATL-PHYS-PUB-2015-023, 2015.

[98] Dilia Maria Portillo Quintero. Impact of Missing Transverse Energy Significance in ATLAS analyses in Run2. Mar 2019.

[99] A. Tnourji. Missing Transverse energy in PbPb collisions in LHC with CMS detector. Technical report, Aug 2019.

[100] J. Alison. *The Road to Discovery: Detector Alignment, Electron Identification, Particle Misidentification, WW Physics, and the Discovery of the Higgs Boson*. PhD thesis, University of Pennsylvania, 2012.

[101] $H \to \tau^+\tau^-$ analysis group (A. Andreazza and M. K. Ayoub and et all). Measurement of the $H \to \tau^+\tau^-$ cross-section in 13 TeV Collisions with the ATLAS Detector. Technical report, CERN, Geneva, 2017.

[102] $H \to \tau^+\tau^-$ analysis group (D. Alvarez Piqueras and A. Andreazza and all). Measurement of the Higgs boson couplings in the $\tau\tau$ final state with the ATLAS detector - Supporting Note. Technical report, CERN, Geneva, 2014.

[103] A. Loginov. Strategies of data-driven estimations of $t\bar{t}$ backgrounds in ATLAS. Technical report, CERN, Geneva, 2010.

[104] W. Buttinger. Background Estimation with the ABCD Method. Technical report, Oct 2018.

[105] ATLAS Collaboration. Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV. *The European Physical Journal C*, 71(3), Mar 2011.

[106] ATLAS Collaboration. Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector. ATLAS-CONF-2014-058, 2014.

[107] $H \to \tau^+\tau^-$ analysis group (F. H. Becherer and P. Bechtle and all). Measurement of the Higgs boson coupling properties in the $H \to \tau\tau$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical report, CERN, Geneva, 2020.

[108] D. Scheirich. AAC (Athena Analysis core): a framework for data processing and event reconstruction with the ATLAS Athena software. `https://gitlab.cern.ch/scheiric/AAC`.

[109] J. Catmore. Introduction to atlas data preparation and analysis. `https://indico.cern.ch/event/472469/contributions/1982677/`. Joint Oslo/Bergen/NBI ATLAS Software Tutorial.

[110] ATLAS Collaboration. Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector. ATLAS-CONF-2015-029, 2015.

[111] M. Baak, G. J. Besjes, D. Côté, A. Koutsman, J. Lorenz, and D. Short. HistFitter software framework for statistical data analysis. *The European Physical Journal C*, 75(4), Apr 2015.

[112] Histfitter: Advanced tutorial outside atlas. `https://twiki.cern.ch/twiki/bin/view/Main/HistFitterAdvancedTutorialOutsideAtlas`.

[113] T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. ROC Analysis in Pattern Recognition.

[114] V. Pleskot and D. Scheirich. Lecture notes on statistical methods in high energy physics, 2021.

[115] K. Cranmer, G. Lewis, and all. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical report, New York U., New York, 2012.

[116] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:289–337, 1933.

[117] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011.

[118] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1):60 – 62, 1938.

[119] ATLAS Collaboration. Observation of an excess of events in the search for the Standard Model Higgs boson in the $\gamma\gamma$ channel with the ATLAS detector. ATLAS-CONF-2012-091, 2012.

[120] LEP Collaboration. Search for the Standard Model Higgs boson at LEP. *Physics Letters B*, 565:61–75, Jul 2003.

[121] CMS Collaboration. Search for an excited lepton that decays via a contact interaction to a lepton and two jets in proton-proton collisions at $\sqrt{s} =$ 13 TeV. *Journal of High Energy Physics*, 2020(5), May 2020.

[122] T. Junk. Confidence level computation for combining searches with small statistics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 434(2-3):435–443, Sep 1999.

[123] Alexander L. Read. Presentation of search results: the CL_S technique. *J. Phys. G*, 28:2693, 2002.

[124] ALEPH Collaboration, DELPHI Collaboration, et al. Search for the Standard Model Higgs Boson at LEP, 2001.

[125] Particle Data Group (P.A. Zyla et al.). Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[126] ATLAS Collaboration. Searches for lepton-flavour-violating decays of the higgs boson in $\sqrt{s} = 13$ TeV pp collisions with the atlas detector. *Physics Letters B*, 800:135069, 2020.

[127] R.K. Ellis, I. Hinchliffe, and all. Higgs decay to $\tau^+\tau^-$ possible signature of intermediate mass higgs bosons at high energy hadron colliders. *Nuclear Physics B*, 297(2):221–243, 1988.

[128] A. Elagin, P. Murat, A. Pranko, and A. Safonov. A new mass reconstruction technique for resonances decaying to $\tau^+\tau^-$. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 654(1):481 – 489, 2011.

[129] M. Schumacher. Higgs boson production via vector boson fusion in atlas. `https://indico.cern.ch/event/122468/contributions/1337597/`. IOP Half day meeting on Vector Boson Fusion.

[130] Michel Trottier-McDonald. My first recorded seminar at LBNL. `https://licollider.wordpress.com/2013/12/14/my-first-recorded-seminar-at-lbnl/`.

[131] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1 – 29, 2012.

[132] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30 – 61, 2012.

[133] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[134] D. de Florian and C. Grojean and all. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. CERN, Geneva, Oct 2016.

[135] LHC Higgs Cross Section Working Group (J. R. Andersen and all). Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. Jul 2013.

[136] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132 – 133, 1964.

[137] P. W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966.

[138] ATLAS Collaboration. Evidence of off-shell higgs boson production and constraints on the total width of the higgs boson in the the $ZZ \to 4l$ and $ZZ \to 2l2\mu$ decay channels with the atlas detector. `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2022-068/`.

[139] A. Bohenic. *Kinematics of Tau Lepton Decays*. PhD thesis, University of Colorado, 2019.

[140] ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the atlas detector using proton–proton collisions at $\sqrt{s}$=13 tev. *The European Physical Journal C*, 78(11):903, Nov 2018.

[141] S. Giagu. WIMP Dark Matter Searches With the ATLAS Detector at the LHC. *Frontiers in Physics*, 7:75, 2019.

[142] CDF Collaboration. Search for anomalous production of events with two photons and additional energetic objects at cdf. *Phys. Rev. D*, 82:052005, Sep 2010.

[143] A. De Maria. Impact of pileup jets to the mass estimation in the couplings analysis. `https://indico.cern.ch/event/856192/contributions/3605830/attachments/1927091/3190526/Slides_HLept_1016.pdf`. HLeptons ATLAS meeting.

[144] ATLAS Collaboration. Object-based missing transverse momentum significance in the ATLAS detector.

[145] Monte Carlo Particle Numbering Scheme. `http://pdg.lbl.gov/current/mc-particle-id`. Accessed: 2022-05-19.

[146] V. Pleskot, D. Scheirich., K. Mihule, and T. Davídek. Search for excited tau-leptons at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical report, CERN, Geneva, 2020.

[147] J. Catmore. A bit about MC truth: Monte Carlo Truth in the xAOD. `https://indico.cern.ch/event/472469/contributions/1982685/attachments/1222751/1789718/truth_tutorial.pdf`, note = Joint Oslo/Bergen/NBI ATLAS Software Tutorial.

[148] M. H˙ `https://bonndoc.ulb.uni-bonn.de/xmlui/bitstream/handle/20.500.11811/9734/6567.pdf?sequence=1&isAllowed=y`.

[149] I. A. D'Souza and C. S. Kalman. *Preons: Models of Leptons, Quarks and Gauge Bosons as Composite Objects*. 1992.

[150] M. Kłusek-Gawenda, A. Szczurek, and all. Limits for anomalous magnetic and electric dipole moments of tau leptons from heavy-ion UPCs. *PoS*, EPS-HEP2021:275, 2022.

[151] CMS Collaboration. Observation of τ-lepton pair production in ultraperipheral lead-lead collisions at $\sqrt{s_{\mathrm{nn}}} = 5.02$ TeV, 2022.

[152] BaBar Collaboration. Measurement of an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays and implications for charged Higgs bosons. *Phys. Rev. D*, 88(7):072012, 2013.

[153] LHCb Collaboration. Measurement of the Ratio of Branching Fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$. *Phys. Rev. Lett.*, 115(11):111803, 2015.

[154] Belle Collaboration. Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle. *Phys. Rev. D*, 92(7):072014, 2015.

[155] Belle Collaboration. Measurement of the branching ratio of $\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method. *Phys. Rev. D*, 94(7):072007, 2016.

[156] Belle Collaboration. Measurement of the τ Lepton Polarization and $R(D^*)$ in the Decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$. *Phys. Rev. Lett.*, 118(21):211801, 2017.

[157] P. Stangl. Lepton flavour universality violation from composite muons. *PoS*, EPS-HEP2015:555, 2015.

[158] U. Baur, M. Spira, and P. M. Zerwas. Excited-quark and -lepton production at hadron colliders. *Phys. Rev. D*, 42:815–824, Aug 1990.

[159] E. N. Argyres, U. Baur, and all. Compositeness. 1990.

[160] ATLAS Collaboration. Search for excited electrons singly produced in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment at the LHC. *The European Physical Journal C*, 79(9), Sep 2019.

[161] CMS Collaboration. Search for excited leptons decaying via contact interaction to two leptons and two jets. Technical report, CERN, Geneva, 2019.

[162] ALEPH Collaboration. Search for evidence of compositeness at LEP I. *Eur. Phys. J. C*, 4:571–590. 36 p, Feb 1998.

[163] Determination of the $e^-e^+ \to \gamma\gamma(\gamma)$ cross-section at LEP 2. *The European Physical Journal C*, 37(4):405–419, Nov 2004.

[164] L3 Collaboration. Search for excited leptons at LEP. *Physics Letters B*, 568(1-2):23–34, Aug 2003.

[165] OPAL Collaboration. *Physics Letters B*, 602(3-4):167–179, nov 2004.

[166] H1 Collaboration. Search for excited electrons in ep collisions at HERA. *Physics Letters B*, 666(2):131–139, Aug 2008.

[167] ZEUS Collaboration. Searches for excited fermions in ep collisions at HERA. *Physics Letters B*, 549(1-2):32–47, Nov 2002.

[168] CDF Collaboration. Search for new particles decaying into dijets in proton-antiproton collisions at sqrts=1.96 tev. *Physical Review D*, 79(11), jun 2009.

[169] D0 Collaboration. Measurement of Dijet Angular Distributions at $\sqrt{s}$ = 1.96 TeV and Searches for Quark Compositeness and Extra Spatial Dimensions. *Physical Review Letters*, 103(19), Nov 2009.

[170] ATLAS Collaboration. Search for new phenomena in dijet events using $37\,\mathrm{fb}^{-1}$ of pp collision data collected at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector. *Phys. Rev. D*, 96:052004, 2017.

[171] ATLAS Collaboration. Search for new resonances in mass distributions of jet pairs using $139\,\mathrm{fb}^{-1}$ of pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector. *JHEP*, 03:145, 2020.

[172] ATLAS Collaboration. Search for new phenomena with photon+jet events in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector. *JHEP*, 03:041, 2016.

[173] ATLAS Collaboration. Search for new phenomena in high-mass final states with a photon and a jet from pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector. *Eur. Phys. J. C*, 78:102, 2018.

[174] ATLAS Collaboration. Search for resonances in the mass distribution of jet pairs with one or two jets identified as b-jets in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector. *Phys. Rev. D*, 98:032016, 2018.

[175] ATLAS Collaboration. Search for resonances in the mass distribution of jet pairs with one or two jets identified as b-jets in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector. *Phys. Lett. B*, 759:229, 2016.

[176] ATLAS Collaboration. Search for the production of single vector-like and excited quarks in the Wt final state in pp collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ with the ATLAS detector. *JHEP*, 02:110, 2016.

[177] ATLAS Collaboration. Search for single b^*-quark production with the ATLAS detector at $\sqrt{s} = 7\,\mathrm{TeV}$. *Phys. Lett. B*, 721:171, 2013.

[178] ATLAS Collaboration. Search for excited electrons and muons in $\sqrt{s}$ = 8 TeV proton–proton collisions with the ATLAS detector. *New J. Phys.*, 15:093011, 2013.

[179] ATLAS Collaboration. Search for new phenomena in events with three or more charged leptons in pp collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ with the ATLAS detector. *JHEP*, 08:138, 2015.

[180] ATLAS Collaboration. A search for an excited muon decaying to a muon and two jets in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector. *New J. Phys.*, 18:073021, 2016.

[181] ATLAS Collaboration. Search for excited electrons singly produced in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS experiment at the LHC. *Eur. Phys. J. C*, 79:803, 2019.

[182] CMS Collaboration. Search for narrow resonances using the dijet mass spectrum in pp collisions at $\sqrt{s} = 8\,\text{TeV}$. *Phys. Rev. D*, 87:114015, 2013.

[183] CMS Collaboration. Search for resonances and quantum black holes using dijet mass spectra in proton–proton collisions at $\sqrt{s} = 8\,\text{TeV}$. *Phys. Rev. D*, 91:052009, 2015.

[184] CMS Collaboration. Search for narrow and broad dijet resonances in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ and constraints on dark matter mediators and other new particles. *JHEP*, 08:130, 2018.

[185] CMS Collaboration. Search for high mass dijet resonances with a new background prediction method in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *JHEP*, 05:033, 2020.

[186] CMS Collaboration. Search for excited quarks in the γ+jet final state in proton–proton collisions at $\sqrt{s} = 8\,\text{TeV}$. *Phys. Lett. B*, 738:274, 2014.

[187] CMS Collaboration. Search for excited quarks of light and heavy flavor in $\gamma +$ jet final states in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Phys. Lett. B*, 781:390, 2018.

[188] CMS Collaboration. Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s} = 8\,\text{TeV}$. *JHEP*, 08:173, 2014.

[189] CMS Collaboration. Search for massive resonances decaying into WW, WZ, ZZ, qW, and qZ with dijet final states at $\sqrt{s} = 13\,\text{TeV}$. *Phys. Rev. D*, 97:072006, 2018.

[190] CMS Collaboration. Search for the production of an excited bottom quark decaying to tW in proton–proton collisions at $\sqrt{s} = 8\,\text{TeV}$. *JHEP*, 01:166, 2016.

[191] CMS Collaboration. Search for excited leptons in pp collisions at $\sqrt{s} = 7\,\text{TeV}$. *Phys. Lett. B*, 720:309, 2013.

[192] CMS Collaboration. Search for excited leptons in proton–proton collisions at $\sqrt{s} = 8\,\text{TeV}$. *JHEP*, 03:125, 2016.

[193] CMS Collaboration. Search for excited leptons in $\ell\ell\gamma$ final states in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *JHEP*, 04:015, 2019.

[194] CMS Collaboration. Search for an excited lepton that decays via a contact interaction to a lepton and two jets in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *JHEP*, 05:052, 2020.

[195] ATLAS collaboration. Search for new phenomena in events with three or more charged leptons in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 2015(8), Aug 2015.

[196] ATLAS Collaboration. Search for new non-resonant phenomena in high-mass dilepton final states with the ATLAS detector. *JHEP*, 11:005, 2020.

[197] CMS Collaboration. Search for contact interactions and large extra dimensions in the dilepton mass spectra from proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 04:114, 2019.

[198] CMS Collaboration. Search for new physics in dijet angular distributions using proton–proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter and other models. *Eur. Phys. J. C*, 78(9):789, 2018.

[199] ATLAS Collaboration. Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector. *Phys. Rev. D*, 96(5):052004, 2017.

[200] Particle Data Group (P.A. Zyla et al.). Searches for Quark and Lepton Compositeness. `https://pdg.lbl.gov/2021/reviews/rpp2021-rev-searches-quark-lep-compositeness.pdf`. PDG updated: 2021, revised: October 2021, accessed: 2022-04-12.

[201] M. Presilla. Unitarity in exotic searches. `https://indico.cern.ch/event/952140/contributions/3999991/attachments/2097893/3526444/unitarity_ATLAS-LPX_7Sept2020.pdf`. Talk at the LPX subgroup meeting on September 7, 2020.

[202] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152, 2001.

[203] D. Scheirich. RAC (ROOR Analysis Core): software for physics analyses with the ROOT package. `https://gitlab.cern.ch/scheiric/RAC`.

[204] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[205] V. Pleskot. PyRooPlotting: an automated tool for plotting of paper quality plots compatible with the official ATLAS style. `https://gitlab.cern.ch/pleskot/PyRooPlotting`.

[206] ATLAS Collaboration. Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data. Technical report, CERN, Geneva, 2022.

[207] ATLAS Collaboration. Reconstruction and identification of boosted di-τ systems in a search for Higgs boson pairs using 13 TeV proton-proton collision data in ATLAS. *JHEP*, 163, 2020.

[208] Public Tau Trigger Plots for Collision Data. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TauTriggerPublicResults`. Efficiency of τ-lepton triggers as a function of the offline τ-lepton transverse momentum pT. Accessed: 2022-08-04.

[209] V. Pleskot and all. Fake Tau Task Force: supporting documentation. `https://gitlab.cern.ch/atlas-perf-tau/fake-tau/-/tree/master/Documentation/support-note-rel21`.

[210] A. Buckley, J. Ferrando, and all. LHAPDF6: parton density access in the LHC precision era. *The European Physical Journal C*, 75(3), Mar 2015.

[211] J. Butterworth, G. Dissertori, and all. Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group Report. 2014.

[212] D. Scheirich. YAW (Yet Another Wrapper): a tool for statistical analysis in experimental particle physics. `https://gitlab.cern.ch/scheiric/YAW`.

[213] ATLAS Collaboration. The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS. *JINST*, 13(07):P07017, 2018.

[214] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} =$ 13 TeV using the ATLAS detector at the LHC. Technical report, CERN, Geneva, Jun 2019.

