Searching for Beyond-the-Standard-Model Physics Using Signatures with Tau Leptons in the Final State at the ATLAS Experiment

mgr Janina Krzysiak

Supervisor
dr hab. Paweł Brückman de Renstrom

Co-supervisor
dr hab. Marcin Wolter

A thesis presented for the degree of
Doctor of Philosophy

Division of Particle and Astroparticle Physics
Henryk Niewodniczański Institute of Nuclear Physics
Polish Academy of Sciences
Kraków, 2021

Acknowledgements

Abstract

The main focus of this thesis is the search for heavy, neutral Higgs bosons decaying to two tau leptons. The existence of such bosons is predicted by certain Beyond-the-Standard-Model (BSM) models such as Two-Higgs-Doublet Models and the Minimal Supersymmetric Standard Model (MSSM). The search was conducted in the mass range between 200 GeV and 2.5 TeV using proton-proton collision data collected by the ATLAS detector at the LHC in the years 2015-2018 at a center-of-mass energy of 13 TeV. The analysis was split into two rounds. In the first round, the analysis group followed the same strategy as the one used in a previous paper analyzing only 2015 and 2016 data. The author's main task within the analysis group was to validate the estimation of background contributions with QCD jets misidentified as tau leptons in events where one of the tau leptons decays to leptons and the other one to hadrons and a neutrino ('$\tau_{\text{lep}}\tau_{\text{had}}$ decay channel'). In this analysis round, no evidence of new bosons was found, and new exclusion limits were set on the BSM Higgs boson production cross-section times the branching ratio to two tau leptons (model-independent limit) and on the phase space of two MSSM benchmark scenarios (model-dependent limits). In the second analysis round, improvements were introduced to the analysis strategy. The author's task was to develop machine learning models for discriminating between signal and background events in the $\tau_{\text{lep}}\tau_{\text{had}}$ decay channel. The goal was to obtain a better performance compared to the standard final discriminating variable, the mass variable m_T^{tot}. Boosted decision trees and mass-parameterized neural networks were trained using mass variables and kinematic variables. Ultimately, the best performance was obtained with boosted decision trees, improving the expected model-independent exclusion limits by up to 67.5% and 28.6% for low-signal-mass events with and without b-jets in the final state, respectively. At the time of submitting this thesis, the second analysis round was ongoing.

Two additional areas of focus are presented in this thesis. The first is a theoretical examination of the possible link between the Higgs mass fine-tuning problem and the cosmological constant problem, taking into consideration Veltman and Pauli divergence-cancelling conditions, as well as the Standard Model effective field theory. The possible role of BSM Higgs bosons is discussed here. This part of the thesis is the result of a collaboration between the author and Steven Bass. The second area of focus is the validation of ATLAS Monte Carlo samples containing tau leptons. Validation studies included polarization measurements, branching ratio and kinematic variable checks, and comparisons of input variables for the ATLAS tau-lepton-identifying boosted decision tree model.

Streszczenie

Głównym tematem niniejszej rozprawy są poszukiwania ciężkich, nienaładowanych bozonów Higgsa rozpadających się na dwa leptony tau. Istnienie takich bozonów jest przewidywane w niektórych modelach spoza Modelu Standardowego, jak na przykład w modelach z dwoma dubletami Higgsa i w Minimalnym Supersymetrycznym Modelu Standardowym (MSSM). Poszukiwania obejmowały przedział mas od 200 GeV do 2.5 TeV i wykorzystywały dane ze zderzeń protonów, które zostały zebrane przez detektor ATLAS na LHC w latach 2015-2018 przy energii środka masy równej 13 TeV. Analiza była podzielona na dwie rundy. W pierwszej rundzie grupa wykonująca analizę korzystała ze strategii używanej we wcześniejszej wersji analizy, w której badano tylko dane z lat 2015 i 2016. Głównym zadaniem autorki wewnątrz grupy była walidacja estymacji tła z dżetami QCD niepoprawnie zidentyfikowanymi jako leptony tau, w zdarzeniach, w których jeden lepton tau rozpada się na leptony, a drugi na hadrony i neutrino ('kanał rozpadu $\tau_{\text{lep}}\tau_{\text{had}}$). W tej rundzie analizy nie stwierdzono obecności nowych bozonów. Wyznaczono nowe limity wykluczające ich istnienie powyżej pewnych wartości iloczynu przekroju czynnego na produkcję bozonu i współczynnika rozgałęzienia bozonu na dwa leptony tau (limit niezależny od modelu) oraz w pewnych obszarach przestrzeni fazowej dwóch wersji modelu MSSM (limit zależny od modelu). W drugiej rundzie analizy wprowadzono ulepszenia strategii. Zdaniem autorki było stworzenie modeli uczenia maszynowego do rozróżniania sygnału i tła w kanale rozpadu $\tau_{\text{lep}}\tau_{\text{had}}$. Celem było uzyskanie poprawy czułości analizy względem używania standardowej zmiennej dyskryminacyjnej, którą była zmienna masowa m_T^{tot}. Wytrenowano zboostowane (wzmacniane) drzewa decyzyjne oraz sieci neuronowe sparametryzowane masą używając zmiennych masowych i kinematycznych. Najlepszą czułość uzyskano stosując zboostowane drzewa decyzyjne; oczekiwany limit niezależny od modelu został poprawiony o maksymalnie 67.5% i 28.6% dla zdarzeń z niskimi masami sygnałowymi, odpowiednio w stanach końcowych z i bez dżetów b. W momencie składania tej pracy analiza jest ciągle w toku.

W rozprawie poruszane są jeszcze dwa tematy. Pierwszym są teoretyczne rozważania o możliwym związku między problemem 'nienaturalnie dostrojonej' masy bozonu Higgsa i problemem stałej kosmologicznej, w kontekście warunków Veltmana i Pauliego na kasowanie się rozbieżności oraz w kontekście Modelu Standardowego jako efektywnej teorii pola. Omówiona jest również możliwa rola dodatkowych bozonów Higgsa. Ta część rozprawy jest skutkiem współpracy między autorką a Stevenem Bassem. Drugim tematem jest walidacja ATLAS-owych próbek Monte Carlo z leptonami tau. W ramach walidacji sprawdzane były: polaryzacja, współczynniki rozgałęzienia i rozkłady zmiennych kinematycznych leptonów tau oraz rozkłady zmiennych wejściowych do zboostowanego drzewa decyzyjnego używanego do identyfikacji leptonów tau w eksperymencie ATLAS.

Contents

Introduction

The development of the Standard Model, a theory describing elementary particles and three of the four fundamental interactions, was one of the great successes of 20th century physics. 2012 marked the year when the last particle predicted by the Standard Model - the elusive Higgs boson - was finally discovered at the Large Hadron Collider. This exciting discovery led to questions about the future of high-energy physics and an intensified focus on extensions of the Standard Model - so-called Beyond-the-Standard-Model (BSM) models. These models seek to explain phenomena which are not explained by the Standard Model, such as the existence of neutrino masses and the - apparently - 'unnaturally fine-tuned' mass of the Higgs boson. Historically, a lot of focus was put on supersymmetric extensions of the Standard Model. In addition to predicting a yet-undiscovered 'superpartner' particle for each Standard Model particle, the simplest supersymmetric model also predicts the existence of an additional Higgs doublet, resulting in a total of five Higgs bosons. These Higgs bosons continue to be the object of searches at the Large Hadron Collider. One of these searches, the search for heavy neutral Higgs bosons decaying to two tau leptons with the ATLAS experiment, will be described in this thesis.

The thesis is split into two parts: Background and Results. Background will cover the basics of theoretical and experimental (high-energy) particle physics, machine learning, and the workings of the Large Hadron Collider and the ATLAS detector. In Results, the author will discuss four areas of focus: theoretical musings about the possible link between the Higgs mass puzzle and the cosmological constant puzzle, validation studies of Monte Carlo samples with tau leptons used by the ATLAS experiment, and two contributions to the ATLAS search for additional neutral Higgs bosons decaying to two tau leptons - the validation of estimated background and the development of a machine learning model for differentiating between signal and background events.

Author contribution statement

The author's contributions to studies described in each chapter of the Results part of this thesis were as follows:

- **Chapter 6.** The author collaborated with Steven Bass on a theoretical paper published in Physics Letters B [1], with a longer discussion in conference proceedings published in Acta Physica Polonica [2].

- **Chapter 7.** The author conducted a series of Monte Carlo validation studies requested by the ATLAS Tau Combined Performance Group and the ATLAS Weak Boson Group, including one study done jointly with Joseph Patton. Results were documented in an internal note written by the author. Some results were also included in a paper written by the Weak Boson Group, which was undergoing internal review in preparation for journal submission at the time of submitting this thesis.

- **Chapter 8.** The author was involved in the first round of the Run 2 ATLAS BSM $H/A \to \tau^+\tau^-$ search. The author's tasks in the analysis group included validating background with QCD jets misidentified as tau leptons, producing ntuples, and contributing to ongoing analysis framework development. The analysis was published in Physical Review Letters [3]. The author also published a conference note about the analysis in Acta Physica Polonica [4] on behalf of the ATLAS Collaboration.

- **Chapter 9.** The author contributed to the second round of the Run 2 ATLAS BSM $H/A \to \tau^+\tau^-$ analysis by developing machine learning classifiers for the $\tau_{\text{lep}}\tau_{\text{had}}$ decay channel. The work involved training and evaluating the performance of boosted decision trees and

neutral networks, investigating different tau lepton 'working points', and analysis framework development. The analysis was ongoing at the time of submitting this thesis.

Notation

Throughout this thesis, particle categories (ex. electrons, quarks, fermions) will refer collectively to particles and their antiparticle counterparts, unless otherwise specified. Since much of the discussion will involve tau leptons and their many possible decay modes, the following notation is introduced to differentiate between various objects:

- $l^{(\pm)} \in \{e^{\pm}, \mu^{\pm}, \tau^{\pm}\}$ - any charged lepton,

- $\ell^{(\pm)} \in \{e^{\pm}, \mu^{\pm}\}$ - light charged lepton (electron or muon),

- $\tau^{(\pm)}$ - tau lepton,

- $\tau_{\mathbf{lep}}$ - tau lepton decaying to leptons,

- $\tau_{\mathbf{had}}$ - tau lepton decaying to hadrons and a neutrino,

- $\tau_{\mathbf{vis}}$ - 'visible' tau lepton decay products: all decay products except for neutrinos, which escape detection,

- $\tau_{\mathbf{had\text{-}vis}}$ - 'visible' products of a hadronic tau lepton decay.

Additionally, the speed of light in a vacuum is set to $c = 1$.

Part I

Background

Chapter 1

Theoretical Background - the Standard Model and Beyond

1.1 The Standard Model of particle physics

The Standard Model of particle physics is a theory developed mainly throughout the 1950's-1970's [5–21] which describes all known elementary particles and three of the four fundamental forces - electromagnetic, weak, and strong - which govern their interactions. Only the gravitational force, which is much weaker compared to the other three, is not included in the model. Elementary particles, in the current understanding, are point-like with no internal structure, but can have properties like mass, electric charge, and spin (a quantum number which is an intrinsic form of angular momentum). Figure 1.1 shows all elementary particles grouped by particle type. There are two main classes: fermions, which have half-integer spin, and are considered to be matter particles, and bosons, which have integer spin and act as force carriers.
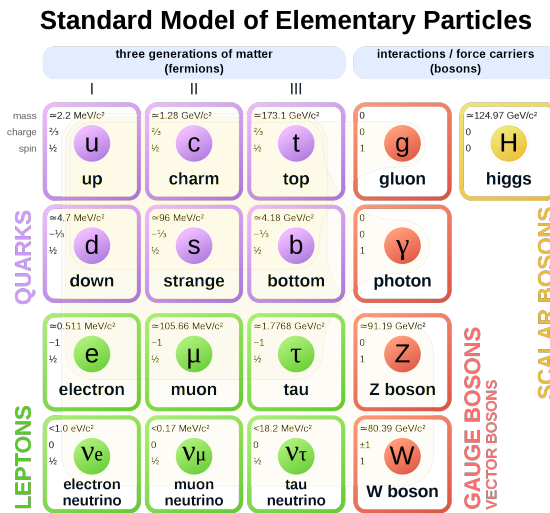
Figure 1.1: The Standard Model of elementary particles. [22]

On the fermion side, there are again two classes of particles: quarks and leptons. Quarks interact via all three forces described by the Standard Model and bind together to form composite (non-fundamental) particles called hadrons, which include the protons and neutrons that make up atomic nuclei. Leptons do not interact via the strong force; electrically-neutral leptons, called neutrinos, only interact via the weak force, while electrically-charged leptons interact weakly and electromagnetically. Both quarks and leptons are arranged in three 'generations' with increasing masses and decreasing lifetimes. In each lepton generation, there is an electrically-charged lepton with -1 charge (electron, muon, or tau lepton) and a neutrino of corresponding type ('flavor'). In

each quark generation, there is one 'up-type' quark with $+2/3$ electric charge and one 'down-type' quark with $-1/3$ electric charge, for a total of six quark flavors: up u, down d, charm c, strange s, top t, and bottom b. In addition, there are 'antimatter' equivalents of fermions, with the same mass but opposite electric charge and chirality - a property related to spin, which will be discussed in Section 1.1.4. With the exception of the electron, u quark, and d quark (and their antimatter equivalents), which are stable, charged fermions decay via the weak interaction, transforming into multiple other, lighter fermions.

On the boson side, there are five types of elementary particles: massless gluons, which carry the strong interaction; massless photons, which carry the electromagnetic interaction; massive Z and $W^{\pm}$ bosons, which carry the weak interaction; and finally massive Higgs bosons, fluctuations of the Higgs field involved in a spontaneous symmetry breaking process which generates masses of Standard Model particles. The Higgs boson is the only scalar (spin-0) boson in the Standard Model; other bosons, called gauge or vector bosons, are spin-1 particles.

Mathematically, the Standard Model is formulated in terms of gauge quantum field theory (QFT), a theory merging quantum mechanics and special relativity. The fundamental objects in QFT are not particles, waves, or wave functions (as in quantum mechanics) but fields - entities which have physical quantities defined for each point in spacetime. There is a field corresponding to each type of fermion shown in Figure 1.1 - an electron field, a u-quark field, and so forth. There is also a Higgs field. For vector bosons, the relationship is more complicated, as will be shown in Section 1.1.4. Since these fields are quantum rather than classical, they are operator-valued: they act on quantum states and generally do not commute. This non-commutativity makes boson self-interactions possible.

The dynamics of these fields and the quantum states they act on are described by the Lagrangian density (called the Lagrangian for short), a function of fields and their derivatives. The Standard Model Lagrangian was formulated using certain assumptions about internal symmetries of the model - degrees of freedom that do not correspond to changes in the physical state. The symmetry group (gauge group) of the Standard Model is $SU(3)_C \times SU(2)_L \times U(1)_Y$, where $SU(3)_C$ describes the symmetry of strong interactions and $SU(2)_L \times U(1)_Y$ describes the symmetry of electroweak interactions. Here, C refers to color charge, L refers to left-handed particles, and Y refers to weak hypercharge; these concepts will be discussed in later sections.
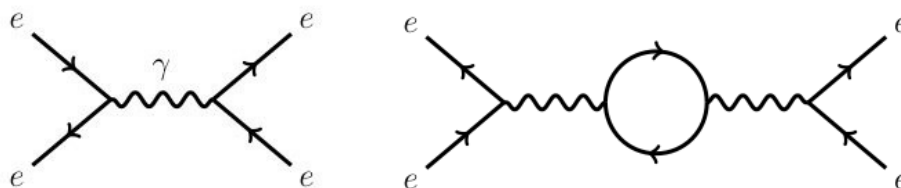
Figure 1.2: Examples of Feynman diagrams. The diagram on the left represents the interaction of an electron and a positron (anti-electron) involving the exchange of a virtual photon. The diagram on the right represents the same process, but with the virtual photon interacting with a virtual fermion-antifermion pair.

The Lagrangian can be used to calculate particle decay rates and cross-sections of different physics processes (probabilities of particle interactions). Both of these depend on two fundamental quantities: the transition amplitude from the initial state to the final state, also called the matrix element, and the available final state phase space. The calculations are done in the framework of perturbation theory, where solutions to problems that are too complex to be solved exactly are approximated with 'perturbed' versions of simpler, solvable problems. This can be formulated as a power series expansion and truncated at some order to give approximate solutions with acceptable accuracy. The recipe for calculating the matrix element is given by Feynman rules [23], encapsu-

lated in schematic drawings called Feynman diagrams. These drawings include lines for particles involved in the interaction (including force carriers) and vertices where the interaction takes place. Each line and vertex contributes an expression to the matrix element calculation. Lines leaving the diagram (not bounded by vertices) represent 'real', observable particles with mass m, energy E, and momentum p values obeying the energy-momentum relation [24]:

$$E^2 = (pc)^2 + (mc^2)^2, \tag{1.1}$$

where c is the speed of light in a vacuum. Internal lines bounded by vertices represent 'virtual' particles, transient quantum fluctuations which do not have to obey this relation, though they still obey the laws of conservation of energy and momentum. Virtual particles are created not only during interactions of real particles, but also spontaneously in the vacuum. A variation of Heisenberg's uncertainty principle [25, 26] states that there is a limit to the accuracy with which time and energy can be simultaneously determined:

$$\Delta t \Delta E \geq \frac{\hbar}{2}, \tag{1.2}$$

where $\hbar = 6.582... \times 10^{-22}$ MeV$\cdot$s is the reduced Planck constant [27] (this constant is determined from exact values and does not have an uncertainty). A consequence of this is that virtual particle-antiparticle pairs with an energy of E can be created out of the vacuum, if their lifetime does not exceed time t given by this relation. These virtual particles can interact with other real and virtual particles, and are represented in Feynman diagrams as loops. An example of a Feynman diagram with and without a loop is shown in Figure 1.2. The number of possible loops in a Feynman diagram for any given physics process is unlimited. Loop-less, tree-level Feynman diagrams are used as the first approximation in the calculation. Diagrams with increasing numbers of loops act as higher-order corrections (also called radiative corrections), increasing the accuracy. However, there is a problem: since virtual particles can carry arbitrarily high momenta, the corrections are logarithmically divergent; this is also referred to as ultraviolet divergence.