[215] K. Petukhova, T. Davidek, M. Huebner, and P. Bechtle. The status of Missing Mass Calculator for Higgs boson mass estimation in the ATLAS $H \to \tau\tau$ analysis. Technical report, CERN, Geneva, 2019.

[216] K. Petukhova. The status of Missing Mass Calculator for Higgs boson mass estimation in the ATLAS $H \to \tau\tau$ analysis, 2021. Higgs Couplings 2019.

[217] K. Mihule. Measurements of Higgs boson production cross sections in the $H \to \tau\tau$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, 2021. Higgs 2021.

[218] S. Ordek. Performance of the ATLAS tau-lepton trigger at the LHC in Run 2. *PoS*, EPS-HEP2019:205, 2020.

[219] ATLAS Collaboration. Search for excited τ-leptons and leptoquarks in the final state with τ-leptons and jets in pp collisions at $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. The paper is not published yet. It is currently in ATLAS approval process and targets publishing in Journal of High Energy Physics.

[220] ATLAS Collaboration. Operation and Performance of the ATLAS Tile Calorimeter in LHC Run-2. The paper is not published yet. It is currently in ATLAS approval process and targets publishing in The European Physical Journal C.

[221] ATLAS Collaboration. Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. Technical report, CERN, Geneva, 2021.

[222] K. Mihule. Time calibration, monitoring and performance of the atlas tile calorimeter in run 2. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1048:167942, 2023.

[223] K. Petukhova. Calibration and Performance of the ATLAS Tile Calorimeter During the LHC Run 2. *PoS*, LeptonPhoton2019, 2019.

[224] K. Petukhova. Upgrade of the ATLAS Tile Calorimeter for the High Luminosity LHC. *PoS*, LeptonPhoton2019, 2019.

[225] K. Petukhova. Calibration and performance of the ATLAS tile calorimeter. *Journal of Physics: Conference Series*, 1390(1):012107, Nov 2019.

[226] ATLAS Collaboration. The ATLAS Tau Trigger in Run 2. ATLAS-CONF-2017-061, 2017.

[227] ATLAS Collaboration. Performance of the ATLAS Level-1 topological trigger in Run 2. *Eur. Phys. J. C*, 82(1):7, 2022.

[228] 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.

[229] ATLAS Collaboration. Trigger Menu in 2017. ATL-DAQ-PUB-2018-002, 2018.

List of Figures

List of Tables

List of Abbreviations

SM - Standard Model
BSM - Beyond the Standard Model
QCD - Quantum Chromodynamics
EWK - electroweak
BEH - R. Brout, F.F. Englert, and P. Higgs (theory)
GUT - Grand Unification Theory
CP - charge-parity
ETL - excited tau-lepton
CERN - European Organisation for Nuclear Research (abbreviated from French name “Conseil Européen pour la Recherche Nucléaire”)
LHC - Large Hadron Collider
HEP - high energy physics
ATLAS - A Toroidal LHC ApparatuS
CMS - Compact Muon Solenoid
ALICE - A Large Ion Collider Experiment
LHCb - Large Hadron Collider beauty (experiment)
TOTEM - TOTal Elastic and diffractive cross section Measurement
LHCf - Large Hadron Collider forward (experiment)
MoEDAL - Monopole and Exotics Detector at the LHC (experiment)
Linac - linear accelerator
pp - proton-proton
TeV - tera-electron volts (1 GeV = 10^{12} eV)
GeV - giga-electron volts (1 GeV = 10^{9} eV)
MeV - mega-electron volts (1 MeV = 10^{6} eV)
keV - kilo-electron volts (1 keV = 10^{3} eV)
eV - electron volt
PDF - parton density function
PDF - probability density function
L1 - Level 1 (applicable to trigger)
HLT - High Level Trigger
EM - electromagnetic
MIP - main interaction point
IBL - Insertable B-Layer
SCT - Semiconductor Tracker
TRT - Transition Radiation Tracker
LAr - Liquid Argon
FCal - Forward Calorimeter
TileCal - Tile Calorimeter
MS - muon spectrometer
EC - End-Cap
LUCID - LUminosity Cherenkov Integrating Detector
ZDC - Zero Degree Calorimeter
ALPHA - Absolute Luminosity For ATLAS
AFP - Atlas Forward Proton
DAQ - Data Acquisition

MC - Monte Carlo
LO -leading order
NLO - next-to-leading order
PMG - Particle Management Group
RDO - real detector output
OLR - overlap removal
MET - missing transverse energy
ID - identification
WP - working point
SF - scale factor
PFlow - Particle Flow
JVT - Jet Vertex Tagger
fJVT - forward Jet Vertex Tagger
MVA - multivariate analysis
BDT - Boosted Decision Tree
RNN - Recurrent Neural Networks
PFO - particle flow object
TST- Track-based Soft Term
CST - Calorimeter Soft Term
FF - Fake Factor
MM - the Matrix Method
GRL - Good Run List
SR - signal region
CR - control region
POI - parameter of interest
ROC - receiver operating characteristics
AUC - area under curve (applicable to the ROC)
NLL - negative logarithm of likelihood
FF - Fake Factor (method)
MM - Matrix Method
AAC - Athena Analysis Common (framework)
RAC - ROOT Analysis Common (framework)
YAW - Yeat Another Wrapper (framework)
NP - nuisance parameter
POI - parameter of interest
AOD - analysis object data
DAOD - derivation analysis object data
CL - confidence level
LEP - Large Electron Positron (collider)
MMC - Missing Mass Calculator
CMA - collinear mass approximation
CDF - Collider Detector at Fermilab
ggH - gluon-gluon fusion
VBF - vector boson fusion
RMS - root mean square
PDG ID - Particle Data Group Identification number
CPU - central processing unit
TF - transfer factor

SS - same sign
OS - opposite sign
TES - tau-lepton energy scale
UL - upper limit

List of publications

I have been participating in several analyses at the ATLAS experiment and contributing to the ATLAS detector calibration and operation. This work resulted in the following publications:

Papers:

1. ATLAS Collaboration. Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 08:175, 2022

2. ATLAS Collaboration. Search for excited τ-leptons and leptoquarks in the final state with τ-leptons and jets in pp collisions at $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. The paper is not published yet. It is currently in ATLAS approval process and targets publishing in Journal of High Energy Physics

3. ATLAS Collaboration. Operation and Performance of the ATLAS Tile Calorimeter in LHC Run-2. The paper is not published yet. It is currently in ATLAS approval process and targets publishing in The European Physical Journal C

Conference notes:

1. ATLAS Collaboration. Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. Technical report, CERN, Geneva, 2021

Conference proceedings:

1. K. Mihule. Time calibration, monitoring and performance of the atlas tile calorimeter in run 2. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1048:167942, 2023

2. K. Petukhova. Calibration and Performance of the ATLAS Tile Calorimeter During the LHC Run 2. *PoS*, LeptonPhoton2019, 2019

3. K. Petukhova. Upgrade of the ATLAS Tile Calorimeter for the High Luminosity LHC. *PoS*, LeptonPhoton2019, 2019

4. K. Petukhova. Calibration and performance of the ATLAS tile calorimeter. *Journal of Physics: Conference Series*, 1390(1):012107, Nov 2019

Being an member of the ATLAS Collaboration, I am a co-author of about 200 publications.

A. Di-tau trigger in events with $\tau_{\text{had}}\tau_{\text{had}}$

Events should be accepted by a di-tau trigger [226], which narrows down the all the massive to events to the events of interest. In studies of the Higgs boson decay in a pair of tau-leptons, the presence of at least one jet is required in order to exclude topologies with extremely boosted di-taus which require a specific reconstruction due to tau-leptons overlap.

Table A.1 cites the names of the triggers used in Run 2 data set analysis.

First, tau-leptons and jets are promptly constructed as Level 1 topological clusters and calorimeter inputs [227]. To be further considered, the tau-lepton candidates should pass the minimal thresholds on p_{T}: $p_{\text{T}}^{\tau 0,L1} > 20$ GeV, $p_{\text{T}}^{\tau 1,\text{L1}} > 12$ GeV. Then, the L1Topo and L1Calo become seeds from which the corresponding High Level Trigger (HLT) can identify di-tau. The lowest ATLAS unprescaled triggers are employed to maximise an event yield and to include low-p_{T} tau-lepton for a proper background assessment.

The di-tau triggers are used for event selection in the considered search for excited tau-lepton as well as the $H \to \tau\tau$ measurement. Application of the di-tau triggers allows for a coherent definition of both signal and control regions. In particular, di-tau triggers keep a maximum of $Z \to \tau\tau$ events, making it easier to validate the MC modelling of this background process.

A given event is accepted if trigger-level (so called online) tau-leptons reach minimal criteria on their transverse momenta: 35 GeV and 25 GeV for leading and subleading tau-lepton, respectively. The threshold corresponds to the τ_{had} p_{T} values where the trigger efficiency reaches a plateau [228]. The tau RNN-based HLT trigger provides an efficiency $\geq 80\%$ for one-prong (three-prong) tau-leptons with $p_{\text{T}} > 40$ GeV ($p_{\text{T}} > 60$ GeV) [208]. The tau identification scores are promptly evaluated, and both tau candidates are required to pass Medium ID (identification) requirements. The Medium criteria are applied to the tau isolation. An event should contain a signature of neutrinos: the missing transverse energy above 12 GeV is expected. In 2017, an additional condition of the di-tau opening angle $\Delta R_{\tau\tau} < 2.8$ was introduced at the level of the related Level 1 trigger.

Due to increased instantaneous luminosity in 2016, di-tau trigger requirements were tighten to accord the limitations of CPU and data size. A presence of at least one L1Calo jet with transverse momenta above 25 GeV and $|\eta| < 3.2$ ($|\eta| < 2.3$ in 2018) was required.

The recorded reconstructed tau-leptons (so called offline tau-leptons) should be geometrically matched to the online objects accepted by di-tau trigger.For the excited tau-lepton search, the trigger matching is was implemented by the Athena `MatchFromCompositeTool` with a matching parameter $\Delta R < 0.4$. Trigger matching was done with the `TriggerDecisionTool` in the $H \to \tau\tau$ coupling measurement.

The lowest unprescaled di-tau trigger fires at very high rates (peaking at $\sim$ 61 Hz [229]) due to multi-jet events production cross section raising with increased instantaneous luminosity. To reduce QCD jet background and keep the

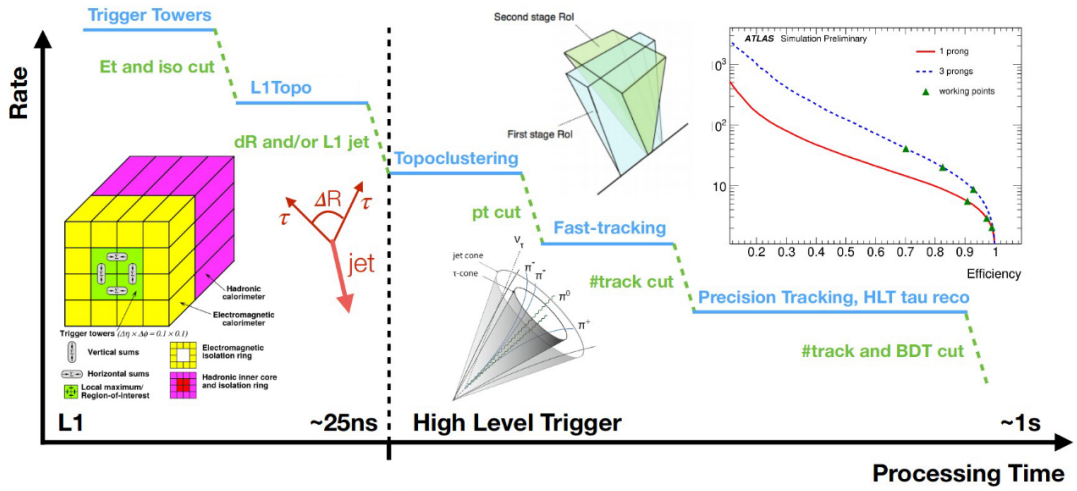


Figure A.1: A schematic representation of tau-lepton triggering during Run 2 ATLAS data-taking [218].

year	di-tau trigger name
2015	HLT_tau35_medium1_tracktwo_tau25_medium1_tracktwo_L1TAU20IM_2TAU12IM
2016	HLT_tau35_medium1_tracktwo_tau25_medium1_tracktwo
2017	HLT_tau35_medium1_tracktwo_tau25_medium1_tracktwo_03dR30_L1DR-TAU20ITAU12I-J25
2018	HLT_tau35_medium1_tracktwoEF_tau25_medium1_tracktwoEF_L1DR-TAU20ITAU12I-J25

Table A.1: The summary of the di-tau triggers used to select events of interest in the Run 2 ATLAS data set.

trigger acceptance manageable by the readout systems, the minimal ΔR criterion on the isolated Level 1 tau objects is imposed: $\Delta R_{\tau\tau} > 0.6$. This condition removes ambiguities between the core parts of the regions of interest for two tau-lepton candidates and keeps the HLT trigger rate at 24-39 Hz [226].

B. Background Monte Carlo modelling

V + jets. QCD production of V + jets is simulated with `Sherpa v2.1.1` [63] Monte Carlo generator. Matrix elements (ME) are calculated with accuracy of next-to-leading order (NLO) for processes with emission of up to two partons (0, 1, 2j@NLO) and accuracy of leading order (LO) for processes involving emission of up to four partons (3, 4j@LO). Parton hard-scatter processes are matched with the parton showering according to the MEPS@NLO prescription [69–72] following the set of tuned parameters. The `NNPDF3.0nnlo` set of PDFs [57] is used. The samples are normalised to a next-to-next-to-leading order (NNLO) prediction [73].

The $Z \to \tau\tau$ process has the final state identical to the final state with considered excited tau-lepton production and Higgs boson production and decay to a pair of tau-leptons. It is a main irreducible background for Higgs boson coupling measurement. For modelling $Z \to \tau\tau$ process, filtered samples in low p_{T}^{Z} region are used to increase event statistics.

Dijet and di-lepton electroweak production. Electroweak production of with dijet and two-lepton final state ($lljj$, $l\nu jj$, $\nu\nu jj$) in VBF is simulated with `Sherpa v2.1.1` [63] Monte Carlo generator. LO matrix elements with up to two additional partons emitted are evaluated and then merged with the `Sherpa` parton shower in the way it is done for V + jets samples. The nominal set of PDFs is `NNPDF3.0NNLO`.

Top quark pair production. The production of a top quark pair is simulated with `PowhegBox v2` [14–16, 55] generator with NLO ME accuracy. The `NNPDF3.0NLO` [57] PDF set is employed. The resummation damping parameter h_{damp} is set to 1.5 of top quark mass [74]. Parton showering, hadronisation, underlying events (UE) are modelled by interfacing the generated hard-scatter process to `Pythia v8.230` [23]. The `Pythia` generator with the A14 tune for UE and the `NNPDF2.3lo` PDF set is used [75]. Heavy-flavour quark decay is simulated with the `EvtGen v1.6.0`. Due to lower statistics in high-p_{T} tail, exclusive modelling $t\bar{t}$ in the high-$H_{\mathrm{T}} := p_{\mathrm{T}}^{j0} + p_{\mathrm{T}}^{j1}$ is considered in addition to inclusive top quark pair samples.

Single top quark production. Single top quark production with s-, t- channel as well as the associated production with W (denoted tW) are considered. These processes are simulated with `Powheg-Box v2` [14–16, 77] generator with NLO accuracy of strong coupling constant α_{S}. Five-flavour scheme is used in the `NNPDF3.0nlo` PDF set. The generated events are interfaced to `Pythia v8.230` [23] using the `A14` tune [58] and the `NNPDF2.3lo` PDF.

Di-boson. Diboson final state is simulated with the `Sherpa` v2.2.1 or v2.2.2 [63]. MEs are calculated with NLO QCD accuracy for up to one additional parton and with LO accuracy for up to three additional parton emissions. The

MEs are matched and merged with the `Sherpa` parton shower based on Catani-Seymour dipole factorisation [64, 68] using the MEPS@NLO prescription [69–72]. The NNPDF3.0NNLO set of PDFs, the set of parton shower tuning parameters, and the virtual QCD corrections [65–67] are employed.

C. Event selection in the $H \to \tau\tau$ analysis with the partial (2015-16) Run 2 ATLAS data set

	$\tau_{\text{lep}}\tau_{\text{lep}}$		$\tau_{\text{lep}}\tau_{\text{had}}$		$\tau_{\text{had}}\tau_{\text{had}}$
	$e\mu$	$ee/\mu\mu$	$e\tau_{\text{had}}$	$\mu\tau_{\text{had}}$	
Pre-selection	# of $e/\mu = 2$, # of $\tau_{\text{had-vis}} = 0$		# of $e/\mu = 1$, # of $\tau_{\text{had-vis}} = 1$		# of $e/\mu = 0$, # of $\tau_{\text{had-vis}} = 2$
	e/μ : p_{T} cut 10–25 GeV, e/μ : Medium, gradient iso.		e/μ : p_{T} cut 21–27.3 GeV, e/μ : Medium, gradient iso.		
			$\tau_{\text{had-vis}}$: $p_{\text{T}} > 30$ GeV, Medium		$\tau_{\text{had-vis}}$: $p_{\text{T}} > 40, 30$ GeV, $\tau_{\text{had-vis}}$: Tight
			1-prong $\tau_{\text{had-vis}}$: eleBDT e-veto		
	Opposite charge		Opposite charge		Opposite charge
(CBA and MVA)	$m_{\tau\tau}^{\text{coll}} > m_Z - 25$ GeV		$m_{\text{T}} < 70$ GeV		
	$30 < m_{e\mu} < 100$ GeV	$30 < m_{\ell\ell} < 75$ GeV			
	# of b-jets = 0		# of b-jets = 0		
					$0.8 < \Delta R_{\tau_{\text{had-vis}}\tau_{\text{had-vis}}} < 2.5$
					$\lvert\Delta\eta_{\tau_{\text{had-vis}}\tau_{\text{had-vis}}}\rvert < 1.5$
	$E_{\text{T}}^{\text{miss}} > 20$ GeV	$E_{\text{T}}^{\text{miss}} > 55$ GeV, $E_{\text{T,HPTO}}^{\text{miss}} > 55$ GeV			$E_{\text{T}}^{\text{miss}} > 20$ GeV
	leading jet $p_{\text{T}} > 40$ GeV		leading jet $p_{\text{T}} > 40$ GeV		leading jet $p_{\text{T}} > 70$ GeV, $\lvert\eta\rvert < 3.2$
Additional cuts (CBA only)	$\Delta R_{\ell\ell} < 2.0$, $\lvert\Delta\eta_{\ell\ell}\rvert < 1.5$		$\Delta R_{\ell\tau_{\text{had-vis}}} < 2.5$, $\lvert\Delta\eta_{\ell\tau_{\text{had-vis}}}\rvert < 1.5$		
			$E_{\text{T}}^{\text{miss}} > 20$ GeV		
	$0.1 < x_1 < 1.0$, $0.1 < x_2 < 1.0$		$0.1 < x_1 < 1.4$, $0.1 < x_2 < 1.2$		$0.1 < x_1 < 1.4$, $0.1 < x_2 < 1.4$
VBF inclusive (CBA and MVA)	sub-leading jet $p_{\text{T}} > 30$ GeV, $m_{jj} > 400$ GeV, $\lvert\Delta\eta_{jj}\rvert > 3$				
VBF CBA (CBA only)	$\eta_{j_1} \times \eta_{j_2} < 0$, ℓ and $\tau_{\text{had-vis}}$ centrality between jets				
Signal regions (CBA only)					high-p_{T}: $p_{\text{T}}^H > 140$ GeV, $\Delta R_{\tau_{\text{had-vis}}\tau_{\text{had-vis}}} < 1.5$
	Tight : $m_{jj} > 800$ GeV		Tight : $p_{\text{T}}^H > 100$ GeV, $m_{jj} > 500$ GeV		Tight : not high-p_{T}, $m_{jj} > (1550 - 250 \times \lvert\Delta\eta_{jj}\rvert)$ GeV
	Loose : otherwise		Loose : otherwise		Loose : otherwise
Boosted inclusive (CBA and MVA)	Not VBF cut, $p_{\text{T}}^H > 100$ GeV				
Boosted CBA (CBA only)	Not VBF additional cut				
Signal regions (CBA only)	Tight : $p_{\text{T}}^H > 140$ GeV, $\Delta R_{\ell\ell} < 1.5$		Tight : $p_{\text{T}}^H > 140$ GeV, $\Delta R_{\ell\tau_{\text{had-vis}}} < 1.5$		Tight : $p_{\text{T}}^H > 140$ GeV, $\Delta R_{\tau_{\text{had-vis}}\tau_{\text{had-vis}}} < 1.5$
	Loose : otherwise		Loose : otherwise		Loose : otherwise

Figure C.1: The event selection in the $H \to \tau\tau$ cross section measurement with the partial Run 2 data set (collected in 2015-16 and corresponding to the integrated luminosity of 36.1 fb^{-1}). Source [101].

There are different p_{T} selection applied, which reflects changes in the trigger settings between years. Another difference comes from the light lepton isolation WP (the `Gradient` one in the 2015-16 analysis in contract to the `Loose` and `Tight` WP applied to electron and muons, respectively, in the 2015-18 analysis). In the `Athena` release 20.7, the `Gradient` isolation criterion required absence of high-p_{T}

tracks in a cone around the object track and no significant energy deposits in a cone around the calorimeter clusters of the given object after correcting for pileup. The `Gradient` WP has an efficiency of 90% (99%) for a real lepton with p_{T} around 60 GeV (25 GeV). More event categories are supported in the $H \to \tau\tau$ analysis of the 2015-16 data set with `Athena` r. 20.7. The $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ channel included three flavour combinations there: $\tau_e\tau_e$, $\tau_\mu\tau_\mu$, and $\tau_e\tau_\mu$. The $H \to \tau\tau$ measurement with the 2015-18 data explored a single $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ decay mode – $\tau_e\tau_\mu$.

D. Auxiliary plots to studies of the MMC performance

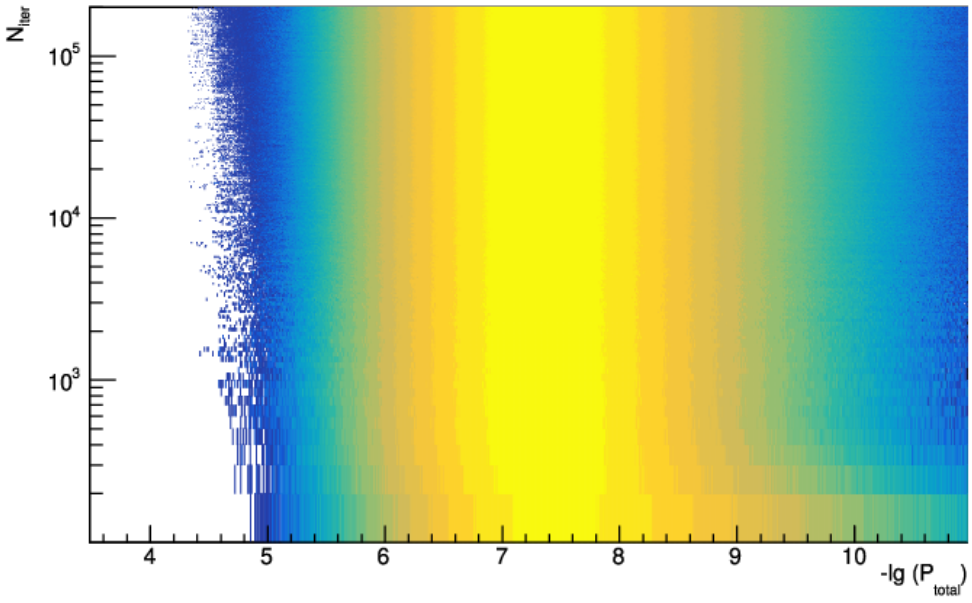


Figure D.1: The total PDF in the current configuration vs. the number of the current iteration in the MMC Markov chain. The scan over the MET values in the $E_{\mathrm{T}}^{\mathrm{miss}} \pm 4\sigma_{\mathrm{MET}}$ range is performed in the $gg \to H \to \tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ events.

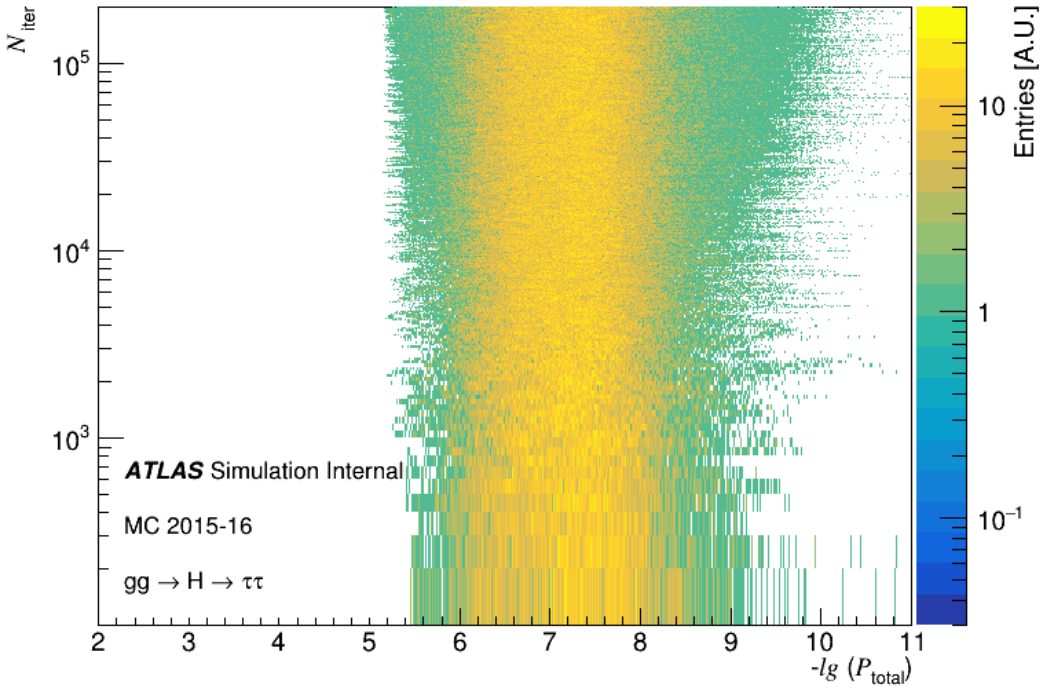


Figure D.2: The total PDF in the current configuration vs. the number of the current iteration in the MMC Markov chain. The scan over the MET values in the $E_{\mathrm{T}}^{\mathrm{miss}} \pm 2\sigma_{\mathrm{MET}}$ range is performed in the $gg \to H \to \tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ events.

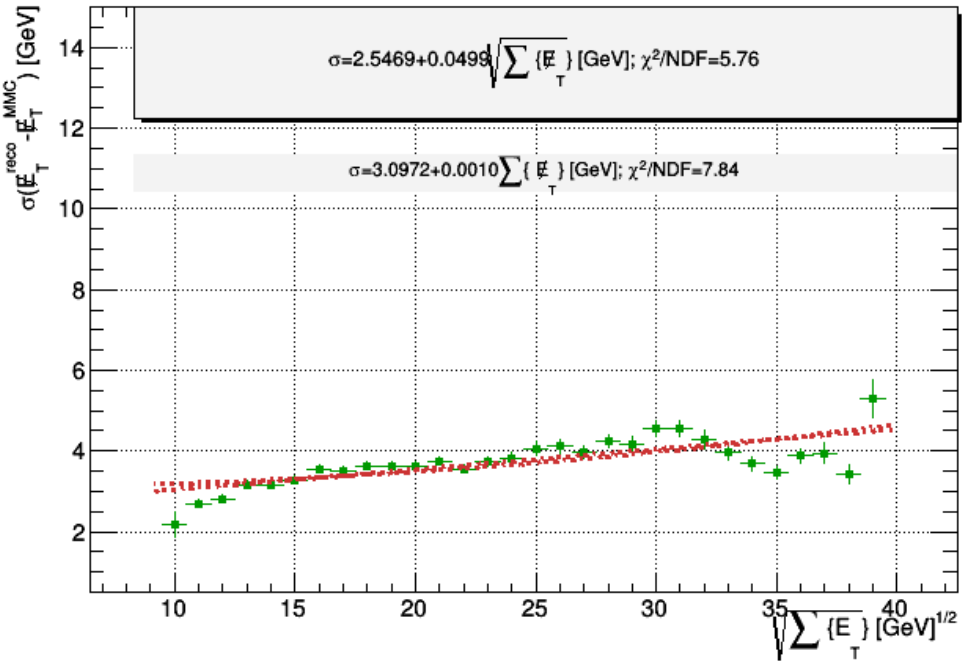


Figure D.3: The $\sigma(E_{\mathrm{T}}^{\mathrm{miss,reco}} - E_{\mathrm{T}}^{\mathrm{miss,MMC}})$ dependence on the sum of the transverse energy in a given event. The $E_{\mathrm{T}}^{\mathrm{miss,MMC}}$ value is computed by the MMC using the MET resolution parametrisation.

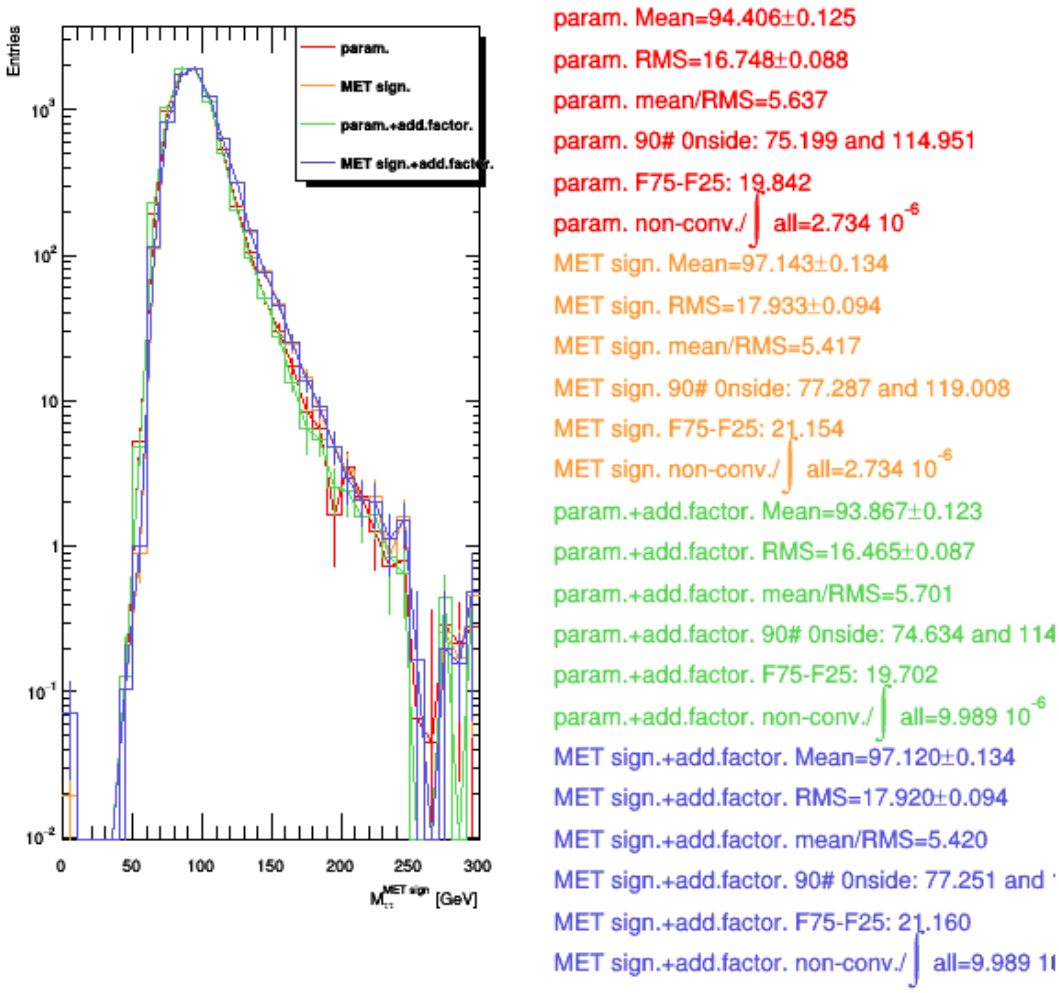


Figure D.4: The $m_{\tau\tau}^{\mathrm{MMC}}$ distribution obtained with a number of alternative settings: where the MET resolution is parametrized (the red curve) and with the additional factorisation on $E_{\mathrm{T}}^{\mathrm{miss,reco}} - E_{\mathrm{T}}^{\mathrm{miss,MMC}}$ (the green line); where MET resolution is calculated through object-level MET significance (the blue line) and with the additional factorisation on $E_{\mathrm{T}}^{\mathrm{miss,reco}} - E_{\mathrm{T}}^{\mathrm{miss,MMC}}$ (the magenta line).

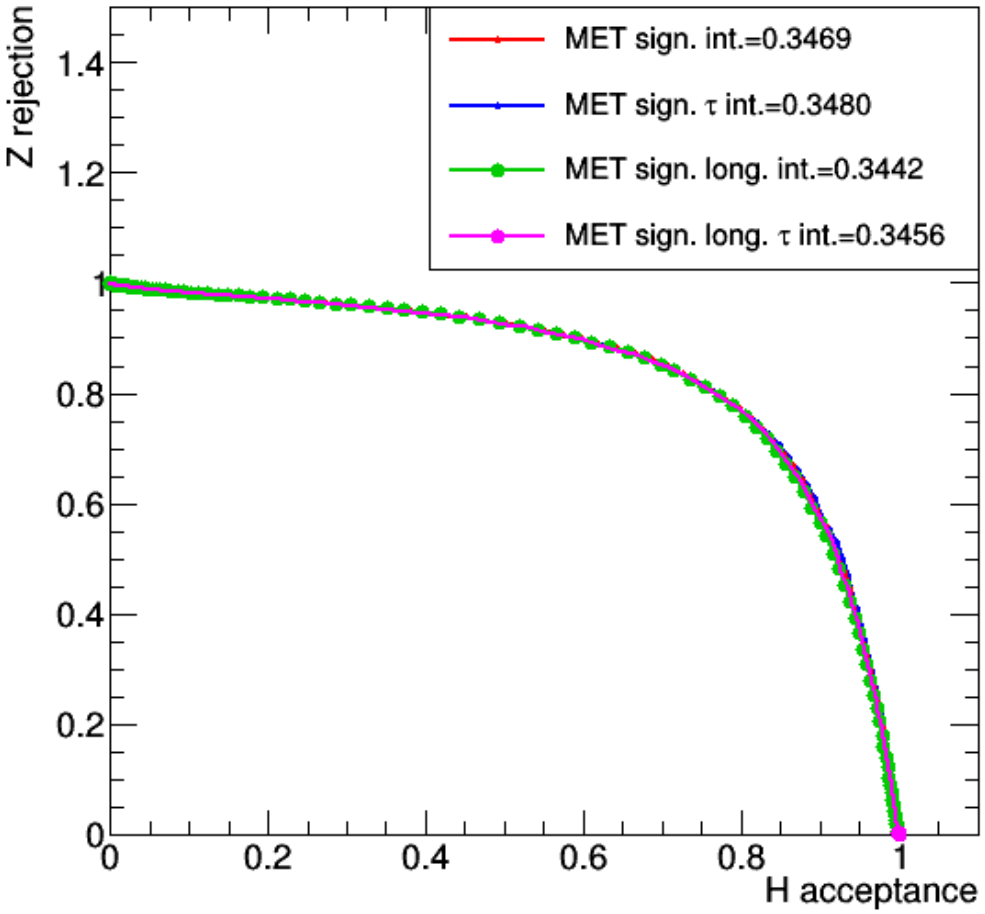


Figure D.5: The signal-to-background discrimination power of the $m_{\tau\tau}^{\mathrm{MMC}}$ variable calculated with varying definitions of MET significance implied.