Luckily, there is a set of techniques for mathematically taming the divergences and obtaining physical results: regularization and renormalization [16–18]. Regularization modifies the loop integrals in such a way that the 'bare' parameters in the original Lagrangian, like coupling constants (parameters which describe the strength of interaction between particles) and particle masses, are finite. This can be accomplished in many ways. Two possibilities are cutoff regularization and dimensional regularization. Cutoff regularization assumes some maximum energy beyond which our understanding of physical laws is insufficient and loop integrals cannot be evaluated. This has a simple physical interpretation but is not Lorentz invariant. In the more widely used dimensional regularization, loop integrals are evaluated in $4 - \epsilon$ rather than four dimensions, where ϵ has a small nonzero value. The result depends on the regularizer, such as the cutoff or the value of ϵ. Renormalization (in some specific scheme) is then used to rewrite the theory in such a way that it depends only on experimentally-measured values of parameters, called physical or pole values, and not on the regularizer. This involves the introduction of counterterms to the bare parameters. Dimensional regularization is usually followed by renormalization in the minimal subtraction (MS) or modified minimal subtraction ($\overline{MS}$) scheme [28, 29]. In the MS scheme, only the divergent part of the radiative corrections is absorbed into the counterterms. In $\overline{MS}$, a universal constant which always appears in divergent Feynman diagram calculations is also absorbed into the counterterms. It should be emphasized that in renormalizable quantum field theories, physical results should not depend on the regularization scheme. A consequence of these procedures is that Standard Model parameters are not true constants, but 'running constants': their value depends on the energy scale at which they are measured. This running is encoded in beta functions (also called renormalization group equations):

$$\beta(g) = \frac{\partial g}{\partial \log(\mu)}, \tag{1.3}$$

where g is some running constant and μ is the energy scale.

1.1.1 The electromagnetic interaction

The electromagnetic interaction describes interactions between electrically-charged particles mediated by a massless, chargeless boson: the photon. This interaction can be described by a rela-

tivistic, Abelian gauge theory with the symmetry group $U(1)_{EM}$, called Quantum Electrodynamics (QED) [30]. The QED Lagrangian describing the interaction between a charged fermion field ψ and the electromagnetic (photon) field A_μ takes the following form:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \qquad (1.4)$$

Here, μ and ν are four-vector spacetime coordinates, $\gamma^\nu = \{\gamma^0, \gamma^1, \gamma^2, \gamma^3\}$ are Dirac (gamma) matrices, $\bar{\psi} = \psi^\dagger\gamma^0$ is the Dirac adjoint of ψ, $D_\mu = \partial_\mu + ieA_\mu + ieB_\mu$ is the gauge covariant derivative, m is the mass of the charged fermion, e is the elementary charge (the electric charge of the electron), A_μ is the covariant four-potential of the electromagnetic field generated by the fermion, B_μ is the external electromagnetic field, and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field tensor. The dimensionless coupling constant describing the strength of the coupling between charged particles and photons is the fine-structure constant α:

$$\alpha = \frac{e^2}{4\pi\hbar c\epsilon_0}, \qquad (1.5)$$

where ϵ_0 is the vacuum permittivity. The fine-structure constant has a value of approximately $\frac{1}{137}$ [27] at very low energy scales.

1.1.2 The strong interaction

The strong interaction gives rise to the strong nuclear force, which binds quarks together in composite particles called hadrons and binds nucleons in atomic nuclei. In this way, it is responsible for holding most ordinary matter together. The strong interaction can be described by a non-Abelian (non-commutative) gauge theory called Quantum Chromodynamics (QCD). This gauge theory is based on the gauge symmetry group $SU(3)_C$. The strong force carrier is the gluon, a massless boson analogous to the photon in QED. Much like electromagnetically-interacting particles carry electric charges, strongly-interacting particles carry 'color' (C) charges - intrinsic properties which govern how these particles interact, but have nothing to do with visual perception. There are a few important differences between electric and color charges. For one thing, color charges come in six types, called red, antired, green, antigreen, blue, and antiblue. Quarks can carry only one color charge at a time: either red, green, or blue. Likewise, antiquarks can carry either antired, antigreen, or antiblue charge. The color of a quark or antiquark is unrelated to its flavor. While the photon does not carry electric charge, the gluon carries a mixture of color charges: one color charge and one anticolor charge. While this gives, in principle, nine possible color-anticolor combinations, gluons actually form a color octet under $SU(3)_C$.

Strongly-interacting particles exhibit color confinement [31]. This means that they form composite states - hadrons - where the total color charge is 'white', in analogy to visible light. The simplest ways for this to occur are to have a bound state of a quark and an antiquark with complimentary color-anticolor charges or three quarks with one of each color charge. The first type of bound state is called a meson, while the second type is called a baryon. Since gluons exhibit self-interaction, they should in principle be able to form bound states called glueballs, but these are difficult to detect experimentally [32]. Quarks and gluons cannot exist in the free state. If separated from bound states or created in physics processes, they rapidly undergo hadronization, forming hadrons with quarks and gluons created from the vacuum. The only exception to this is the t quark, which decays via the weak force before it can hadronize. The process of hadronization is non-perturbative (it cannot be described by perturbation theory) but can be modeled fairly accurately with phenomenological models.

Another important property of QCD is asymptotic freedom [19, 20]: the strength of the strong interaction decreases asymptotically as the length scale decreases and the energy scale increases. This allows for the confinement of quarks and gluons within hadrons at low energies; however, at high energies quarks interact weakly enough that perturbative calculations are possible. Hadrons can be thought of as containing valence quarks (the quark composition that gives a hadron its 'identity', such as two u quarks and one d quark for the proton) and a sea of gluons and quark-antiquark pairs created due to gluon splitting. These pairs can also annihilate back into gluons, and the two opposing processes are in constant flux. Non-valence quarks are referred to as sea

quarks.

The QCD Lagrangian takes the following form:

$$\mathcal{L}_{QCD} = \bar{\psi}_i (i(\gamma^\mu D_\mu)_{ij} - m\delta_{ij})\psi_j - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a. \tag{1.6}$$

Here, ψ_i are quark fields and $(D_\mu)_{ij}$ is the covariant derivative:

$$(D_\mu)_{ij} = \delta_{ij}\partial_\mu - ig_s\lambda^a_{ij}G^a_\mu, \tag{1.7}$$

where λ^a_{ij} are Gell-Mann matrices, g_s is related to the strong coupling constant α_s via $g_s = \sqrt{4\pi\alpha_s}$, and G^a_μ are gluon fields with color-anticolor combination a. Finally, $G^a_{\mu\nu}$ is the gluon field strength tensor:

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^a_{bc} G^b_\mu G^c_\nu, \tag{1.8}$$

where f^{abc} are structure constants of the $SU(3)$ symmetry group.

1.1.3 The weak interaction

The weak interaction is responsible for the radioactive decay of atoms, including nuclear fission. It affects all fermions - particles with 'flavor' - and is mediated by the massive Z and $W^\pm$ bosons. This interaction can be described by a non-Abelian gauge theory called Quantum Flavordynamics (QFD), but is better understood in the framework of electroweak theory, which will be discussed in Section 1.1.4.

There are two types of weak interaction: neutral-current and charged-current interaction. The first type is mediated by the electrically-neutral Z boson, which does not change the electric charge of the fermions. The second type is mediated by electrically-charged $W^\pm$ bosons, which allows the interacting fermions to change their electric charge and flavor. The charged-current weak interaction is the only type of Standard Model interaction which allows the flavor of fermions to change; flavor-changing neutral currents are forbidden at tree level in the Standard Model. In a scenario where fermion mass eigenstates and flavor eigenstates are the same, flavor changes would only be allowed to occur within fermion generations, ex. a u quark (in a bound state) changing to a d quark through the emission of a $W^\pm$ boson, or an electron changing to an electron neutrino. In reality, the mass and flavor eigenstates of quarks are different, causing quarks to exhibit flavor mixing, which is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix V [33, 34]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{1.9}$$

The vector on the left-hand side of the equation contains flavor eigenstates, while the vector on the right-hand side contains mass eigenstates. The usage of down-type quarks in this equation is a convention; a formulation with up-type quarks would be equally valid. The mixing is quite small. Measurements currently put the magnitudes of CKM matrix elements at the following values [27]:

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.0410 \pm 0.0014 \\ 0.0080 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.030 \end{pmatrix}. \tag{1.10}$$

The squares of these values correspond to transition probabilities.

A similar mixing occurs for neutrinos. This was first postulated by Pontecorvo [35] and later discovered by the Super-Kamiokande experiment [36], which observed neutrino oscillations: spontaneous flavor changes in neutrinos propagating through space [37]. Neutrino mixing is described by the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix U [38], which links flavor eigenstates $\nu_{e,\mu,\tau}$ and mass eigenstates $\nu_{1,2,3}$:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \tag{1.11}$$

The implication is that neutrinos have non-zero masses, though they must be very small - experimental upper limits have been set at sub-eV levels [27]. The magnitudes of PMNS matrix elements are quoted as 3σ (99.7%) confidence level ranges [39]:

$$\begin{pmatrix} |U_{e1}| & |U_{e2}| & |U_{e3}| \\ |U_{\mu1}| & |U_{\mu2}| & |U_{\mu3}| \\ |U_{\tau1}| & |U_{\tau2}| & |U_{\tau3}| \end{pmatrix} = \begin{pmatrix} 0.801...0.845 & 0.513...0.579 & 0.143...0.156 \\ 0.233...0.507 & 0.461...0.694 & 0.631...0.778 \\ 0.261...0.526 & 0.471...0.701 & 0.611...0.761 \end{pmatrix}. \qquad (1.12)$$

The degree of mixing is larger compared to quarks.

The weak interaction is also the only fundamental interaction that violates parity (P) symmetry, the symmetry under an inversion of spacial coordinates:

$$\psi(t, x, y, x) \rightarrow \psi(t, -x, -y, -z). \qquad (1.13)$$

This was first discovered in the Wu experiment [40], where electrons from the beta decay of cobalt nuclei were found to be emitted preferentially in the direction of the nucleus spin vector. Since the nucleus spin vector is invariant under the P transformation and the direction of an emitted electron is not, in a scenario with conserved P symmetry the electrons would be emitted isotropically. The implication is that $W^{\pm}$ and Z bosons couple differently to fermions based on their chirality: a Lorentz-invariant property related to particle spin. For massless particles, which move at the speed of light, chirality is the same as helicity: it is the sign of the projection of a particle's spin vector onto its momentum vector. Particles are right-handed (have positive chirality) if their spin is in the same direction as their motion, and left-handed (have negative chirality) if the opposite is true. For massive particles, chirality and helicity must be distinguished, since changes in the observer's frame of reference can reverse a particle's direction of motion, changing the helicity, but not the Lorentz-invariant chirality.

Experimentally, it appears that P symmetry is maximally violated in charged-current interactions, with $W^{\pm}$ bosons coupling exclusively to left-handed fermions and right-handed antifermions. Since neutrinos are produced only in weak interactions, including in weak decays of charged leptons, this means that only left-handed neutrinos and right-handed antineutrinos can exist. This is problematic, because it precludes the construction of a renormalizable (dimension-4) neutrino mass term in the Standard Model Lagrangian [41]. The lack of this term renders neutrinos massless in the theory, in conflict with neutrino oscillation observations. The neutrino mass puzzle remains an open problem in particle physics. Charged leptons and quarks, on the other hand, come in both right-handed and left-handed varieties. In the mathematical description, this is handled by stating that left-handed fermions form doublets under $SU(2)$ symmetry group transformations, while right-handed fermions form singlets; conversely, left-handed antifermions form singlets, while right-handed antifermions form doublets. In neutral current interactions, the Z boson can couple to both left-handed and right-handed fermions, but the strength of the coupling is not the same, as will be shown in the next section.

In addition to violating P symmetry, weak interactions also violate charge-parity (CP) symmetry, the symmetry under a simultaneous inversion of spacial coordinates and the electric charge. This was first discovered in experiments involving neutral kaons and antikaons (mesons containing an s quark or antiquark) [42]. These two types of particle can oscillate into each other. If CP symmetry were conserved, the probabilities of a kaon oscillating into an antikaon and vice versa would be the same. However, it was found that the probabilities are slightly different. Mathematically, the CP violation can be accounted for by adding a complex phase to the CKM matrix. The CP symmetry violation in the Standard Model is small, but it does mean that matter and antimatter (which are CP conjugates) are not treated the same way in the theory.

1.1.4 The electroweak interaction

The electromagnetic and weak interactions can be unified under the Glashow-Weinberg-Salam theory [6,7,14], a non-Abelian gauge theory with the $SU(2)_L \times U(1)_Y$ symmetry group. The subscript L refers to left-handed particles, which transform as doublets under $SU(2)$ transformations; right-handed particles form singlets. The generators of this group are called weak isospin T and weak

hypercharge Y. They are related to the electric charge Q via the following relation:

$$Q = T_3 + \frac{1}{2}Y, \tag{1.14}$$

where T_3 is the third component of the weak isospin. These generators give rise to four massless gauge boson fields: W_1, W_2, W_3, and B. Through the process of spontaneous symmetry breaking, part of the Higgs mechanism described in Section 1.1.5, these massless fields can combine to create the physical massive bosons Z and $W^\pm$ and the physical massless boson γ (the photon). The electrically-neutral physical bosons are combinations of W_3 and B fields:

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_3 \\ B \end{pmatrix}, \tag{1.15}$$

where θ_W is the Weinberg angle or weak mixing angle. The electrically-charged physical bosons are combinations of the W_1 and W_2 fields:

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2). \tag{1.16}$$

The (degenerate) masses of the $W^\pm$ bosons and the mass of the Z boson, created during spontaneous symmetry breaking, are related at tree level through the Weinberg angle:

$$m_W = m_Z \cos\theta_W. \tag{1.17}$$

The Lagrangian of the electroweak interaction has the following form:

$$\mathcal{L}_{EW} = g\vec{J}_\mu \cdot \vec{W}_\mu + g' J_\mu^Y B_\mu, \tag{1.18}$$

where $\vec{J}_\mu$ and J_μ^Y are currents carrying the weak isospin and weak hypercharge, respectively, while g and g' are coupling constants related to each other and to the electromagnetic coupling constant via

$$g \sin\theta_W = g' \cos\theta_W = \alpha. \tag{1.19}$$

Following from Equation 1.14, J_μ^Y can be rewritten in terms of the electromagnetic current and the third component of $\vec{J}_\mu$:

$$J_\mu^Y = 2(J_\mu^{EM} - J_\mu^{(3)}). \tag{1.20}$$

Then, $\mathcal{L}_{EW}$ can also be rewritten in terms of physical currents:

$$\mathcal{L}_{EW} = \frac{g}{\sqrt{2}}(J_\mu^- W_\mu^+ + J_\mu^+ W_\mu^-) + \frac{g}{\cos\theta_W}(J_\mu^{(3)} - \sin^2\theta_W J_\mu^{EM})Z_\mu + g\sin\theta_W J_\mu^{EM} A_\mu, \tag{1.21}$$

where $J_\mu^\pm = J_\mu^{(1)} \pm iJ_\mu^{(2)}$ are currents associated with $W^\pm$ exchange. Here, the first term describes the couplings of the $W^\pm$ bosons with strength g, the last term describes the coupling of the photon with strength α, and the middle term describes the couplings of the Z boson, which has both weak and electromagnetic components. The left-handed and right-handed couplings of the Z boson can be written in the following way:

$$g_L = T_3 - Q\sin^2\theta_W, \ \ g_R = -Q\sin^2\theta_W, \tag{1.22}$$

but more commonly vector V and axial-vector (pseudovector) A couplings are used instead:

$$c_V = g_L + g_R = T_3 - 2Q\sin^2\theta_W, \ \ c_A = g_L - g_R = T_3. \tag{1.23}$$

The dependence on T_3 and Q means that the Z boson couples differently to different fermion types. Since left-handed fermions and right-handed antifermions form doublets, their weak isospin values are $T = \frac{1}{2}$ and $T_3 = \pm\frac{1}{2}$; right-handed fermions and left-handed antifermions form singlets with $T = 0$, $T_3 = 0$. The couplings following from this are listed in Table 1.1. The effective value of the Weinberg angle, which includes radiative corrections not present in the tree-level formula in Equation 1.17, can be accessed experimentally at particle colliders by measuring the forward-backward and left-right asymmetries of Z boson decay products [43].

Table 1.1: Couplings of the Z boson to different fermion types.

(anti)fermion type	$2c_V$	$2c_A$
neutrino	1	1
antineutrino	1	-1
charged lepton	$-1 + 4\sin^2\theta_W$	-1
charged antilepton	$-1 + 4\sin^2\theta_W$	1
up-type quark	$1 - \frac{8}{3}\sin^2\theta_W$	1
up-type antiquark	$1 - \frac{8}{3}\sin^2\theta_W$	-1
down-type quark	$-1 + \frac{4}{3}\sin^2\theta_W$	-1
down-type antiquark	$-1 + \frac{4}{3}\sin^2\theta_W$	1

1.1.5 The Higgs mechanism and the Higgs boson

The problem of mass generation in theoretical particle physics is complicated for several reasons. One is that mass mediates the gravitational interaction between bodies, and theories of gravity are non-renormalizable and have yet to be reconciled with quantum theory and the Standard Model. Another is that explicit mass terms in the Lagrangian are manifestly non-gauge invariant. This problem was solved in the Standard Model by introducing the Higgs mechanism, also called the Englert–Brout–Higgs–Guralnik–Hagen–Kibble mechanism in honor of all researchers in the three groups that independently proposed it in 1964 [10–13]. This is a gravity-free process in which gauge bosons can dynamically acquire mass, a property they are experimentally known to have. In the simplest formulation, the Higgs mechanism adds a single scalar field (the Higgs field), which develops a non-zero vacuum expectation value below a certain (very high) critical energy. This process breaks the $SU(2)_L \times U(1)_Y$ gauge symmetry associated with the electroweak force, splitting up the electromagnetic and weak forces, and is an example of spontaneous symmetry breaking, as it does not involve an external force.

The Higgs field is a weak isospin doublet $\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ with electrically-charged and electrically-neutral complex components, which acts as a scalar under Lorentz transformations and has the following Lagrangian:

$$\mathcal{L}_\phi = |D_\mu \phi|^2 - V(\phi). \tag{1.24}$$

Here, D_μ is the covariant derivative:

$$D_\mu = \partial_\mu + \frac{ig}{2}\sigma^a W_\mu^a + \frac{ig'}{2}YB_\mu, \tag{1.25}$$

where σ^a, $a = 1, 2, 3$ are Pauli matrices, g and g' are coupling constants associated with $SU(2)_L$ and $U(1)_Y$ gauge symmetries (respectively), Y is the weak hypercharge, and W^a and B are the massless gauge boson fields mentioned in Section 1.1.4, while $V(\phi)$ is the energy potential:

$$V(\phi) = \mu^2 |\phi|^2 + \lambda |\phi|^4 \tag{1.26}$$

with real constants $\mu^2 < 0$ and $\lambda > 0$. This potential function has four degrees of freedom. When visualized in three dimensions (two dimensions in polar coordinates) as in Figure 1.3, it is shaped like a sombrero, and thus frequently referred to as the 'Mexican hat potential'. The brim of the sombrero corresponds to an infinite number of minima, all at the same nonzero value. Choosing one of these minima - necessary for calculating perturbative expansions - breaks the symmetry of the ground state (the vacuum), but does not affect the symmetry of the Lagrangian. This is equivalent to a breaking of the electroweak gauge symmetry:

$$SU(2)_L \times U(1)_Y \rightarrow U(1)_{EM}. \tag{1.27}$$

When the symmetry of the vacuum is broken, it can be said that the Higgs field acquires a nonzero vacuum expectation value v. This value is related to the Fermi constant G_F (a constant denoting the strength of weak four-fermion interaction in radioactive beta decay), the mass of the $W^\pm$ boson and the weak coupling constant g through the following expressions:

$$v^2 = \frac{(\hbar c)^3}{\sqrt{2} G_F} = \frac{4 m_W^2}{g^2}. \tag{1.28}$$

Since the value of the Fermi constant has been measured with high accuracy in muon lifetime experiments [44], it is used to calculate the value of v: about 246 GeV [27].
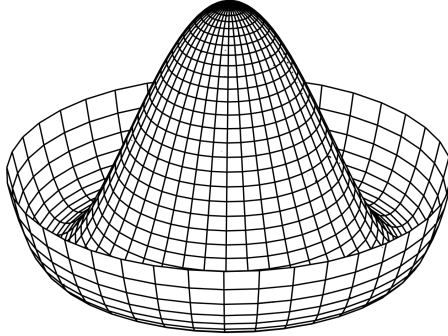
Figure 1.3: A three-dimensional Mexican hat or sombrero potential function.

According to Goldstone's theorem [8], instances of the spontaneous breaking of continuous symmetries lead to the appearance of massless bosons called Goldstone bosons - one for each generator of the symmetry group. In theories with a local gauge symmetry, the degrees of freedom corresponding to Goldstone bosons are absorbed by gauge bosons, becoming new longitudinal polarization states and leading the gauge bosons to acquire mass. In the Higgs sector of the Standard Model, the vacuum expectation value can be rotated, without losing generality, to take the following form:

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \; v = \sqrt{-\frac{\mu^2}{\lambda}}. \tag{1.29}$$

To develop the theory perturbatively around this minimum, four new fields can be introduced: $\xi_{1,2,3}$ and H. The Higgs field can then be written in the following form:

$$\phi = \begin{pmatrix} \xi_1 + i\xi_2 \\ \frac{1}{\sqrt{2}}(v+H) - i\xi_3 \end{pmatrix} = e^{\frac{i\xi_a \sigma^a}{2v}} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v+H) \end{pmatrix}. \tag{1.30}$$

The three fields $\xi_{1,2,3}$ correspond to three Goldstone bosons, which give rise to mass terms for the $W^{\pm}$ and Z bosons in $\mathcal{L}_\phi$. The acquired masses have the following form:

$$m_W^2 = \frac{g^2 v^2}{4}, \; m_Z^2 = \frac{(g^2 + g'^2)^2 v^2}{4}, \tag{1.31}$$

which, combined with the relation

$$\alpha = g \sin \theta_W = g' \cos \theta_W, \tag{1.32}$$

leads to the mass relation in Equation 1.17. The $U(1)_{EM}$ symmetry is conserved, and the photon does not acquire a mass. Instead, the fourth degree of freedom, H, corresponds to a excitation of the Higgs field: the Higgs boson, which does acquire a mass. The Lagrangian of the Higgs field describing its mass and self-interaction is then of the following form:

$$\mathcal{L}_H = \frac{1}{2} \partial_\mu H \partial^\mu H - \lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4} H^4. \tag{1.33}$$

The value of λ, and consequently the Higgs boson mass, is not predicted by the Standard Model. However, in 2012 a particle consistent with the Standard Model Higgs boson was discovered by the ATLAS and CMS experiments at the LHC with a mass of around 125 GeV [45, 46]. This value is interesting, because - combined with t quark mass measurements - it places the electroweak vacuum at the border between stability and metastability. This will be discussed in more detail in

Section 1.1.6 below.

The Higgs mechanism can also give mass to electrically-charged fermions through the Yukawa coupling [14, 47]:

$$\mathcal{L}_{\text{Yukawa}} = -\sqrt{2} m_f (\bar{\psi}_R \phi^\dagger \psi_L + \bar{\psi}_L \phi \psi_R), \tag{1.34}$$

where ψ_L and ψ_R are left-handed and right-handed fermion fields, respectively, and m_f is the fermion mass. Because all neutrinos (neutral fermions) seem to be left-handed, as mentioned in Section 1.1.3, they cannot obtain mass through this mechanism, unless some yet-unobserved right-handed neutrinos also exist [48].

1.1.6 The electroweak vacuum

In the Mexican hat potential function shown in Figure 1.3, the brim of the hat represents an infinity of states with the lowest possible energy - a true minimum. However, this shape depends on the Higgs parameter λ remaining positive at all energy scales. If it were to become negative at some energy scale, the shape of the Higgs potential would start to resemble an inverted parabola, with lower energy values located past the local minimum of the brim. This would mean that the electroweak vacuum, residing in what used to be the brim, is not a true vacuum - it could decay to an even lower energy state via quantum tunneling or due to some high-energy event. This is called a false vacuum decay and would have catastrophic consequences, destroying the Universe as we know it.
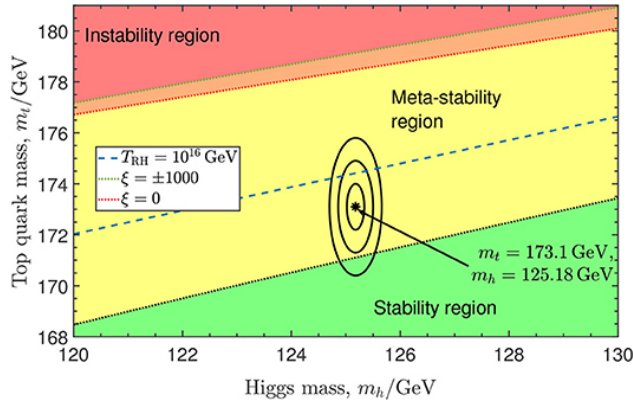
Figure 1.4: An electroweak vacuum stability diagram showing 1σ (63.3%), 2σ (95.5%), and 3σ (99.7%) confidence intervals based on m_t and m_H measurements. The contour center, corresponding to central mass values, is in the metastability region; a stable vacuum is still allowed. [49]

The renormalization group equation running of λ depends very strongly on the Yukawa coupling of the t quark y_t [50]:

$$(4\pi)^2 \frac{d\lambda}{d\ln\mu} = -6y_t^4 + \frac{9}{8}g^4 + \frac{27}{200}g'^4 + \frac{9}{20}g^2 g'^2 + \lambda(12y_t^2 - 9g^2 + \frac{9}{5}g'^2) + 24\lambda^2 + \text{higher-order corrections}. \tag{1.35}$$

Depending on the value of the t quark Yukawa coupling in relation to other coupling constants, λ can be positive at all energy scales (stable vacuum), negative at all energy scales (unstable vacuum), or switch from positive to negative at some energy scale (metastable vacuum). As it turns out, the experimentally measured Higgs boson and t quark masses put the electroweak vacuum right at the border of stability and metastability [49, 51], as shown in Figure 1.4. If the vacuum is indeed metastable, it will eventually decay via bubble nucleation [52]: false vacuum decay in any one region of the Universe will manifest itself as the creation of a bubble of true vacuum, which will expand throughout the Universe at the speed of light. In this scenario, it is unclear why the vacuum did not already decay during the early stages of the Universe or in response to high-energy cosmic rays.

14

1.1.7 Physics of tau leptons

Tau leptons have the richest physics out of the three lepton generations. As the heaviest leptons, they have have $290.3 \pm 0.5 \times 10^{-15}$ s [27] and the largest selection of decay modes. Since their mass of 1776.86 ± 0.12 MeV [27] is above the masses of light hadrons such as pions π and kaons K (mesons composed of combinations of u, d, and s quarks and antiquarks), they can decay to these hadrons as well as to lighter leptons. Decays to light leptons are called leptonic decays, while decays to a tau antineutrino and hadrons are called hadronic decays. Both decay types involve the $W^{\pm}$ boson, as shown in Figure 1.5, and thus violate P symmetry. The main tau lepton decay modes are listed in Table 1.2. Hadronic decays can proceed via the creation of a meson resonance (a short-lived meson) which decays to more stable mesons. These 'intermediate' mesons are omitted from Table 1.2; for example, the decay $\tau^+ \to \rho^+(770)\nu_\tau \to \pi^+\pi^0\nu_\tau$ is classified as simply $\tau^+ \to \pi^+\pi^0\nu_\tau$. The properties of selected mesons (including resonances) are given in Table 1.3.
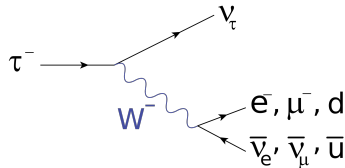
Figure 1.5: Feynman diagram showing the decay of a tau lepton to a tau antineutrino and two additional leptons or quarks. The d quark can mix to s and b quarks with probabilities defined by the CKM matrix in Equation 1.9.

The short lifetime and P-symmetry-violating decays are convenient for polarization measurements at particle colliders [53]. The polarization P_τ is the asymmetry between the cross-sections σ for left-handed and right-handed tau lepton production in some process:

$$P_\tau = \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L}. \tag{1.36}$$

Information about the polarization of tau leptons is encoded in the spin correlations of the decay products, with the possibility to define various observables. A typical processes of interest is Z boson decay to two tau leptons; the polarization of tau leptons from such decays can be used to calculate the value of the Weinberg angle, as described in Section 1.1.4. The polarization of fermions P_f from Z boson decays is related to the vector and axial-vector couplings c_V and c_A in the following way [43]:

$$P_f = \frac{2c_V c_A}{c_V^2 + c_A^2}. \tag{1.37}$$

Polarization measurements can also be used to distinguish between the decays of Z bosons and Standard Model Higgs bosons, which decay to unpolarized tau leptons, or to investigate the properties of new particles.

1.1.8 Free parameters in the Standard Model

The Standard Model has a number of free parameters: charged fermion and Higgs boson masses, quark and neutrino mixing angles, gauge couplings, and the Higgs vacuum expectation value. These values of these parameters are not predicted by the model and must be determined experimentally, which always involves an amount of uncertainty. In order to calculate Feynman diagrams, accurate values must be used as 'input parameters'. Thanks to relationships between various (running) constants of the Standard Model, there is some freedom in choosing the set of these input parameters. A popular choice is the set of fermion and Higgs boson masses, mixing angles, and three of five electroweak parameters: the Fermi constant G_F, the fine structure constant α, the Weinberg angle θ_W (usually in the form $\sin^2\theta_W$), and the Z and $W^{\pm}$ boson masses. Typically, the set $\{G_F, \alpha, m_Z\}$ is used [27], as the values of these parameters have been measured with the largest accuracy. Table 1.4 lists the values of selected parameters at specified energies. The running of gauge couplings and masses will be discussed in Chapter 5.

Table 1.2: Particle Data Group values of branching ratios of selected tau lepton decay modes. The decay modes are sorted by the number of prongs (charged decay products - leptons or hadrons), hadronic and leptonic modes, the two most common leptonic decay modes, and hadronic modes in $Nh^\pm\ Mh^0$ format, where $h^\pm$ represents charged pions and/or kaons and h^0 represents neutral pions and/or kaons. [27]

decay mode	branching ratio
1-prong	84.71±0.07%
3-prong	115.20±0.08%
5-prong	0.102±0.004%
leptonic	35.21±0.08%
hadronic	64.8±0.2%
$\tau^\pm \to e^\pm \nu_e \nu_\tau$	17.82±0.04%
$\tau^\pm \to \mu^\pm \nu_\mu \nu_\tau$	17.39±0.04%
$1h^\pm\ 0h^0$	12.03±0.05%
$1h^\pm\ 1h^0$	25.93±0.09%
$1h^\pm\ 2h^0$	9.48±0.10%
$1h^\pm\ 3h^0$	1.18±0.07%
$3h^\pm\ 0h^0$	9.46±0.05%
$3h^\pm\ 1h^0$	4.76±0.05%
$3h^\pm\ 2h^0$	0.51±0.03%
$3h^\pm\ 3h^0$	0.21±0.03%

1.2 Beyond the Standard Model

The Standard Model has been remarkably successful, making many correct, testable predictions. So far, no statistically-significant discrepancies have been found between values predicted by the Standard Model and those measured in experiments. However, it is clear that the Standard Model cannot be the whole story of particle physics, as there are certain phenomena that it simply does not account for. One example is the neutrino mass problem, which was discussed in Section 1.1.3. Some other examples are discussed below.

- **Matter-antimatter asymmetry**
  Also called the baryon asymmetry problem, this is the observation that the Universe is composed almost exclusively of matter, despite the expectation that equal amounts of matter and antimatter should have been produced in the Big Bang (and subsequently annihilated) [54]. While the Standard Model does treat matter and antimatter on a slightly unequal footing due to a small CP symmetry violation (discussed in Section 1.1.3), this effect is not sufficient to explain the observed asymmetry [55, 56]. In addition, the Standard Model strictly conserves the baryon number (also called the quark number) $B = \frac{1}{3}(n_q - n_{\bar{q}})$ [43]; this conservation would need to be violated in order for baryon asymmetry to develop.

- **Gravity**
  The only fundamental interaction not described by the Standard Model is gravity. It is by far the weakest fundamental interaction - 38 orders of magnitude weaker than the strong force, 36 orders of magnitude weaker than the electromagnetic force, and 29 orders of magnitude weaker than the weak force. The currently accepted theory of gravity is Einstein's general relativity [57,58]. This theory states that gravity is not a force, but the result of the warping of spacetime by objects possessing mass and/or momentum. Objects moving through spacetime move along geodesic lines - straight lines in the local inertial frames of these objects. While general relativity has been very successful, making accurate predictions about the existence of time dilation, gravitational redshift, and gravitational waves (first observed in 2016 [59]), it is non-renormalizable and cannot be reconciled with quantum mechanics or the Standard Model [60]. There are ongoing attempts to construct a theory of quantum gravity. One of the difficulties is the lack of experimental results to guide theorists, as quantum gravitational effects are expected to only appear close to Planck energy scale of 10^{19} GeV, some 16 orders of magnitude above energies available at particle colliders. The Planck energy E_{Pl} is a 'natural unit' of energy derived in terms of three universal physical constants - the speed of light in

Table 1.3: Valence quark compositions, electric charges, masses, lifetimes, and main decay modes of selected mesons [27]. Charged mesons have antimeson (CP conjugate) counterparts with opposite electric charge and switched quark-antiquark content (ex. $\bar{u}d$ instead of $u\bar{d}$), but the same mass and lifetime. Antimesons have the same decay modes as mesons, but with CP conjugates of daughter particles. The neutral kaon K^0 does not have a listed lifetime or decay mode, because it is not a weak eigenstate. In this case, the weak eigenstates (states which have definite lifetimes when decaying via the weak force) are combinations of K^0 and $\bar{K}^0$, called K_L (long-lived) and K_S (short-lived), which can decay into 2-3 pions or a pion, a charged lepton, and a neutrino.

meson	charge	comp.	mass [MeV]	lifetime [s]	main decay modes
π^+	+1	ud	139.57018±0.00035	$(2.6033\pm0.0005)\times10^{-8}$	$\mu^+\nu_\mu$
π^0	0	$\frac{u\bar{u}-d\bar{d}}{\sqrt{2}}$	134.9766±0.0006	$(8.52\pm0.18)\times10^{-17}$	$\gamma\gamma$
K^+	+1	$u\bar{s}$	493.677±0.016	$(1.2380\pm0.0021)\times10^{-8}$	$\mu^+\nu_\mu,\ \pi^+\pi^0$ $\pi^0 e^+\nu_e,\ \pi^+\pi^+\pi^-$
K^0	0	$d\bar{s}$	497.614±0.024	-	-
$\rho^+(770)$	+1	$u\bar{d}$	775.11±0.34	$(4.41\pm0.02)\times10^{-24}$	$\pi^+\pi^0$
K^{*+}	+1	$u\bar{s}$	891.66±0.26	$(3.26\pm0.06)\times10^{-23}$	$K^+\pi^0$

Table 1.4: Particle Data Group values of gauge boson masses, the Higgs boson mass, the fine-structure constant, the strong coupling constant, the Fermi constant, and $\sin^2\theta_W$, quoted at particular energy scales μ. [27]

parameter	value	μ
m_Z	91.1876 ± 0.0021 GeV	m_Z
m_W	80.379 ± 0.012 GeV	m_W
m_H	125.10 ± 0.14 GeV	m_H
α	$1/(137.035999084 \pm 0.000000021)$	0 GeV
g_s	0.1179 ± 0.0010	m_Z
$G_F/(\hbar c)^3$	$(1.1663787 \pm 0.0000006) \times 10^{-5}$ GeV^{-2}	0 GeV
$\sin^2\theta_W$	0.23121 ± 0.00004	m_Z

a vacuum c, the reduced Planck constant $\hbar$, and the gravitational constant G [61]:

$$E_{Pl} = \sqrt{\frac{\hbar c^5}{G}}. \tag{1.38}$$

- **Dark matter**
  Dark matter is a postulated unknown form of non-luminous matter which could account for approximately 85% of matter in the Universe [62–64]. Experimental evidence for the existence of this matter includes observations of galaxy rotation curves, velocity dispersion in galaxies, gravitational lensing, and the distribution of Cosmic Microwave Background, which cannot be explained by visible matter alone in the framework of general relativity. An alternative approach to explaining these discrepancies is the modification of the theory of gravity [65, 66].