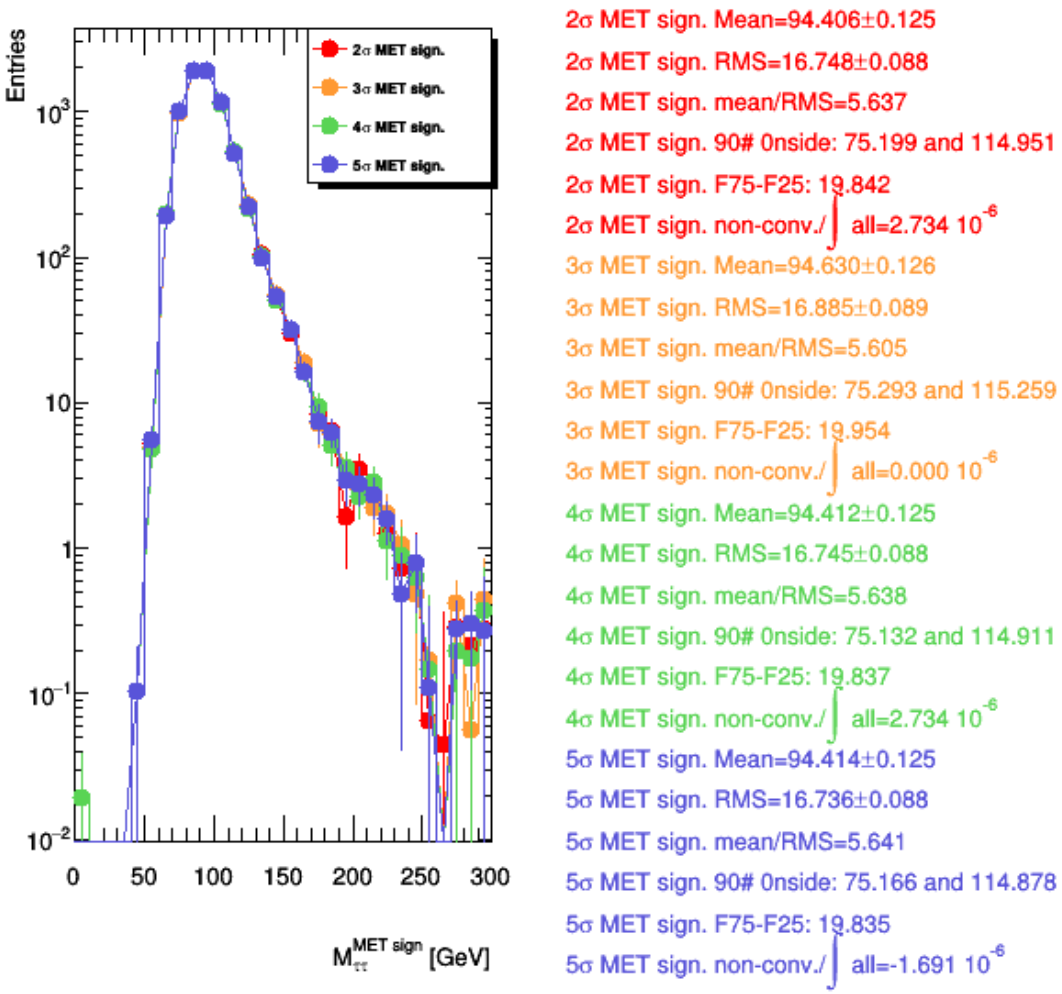


Figure D.6: The $m_{\tau\tau}^{\text{MMC}}$ spectra measured with varying settings of the $E_{\text{T}}^{\text{miss}}$ scan range. The MET resolution is estimated with the object-level MET significance.

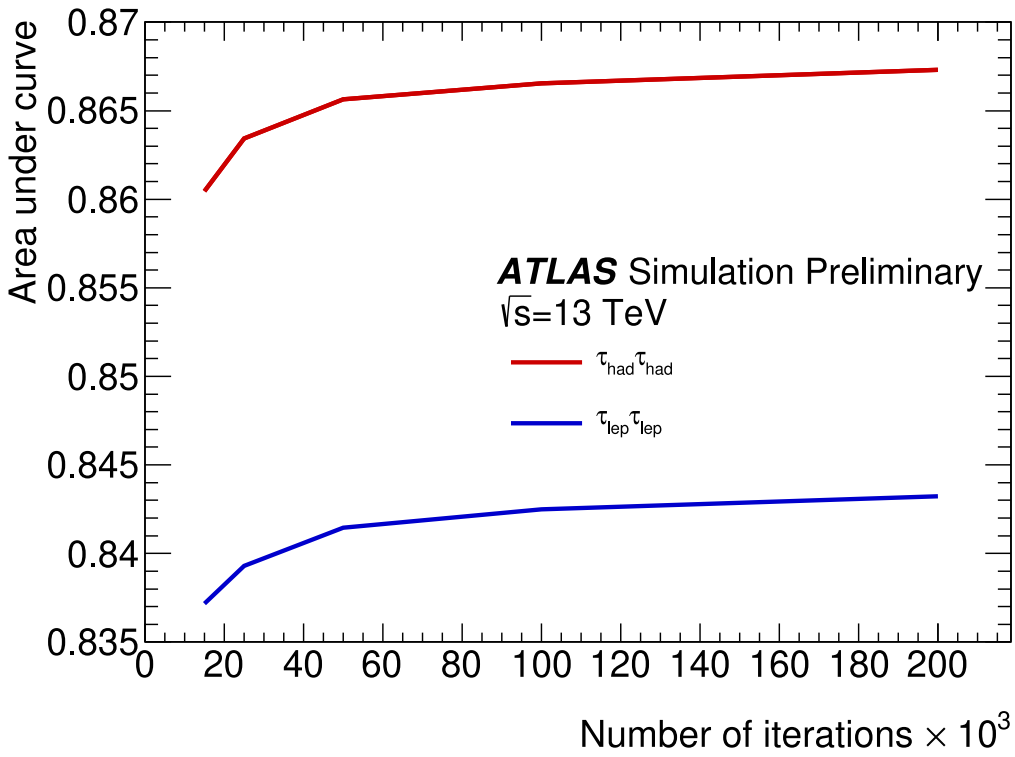


Figure D.7: The AUC of the $H \to \tau\tau$ signal to the $Z \to \tau\tau$ background separation as a function of N_{iter}. The AUC values for τ_{had} τ_{had} and τ_{lep} τ_{lep} events are presented.

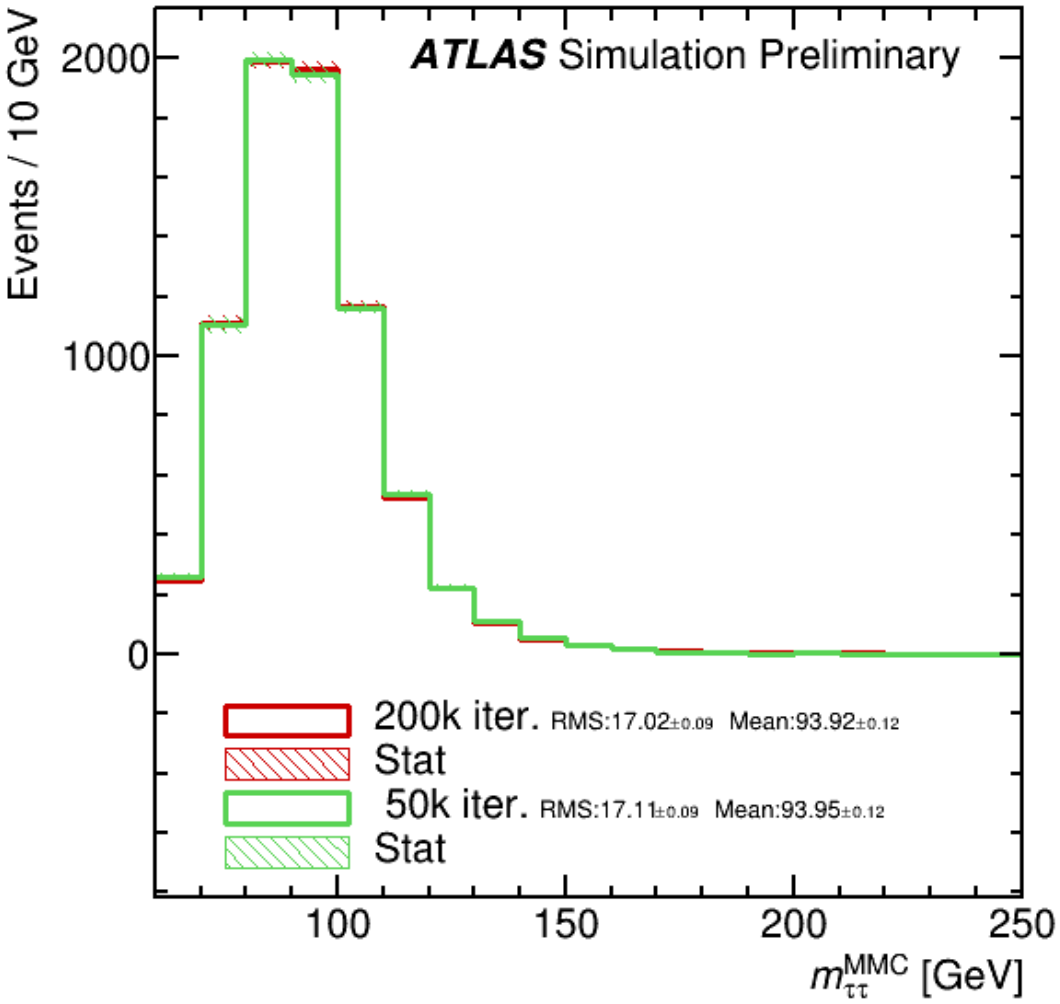


Figure D.8: Comparison of the $m_{\tau\tau}^{\mathrm{MMC}}$ distributions obtained with the nominal (200 K) and reduced (50 K) number of the scan iterations.

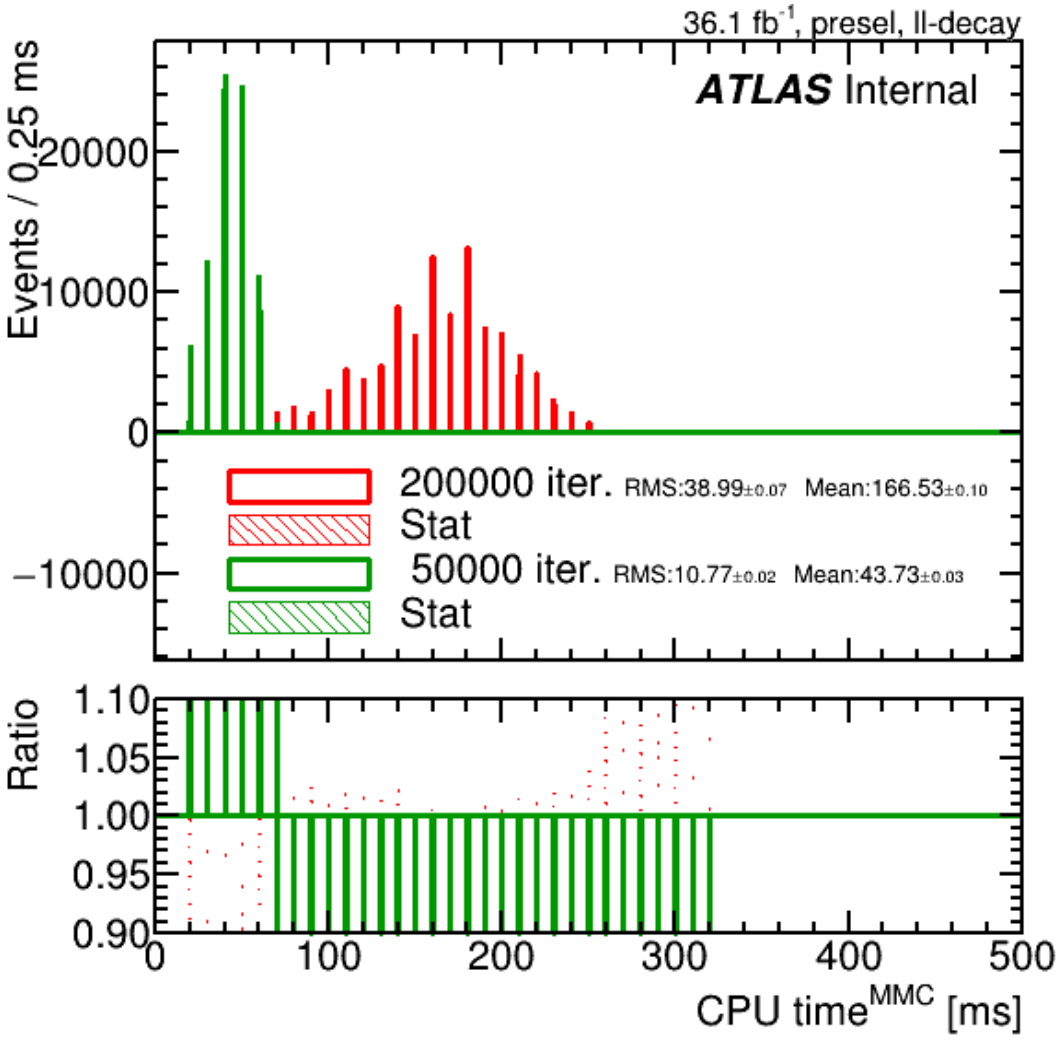


Figure D.9: The CPU time required for $m_{\tau\tau}^{\mathrm{MMC}}$ computation per event with the nominal (200 K) and reduced (50 K) number of the scan iterations in the Missing Mass Calculator.

E. Discussion of the MET significance usage in the $H \to \tau\tau$ analysis

Applicability of the MET significance in the $H \to \tau\tau$ coupling measurement was farther scrutinized.

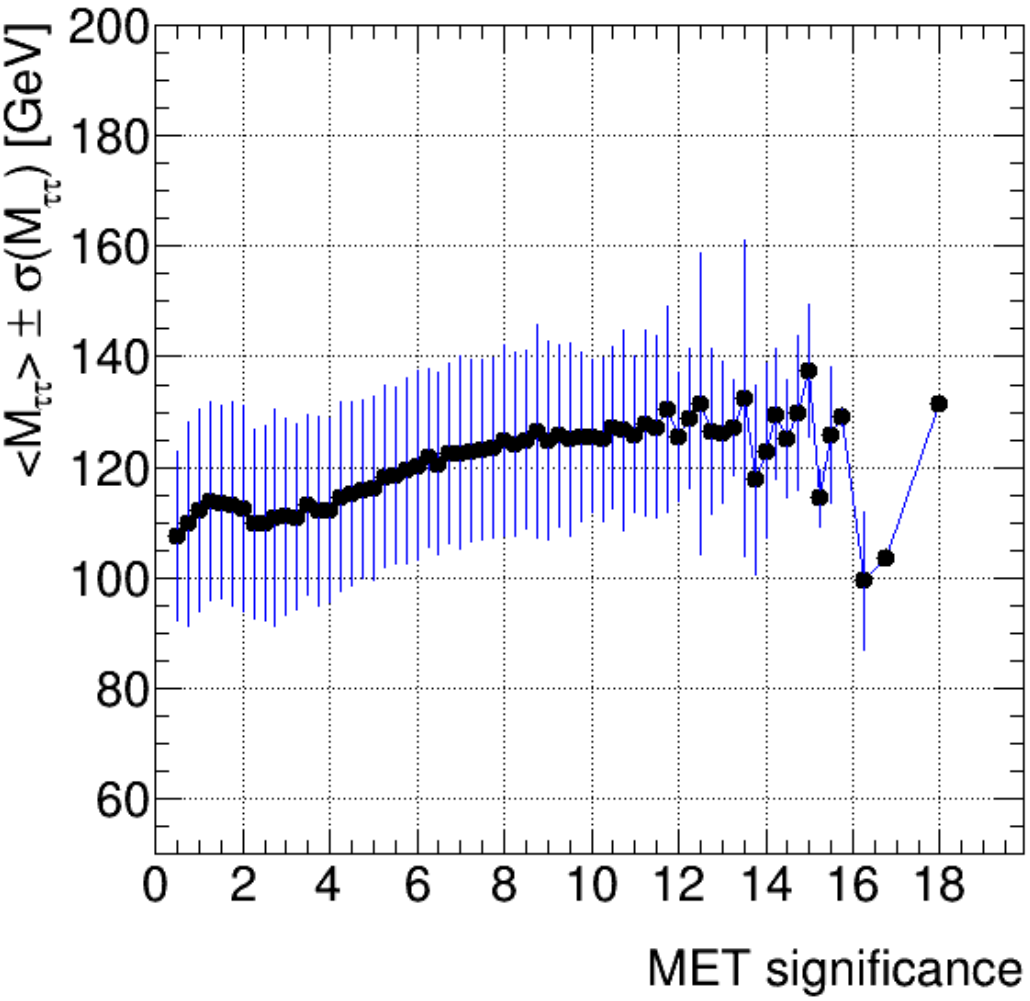


Figure E.1: The mean reconstructed $m_{\tau\tau}^{\mathrm{MMC}}$ in Higgs boson decays ($H \to \tau\tau$) as a function of the object-level MET significance. The error bands depict the RMS of the $m_{\tau\tau}^{\mathrm{MMC}}$ distribution.