- **Hierarchy problems**
  Hierarchy problems occur when there are large discrepancies between parameter values that cannot be explained by the model or when a surprising ('unnatural') amount of parameter fine-tuning is needed to produce values consistent with experimental results. One hierarchy problem is the vast difference between the strength of the gravitational force and the other three fundamental forces. Another, in cosmology, is the cosmological constant problem: the difference between the very small value of the vacuum energy density required to explain the accelerating expansion of the Universe and the very large theoretical value of zero-point energy (the energy of the vacuum associated with quantum fluctuations). In the Standard Model, the most significant hierarchy problem is the Higgs mass puzzle. Because of large quadratic radiative corrections, the mass of the Higgs boson would be expected to be very large, possibly near the Planck scale; however, the measured mass is only about 125 GeV, implying a fine-tuned cancellation between the corrections and the bare mass of the boson

[43]. This problem is central to the activities at ATLAS and will be discussed in more detail in Section 1.2.2. It remains an open question whether the Higgs and cosmological constant hierarchy problems are related.

Many mathematical extensions of the Standard Model, called Beyond-the-Standard-Model (BSM) models, have been proposed to address these issues. A general assumption is that the Standard Model is an effective field theory - a field theory which is valid up to some energy scale (called a cutoff), where a more fundamental theory becomes apparent. This assumption has some mathematical consequences, which will be discussed in Section 1.2.1. The largest possible cutoff is considered to be the Planck scale, as this is the scale at which quantum gravitational effects should become significant. However, new physics - such as supersymmetry or additional Higgs bosons, described in Sections 1.2.3 and 1.2.4 - might appear well below this scale.

1.2.1   The Standard Model as an effective field theory

The effective field theory formalism was first developed in condensed matter physics by Wilson, and later successfully applied to particle physics [60, 67]. Effective field theories are approximations of underlying theories, which work well for energy scales much lower than the cutoff Λ. The Lagrangian of an effective field theory can be written as an expansion containing an infinite number of increasingly higher-dimensional operators $\mathcal{O}$. These operators have coefficients of the form $c\Lambda^{4-d}$, where c is a Wilson coefficient and d is the number of dimensions. This means that operators with a dimension higher than four are suppressed by increasingly large powers of Λ. In the case of the Standard Model, the basic Lagrangian already contains all operators of dimension four and below that can be constructed from Standard Model fields [68]. Operators with a dimension above four are non-renormalizable, but this is not an issue if an energy cutoff exists - infinities will not appear when calculating physical quantities, though these quantities will depend on the cutoff. The Standard Model effective field theory Lagrangian is then the basic Standard Model Lagrangian plus all the possible higher-dimensional terms than can be constructed using Standard Model fields:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i^{(5)}}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i \frac{c_i^{(6)}}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_i \frac{c_i^{(7)}}{\Lambda^3} \mathcal{O}_i^{(7)} + \sum_i \frac{c_i^{(8)}}{\Lambda^4} \mathcal{O}_i^{(8)} + ... \qquad (1.39)$$

This expansion can be truncated at some order, depending on the desired accuracy, to give testable predictions.

The lowest-dimensional new operator is the dimension-5 Weinberg operator [69]:

$$\mathcal{O}^5 = (LH)^2, \qquad (1.40)$$

where L refers to left-handed lepton doublets and H to the Higgs doublet. After electroweak symmetry breaking, this operator generates masses for left-handed neutrinos: $m_\nu \sim \frac{c^{(5)}v^2}{\Lambda}$, where v is the Higgs vacuum expectation value. In this scenario, neutrinos must be Majorana fermions - particles that are their own antiparticles [70]. This operator also triggers processes which violate lepton number conservation, such as $ll \leftrightarrow H^*H^*$ and $\bar{l}H^* \leftrightarrow lH$, where $\bar{l}$ and H^* are CP conjugates of leptons l and the Higgs boson H. Lepton number violation has so far not been observed in experiments; the most stringent limits come from searches for neutrino-less double beta decay [71] and imply an energy cutoff scale of $\Lambda \geq 10^{15}$ GeV, assuming a 'natural' Wilson coefficient [67, 72].

While there is only one dimension-5 operator in the Standard Model effective field theory, many more can be constructed at dimensions 6 and up [73–75]. Dimension-6 operators can violate baryon number conservation and thus explain the baryon asymmetry of the Universe. As with lepton number conservation, this effect has not been experimentally observed. The most stringent limits come from proton lifetime experiments [76] (with baryon number violation, the proton would be expected to decay), and also imply $\Lambda \geq 10^{15}$ GeV [67, 77]. Dimension-7 operators (and generally all odd-dimensional operators from dimension 5 up) violate $B - L$ conservation: the conservation of the difference between the baryon and the lepton numbers. At dimension four and below, only operators conserving lepton and baryon number (and hence also $B - L$) can be constructed. This is an 'accidental' symmetry: a symmetry which appears in a renormalizable field theory only because terms breaking this symmetry have dimensions which are too high to appear in the Lagrangian [78].

1.2.2 The Higgs mass fine-tuning problem

After regularization and renormalization using an energy cutoff, the bare and renormalized Standard Model Higgs boson masses are related via a counterterm δm_h^2 [79]:

$$m_{h\,\text{bare}}^2 = m_{h\,\text{renorm}}^2 + \delta m_h^2. \tag{1.41}$$

The counterterm diverges quadratically with growing energy due to interactions between the Higgs boson and virtual particles arising from the vacuum. Since the measured Higgs boson mass is relatively low, at around 125 GeV, this implies a very delicate fine-tuning between the counterterm and the bare mass. This is called the Higgs mass fine-tuning problem or the Higgs hierarchy problem.

The fine-tuning problem can be 'hidden' by using dimensional regularization instead of cutoff regularization [79]. In this scheme, quadratic and logarithmic divergences are treated on the same footing, and only $\frac{1}{\epsilon}$ divergences appear; these are absorbed into counterterm definitions. However, this only obscures the problem rather than solving it, as the same real-world conclusions should be reached regardless of the regularization scheme.

The fine-tuning problem can also be formulated in terms of the Standard Model effective field theory [80]. Almost all operators in the basic Standard Model Lagrangian are of dimension 4, and do not have the cutoff Λ in the coefficient of the effective field theory expansion. The one exception is the two-dimensional and three-dimensional Higgs boson operators $|H|^2$ and $|H|^3$, which would be multiplied by positive powers of Λ. In the more problematic dimension-2 operator, the Higgs mass would correspond to $c\Lambda^2$. Since Λ can potentially be the Planck scale, c would need to be fine-tuned to a very small value in order to obtain the experimentally-measured Higgs boson mass. To avoid fine-tuning, new physics would have to appear at an energy scale of around a few TeV.

1.2.3 Supersymmetry

One popular solution to the Higgs mass puzzle is the introduction of supersymmetry [81–84], a type of spacetime symmetry postulating that each Standard Model particle has a yet-undiscovered super-symmetric partner (superpartner). Superpartners of Standard Model bosons would be fermionic, while superpartners of Standard Model fermions would be bosonic. In the simplest supersymmetric models with unbroken supersymmetry, particles and their superpartners would have the same mass and internal quantum numbers (except chirality), much like particles and antiparticles. However, if that were the case, it should be simple to find superpartners using present-day technology. Since to date no superpartners have been found, the supersymmetry would have to be spontaneously broken, allowing particles and their superpartners to vary in mass.

Supersymmetry breaking spontaneously near the electroweak scale would solve the Higgs fine-tuning problem. Within the Standard Model effective field theory, the fine-tuning would be eliminated by the relatively low cutoff energy. Above this energy, ultraviolet corrections would cancel naturally, since bosonic and fermionic loops have opposite signs. Supersymmetry could also provide a dark matter candidate, since the lightest supersymmetric particle can be arranged to be massive, electrically neutral, weakly interacting, and stable; it would also allow for the electroweak and strong interaction to be unified below the Planck scale.

The most basic version of supersymmetry which accounts for mass differences between superpartners is called the Minimal Supersymmetric Standard Model (MSSM) [85]. The MSSM considers only the minimum required number of new particles: the Standard Model superpartners and an additional Higgs boson doublet (an extension of the Higgs sector which will be described in Section 1.2.4 below). A more complex model is the Next-to-Minimal Supersymmetric Standard Model, which adds an additional singlet [86].

1.2.4 Two-Higgs-Doublet Models

In BSM scenarios, the Higgs sector of the Standard Model can be extended in many ways. Some possibilities are: composite Higgs models, where the Higgs boson is a bound state of some new strong interaction [87, 88]; twin Higgs models, with a $SU(4)$ symmetry spontaneously broken to $SU(3)$ [89]; Higgs portal models, where dark matter couples only to the Higgs sector of the Standard

Model, allowing for 'invisible' Higgs boson decays to dark matter particles [90]; and a variety of Higgs singlet, doublet, triplet, and even quadruplet models which predict additional Higgs bosons. Of particular interest is the family of Two-Higgs-Doublet Models (2HDMs) [91, 92]. This family is considered to be one of the simplest and most natural extensions of the Standard Model Higgs sector, adding just one additional Higgs doublet to the one already predicted by the Standard Model. Perhaps more importantly, one of the models in this family, the type-II 2HDM, corresponds to the Higgs sector of the MSSM. Finding additional Higgs bosons predicted by this model would thus be a strong hint for the existence of supersymmetry and potentially solve the Higgs mass fine-tuning problem.

2HDMs assume the existence of two Higgs doublets ϕ_1 and ϕ_2, and thus eight degrees of freedom. After spontaneous symmetry breaking, three of these degrees of freedom produce the masses of the $W^\pm$ and Z bosons, like in the Standard Model, and the remaining ones lead to the existence of five physical scalar states: two neutral CP-even bosons h and H (with $m_h < m_H$), one neutral CP-odd boson A, and two charged bosons $H^\pm$. The Lagrangian of the Higgs sector of 2HDMs takes the following form:

$$\mathcal{L}_{\text{2HDM}} = \sum_{i=1,2} |D_\mu \phi_i|^2 - V(\phi_1, \phi_2) + \mathcal{L}_{\text{Yukawa}}, \tag{1.42}$$

where D_μ is the covariant derivative, V is the 2HDM Higgs potential, and $\mathcal{L}_{\text{Yukawa}}$ describes Yukawa couplings of the Higgs bosons to Standard Model fermions. The most general 2HDM, in which quarks and charged leptons couple to both doublets, introduces problematic tree-level flavor-changing neutral currents, which are strongly constrained by experimental results. This problem was solved by Glashow and Weinberg [93] and independently by Paschos [94] by imposing a discrete Z_2 symmetry, $\phi_1 \to -\phi_1$, $\phi_2 \to \phi_2$. Z_2 parity can then be assigned to Standard Model fermions in one of four ways, leading to four types of commonly-considered 2HDMs:

- Type-I (fermiophobic): quarks and charged leptons couple only to ϕ_2.

- Type-II (MSSM-like): up-type quarks couple to ϕ_2, down-type quarks and charged leptons couple to ϕ_1.

- Type-X (lepton-specific): quarks couple to ϕ_2, charged leptons couple to ϕ_1.

- Type-Y (flipped): up-type quarks and charged leptons couple to ϕ_2, down-type quarks couple to ϕ_1.

By convention, Φ_2 is the doublet which couples to up-type quarks. 2HDMs without Z_2 symmetry are referred to as type-III.

The most general, renormalizable potential V that can be constructed is of the form

$$V(\phi_1, \phi_2) = m_{11}^2 \phi_1^\dagger \phi_1 + m_{22}^2 \phi_2^\dagger \phi_2 - (m_{12}^2 \phi_1^\dagger \phi_2 + \text{h.c.}) +$$
$$+ \frac{1}{2}\lambda_1(\phi_1^\dagger\phi_1)^2 + \frac{1}{2}\lambda_2(\phi_2^\dagger\phi_2)^2 + \frac{1}{2}\lambda_3(\phi_1^\dagger\phi_1)(\phi_2^\dagger\phi_2) + \frac{1}{2}\lambda_4(\phi_1^\dagger\phi_2)(\phi_2^\dagger\phi_1) +$$
$$+ [\frac{1}{2}\lambda_5(\phi_1^\dagger\phi_2)^2 + \lambda_6(\phi_1^\dagger\phi_1)(\phi_1^\dagger\phi_2) + \lambda_7(\phi_2^\dagger\phi_2)(\phi_1^\dagger\phi_2) + \text{h.c.}], \tag{1.43}$$

where h.c. is the Hermitian conjugate of the preceding expression, m_{11}, m_{22}, and $\lambda_{1,2,3,4}$ are real parameters, and m_{12} and $\lambda_{5,6,7}$ are - in general - complex. Imposing a softly-broken Z_2 symmetry (to allow for small CP violation, such that is seen experimentally) and CP-conserving vacuum expectation values of the form

$$\langle 0|\phi_1|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \langle 0|\phi_2|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v_2 \end{pmatrix}, \tag{1.44}$$

where $v = \sqrt{v_1^2 + v_2^2} \approx 246$ GeV (the Standard Model Higgs field vacuum expectation value), the potential simplifies to

$$V(\phi_1, \phi_2) = m_{11}^2 \phi_1^\dagger \phi_1 + m_{22}^2 \phi_2^\dagger \phi_2 - m_{12}^2(\phi_1^\dagger\phi_2 + \phi_2^\dagger\phi_1) + \frac{1}{2}\lambda_1(\phi_1^\dagger\phi_1)^2 + \frac{1}{2}\lambda_2(\phi_2^\dagger\phi_2)^2 +$$
$$+ \lambda_3(\phi_1^\dagger\phi_1)(\phi_2^\dagger\phi_2) + \lambda_4(\phi_1^\dagger\phi_2)(\phi_2^\dagger\phi_1) + \frac{1}{2}\lambda_5[(\phi_1^\dagger\phi_2)^2 + (\phi_2^\dagger\phi_1)^2], \tag{1.45}$$

20

Table 1.5: Yukawa couplings y of neutral 2HDM Higgs bosons h, H, and A to pairs of Standard Model up-type quarks ($u\bar{u}$), down-type quarks ($d\bar{d}$), and charged leptons ($l\bar{l}$) in each of the Z_2 symmetry-preserving 2HDM types, relative to Standard Model Higgs boson couplings.

	type-I	type-II	type-X	type-Y
$y_{u\bar{u}}^{h}$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$
$y_{d\bar{d}}^{h}$	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$
$y_{l\bar{l}}^{h}$	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\sin\beta$
$y_{u\bar{u}}^{H}$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$
$y_{d\bar{d}}^{H}$	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$
$y_{l\bar{l}}^{H}$	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$	$\cos\alpha/\cos\beta$	$\sin\alpha/\sin\beta$
$y_{u\bar{u}}^{A}$	$\cot\beta$	$\cot\beta$	$\cot\beta$	$\cot\beta$
$y_{d\bar{d}}^{A}$	$-\cot\beta$	$\tan\beta$	$-\cot\beta$	$\tan\beta$
$y_{l\bar{l}}^{A}$	$-\cot\beta$	$\tan\beta$	$\tan\beta$	$-\cot\beta$

where the m_{12}^2 term softly breaks the Z_2 symmetry.

2HDMs can be described in terms of the Standard Model gauge boson masses m_Z and m_W, the five Higgs boson masses, the ratio of the doublets' vacuum expectation values $\tan\beta = \frac{v_1}{v_2}$, and the angle of mixing between h and H states α. At tree level, the Higgs boson masses obey the following relations:

$$m_{H^\pm}^2 = m_A^2 + m_W^2, \tag{1.46}$$

$$m_{h,H}^2 = \frac{1}{2}\left(m_A^2 + m_Z^2 \mp \sqrt{(m_A^2 + m_Z^2)^2 - 4m_A^2 m_Z^2 \cos^2 2\beta}\right), \tag{1.47}$$

while the mixing angle α obeys a more complicated relation:

$$\tan\alpha = \frac{-(m_A^2 + m_Z^2)\sin 2\beta}{(m_Z^2 - m_A^2)\sin 2\beta + \sqrt{(m_A^2 + m_Z^2)^2 - 4m_A m_Z^2 \cos^2 2\beta}}. \tag{1.48}$$

Finally, Yukawa couplings of neutral 2HDM Higgs bosons to pairs of Standard Model fermions, relative to the corresponding Standard Model couplings, are given in Table 1.5. Yukawa couplings of charged 2HDM Higgs bosons have more complicated expressions and are omitted here. Since 2HDM parameters are related, at tree level it is sufficient to only use two non-Standard-Model parameters: these are usually chosen to be $\tan\beta$ and m_A. The LHC discovery of a neutral, CP-even boson with a mass of 125 GeV and couplings consistent with those predicted for the Standard Model Higgs boson implies that either h or H in 2HDMs would need to be Standard-Model-like. This can be achieved, respectively, in the limits $\cos(\beta - \alpha) \to 0$ or $\sin(\beta - \alpha) \to 0$, collectively called the alignment limit.