As shown in Figures E.1, there is a clear trend of $m_{\tau\tau}^{\mathrm{MMC}}$ biasing in events with significant MET contribution. The calculated $m_{\tau\tau}^{\mathrm{MMC}}$ tend to bias towards higher values with increasing S_{MET}.

The Z boson mass was found to be optimally reconstructed in events with MET significance between 2 and 5 (Figure E.2). This corresponds to realistic scenario in the $Z \to \tau\tau$ decay: there is a real missing transverse energy which does not reach high values assuming there is no other MET sources.

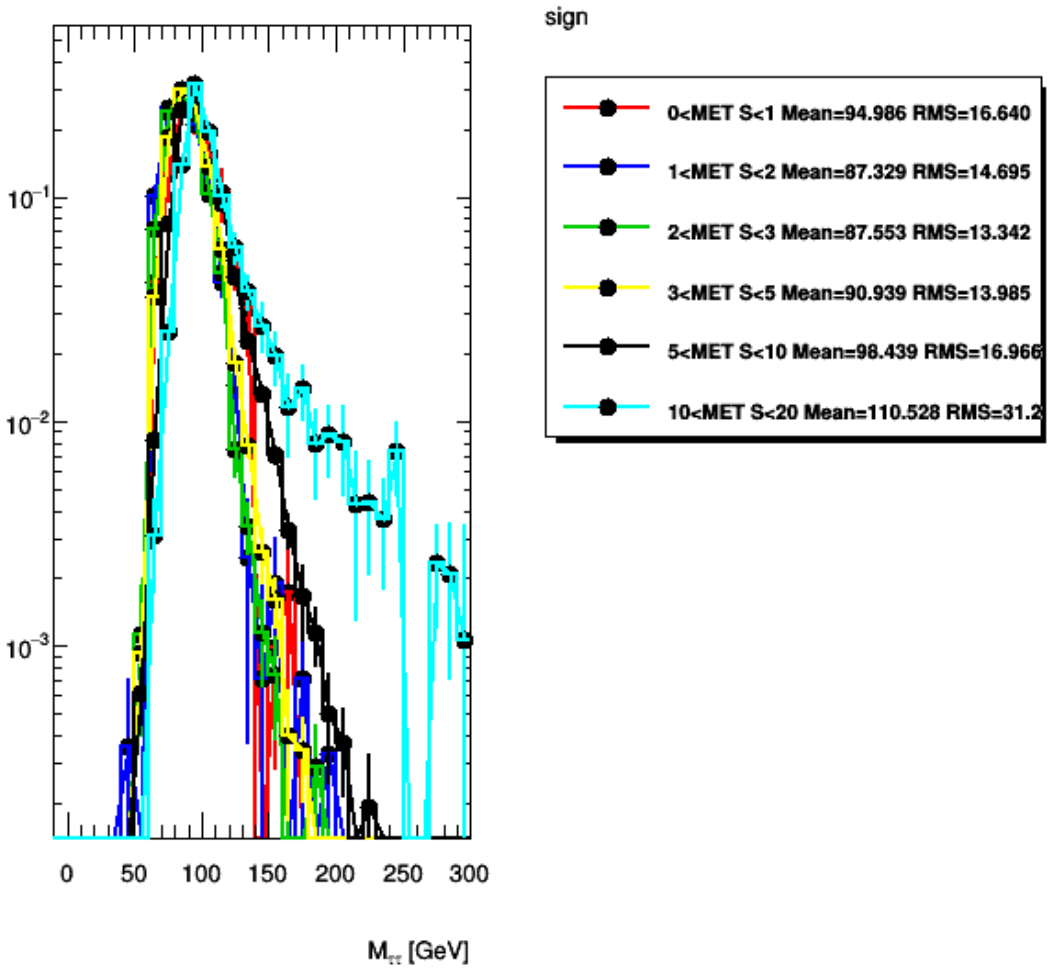


Figure E.2: The $m_{\tau\tau}^{\mathrm{MMC}}$ spectra measured in ranges of the object-level MET significance.

Overall good performance at $m_{\tau\tau}^{\mathrm{MMC}}$ calculation is observed at the range $S_{\mathrm{MET}} \in (0, 10)$. This finding is confirmed by comparison of signal-to-background discrimination in several ranges of S_{MET} (Figure E.3). The rejection of the $Z \to \tau\tau$ background is less effective for topologies with $S_{\mathrm{MET}} > 10$ (the magenta line in Figure E.3). The Z boson mass is overestimated by more than 50 GeV in about 20% of events with the MET significance values above 10 (the cyan line in Figure E.2). Based on these observations, employing events with $S_{\mathrm{MET}} < 10$ seems to be promising.

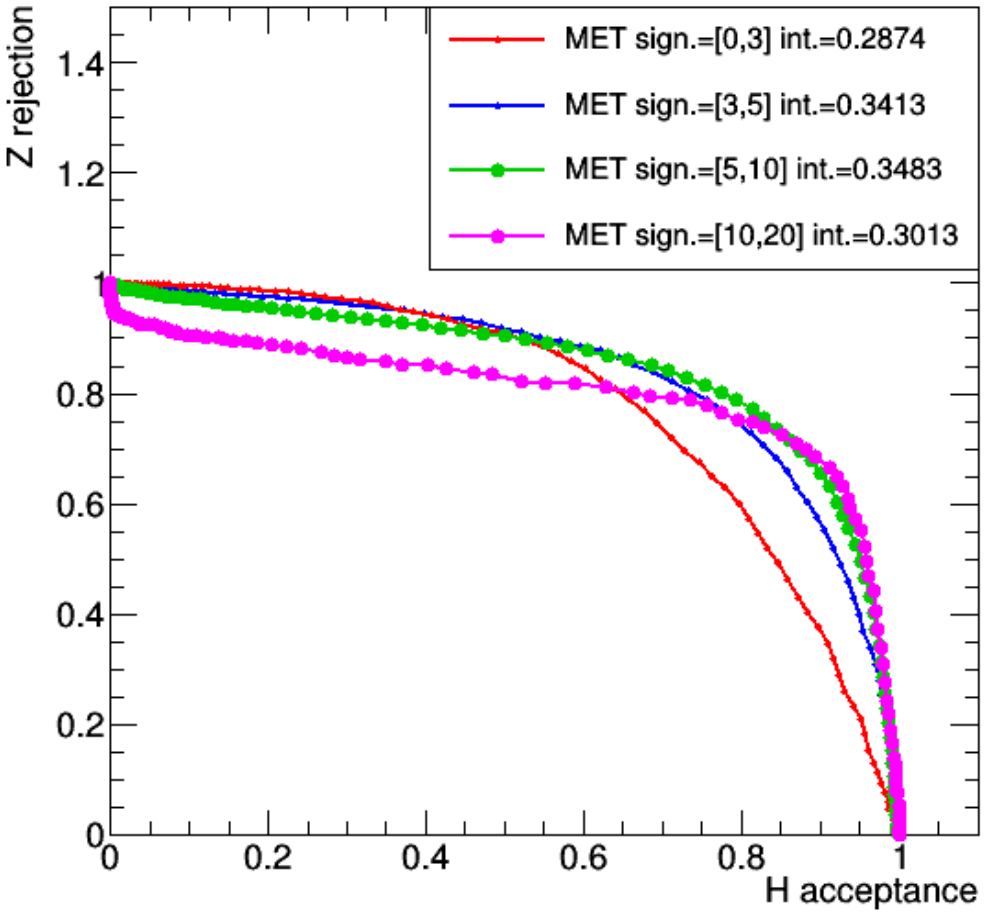


Figure E.3: The separation of the $H \to \tau\tau$ signal from the $Z \to \tau\tau$ background measured in ranges of the object-level MET significance.

F. Selection of MC events in $\tau_{\text{lep}}\tau_{\text{lep}}$: fake estimation

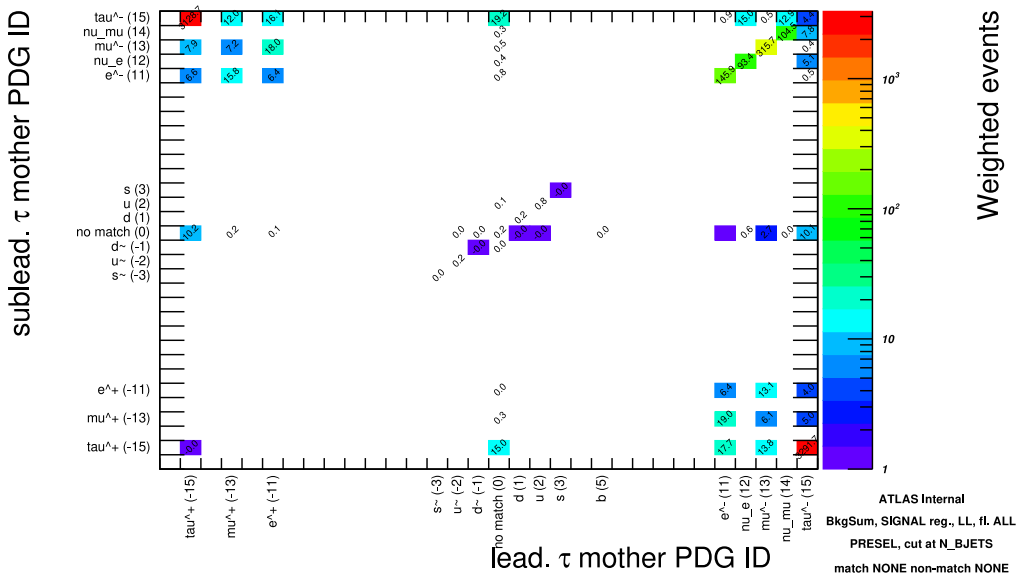


Figure F.1: The leading tau-lepton mother PDG ID versus subleading tau-lepton mother PDG ID in preselected background events.

In addition, the PDG IDs of generator-level mother particles was checked. The correspondence between mother PDG ID for leading and subleading tau-leptons is shown in Figure F.1. There is a negligible fraction of events (below 0.5%) where one of leptons originate from quark or gluon. The analysis of individual background processes showed that they mostly arise in di-boson events.

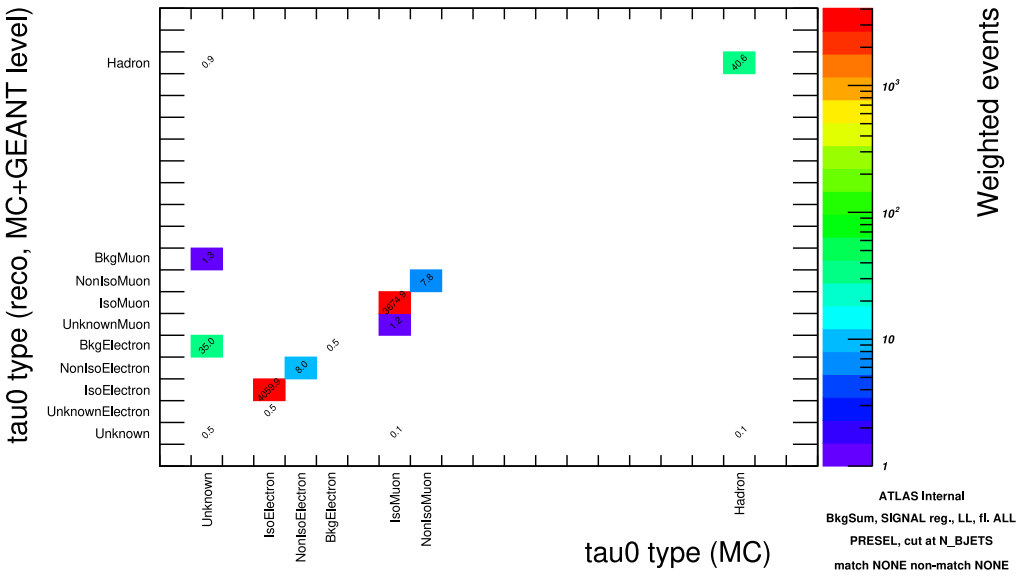


Figure F.2: The relation between the “MC" and “MC + Geant 4" type variables for leading tau-lepton.

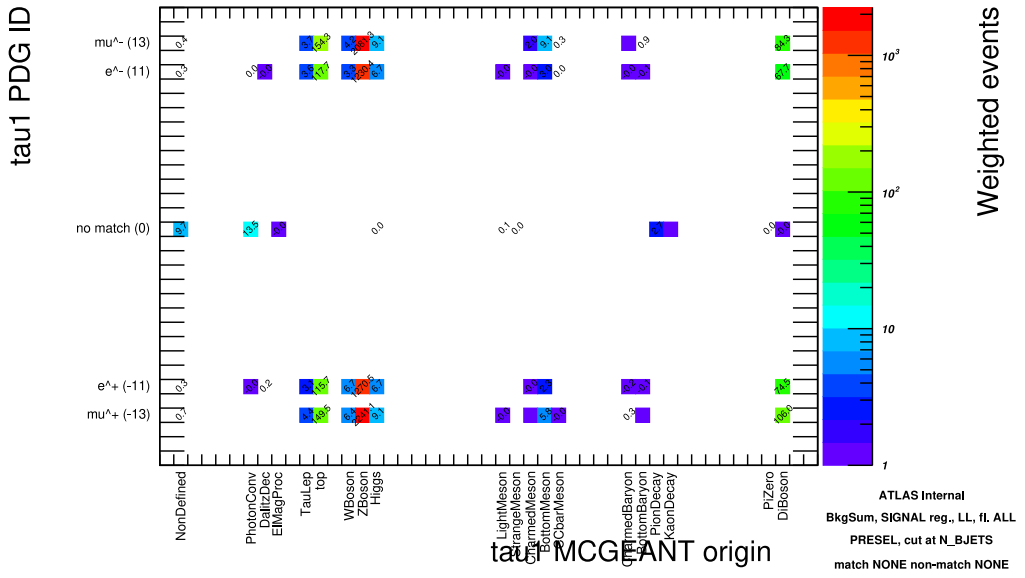


Figure F.3: The origin of matched particle-level counterpart of the leading tau-lepton in relation to its PDG ID.

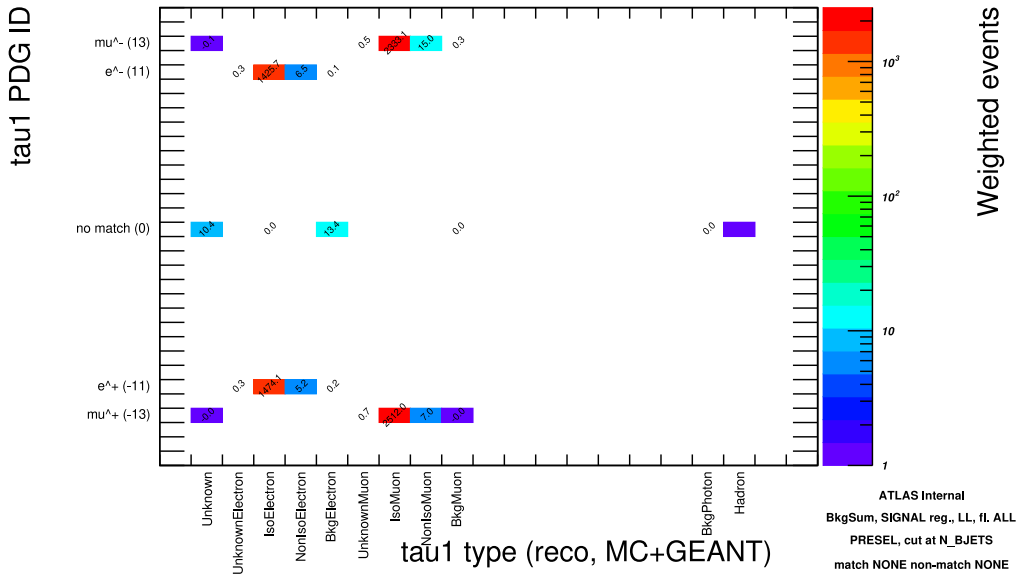


Figure F.4: The type of matched particle-level counterpart of the leading tau-lepton in relation to its PDG ID.

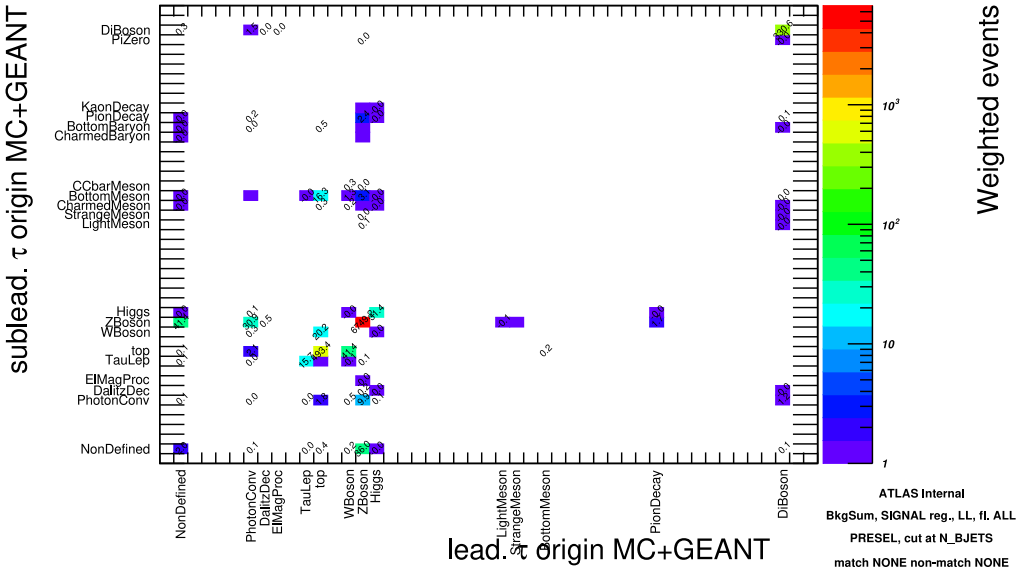


Figure F.5: The origin of matched particle-level counterpart of the leading tau-lepton and subleading tau-lepton. The preselected MC-modelled fake background events are analysed.