The type-II 2HDM with supersymmetry

Experimental constraints on the MSSM Higgs sector come not only from the discovery of the 125 GeV boson and from limits set by searches for extra Higgs bosons, but also from the lack of observed superpartners at LHC energies. The alignment limit can be achieved in two scenarios, with and without so-called decoupling:

- $m_A \gg m_Z$. This is called the decoupling limit and corresponds to the alignment limit $\cos(\beta - \alpha) \to 0$, with h being the Standard-Model-like boson. In this limit, the tree-level masses of bosons H, A, and $H^\pm$ become approximately degenerate, and the tree-level mass of h cannot exceed the mass of the Z boson:

$$m_h = m_Z |\cos 2\beta|. \tag{1.49}$$

  However, radiative corrections can push this mass up to be consistent with 125 GeV.

- $m_A \lesssim m_Z$. In this case, either h or H can be Standard-Model-like depending on the precise values of m_A and $\tan\beta$; this is called alignment without decoupling.

Since small values of m_A are already tightly constrained by experiments [68], decoupling is usually assumed. The radiative corrections to m_h depend on supersymmetric parameters; the dominant contribution is from loops of top quarks and their superpartners, stop squarks. In order to avoid making assumptions about these parameters (beyond the assumption that supersymmetry breaking occurs at high energy scales), the hMSSM approach was developed [68, 95, 96]. In this approach, the mass matrix of neutral CP-even Higgs bosons is decomposed into a tree-level part and a radiative-corrections part:

$$\mathcal{M}^2 = \begin{pmatrix} m_A^2 \sin^2\beta + m_Z^2 \cos^2\beta & -(m_A^2 + m_Z^2)\sin\beta\cos\beta \\ -(m_A^2 + m_Z^2)\sin\beta\cos\beta & m_A^2\cos^2\beta + m_Z^2\sin^2\beta \end{pmatrix} + \begin{pmatrix} \Delta\mathcal{M}_{11}^2 & \Delta\mathcal{M}_{12}^2 \\ \Delta\mathcal{M}_{21}^2 & \Delta\mathcal{M}_{22}^2 \end{pmatrix} \quad (1.50)$$

with the assumption that only the element $\Delta\mathcal{M}_{22}^2$ needs to be taken into account. This radiative correction can then be related to the experimentally-measured 125 GeV mass. Assuming that superpartners are heavy enough that they do not cause a significant modification of Higgs sector parameters (decay widths, couplings) other than affecting the mass matrix, expressions for m_H and $\tan\alpha$ then become:

$$m_H^2 = \frac{m_h^2(m_A^2 + m_Z^2 - m_h^2 - m_A^2 m_Z^2 \cos^2 2\beta)}{m_Z^2 \cos^2\beta + m_A^2 \sin^2\beta - m_h^2}, \quad (1.51)$$

$$\tan\alpha = -\frac{(m_A^2 + m_Z^2)\sin\beta\cos\beta}{m_Z^2 \cos^2\beta + m_A^2 \sin^2\beta - m_h^2}. \quad (1.52)$$

The hMSSM approach only works well in regions of m_A and $\tan\beta$ where the denominator of Equations 1.51 and 1.52 is greater than zero.

An orthogonal approach is to find a set of MSSM parameters which give the correct value of m_h in most of the m_A-$\tan\beta$ phase space, rather than using $m_h = 125$ GeV as an input. Here, a number of scenarios are possible. The most basic is the m_h^{125} GeV scenario [97, 98], which assumes that all superpartners are heavy. Scenarios where some of the superpartners are light or where some special properties are enforced can also be used.

Chapter 2

Experimental Background - Searching for New Particles

In order to discover or study particles with short lifetimes, it is necessary to produce them in sufficient quantity and in a suitable place. The production of heavy particles such as the Higgs boson requires large energies, which only occur naturally in violent cosmic events. On Earth, we can replicate these conditions in particle colliders: machines which accelerate particles to high energies using electromagnetic fields and then collide them inside particle detectors. The largest particle collider in the world is the LHC (Large Hadron Collider), which is capable of colliding protons at a 13 TeV center-of-mass energy.

Two widely used types of particle colliders are electron-positron colliders and hadron colliders. The first type is often used for producing large quantities of a particle with a known mass to study its properties. Because leptons are point particles, their interactions are clean and can be tuned to a specific energy. Hadrons, on the other hand, are composed of partons: quarks and gluons, which can interact in numerous ways during collisions and carry different momentum fractions. Because of this, hadron collisions produce many collision products with different energies, causing hadron colliders to be dubbed 'discovery machines'. In addition, it is easier to accelerate hadrons to large energies, because they do not lose as much energy via synchrotron radiation (electromagnetic radiation emitted when charged particles are accelerated radially) as electrons do [99]. While the large range of collision products is useful, the downside is that many produced particles are not of interest and constitute a large background.

A typical hadron collider 'event' (collision) with a high momentum transfer is shown in Figure 2.1. The two dark green ovals with three incoming arrows represent two colliding protons. Two parton interactions take place in this example. The red loopy lines correspond to two gluons engaging in a 'hard scattering', a high momentum transfer process which can result in the creation of a high-mass resonance (short-lived) particle. Purple lines originating from the colliding protons correspond to four other partons taking part in a 'soft' (lower-energy) interaction. Small red dots are the high-mass resonance decay products, which happen to be unstable and decay further (red lines), while also radiating gluons (blue loopy lines) - the last part is also called the parton shower. Gluons and quarks created in the collision hadronize, coalescing into hadrons represented by light green ovals. These hadrons can also decay further, with decay products represented by dark green circles. Hadrons produced in the hadronization of a single parton travel in a collimated cone and are collectively called a QCD jet or simply a jet. Additionally, photons can be radiated off throughout the process, and are shown as yellow wavy lines in the schematic. Partons and photons radiated from incoming protons are referred to as initial state radiation, while similar radiation from collision products is referred to as final state radiation.

Heavy resonance particles typically decay before they can enter the detector volume, and thus their presence must be inferred from decay products, called a signature. An additional complication is that the cross-sections of heavy resonance production processes are typically very small, resulting in a huge imbalance between events of interest (the signal) and background events. This chapter will cover a few important ingredients for making high-energy particle searches possible: collider
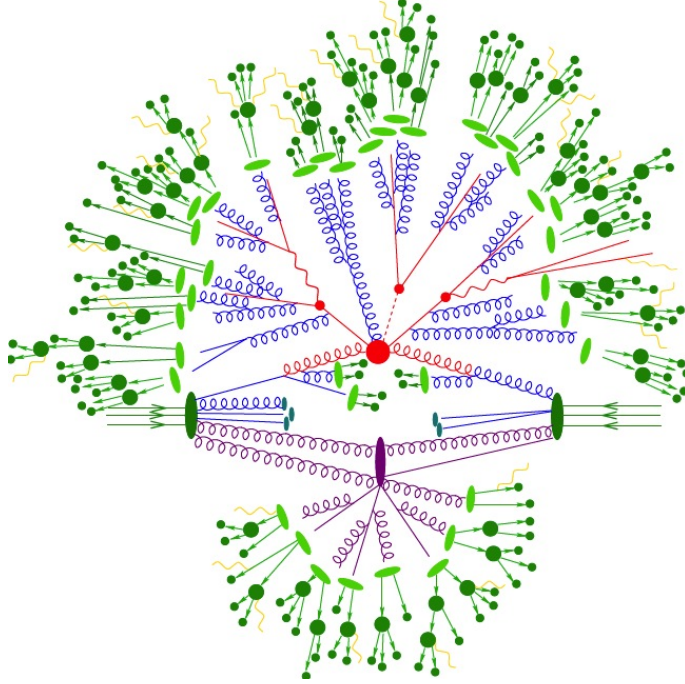
Figure 2.1: Schematic representation of a high-energy proton-proton collision. [100]

luminosities, Monte Carlo event generators, and the bump-hunt approach.

2.1 Luminosity

A key parameter of particle colliders, other than the type and energy of colliding particles, is a quantity called luminosity L [101]. This parameter is defined as the ratio of the number of events N detected in a certain time t to the interaction cross-section σ, which depends on the dynamics of the collision process:

$$L = \frac{1}{\sigma}\frac{dN}{dt}.$$
(2.1)

In particle collision experiments, this can be rephrased as the collision rate divided by the scattering cross-section for the particles being collided. As such, luminosity depends on particle beam parameters like beam width, beam energy, and the density of particles in the beam. Particle colliders generally seek to maximize the luminosity, thus increasing the collision rate and the probability of detecting very rare processes. A commonly-used, related quantity is integrated luminosity:

$$L_{\text{int}} = \int L dt,$$
(2.2)

which gives a sense of how many particle collisions took place in a period of time, such as the lifetime of a particle collider. Integrated luminosity is expressed in units of inverse area, such as inverse barns.

2.2 Monte Carlo event generators

For any type of high-energy analysis, it is necessary to have detailed knowledge of what the data collected by a detector should look like for different physics hypotheses. This is a challenging endeavor for two reasons. One, even though distributions of physics observables are derived from quantum physics and are - in principle - quite simple and elegant, their simplicity is muddled by messy effects like higher-order perturbative calculations and hadronization. Two, the raw output from a modern detector experiment is simply a collection of electric signals from different detector components, caused by interactions between particles and detector material. To become human-interpretable, these signals are converted into most-probable physics objects (such as electrons,

photons, and QCD jets) by using various reconstruction and identification algorithms. Due to detector noise and the sheer number of particles interacting with the detector, these algorithms are not fully accurate or efficient.

This is where Monte Carlo event generators come in. Using statistical methods and theoretical physics knowledge, it is possible to simulate complex particle interactions including the hard scattering, parton shower, hadronization, and interactions of particles with detector material, thus building a model for statistical analysis [102]. These simulated physics samples can then be compared to real data, and statistical tests can be carried out to assess the plausibility of different hypotheses. An important part of most physics analyses is making sure that the Monte Carlo being used accurately models real processes.

Monte Carlo event generators use a combination of perturbative and non-perturbative calculations. The distributions of partons in colliding protons are described with parton distribution functions (PDFs), which define the probability densities of finding partons with a certain momentum fraction at a given energy scale. For the hard scattering, Feynman-diagram matrix elements are calculated at a fixed order in perturbation theory (leading order LO, next-to-leading order NLO, or next-to-next-to-leading-order NNLO). Perturbative calculations are also used for parton showers. Non-perturbative, phenomenological models are used for hadronization, hadron decays, and the underlying event: soft interactions of other partons in the colliding protons. These models typically have multiple tunable parameters. Recommended parameter values are given by sets of 'parameter tunes' based on real data.

Not all physics processes can be reliably modeled with Monte Carlo. A prominent example is the production of multiple QCD jets; these jets can be produced with large multiplicities, and the number of Feynman diagrams that have to be evaluated grows asymptotically factorially with the number of external particles. In these cases, background contributions are usually estimated with data-driven methods, combining real data collected by the detector and Monte Carlo samples of processes that can be reliably modeled.

2.3   Bump hunts

A commonly-used, robust method of searching for new particles is to look for excesses, or 'bumps' in invariant mass spectra (calculated using the four-momenta of decay products) over smoothly-falling background distributions. This works because resonance particles have reasonably well-defined masses, with a natural width Γ which is inversely proportional to the resonance half-life τ:

$$\Gamma = \frac{\hbar}{\tau} \tag{2.3}$$

and thus show up as localized excesses if the signal-to-background ratio is large enough. A famous example of this - an excess in a photon-photon invariant mass distribution which resulted in the discovery of a new 125 GeV boson, considered to be the Standard Model Higgs boson - is shown in Figure 2.2.

In order to get a favorable signal-to-background ratio for a rare process, selection criteria ('cuts') are typically applied to events. These cuts exploit knowledge about event characteristics in order to remove a large portion of the background while keeping as much signal as possible. They might involve the types and multiplicities of particles and jets in the final state, their energies and/or momenta, relative angles between them, and their electric charges. An analysis making use of such rectangular cuts is called a cut-based analysis. Alternatively, multivariate methods can be used to combine various discriminating variables - which can include the invariant mass - in one classifier. This is a growing area of research in high-energy particle physics and will be discussed in the next chapter. It is important no note that the lack of a statistically-significant excess of events in any particular analysis does not necessarily mean that the particle being searched for does not exist. Because of this, physics searches set upper limits on the cross-sections for the production of new particles rather than rejecting their existence altogether.

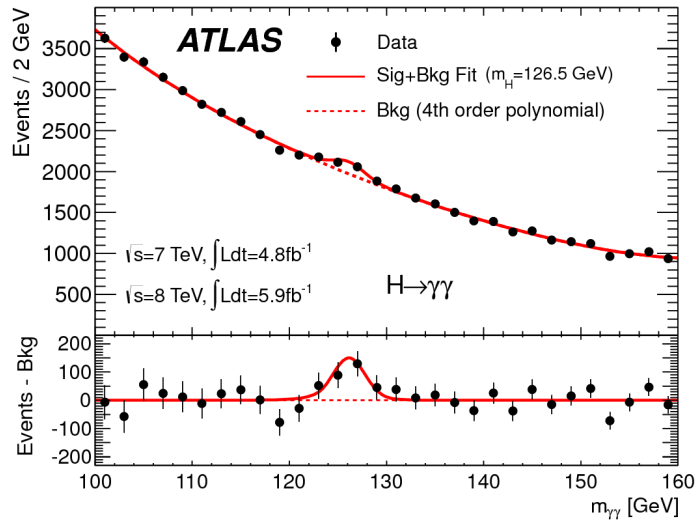Figure 2.2: A bump in the $m(\gamma\gamma)$ spectrum seen by the ATLAS experiment in data collected at 7 and 8 TeV center-of-mass energy [45]. The dashed line corresponds to the background-only expectation (from Monte Carlo) and black points correspond to real data. The resonance mass was measured to be around 126.5 GeV in this analysis, but subsequent, more accurate measurements brought this value down to 125.10 ± 0.14 GeV [27].

Chapter 3

Machine Learning in High-Energy Physics

Machine learning, the field of inducing computers to learn patterns, has found countless uses in industry and academia since its advent in the second half of the 20th century. High-energy physics also benefits extensively from machine learning techniques, including classification (predicting class membership), regression (predicting continuous values), and more advanced methods like generative models. Popular use cases include real-time data selection algorithms; object reconstruction and identification; developing fast, non-Monte-Carlo event simulation with generative adversarial networks; finding anomalies with variational autoencoders; and - finally - separating signal and background events in physics analyses. There is even potential to eventually use quantum machine learning - machine learning done on quantum computers - for simulating complex quantum systems. A good overview of the machine learning in high-energy physics landscape can be found in the 2018 community white paper [103].

Searches for new particles in high-energy physics are generally done through a bump hunt using some mass distribution, as described in Section 2.3. While mass distributions of resonance particles are usually narrow, and thus can easily show up as a bump on a wider background mass distribution, there is other information that could be used to distinguish between signal and background, such as kinematic variables, polarization variables, or jet multiplicities in an event. Constructing a machine learning model which takes all these variables as inputs, and in turn outputs a single score - signifying how signal- or background-like a particular event is - can be a very effective dimensionality reduction technique. The full machine learning score distribution, which should have better discriminating power than any one input variable, can then be used as the final discriminant for conducting a statistical analysis.

3.1 Basics

The basic idea of machine learning is to present a computer with a set of training data and have it devise a mathematical model which makes predictions or carries out tasks when presented with new data of the same type. There are three broad approaches to machine learning: supervised, unsupervised, and reinforcement learning [104, 105]. In supervised learning, the training data is labeled with either class labels or target values of some output variable. The goal of the model is then to predict the labels as accurately as possible. In unsupervised learning, data is unlabeled, and the goal of the model is to cluster the data or otherwise find hidden patterns. In reinforcement learning, the computer is presented with a dynamical environment (ex. a computer game), and must learn what actions to perform in order to reach a specified goal (ex. win the game). Here, only supervised learning will be discussed.

Computers do not build machine learning models from scratch. The human must present, along with training data, a recipe for building the model: explicit rules and possibly a ready-made skeleton with modifiable parameters. During supervised training, this recipe will be followed with the goal of minimizing some loss function, which measures the difference between the output gen-

erated by the model when given a set of inputs and the expected output (i.e. the data label). However, it is not enough for the model to memorize the examples it is given during training - it must learn general rules rather than dataset fluctuations. To ensure this, a second set of data is used for validating a model. If the model performs significantly better on training data than on validation data - which was not used during training - this means the model is overfitting and must be modified by in some way, for example by decreasing its complexity or by stopping training at an earlier stage. The concept of overfitting is illustrated in Figure 3.1. When tasked with coming up with a rule for separating blue and red dots, a well-generalizing model might come up with the black line, while an overfitted one might come up with the green line, which accounts for all the fluctuations in this particular set of dots.
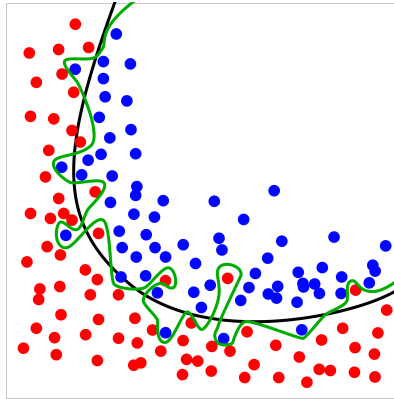
Figure 3.1: An example of overfitting (green line) vs learning a general rule (black line). [106]

One metric for measuring the goodness of a machine learning model is simply the accuracy - that is, what percentage of the data is correctly labeled or how close predicted values are to the target values. However, this approach can backfire spectacularly for datasets with very unequal classes, such as physics samples with a large number of background events and a very small number of signal events. A model that simply labels all event

Figure 3.2: An example of a ROC curve and its relationship to true positive (TP), false positive (FP), true negative (TN), and false negative (FN) event rates. [107]

28

This concept is illustrated in Figure 3.2. The two curves on the top left correspond to model response distributions of data belonging to two different classes. A threshold, represented by 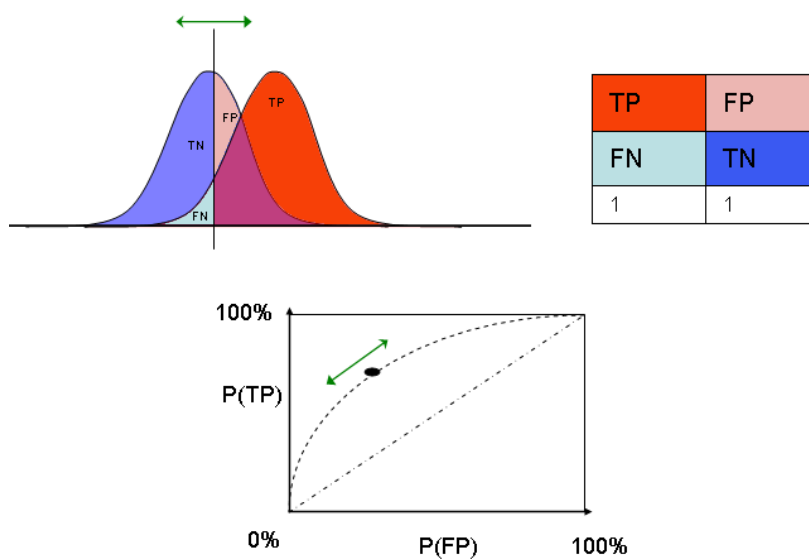the black vertical line, can be chosen: events to the left of the line are considered to be background (negatives), while events to the right of the line are considered to be signal (positives). However, since the distributions overlap, some events are misclassified. In choosing where to place the threshold, a trade-off must be made between keeping the true positive (TP) rate high and the false positive (FP) rate low. In the ROC curve at the bottom of the figure, which has the probability of TP on one axis and the probability of FP on the other, the black dot and green arrow correspond to the chosen threshold.