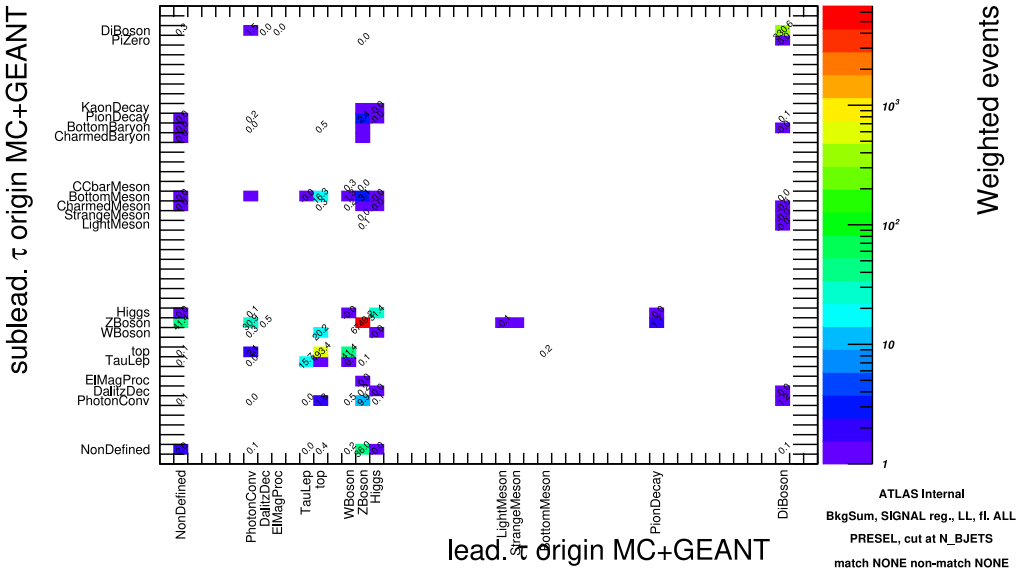


Figure F.6: The origin of matched particle-level counterpart of the leading tau-lepton and subleading tau-lepton. The preselected MC-modelled fake background events are analysed, with the selection removing reconstructed leptons matched to the quarks or gluons (based on PDG ID of the generator-level particle), and removing leptons unmatched to any generator-level particle.

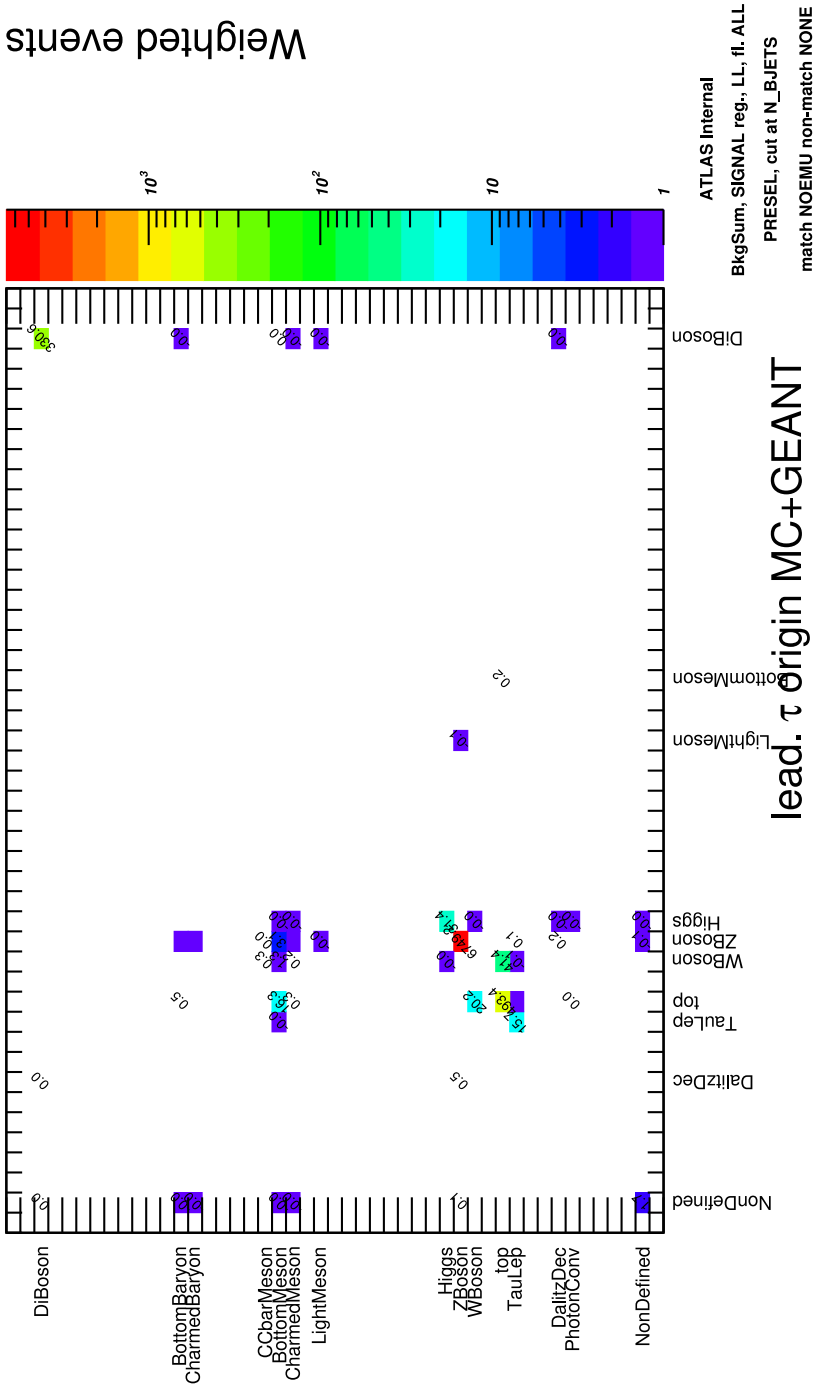


Figure F.7: The origin of matched particle-level counterpart of the leading tau-lepton and subleading tau-lepton. The preselected MC background events are analysed. In addition, only events where final-state light leptons are matched to particle-level light leptons are selected.

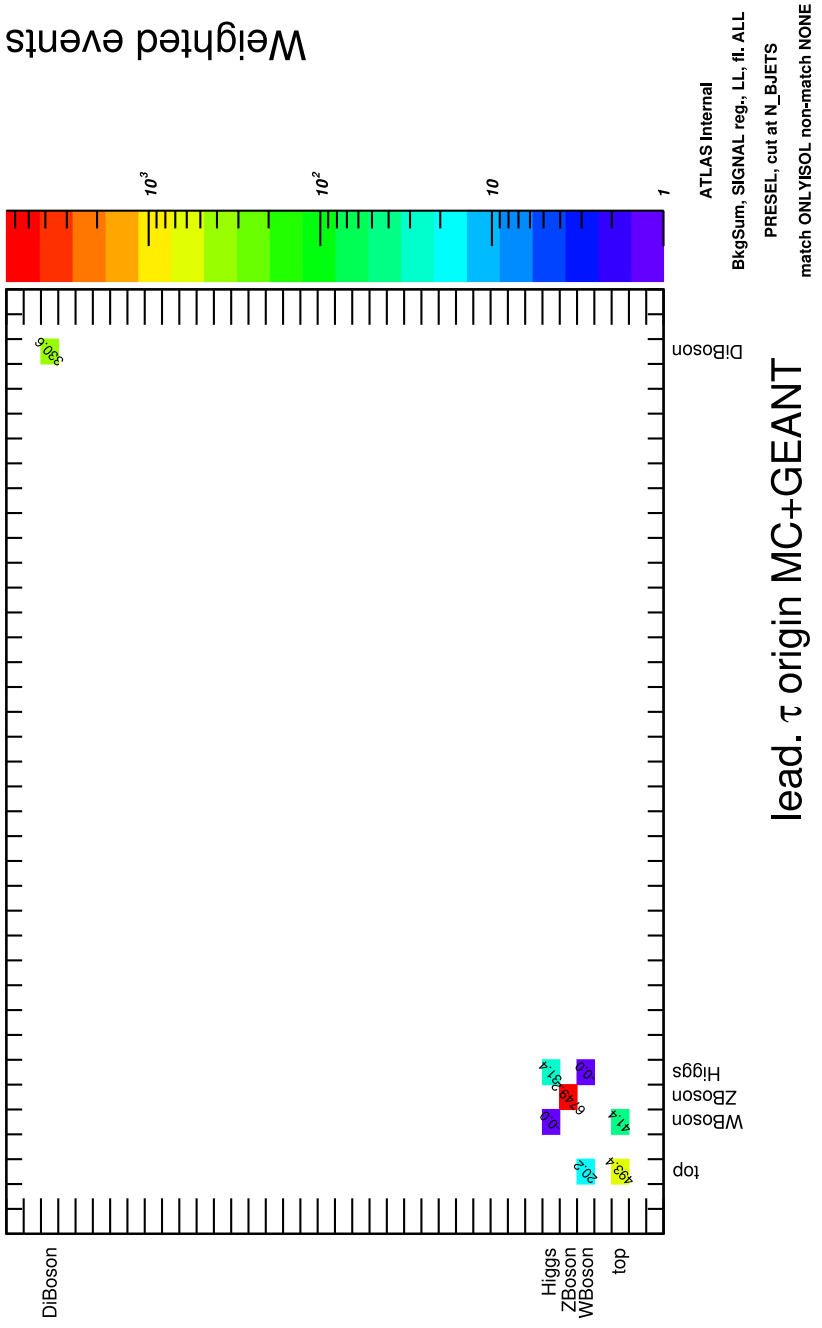


Figure F.8: The origin of matched particle-level counterpart of the leading tau-lepton and subleading tau-lepton. The preselected MC background events are analysed. In addition, reconstructed light leptons are required to be matched to particle-level objects assigned to the isolated lepton type.

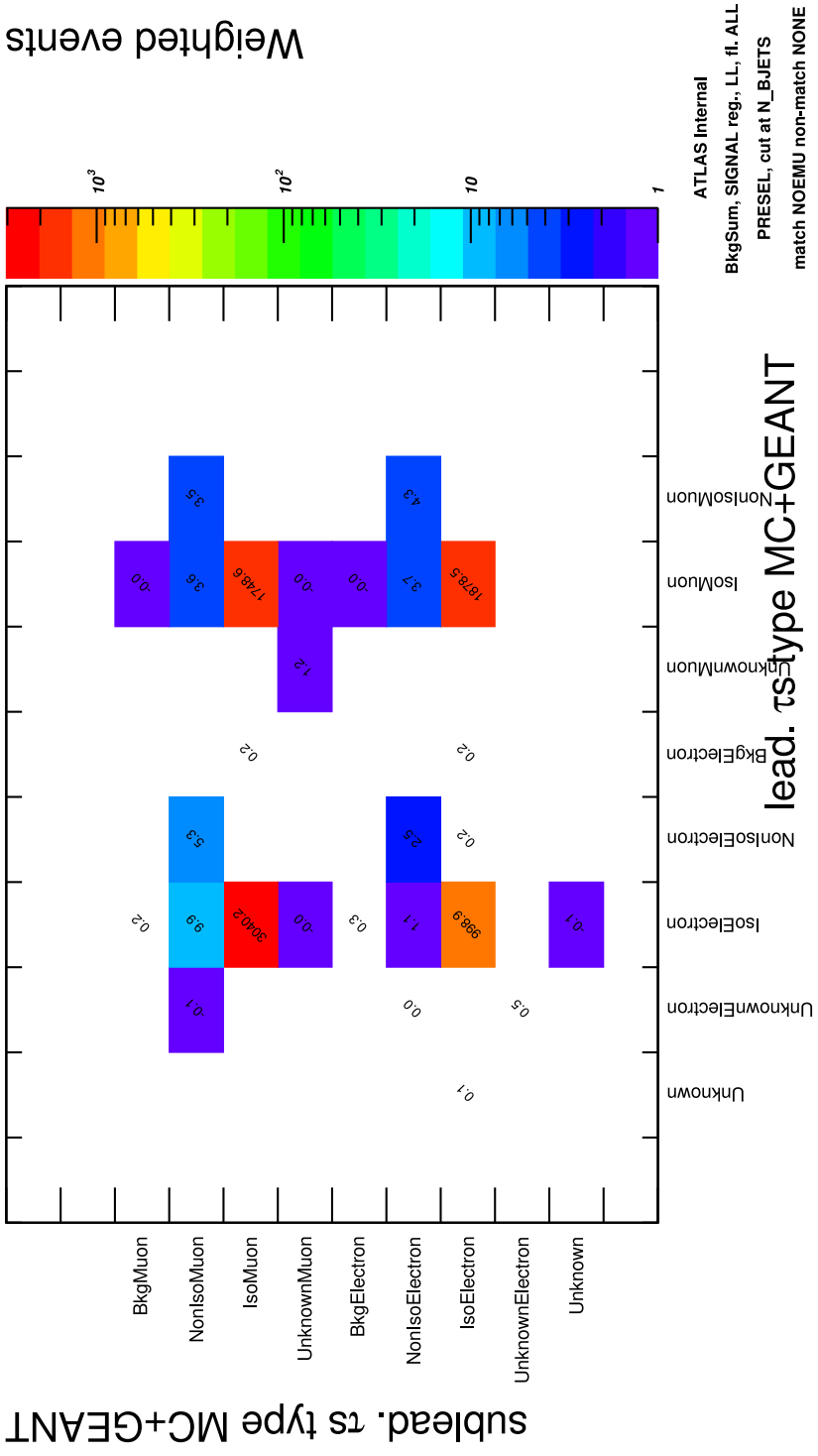


Figure F.9: The type of matched particle-level counterpart of the leading tau-lepton and subleading tau-lepton. The preselected MC background events are analysed. In addition, only events where final-state light leptons are matched to particle-level light leptons are selected.

G. Excited tau-lepton production in contact interaction only

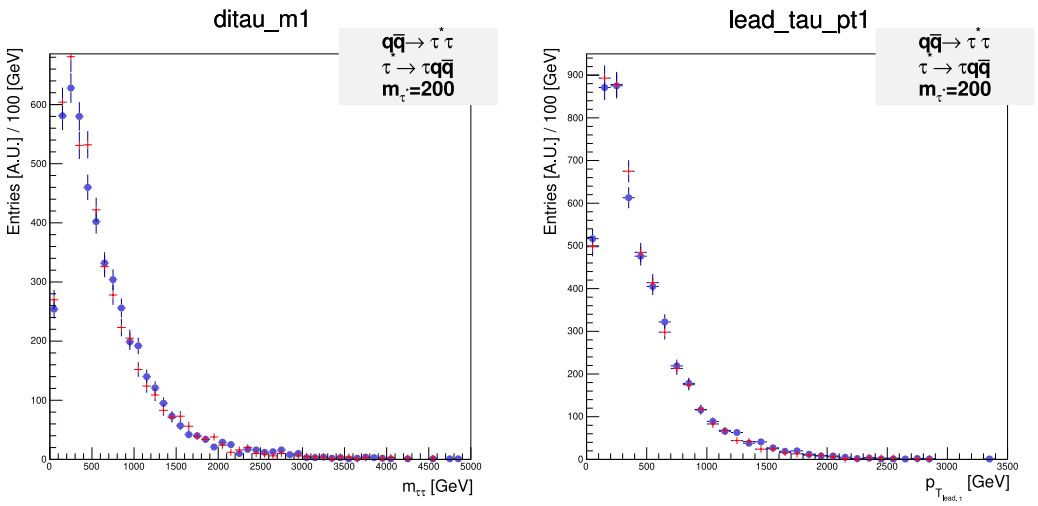
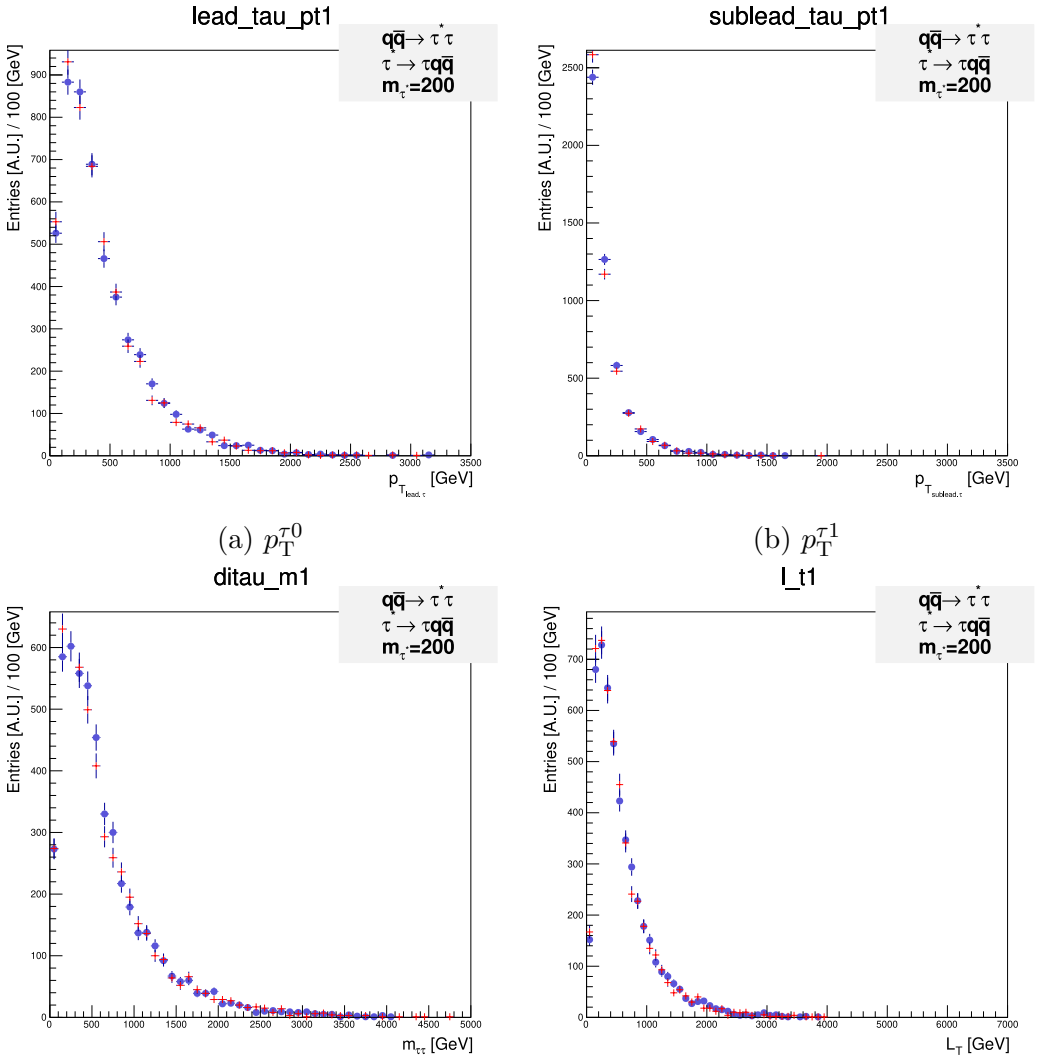


Figure G.1: Distributions of $m_{\tau\tau}$ and $p_{T\,lead:\tau}$ in assumptions of any ETL decay mode and decay to exclusively $tau^* \to \tau jj$ contact-interaction mode.

H. Rescaling compositeness scale

This appendix discusses the choice of the excited lepton decay mode. In general, the model assumes excited lepton decay can decay via contact and electroweak interactions. The presented search focuses on the decay through contact interaction such as $tau^* \to tauq\bar{q}$. Figure H.1 shows comparison of kinematic distributions in assumption of decay to all possible decay modes and the contact-interaction decay mode exclusively. The identity of the distributions justifies the choice of explicit $tau^* \to \tau jj$ mode for the presented search.

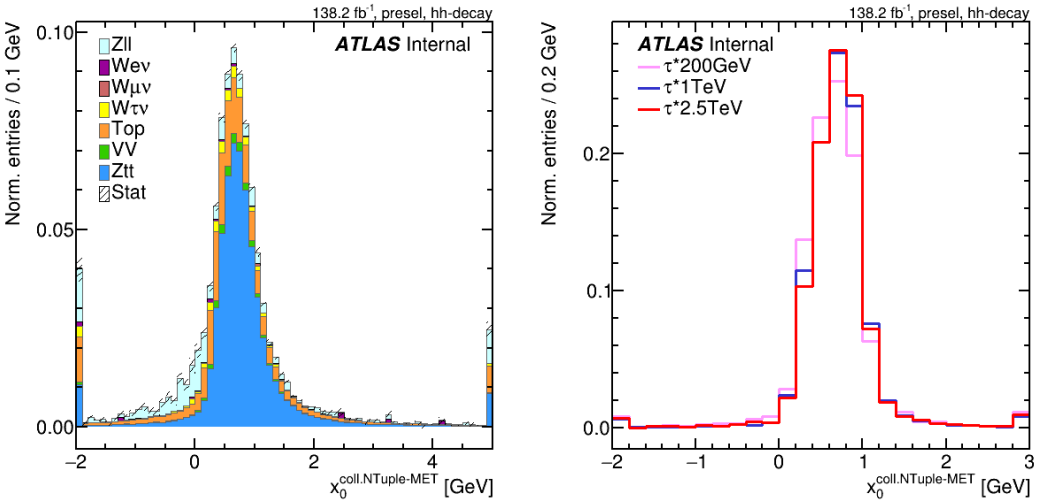


(a) $p_{\mathrm{T}}^{\tau 0}$ (b) $p_{\mathrm{T}}^{\tau 1}$

(c) $m_{\tau\tau}$ (d) L_{T}

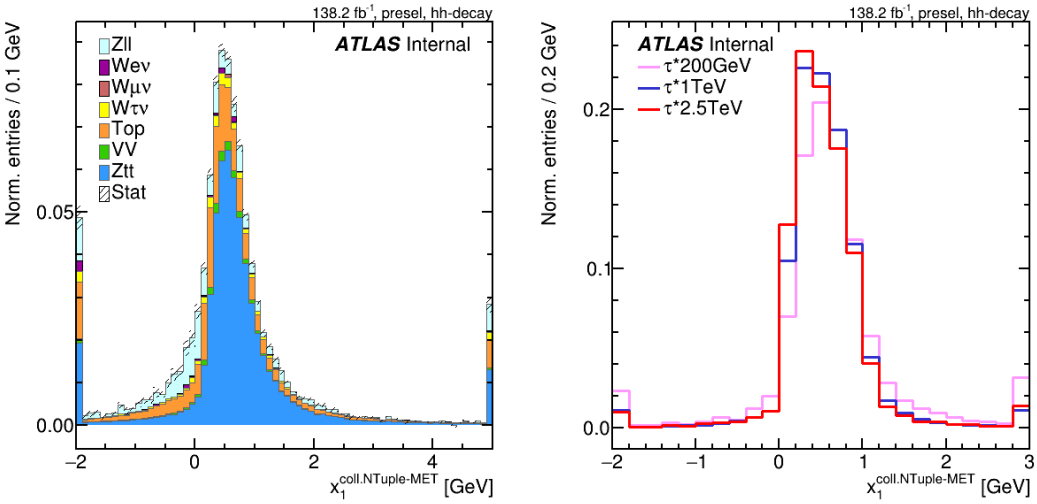
Figure H.1: Comparison of basic kinematic distributions in events with excited tau-leptons ($m_{\tau *}$ = 200 GeV picked up as a representative) generated at the compositeness scale Λ = 3 TeV (depicted in red) and Λ = 10 TeV (depicted in blue).

I. Search for excited tau-leptons: selection on collinear fractions x_0^{coll} and x_1^{coll}



(a) (b)

Figure I.1: The x_0^{coll}, momentum fraction carried by the visible decays of the leading tau-lepton, distributions for SM background (a) and excited tau-leptons (b) processes.



(a) (b)

Figure I.2: The x_1^{coll}, momentum fraction carried by the visible decays of the sub-leading tau-lepton, distributions for SM background (a) and excited tau-leptons (b) processes.

J. Search for excited tau-leptons: fit variable choice

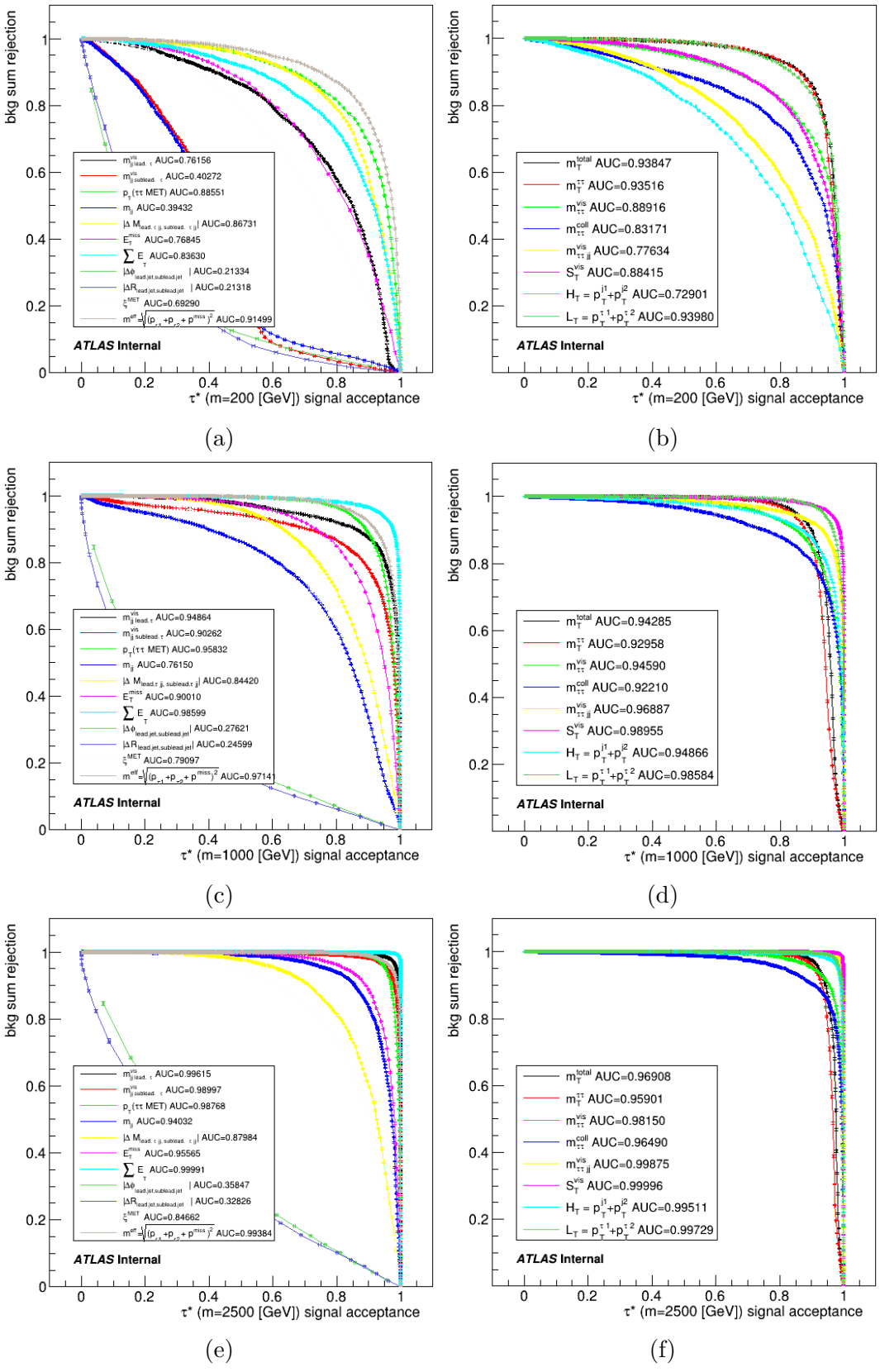
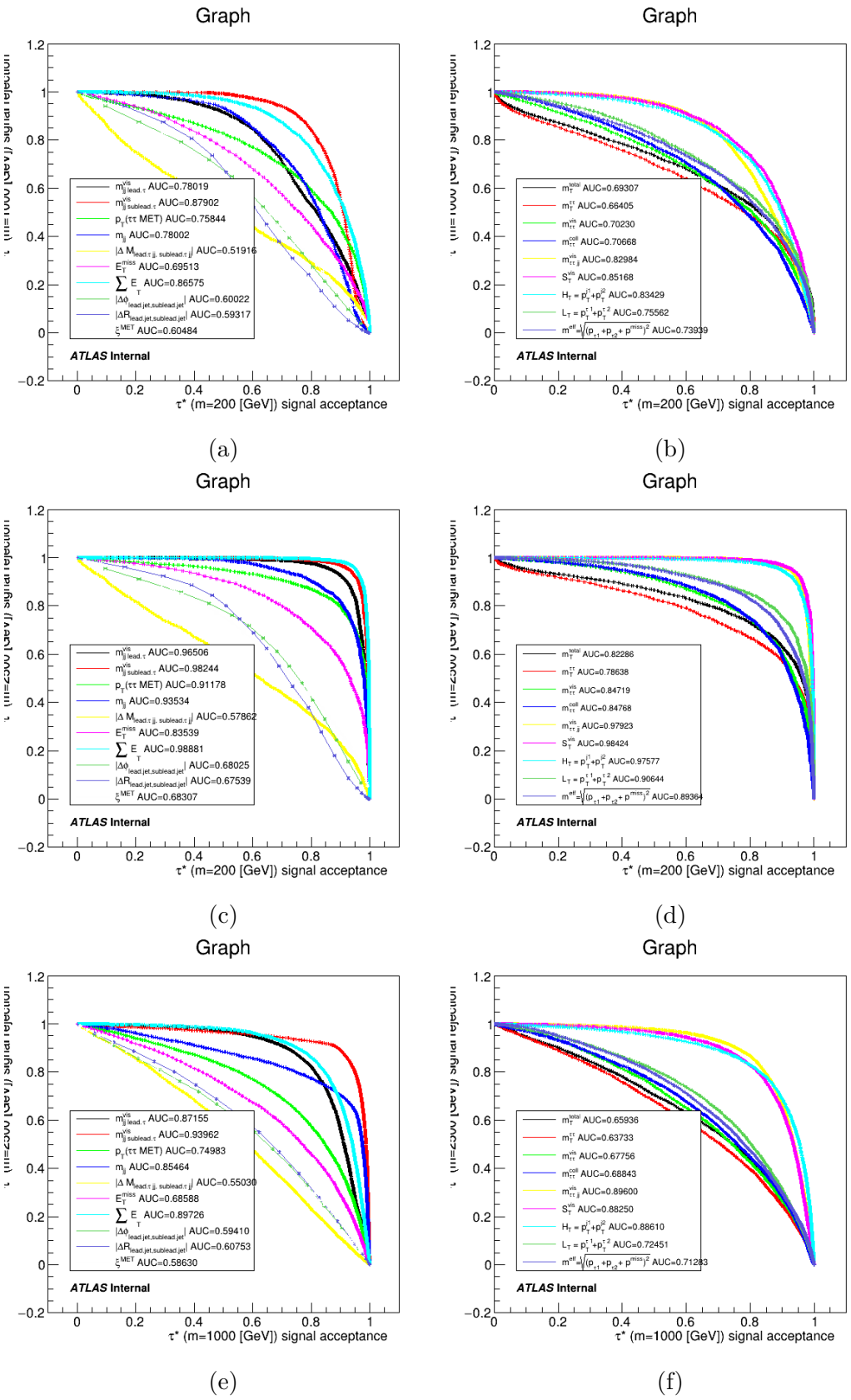


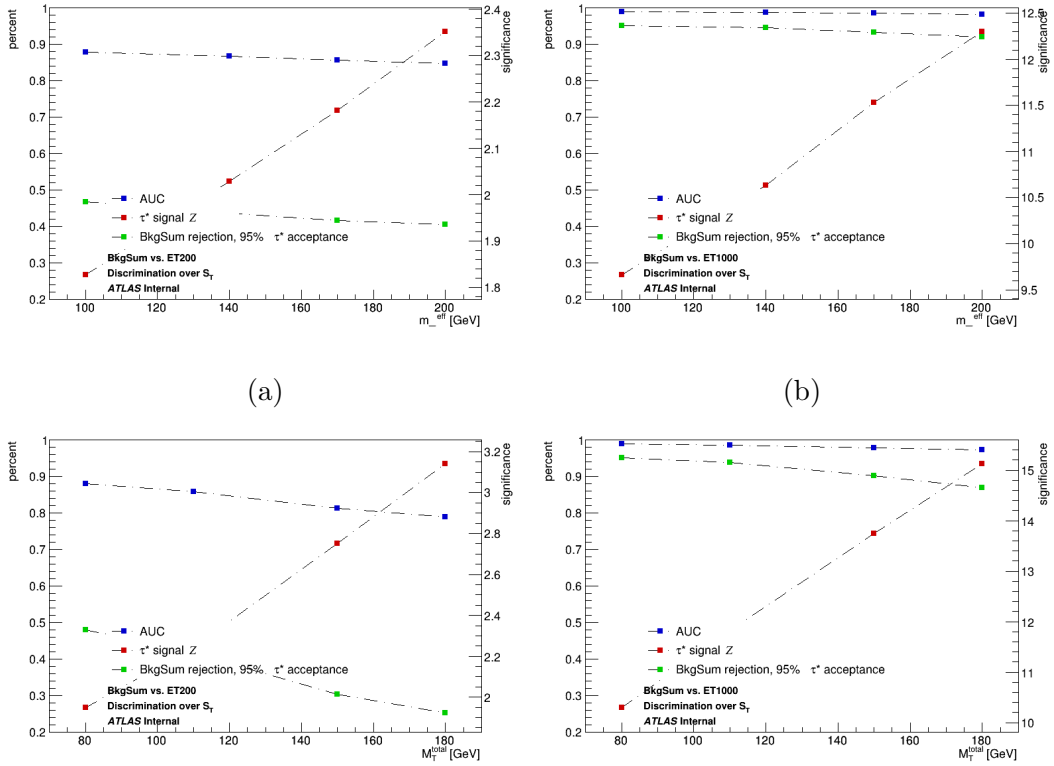
Figure J.1: ROC curves for signal-to-background separation for the following ETL signal mass points: 200 GeV (a), (b); 1 TeV (c), (d); 2.5 TeV (e), (f).