The area under the ROC curve (ROC AUC) is a useful single-number metric for the goodness of a model. For models with no discriminating power, the ROC curve is a straight line and the ROC AUC is 0.5. A perfectly discriminating model, on the other hand, will have a vertical line at $P(FP) = 0$ and a horizontal line at $P(TP) = 1$, with an AUC of 1. ROC curves are typically used in the context of binary classification, but they can be generalized for multiclass classification by constructing 1-vs-rest curves or calculating the volume under the ROC surface [108, 109].

Another issue that can arise when working with highly skewed datasets or small datasets is an inadequate number of examples to train and sufficiently validate a model. Sometimes, it is possible to obtain or generate more data; however, in the case of high-energy physics, the generation of (labeled) Monte Carlo events is usually costly in terms of CPU usage and time, and the size of physics samples, particularly of very specific new signal processes, are limited. A way of maximizing the available data is to use k-folding, also known as k-fold cross-validation [104]. In this method, the dataset is divided into k equal parts, or folds, and k models are trained. Each model is trained on a different set of $(k-1)$ folds and evaluated on the remaining one fold; this is illustrated in Figure 3.3. In this way, all available data is used for both training and validation in a statistically independent manner. If the variance of the performance of the different models is small, then one can be confident that well-generalizing models have been produced.
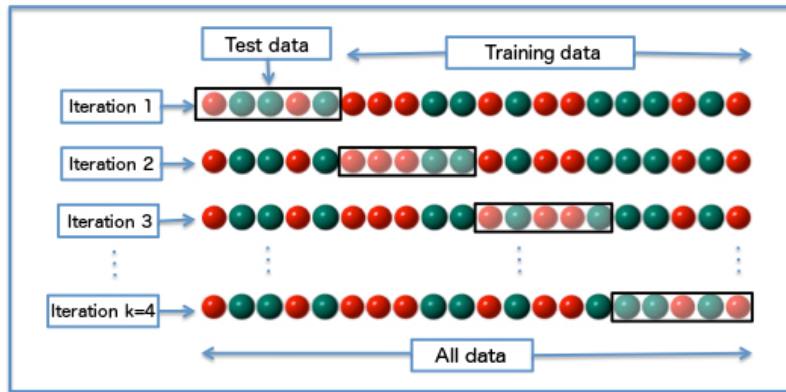
Figure 3.3: k-folding procedure. [110]

Two classes of supervised learning algorithms will be discussed below: boosted decision trees (BDTs) and artificial neural networks (ANNs).

3.2 Boosted Decision Trees

Decision trees are branching, tree-like flowcharts encoding decisions and their possible outcomes. In the case of high-energy physics, these decisions would frequently be cuts on physics observables, and the outcomes would be probabilities of obtaining signal or background events. An example of a simple decision tree is shown in Figure 3.4.

Decision trees can be used as simple machine learning algorithms for classification or regression problems. At each node, a cut on any one input variable is chosen to maximize the splitting of

Figure 3.4: Simple decision tree with cuts on the $p_{\mathrm{T}}(\ell)$ variable terminating in nodes (leaves) with different fractions of signal (S) and background (B) events.

the training sample into two predefined categories according to some metric, typically the Gini index [111]:

$$G = (\sum_{i=1}^{n} W_i) P(1 - P), \tag{3.1}$$

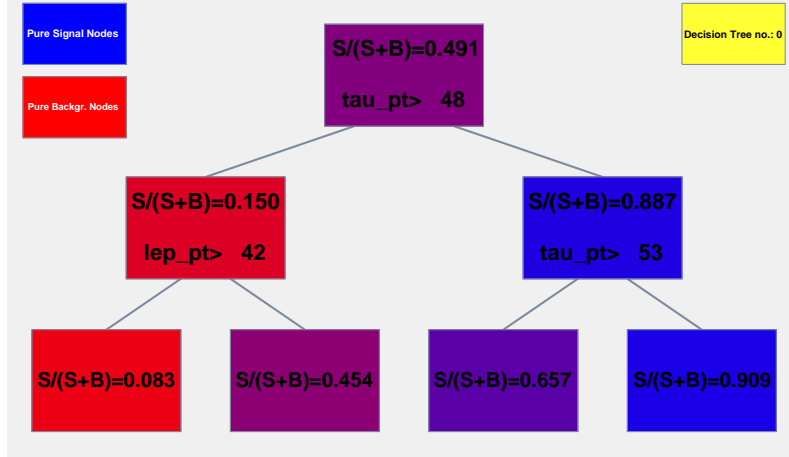where W_i are event weights and P is the purity,

$$P = \frac{\sum_s W_s}{\sum_s W_s + \sum_b W_b}. \tag{3.2}$$

Here, $\sum_s$ is the sum over signal events and $\sum_b$ is the sum over background events. This is an example of a greedy algorithm: an algorithm that makes the locally optimal choice at each step. Though easy to interpret by humans, such algorithms do not usually produce globally optimal solutions and are quite sensitive to statistical fluctuations in the training sample. Luckily, there is a way to make a strong classifier out of a set of weak classifiers like decision trees: ensemble learning [112, 113]. Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. One of the most commonly used ensemble learning algorithms for BDTs is called boosting.

Boosting as a method of improving weak classifiers was first proposed in the 1980's [114, 115] and developed in the 1990's [116, 117]. Though many boosting algorithms are now in use, their common approach is to train a sequence of weak classifiers, reweighting (boosting) misclassified events at each point in the sequence. This can be interpreted as the optimization of a loss function. The final classifier response is obtained by getting the weighted average of the responses of all the decision trees in the ensemble. Two popular types of boosting algorithms are AdaBoost (adaptive boosting) [118] and gradient boosting [119]. Their implementation in the TMVA toolkit used for the analysis described in this thesis is discussed in Section 3.4.

3.3 Artificial Neural Networks

Artificial neural networks (often called simply neural networks, NNs) [120] are artificial computing structures inspired by biological brains. Much as a brain - in simplified terms - is composed of a network of neuronal cells linked by axon-synapse-dendrite connections, through which electrical impulses of varying strength can pass, a neural network is composed of nodes (artificial neurons) connected by links with assigned weights. Each node can receive an input from numerous other nodes and produces a single output, which can in turn be passed on to other nodes. Unlike in a brain, neurons in an artificial neural network are typically organized in layers, with neurons in one layer able to communicate only with neurons in adjacent layers. The first and last layers are

called the input and output layers, respectively. The input layer receives external inputs, normally a vector of variables. The output layer produces the final result, which might be a class label (in the case of a classification problem) or the value of a function (in the case of a regression problem). Between the input and output layers, there might be one or more hidden layers. A schematic of a simple neural network is shown in Figure 3.5.
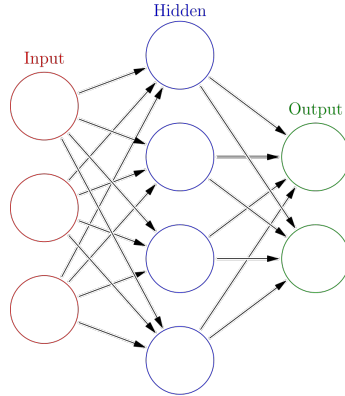

Figure 3.5: Schematic of an artificial neural network with one hidden layer, showing artificial neurons and connections between them. [121]

The output of a single neuron is calculated by taking the sum of all the inputs to this neuron, weighted by weights assigned to the links, and adding a bias term. This value is then passed to a (usually) non-linear activation function, which allows the network to perform non-trivial, non-linear calculations. A few examples of commonly-used activation functions are listed in Table 3.1.

Table 3.1: Examples of activation functions for artificial neurons.

name	function $f(x)$
ReLU (rectified linear unit) [122]	$\begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases}$
sigmoid (also called logistic, soft step)	$\dfrac{1}{1 + e^{-x}}$
tanh	$\dfrac{e^x - e^{-x}}{e^x + e^{-x}}$
softmax	$\dfrac{e^{x_i}}{\Sigma_{j=1}^{J} e^{x_j}}$ for $i = 1, ..., J$

The choice of activation function depends on the problem at hand. For instance, in classification problems the result is usually given in the form of probabilities of each class label being the correct one. Thus, an activation function with a range of (0,1) for neurons in the output layer would be appropriate. If class labels are mutually exclusive, the probabilities should add up to 1, and this can be enforced by the softmax function, for example. In regression tasks, a linear function is used in the output layer to allow for a wide range of output values.

The most basic type of neural network layer - the one shown in Figure 3.5 - is the dense or fully-connected layer. In this type of layer, each neuron is connected to all the neurons in adjacent layers. More sophisticated layers can be constructed for specialized uses, such as computer vision. Convolutional neural networks, for example, employ convolutional and pooling layers, in which neurons are selectively connected in order to scan and identify features in images [123].

During supervised training, a neural network is initialized with (usually) random weights and will then seek to update weights in order to minimize a loss function. This is known as optimiza-

tion. The updating of weights typically takes place through a process called backpropagation, short for backward propagation of errors. During the process, the network calculates the gradient of the loss function with respect to each weight of the network via the chain rule, iterating backwards from the last layer. The weights are then updated in the direction of the gradient in order to find a minimum. A network typically has to perform numerous passes through the entire dataset (such passes are called epochs) before a minimum is found.

In standard gradient descent, weights are updated as follows:

$$w^{\text{new}} = w^{\text{old}} - \eta \frac{\delta L}{\delta w}, \tag{3.3}$$

where η is a constant parameter known as the learning rate. Gradient descent can be done in batches, minibatches, or stochastically. In batch gradient descent, weights are updated once per epoch, which can be slow. A better approach is usually to use stochastic gradient descent, which updates weights after every event. This is faster, but noisy. A compromise between the two is using minibatches of some size N, with weights updated after each minibatch is processed. The learning rate in gradient descent must be chosen carefully. If too low, training will be very slow; if too large, the algorithm can overshoot and the model might not converge.

Neural networks can roughly be divided into two classes: feed-forward and recurrent. In feed-forward neural networks, information moves only forward from the input layer to the output layer, with no cycles or loops. Fully connected feed-forward networks with at least one hidden layer are called multi-layer perceptrons (MLPs). MLPs are sometimes referred to as vanilla networks due to their simplicity. Recurrent networks, on the other hand, do include cycles - their architecture is discussed below.

3.3.1 Recurrent Neural Networks

In recurrent neural networks (RNNs) [124], cyclical connections between neurons form a temporal sequence, allowing for the encoding of sequential information such as time series. In high-energy physics, RNNs have been used for jet flavor tagging and identifying tau leptons from sequences of detector tracks or calorimeter clusters.

RNNs are composed of a chain of repeating modules, which can be thought of as multiple copies of one feed-forward network. These modules can be as simple as a single dense layer with a tanh activation function, though in practice are usually more complex. Each module passes information to the next module via the hidden state vector, which is a representation of all previous inputs acting as the network's memory. A schematic of an RNN with its loop unfolded is shown in Figure 3.6; in this example, each module takes one external input vector and outputs one external response vector. In practice, some modules can have only an input or only an output. Depending on the intended function, RNNs can be set up for one-to-many, many-to-one, or many-to-many processes.
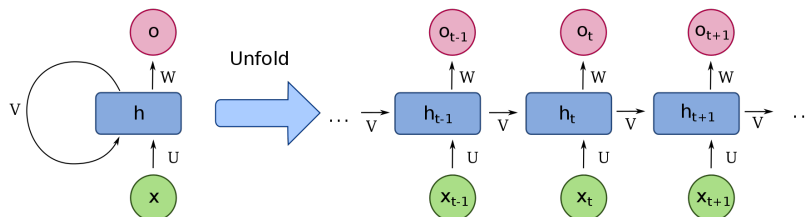

Figure 3.6: A recurrent neural network with a many-to-many setup, before and after unfolding. Here, $\vec{x}$ is an input vector, $\vec{o}$ is an output vector, and $\vec{h}$ is the hidden state vector; u, v, and w are functions, and t refers to the time step. [125]

Vanilla RNNs - RNNs with small, simple modules - usually have trouble learning dependencies between elements that are far from each other in the sequence (long-term dependencies). A more sophisticated type of RNN geared toward handling these dependencies is the Long Short Term Memory network (LSTM) [126]. The architecture of an LSTM module is shown in Figure 3.7. It contains three dense layers with the sigmoid activation function, one dense layer with the tanh activation function, and five places where pointwise mathematical operations are applied. Diverging arrows represent vector duplication, and converging arrows represent vector concatenation. There are two vectors being passed between modules: the hidden state h and the cell state c. The LSTM can add information to the cell state or remove it by two gates: the forget gate (the first sigmoid layer, which updates the cell state through pointwise multiplication) and the remember gate (the sigmoidal input gate layer and a tanh layer, which update the cell state through pointwise multiplication and addition). The updated cell state is then used to update the hidden state, which passes through another sigmoidal layer, via two pointwise functions: tanh and multiplication. Unlike the hidden state, the cell state can hold (remember) information indefinitely, if no information is passed through the forget gate.
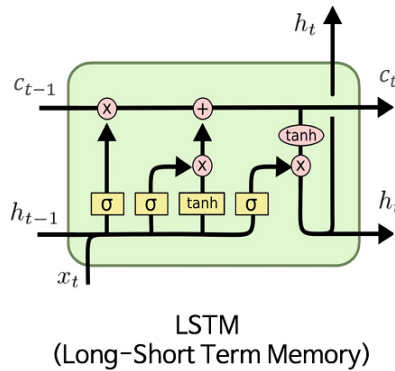

Figure 3.7: A single LSTM module. [127]

3.4 TMVA

TMVA, the Toolkit for Multivariate Analysis [128], is an open-source toolkit integrated into the ROOT computing library [129] for using multivariate analysis techniques in high-energy physics. It supports a number of supervised learning algorithms including BDTs, ANNs, rectangular cut optimization, and various likelihood estimations. It also provides automatic preprocessing of physics inputs, such as normalizing the number of signal and background events to the same value (as typically the two classes are highly unbalanced) and transforming variables via normalization, decorrelation, or transformation into uniform or Gaussian distributions.

3.4.1 BDTs

TMVA uses AdaBoost and gradient boosting algorithms. In the AdaBoost algorithm for classification, events misclassified by a decision tree are weighted by a common boost weight α before being fed into the next tree in the sequence. The boost weight is defined as follows:

$$\alpha = \frac{1 - \epsilon}{\epsilon},\tag{3.4}$$

where ϵ is the misclassification rate of the tree. Event weights of the entire training sample are then renormalized so that the sum of weights remains constant. The final classification for an event with a tuple of input variables δx is given by y:

$$y(\delta x) = \frac{1}{N_{\text{trees}}} \cdot \sum_{i=1}^{N_{\text{trees}}} \ln(\alpha_i) \cdot h_i(\delta x),\tag{3.5}$$

where $h_i(\delta x)$ is the result of classifier i, defined as $+1$ for signal events and -1 for background events. Small values of $y(\delta x)$ indicate background-like events, while large values indicate signal-like

events. This approach corresponds to minimizing the exponential loss function $L(F, y) = e^{-F(\vec{x})y}$. The learning rate of the algorithm can be modified by giving an exponent β to the boost weight: $\alpha \to \alpha^{\beta}$. A slower learning rate with a larger number of boost steps can enhance the performance of the ensemble.

While AdaBoost is a popular algorithm giving good out-of-the-box results, its performance can degrade significantly in noisy environments with many outliers. This shortcoming is addressed in the gradient boost algorithm by using a more robust loss function, the binomial log-likelihood loss:

$$L(f, y) = \ln\left(1 + e^{-2F(\vec{x})y}\right). \tag{3.6}$$

The minimization of this loss function is done via steepest gradient descent. Like with AdaBoost, the performance of this classifier can be improved by lowering the learning rate (by introducing a shrinkage parameter which controls the weights of individual trees) and increasing the number of boost steps.

3.4.2 MLPs

The recommended neural network to use with TMVA is the TMVA implementation of the MLP, which can have an arbitrary number of hidden dense layers of arbitrary size. This method supports three training algorithms: backpropagation, the Broyden-Fletcher-Goldfarb-Shannon (BFGS) method, and the genetic algorithm. Only backpropagation will be discussed here.

The network is trained on N events, each of which is a vector of n_{var} input variables: $\vec{x}_a = (x_1, ..., x_{n_{\mathrm{var}}})_a$, $a = 1, ..., N$. The loss function L in TMVA is then defined as follows:

$$L(\vec{x}_1, ..., \vec{x}_{n_{\mathrm{var}}} | \vec{w}) = \sum_{a=1}^{N} L_a(\vec{x}_a | \vec{w}) = \sum_{a=1}^{N} \frac{1}{2}(\hat{y}_a - y_{\mathrm{NN},a})^2, \tag{3.7}$$

where $\vec{w}$ is a vector of the adjustable network weights, $\hat{y}_a \in \{1, 0\}$ is the class label, and $y_{\mathrm{NN},a}$ is the network response for training event a. Weights are updated via stochastic gradient descent with a constant learning rate. The recommended variable transformation for the MLP method is TMVA normalization, which normalizes the input to the interval [-1,1].