(a) (b) (c) (d) (e) (f)

Figure J.2: ROC for 200 GeV τ^* vs. 1 TeV τ^* (a), (b). ROC for 200 GeV τ^* vs. 2.5 TeV τ^* (c), (d). ROC for 1 TeV τ^* vs. 2.5 TeV τ^* (e), (f).

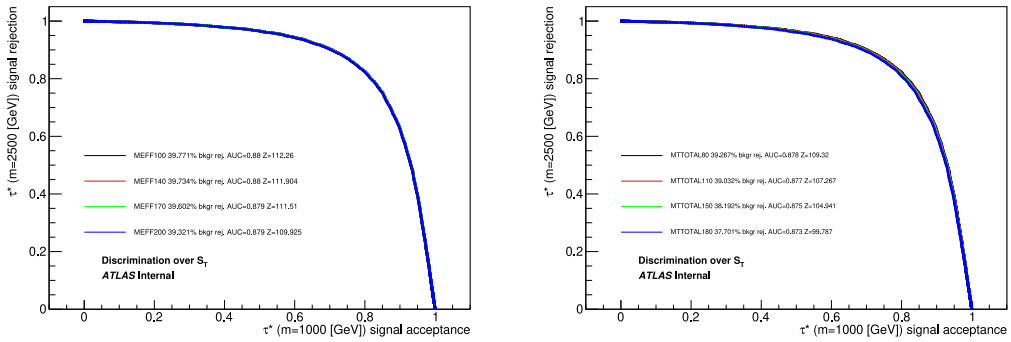
K. Search for excited tau-leptons: optimization of the selection on L_{T}



(a) (b)

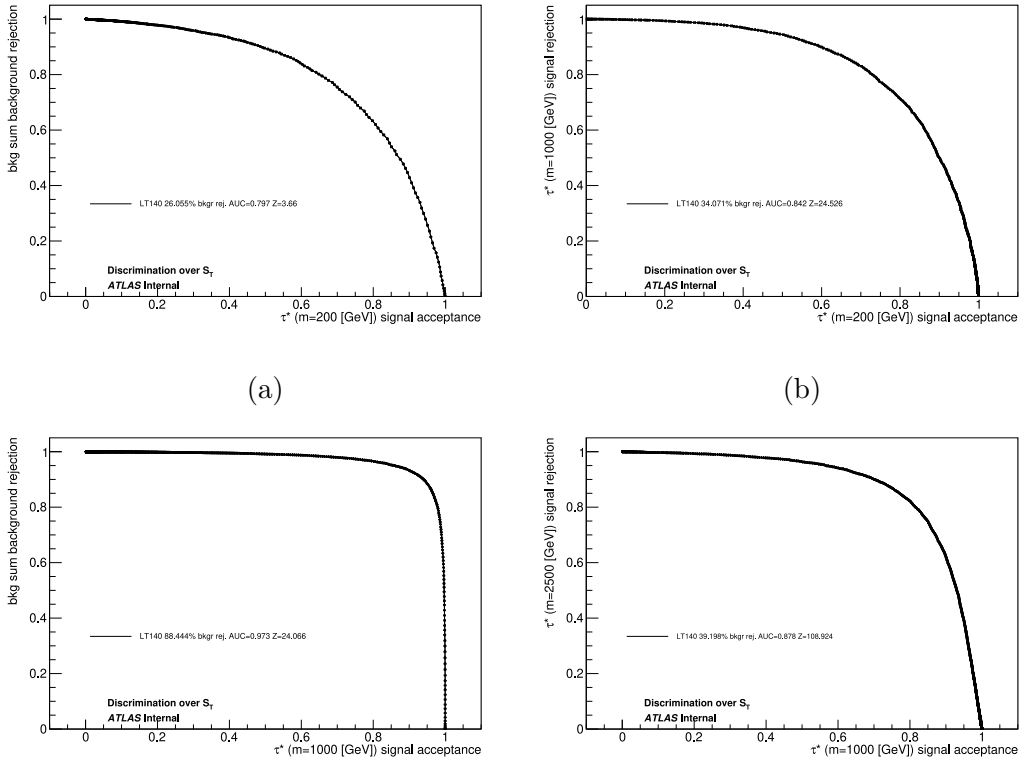
(c) (d)

Figure K.1: The graphs showing effectiveness of discrimination between excited tau-lepton signal of 200 GeV (1 TeV) mass on the left (right) plots and the total SM background. The values are obtained in the scan over the S_{T} variable in the signal region. Several selections intending to increase signal presence are tested: additional criteria imposed on one the variables m_{eff} (the upper plots) and $m_{\mathrm{T}}^{\mathrm{total}}$ (the lower plots). The integral significance and AUC are depicted with red and blue, correspondingly, as a function of the varied value of the given cut. The background rejection level at the 95% acceptance of excited tau-lepton events is shown as well (depicted with green).



(a) (b)

Figure K.2: The ROC curves for separation between excited tau-lepton signal of masses $m = 1$ TeV and $m = 2.5$ TeV when a given selection on m_{eff} (a) and $m_{\text{T}}^{\text{total}}$ is applied.



(a) (b)

(c) (d)

Figure K.3: The ROC curves for separation between signal and background (signal) obtained with for the signal regions with the definitive selection $L_{\text{T}} > 140$ GeV.

L. Fit for setting upper limits with the Asimov data set

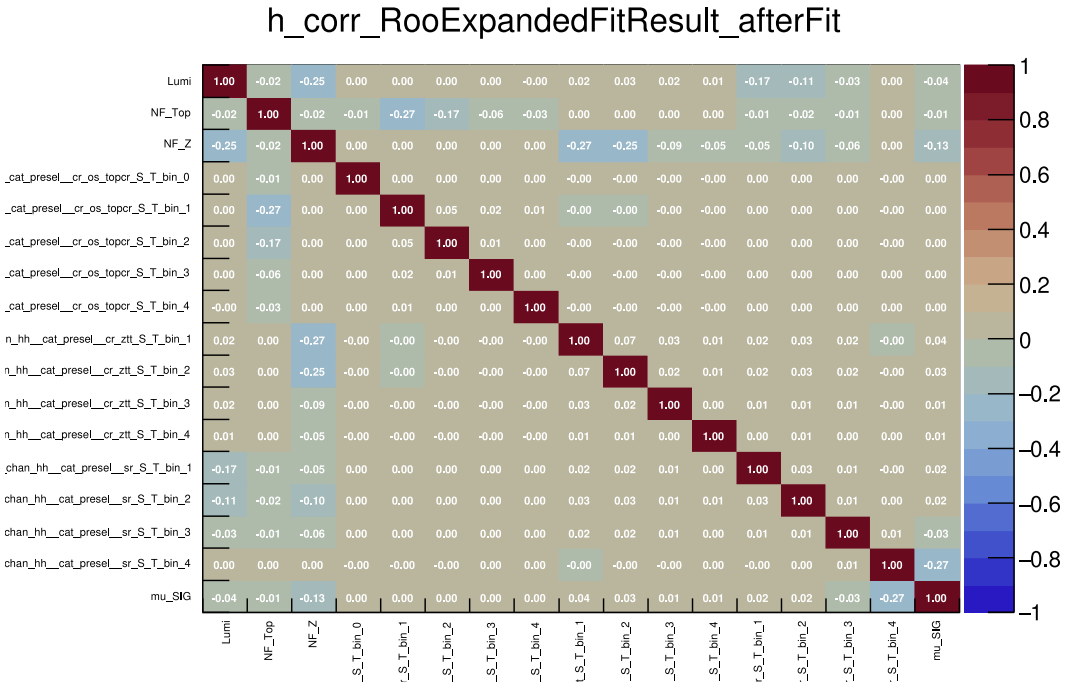


Figure L.1: The correlation between NPs in the fit to real (Asimov) data in the CRs (SR). The result for the fit in the ETL with mass of 1 TeV is shown.

Process, N_{events}	$Z \to \tau\tau$ CR	Top CR	SR
Observed events	92395	3579	4287
Fitted bkgr events	92395.19 ± 303.69	3579.05 ± 57.39	4287.97 ± 43.68
Fitted Z events	37.51 ± 0.83	2872.22 ± 59.10	1185.17 ± 26.64
Fitted Top events	92276.80 ± 303.73	113.86 ± 0.90	1165.56 ± 9.81
Fitted Fakes events	0.00 ± 0.00	501.86 ± 8.93	1833.81 ± 30.17
Fitted Other events	80.87 ± 1.29	91.11 ± 1.60	103.43 ± 1.71
Fitted ET1000 events	0.00 ± 0.22	0.00 ± 0.06	$0.00^{+17.86}_{-0.00}$
MC exp. SM events	92395.18	3579.06	4287.97
MC exp. Z events	37.51	2872.23	1185.17
MC exp. Top events	92276.80	113.86	1165.56
MC exp. Fakes events	0.00	501.86	1833.81
MC exp. Other events	80.87	91.11	103.43
MC exp. ET1000 events	1.84	0.47	150.76

Table L.1: The event yields in the fit to the Asimov data set (B-only hypothesis) in the $Z \to \tau\tau$ and Top CRs, and the signal region. The signal of 1 TeV excited tau-lepton is considered. The Z background combines $Z \to \tau\tau$ and $Z \to ll$ processes. The Top background includes $t\bar{t}$ and Wt. The "Other" background category includes W + jets, and VV events.

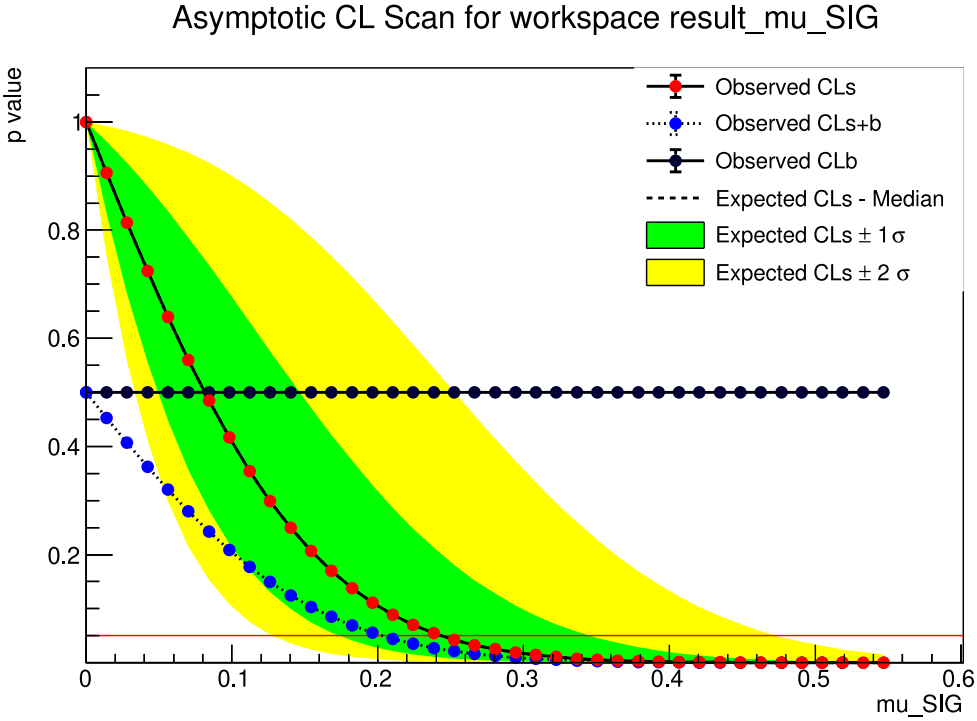
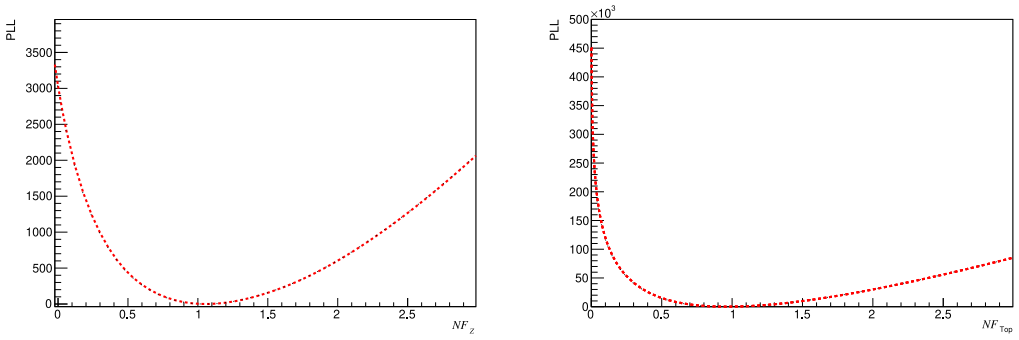


Figure L.2: The dependence of the p_0-value on the excited tau-lepton (1 TeV) signal strength μ. Here the CL_b and CL_{s+b} depict the level of confidence in background (signal+background) model.

M. Profile scans of nuisance parameters



(a) (b)

Figure M.1: The profile logarithmic likelihood scan for the NF_Z (a) and NF_{Top} (b) nuisance parameters.

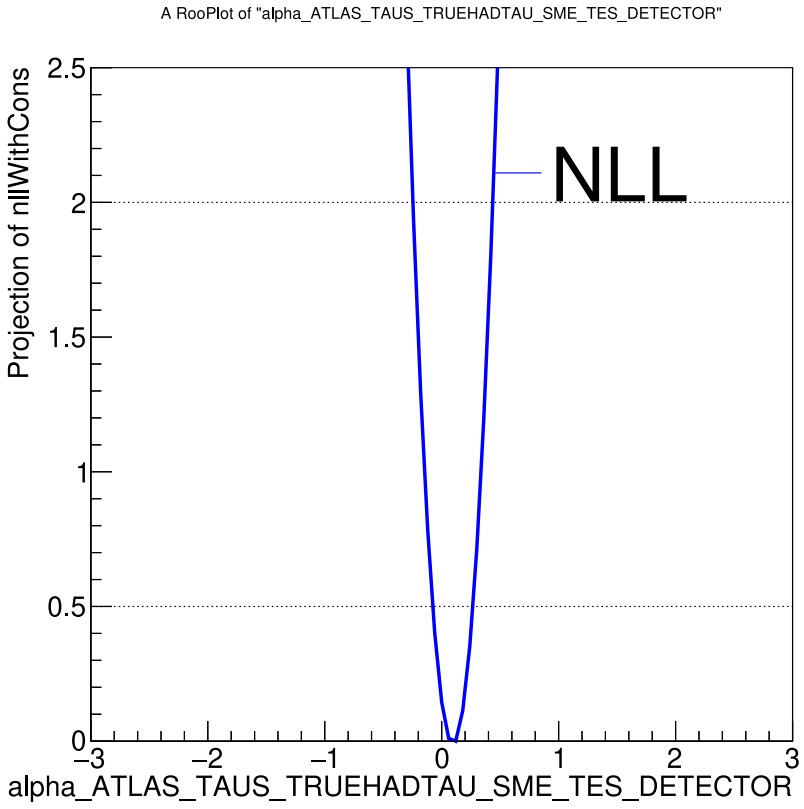


Figure M.2: The negative log-likelihood scan of the tau-lepton energy scale uncertainty (specifically, its detector component).

N. Search for excited tau-leptons: fit to $S + B$ Asimov data set

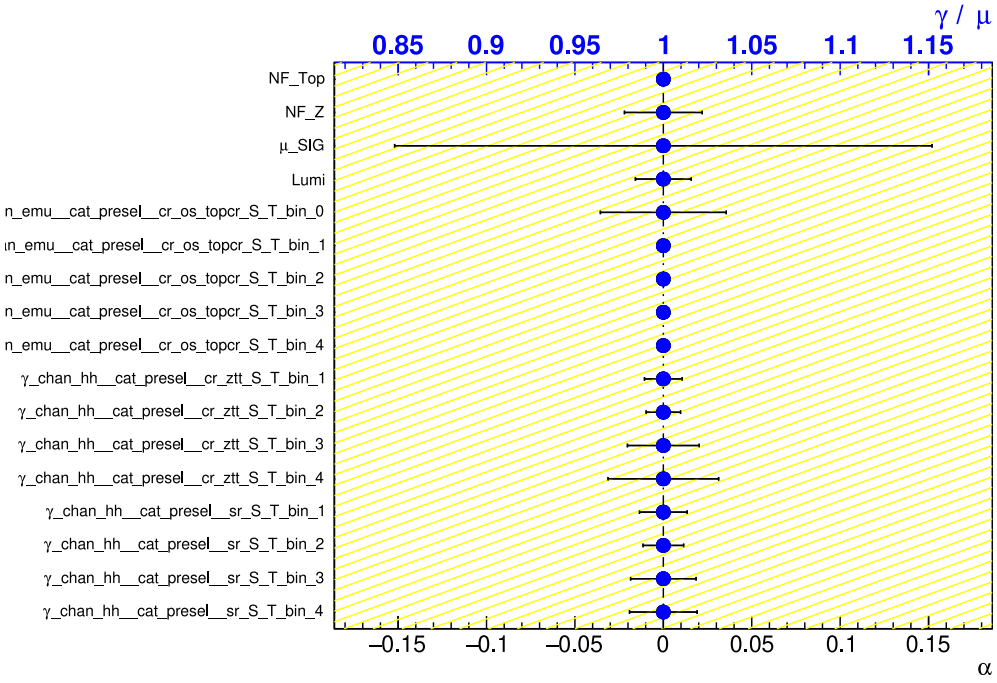


Figure N.1: The pulls and constraints on the nuisance parameters in the fit to $S + B$ Asimov data set in the considered CRs and the SR. This result is obtained in the fit for excited tau-lepton of 1 TeV mass.