Chapter 4

CERN, the LHC, and the ATLAS Detector

CERN, short for Conseil Européen pour la Recherche Nucléaire (the European Council for Nuclear Research) and also known as the European Organization for Nuclear Research, was established in 1954 by twelve European nations. Currently, it comprises 23 member states - including Poland, which joined in 1991 - and operates the Large Hadron Collider (LHC), the world's biggest and highest-energy particle collider, among other endeavors. Located in a 27-kilometer-long circular tunnel under the Swiss-French border near Geneva, the LHC was designed to collide proton beams at a center-of-mass energy of $\sqrt{s} = 14$ TeV and a peak instantaneous luminosity of 10^{34} cm^{-2}s^{-1}. In addition to proton-proton collisions, which make up the bulk of the physics program, the LHC is capable of colliding heavy nuclei, such as lead, argon, and xenon ions, at a center-of-mass energy of $\sqrt{s} = 2.76$ TeV per nucleon.
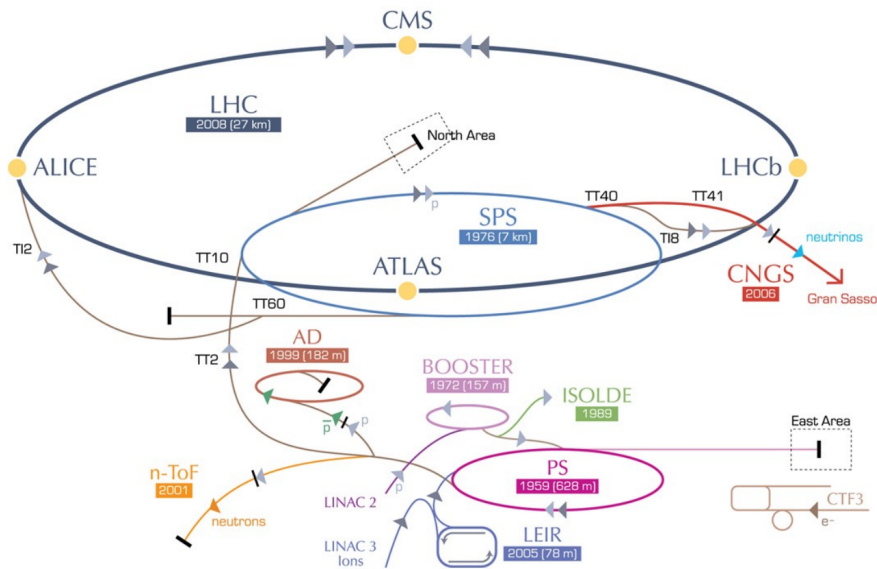
Figure 4.1: A schematic of the LHC complex. [130]

A schematic of the full LHC complex is shown in Figure 4.1. The collider has four beam-crossing (interaction) points where particle beams accelerated in opposite directions can be made to collide. Each of these points is encased in one of four large detectors. These are ATLAS [131] and CMS [132], two general-purpose detectors; ALICE [133], a detector dedicated to studies of heavy-ion collisions; and LHCb [134], an asymmetric detector designed primarily for studying b-meson decays. Additionally, three smaller detectors are installed around the accelerator ring: TOTEM [135], MoEDAL [136], and LHCf [137]. TOTEM shares an interaction point with CMS and is used for measuring cross-sections of elastic and diffractive processes. MoEDAL is located

near LHCb and is specialized for magnetic monopole searches. LHCf shares an interaction point with ATLAS and collects information about neutral pions to complement high-energy cosmic ray measurements. A fourth small detector, FASER [138], is scheduled to go online in 2022 with the purpose of searching for light, weakly-interacting exotic particles like dark photons, axion-like particles, and sterile neutrinos.

LHC operations are based on multi-year running and data-taking periods ('runs') separated by multi-year shutdowns for completing maintenance work and upgrades to the accelerator and detectors. Since becoming operational, the LHC has completed two runs: Run 1 in 2010-2013 and Run 2 in 2015-2018, during which the collision energies and luminosities were gradually ramped up. Run 2 proton-proton collisions were conducted at 13 TeV center-of-mass energy and with a peak instantaneous luminosity of over two times the nominal design value, 2.2×10^{34} cm^{-2}s^{-1}. As of the second half of 2021, the LHC is undergoing the Long Shutdown 2 in preparation for Run 3, which is scheduled to start in 2022 after delays caused by the COVID-19 pandemic. After Run 3 ends in 2025, the collider will undergo a major upgrade to become the High Luminosity LHC (HL-LHC) [139], with the goal to increase the instantaneous luminosity by a factor of 10 beyond the LHC's original design value. The HL-LHC is scheduled to become operational in 2027 and continue to provide collisions throughout the 2030's.

4.1   Particle acceleration at the LHC

Protons used in collisions come from a small hydrogen gas bottle, pictured in Figure 4.2. Hydrogen atoms from the bottle are injected into a duoplasmatron, ionized with an electric field, then accelerated to 0.014c. After leaving the duoplasmatron, they pass through a radio-frequency (RF) quadrupole cavity, which further accelerates and focuses the beam, and enter the linear accelerator LINAC 2. Here, they are accelerated to 50 MeV by a series of RF cavities and subsequently injected into the Proton Synchrotron Booster, which ramps their energy up to 1.4 GeV. From the booster, protons enter first the Proton Synchrotron (PS), accelerating to 25 GeV, and later the Super Proton Synchrotron (SPS), where they reach 450 GeV. Finally, the beams are injected, in opposite directions, into two beampipes in the main ring of the LHC, where they are accelerated to the collision energy. At interaction points the two beampipes converge, and magnets can focus and align the beams so that a collision occurs. Beams can circulate inside the LHC for up to 10 hours before they degrade below the necessary intensity and are dumped on a dedicated target. The cycle, called a fill, is then repeated. The LHC is designed to complete 20,000 fills over 20 years.

Figure 4.2: LHC protons originate from a small hydrogen bottle, pictured here with the duoplasmatron. [140]

LHC beams are not uniform, but made up of discrete 'bunches' - packets of around 10^{11} protons with a typical spacing of 25 ns. When two bunches traveling in opposite directions collide inside an interaction point, it is common for multiple proton-proton interactions to occur, causing what is referred to as in-time pile-up. During Run 2, the average number of interactions per bunch crossing at ATLAS was 33.7. Figure 4.3 shows yearly and total luminosity-weighted distributions

of the mean number of interactions per bunch crossing, μ, during Run 2. Values of μ correspond to the mean of the Poisson distribution of the number of interactions per crossing calculated for each bunch using the formula

$$\mu = \frac{L_{\mathrm{bunch}} \times \sigma_{\mathrm{inel}}}{f_r}, \qquad (4.1)$$

where L_{bunch} is the instantaneous per-bunch luminosity, σ_{inel} is the inelastic cross-section (80 mb for 13 TeV collisions), and f_r is the LHC revolution frequency. The average referred to above is the mean of the total μ distribution. In addition to in-time pile-up, experiments also experience out-of-time pile-up: signals from consecutive bunch crossings overlapping due to detector response latency. Both types of pile-up can pose a challenge when reconstructing and identifying physics objects.
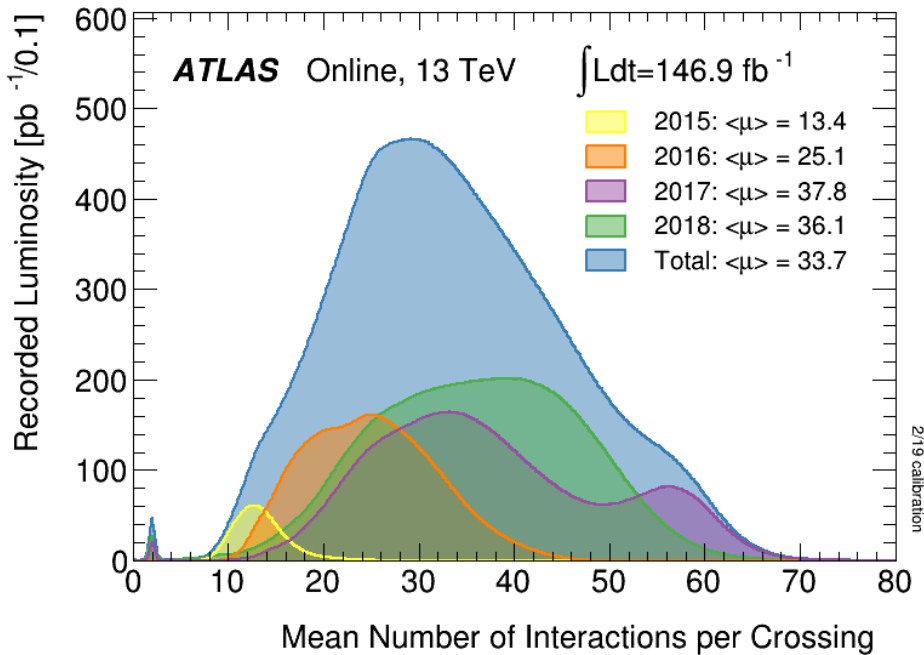

Figure 4.3: Yearly and total luminosity-weighted distributions of the mean number of interactions per bunch crossing during Run 2. [141]

The luminosity is measured independently by each experiment. The cumulative (integrated) luminosity measured by the ATLAS experiment during Run 2 is shown in Figure 4.4. During this time period, the total integrated luminosity delivered to the detector (with stable beams) was 156 fb^{-1}. 147 fb^{-1} of data were recorded by the detector, due to inefficiencies in the data acquisition system, and 139 fb^{-1} of data passed quality criteria and were deemed 'good for physics'.

4.2 Data storage and computing at the LHC

With up to 600 million collisions per second and millions of detector readout channels, the LHC experiments must cope with vast amounts of data to be processed and stored. Since processes of interest to physics analyses typically have very small cross-sections, most of this data can - thankfully - be discarded. This is done by employing detector-specific triggers: hardware- and/or software-based systems which employ certain criteria to select interesting events in real time. Typically, this reduces the event rate by several orders of magnitude. However, even after the drastic data reduction performed by triggers, an average of one petabyte of data per day remains to be processed at the LHC during data-taking runs.

The LHC infrastructure designed to cope with this data is called the Worldwide LHC Computing Grid [142]. It is made up of four levels, or tiers, numbered 0-3. Tier 0 is the CERN Data
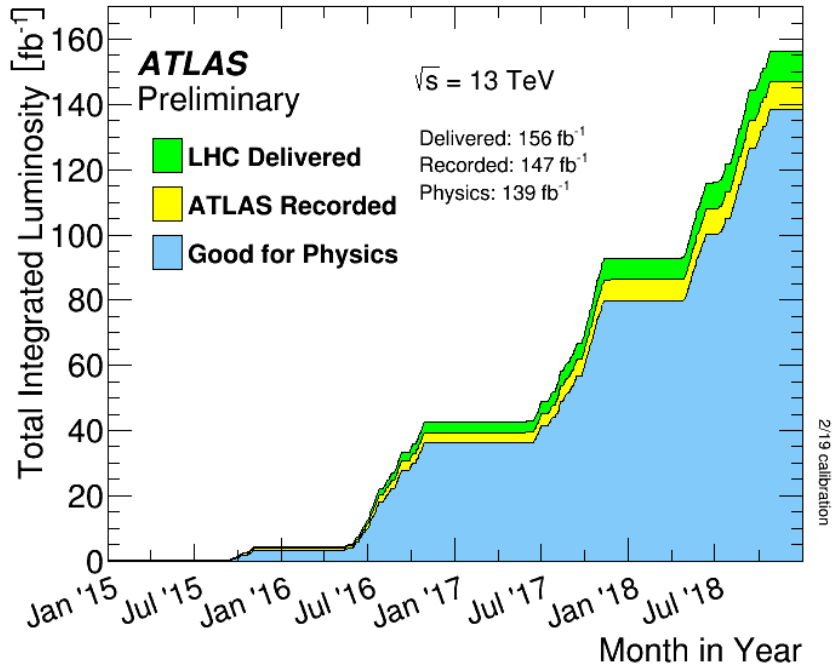
Figure 4.4: Total integrated luminosity and data quality vs time for stable-beam proton-proton collisions at $\sqrt{s} = 13$ TeV during Run 2 (2015-2018). Separately shown is the cumulative luminosity delivered by the LHC to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue). [141]

Centre, an on-site facility with 90 000 processor cores and 10 000 servers, plus a remote extension at the Wigner Research Centre for Physics in Hungary. This tier serves as the central hub through which all raw data from LHC detectors passes, though it provides less than 20% of the Grid's total computing capacity. The raw data undergoes initial reconstruction and is transferred to Tier 1 facilities for further processing. Additionally, the CERN Data Centre performs data archiving, with hundreds of petabytes of data stored long-term on magnetic tapes.

Tier 1 consists of thirteen large computing centers located around the world, connected to CERN via the LHC Optical Private Network, which has optical-fibre links working at 10 gigabits per second. These centers are responsible for storing raw and reconstructed data, performing large-scale data reprocessing and storing its output, distributing stored data to Tier 2 facilities, and storing some Monte Carlo samples (simulated LHC events) produced in Tier 2. Tier 2 consists of over 150 computing sites, typically universities and other scientific institutes, connected to Tier 1 centers via general-purpose national research and education networks. These sites are mostly responsible for producing and processing Monte Carlo samples. Finally, Tier 3 refers to local computing resources: computing clusters in individual university departments or personal computers.

4.3 The anatomy of the ATLAS detector

The ATLAS (A Toroidal LHC ApparatuS) detector has an onion-like structure, with cylindrical layers of different sub-detectors surrounding the interaction point. A cut-away of the detector, with human figures for scale, is shown in Figure 4.5. Starting from the innermost one, the subdetectors are the Inner Detector, which tracks charged particles; the electromagnetic calorimeter, which measures the energy of electromagnetically-interacting particles; the hadronic calorimeter, which measures the energy of hadronically-interacting particles; and the Muon Spectrometer, which tracks muons. Additionally, the detector is equipped with two large superconducting magnet systems to bend the tracks of charged particles in order to measure their momenta. Along the beampipe direction, the detector is divided into a cylindrical barrel region, which envelops the interaction

point, and two endcaps located on each end of the barrel. Between the barrel and the endcaps is the so-called crack region, where many of the readout cables pass through. Because of this, the detector performance around the crack region is low.
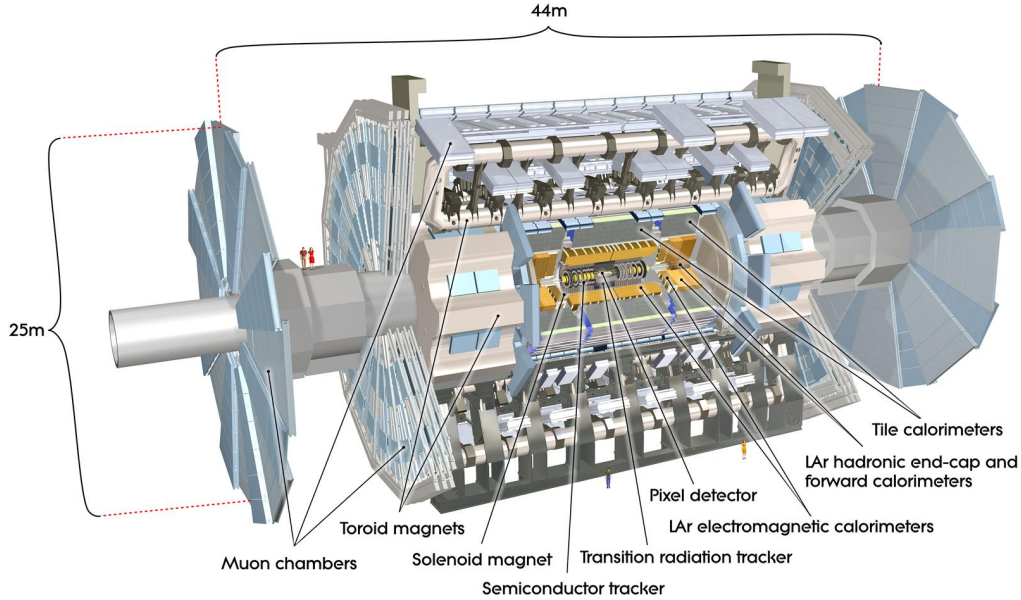

Figure 4.5: The ATLAS detector. [131]

4.3.1 Coordinate system

The coordinate system used for ATLAS has its origin at the interaction point in the middle of the detector, with the z axis pointing along the beampipe. The x axis points toward the center of the collider ring, and the y axis points upwards. Particle trajectories within the detector are usually described using polar coordinates (r, ϕ, θ), with the azimuthal angle ϕ ranging from $-\pi$ to π ($\phi = 0$ is on the positive x axis), and the polar angle θ ranging from 0 to π ($\theta = 0$ is on the positive z axis). Pseudorapidity, defined as

$$\eta = -\ln \tan(\theta/2), \tag{4.2}$$

is often used instead of the polar angle. Pseudorapidity is the Lorentz-invariant rapidity

$$y = \frac{1}{2}\left(\frac{E + p_z}{E - p_z}\right) \tag{4.3}$$

in the limit of massless particles, where E is the particle energy and p_z is the momentum component along the z axis. The $|\eta| < 1.37$ region corresponds to the barrel of the ATLAS detector, $1.37 < |\eta| < 1.52$ covers the crack region, the endcaps extend to $|\eta| = 2.5$, and the area past the endcaps is referred to as the forward region. Angular separation between physics objects within the detector is usually expressed by the distance parameter

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{4.4}$$

Other commonly-used observables at ATLAS are the transverse energy E_T and transverse momentum p_T: the energy and momentum of a physics object measured in the direction perpendicular to the beampipe. The reason for this is that partons in colliding protons can carry different longitudinal momentum fractions, causing the longitudinal momentum component of the interacting system to be unknown. The total transverse momentum, on the other hand, can be considered to be 0 with a good accuracy. In the transverse plane, the conservation of momentum can be used to infer the presence of neutrinos, which do not typically interact with the detector material.

4.3.2 Inner Detector

The Inner Detector [143, 144], pictured in Figure 4.6, is a multi-layer particle tracker capable of measuring the trajectories and momenta of charged particles. It is composed of four parts: the Insertable B Layer (IBL) [145], the Pixel detector [146, 147], the silicon microstrip SemiConductor Tracker (SCT) [148, 149], and the Transition Radiation Tracker (TRT) [150, 151]. Charged particles passing though the Inner Detector interact with the detector components leaving hits, which can later be used to reconstruct 'tracks' - particle trajectories. Surrounding the Inner Detector is a superconducting solenoid magnet, which causes charged particles trajectories to bend due to the Lorentz force, allowing for a measurement of their momenta.

The IBL is situated closest to the interaction point and was installed between the Pixel detector and the beampipe as part of the detector upgrade during Long Shutdown 1. The IBL covers the pseudorapidity range $|\eta| < 3.0$ and comprises 12 million silicon pixels with a typical size of $50 \times 250 \ \mu\mathrm{m}^2$ arranged in a single layer. This upgrade mitigates issues expected with higher luminosity in Run 2, i.e. the possible loss of hits in the three other layers due to high pile-up and radiation damage. It also improves tracking by providing additional measurement points.

The Pixel detector is composed of three layers of pixels with a typical size of $50 \times 400 \ \mu\mathrm{m}^2$. The number of pixels is over 80 million, accounting for roughly half of all the readout channels in the ATLAS detector. Next from the interaction point is the SCT, which is composed of four layers of double-sided silicon strips with a typical size of 80 μm in the $x - y$ direction and 12 cm in the z direction. The SCT has over 6 million readout channels. Pixel and SCT cover the pseudorapidity region $|\eta| < 2.5$ and can be used to locate secondary vertexes from decays of b quarks and hadronically-decaying tau leptons. Furthest from the interaction point is the TRT, a straw detector with multiple layers of proportional-mode drift tubes (straws) embedded in a radiator material generating transition radiation. The straws are 4 mm in diameter and filled with an Xe-CO_2-O_2 gas mixture, which becomes ionized when charged particles pass through the volume. In the center of each straw is a gold-plated tungsten wire with a 31 μm diameter, which collects charge for readout.

4.3.3 Electromagnetic and hadronic calorimeters

Beyond the magnetic coil enveloping the inner detector are the electromagnetic [153] and hadronic [154] calorimeters, pictured in Figure 4.7, which are capable of measuring the energies of electromagnetically and strongly-interacting particles, respectively. This happens through particles initiating electromagnetic or hadronic particle showers upon entering the calorimeter volume; the energy of the particles created in the shower is then deposited in the calorimeter and measured.

Photons, electrons, and positrons can initiate electromagnetic showers. At energies above a few MeV, electrons interact with matter primarily through pair production - that is, through converting to an electron-positron pair in the presence of an atomic nucleus or electrons, which is necessary for conserving momentum. High-energy electrons and positrons, in turn, emit photons via bremsstrahlung. In an electromagnetic shower, pair production and bremsstrahlung continue until photon energies fall below the pair production threshold. The length of the shower depends on the characteristic radiation length of the material hosting the shower, the pair production threshold, and the energy of the particle initiating the shower.

Hadrons like pions, kaons, protons, and neutrons can initiate hadronic showers, which develop through nuclear processes including hadron production, nuclear excitation, nuclear breakup, and decays of secondaries like pions and muons. Hadronic showers also include an electromagnetic component. The length of the shower scales with the nuclear interaction length, i.e. the mean distance travelled by a hadronic particle before undergoing an inelastic nuclear interaction.

The charged particles in a shower can ionize or excite the calorimeter medium, leading to the release of ionization electrons or the production of visible photons via scintillation. These electrons or photons can be detected with anodes or photodetectors, allowing for the measurement of the energy deposited in the detector by the shower. Calorimeters should be deep enough to contain most showers in their entirety, in order to not lose information about the energy of incoming par-
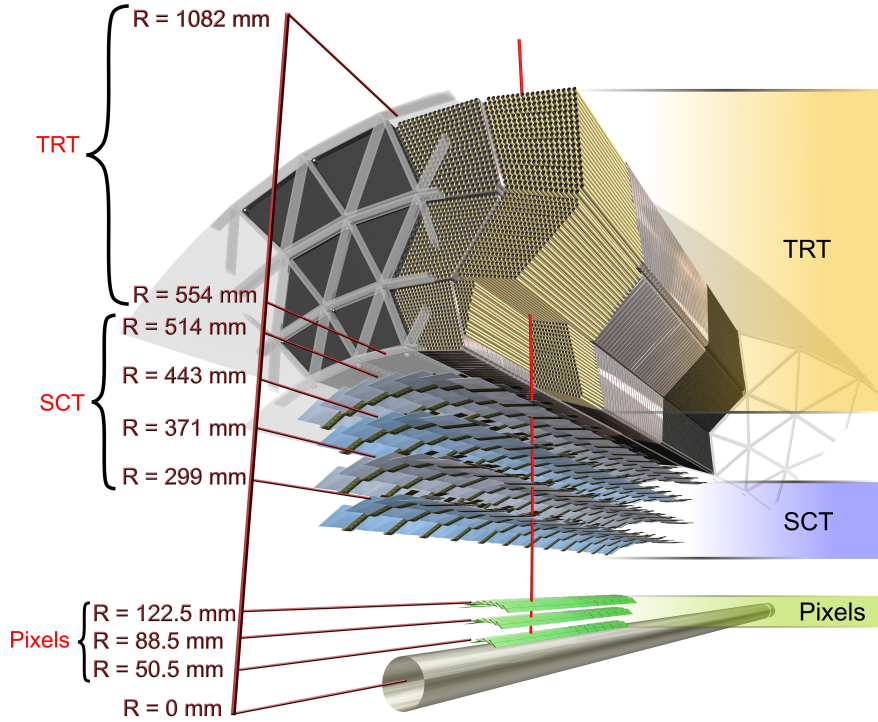
Figure 4.6: The structure of the Inner Detector; not to scale. [152]

ticles. Calorimeter designs can be divided into two categories: homogeneous calorimeters, where a single material is used both for initiating the particle showers and for detection, and sampling calorimeters, where layers of passive absorber material are alternated with layers of active detecting material. Sampling calorimeters indeed 'sample' only a portion of the shower energy, but they can be more compact and cost-effective, as there is greater freedom to choose a high-density absorbing medium for keeping shower lengths relatively short. In addition, segmented designs allow for a better spatial resolution of showers, which can be useful for identifying particles initiating the shower. The electromagnetic and hadronic calorimeters at ATLAS are both sampling calorimeters.

The ATLAS electromagnetic calorimeter uses liquid argon (LAr) as the active medium and lead as the absorber. It employs an accordion geometry in the radial direction and covers the pseudorapidity range $|\eta| < 4.9$ with three layers of different granularity. The typical cell size in the middle layer, which collects most of the energy, is $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$.

The hadronic calorimeter has three distinct, overlapping parts. In the central region ($|\eta < 1.7|$), it employs scintillating plastic tiles and iron absorbing plates. In the endcaps ($1.5 < |\eta| < 3.2$), LAr is used as the active medium and copper as the absorber. Finally, in the forward region ($3.1 < |\eta| < 4.9$), LAr is layered with tungsten. The cell sizes in the hadronic calorimeter are between $\Delta\eta \times \Delta\delta = 0.1 \times 0.1$ and 0.2×0.2.

The calorimeters are housed in three cryostats - one for the detector barrel and two for the endcaps - with a total volume of 83 m^3. The liquid argon in the calorimeters is cooled to approximately 87 K to avoid gas bubble formation, which is detrimental to the functioning of the detector. This is done with saturated liquid nitrogen heat exchangers installed in the cryostats, which also keep the temperature gradients within 0.6 K across the calorimeter volume.
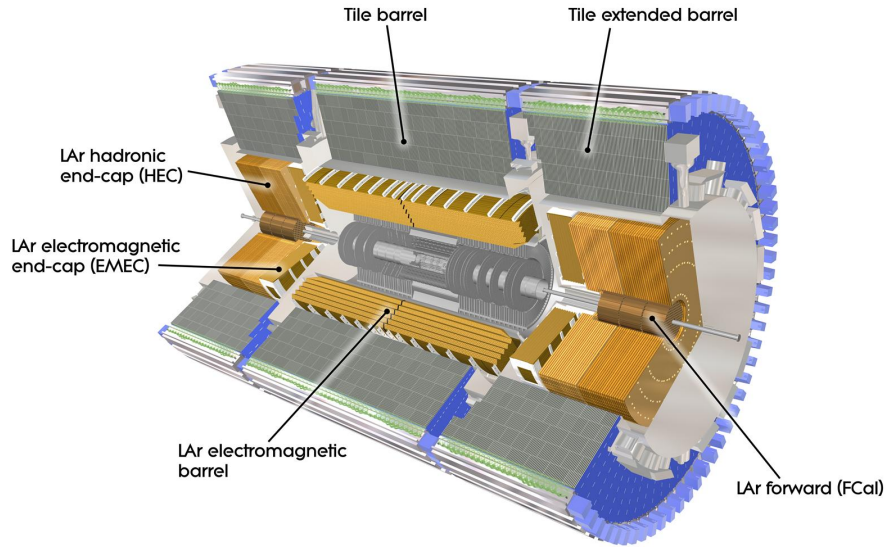
Figure 4.7: The ATLAS calorimeters. [131]

4.3.4 Muon Spectrometer

The Muon Spectrometer [155], pictured in Figure 4.8, is a tracker designed to measure the momenta of charged particles penetrating the calorimeters, which in practice are almost exclusively muons, in the pseudorapidity range $|\eta| < 2.7$ and to trigger on those particles in the range $|\eta| < 2.4$. Like the Inner Detector, the Muon Spectrometer is immersed in a magnetic field generated by a 0.5 T toroidal magnet system with a bending power of $2.0 - 7.5$ Tm. The Muon Spectrometer has four types of gas detectors: Thin Gap Chambers (TGCs), Resistive Plate Chambers (RPCs), Monitored Drift Tubes (MDTs), and Cathode Strip Chambers (CSCs).

Figure 4.8: The ATLAS Muon Spectrometer. [131]

The TGCs and RPCs have a high time resolution and are used for triggering on muons in the

endcap and barrel, respectively, in addition to providing a so-called second coordinate (in the $r - \phi$ non-bending direction) for the tracker. Each TGC is composed of a plane of closely-spaced wires at a high positive voltage, sandwiched between resistive grounded cathode planes. The distance (gap) between anode wires and cathode planes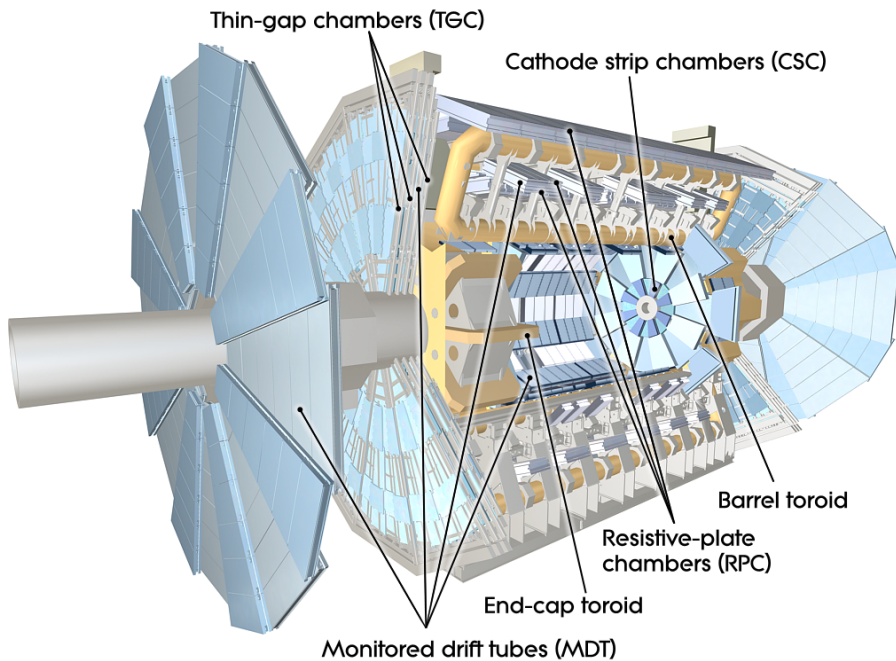 is smaller than the spacing between wires. The gas mixture is composed of CO_2 and n-Pentane. The time resolution of TGCs is approximately 25 ns, which corresponds to the bunch spacing at the LHC. RPCs are composed of closely-spaced phenolic resin plates with graphite electrodes and copper readout strips. They use a $C_2H_2F_4$-C_4H_{10}-SF_6 gas mixture and have a very high time resolution of around 1 ns. The number of readout channels is 440,000 for the TGCs and 380,000 for the RPCs.

The CSTs and MDTs provide precision coordinates for tracks in the endcaps and barrel, respectively, covering $|\eta| < 2.7$. The MDTs are much like the straws in the Inner Detector Transition Radiation Tracker; they are between 85 cm and 6.5 m long, have a 3 mm diameter, and are filled with a Ar-CO_2-H_2O gas mixture. The CSTs are multi-wire proportional chambers with segmented cathodes, using an Ar-CO_2 gas mixture. The MDTs have over 350,000 readout channels and a spatial resolution of 80 μm, while the CSTs have 70,000 channels and a spatial resolution of 70 μm.

4.3.5 Magnet system

The purpose of the ATLAS magnet system, pictured in Figure 4.9, is to bend the trajectories of charged particles so that their momenta can be measured. The magnet system comprises three main parts: the Central Solenoid Magnet, the Barrel Toroid, and the Endcap Toroids. All of these are powerful superconducting magnets made out of Nb-Ti/Cu Rutherford (many-stranded helix) cable encased in aluminum, which acts as a magnetic field stabilizer. The magnets must be cooled below the critical temperature of the cable material in order to enter the superconductivity regime, i.e. a state in which the wire in the cable has no electrical resistance and can conduct much larger electric currents than ordinary resistive wire, resulting in more powerful magnetic fields. At ATLAS, the magnets are kept at a temperature of 4.7 K, which is achieved with a liquid helium cooling system. The Central Solenoid wraps around the Inner Detector and provides a 2 Tesla field. To keep material build-up between detector subsystems to a minimum, the magnet shares a cryostat with the LAr calorimeter. The air-cored toroidal magnets coil around the Muon Spectrometer and produce a non-uniform magnetic field with a bending power of $2.0 - 8.0$ Tm. The electrical current passing through the magnets is $7,600$ Amperes for the solenoid and $20,500$ Amperes for the toroids.
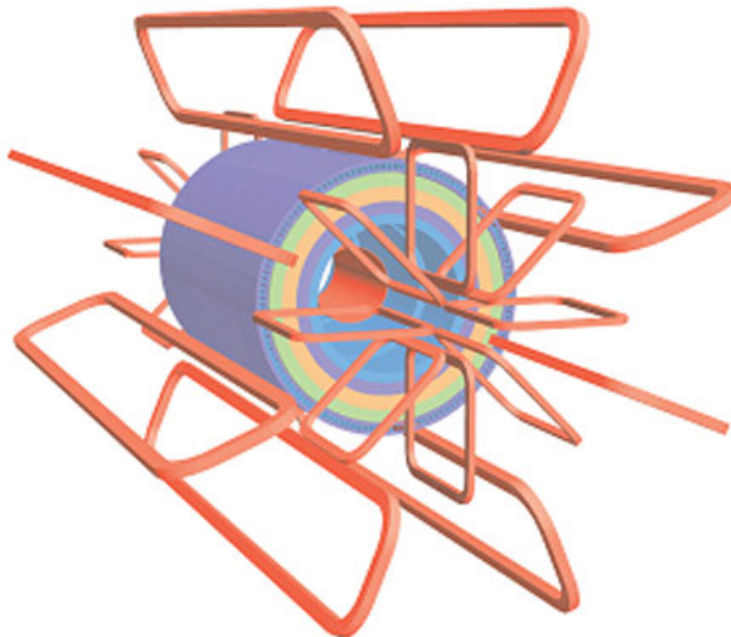

Figure 4.9: The ATLAS magnet system. [131]

4.3.6 Trigger system

ATLAS uses a specialized two-level trigger system for selecting potentially interesting events while keeping the flow of data manageable. The Level 1 trigger (L1) is the first to fire. This is a very fast, hardware-based system which looks at only a subset of available detector data from the calorimeters and the Muon Spectrometer. The calorimetric part of the trigger selects events with high-E_T objects, including events with large missing transverse energy from neutrinos. The muonic subsystem triggers on muons exceeding a predefined p_T threshold. Next, a topological subsystem combines the two types of data into topological variables, based on geometric properties of the event. With this information, the L1 can determine 'regions of interest' (RoIs) in η and ϕ for the next-level trigger, and reduce the event rate from approximately 30 MHz to 100 kHz. The processing time is less than 2.5 ms, during which time events are stored in pipelined buffers.

Events passing the L1 are passed to the software-based high level trigger (HLT) via the RoI Builder, which assembles a full record of L1 RoIs. The HLT can then examine these RoIs with full-granularity information from the detector and use offline-like reconstruction algorithms. If the reconstructed objects pass predefined requirements ('triggers' requested by different analysis groups), the event is transferred to local storage and exported to the Tier 0 Data Centre for offline reconstruction. The event rate is reduced to approximately 1 kHz, with an HLT processing time of about 200-300 ms.

4.4 Particle reconstruction and identification at ATLAS

Despite the high pile-up experienced by experiments at the LHC, many high-energy particles traversing the ATLAS detector can be reconstructed and accurately identified with the help of specialized algorithms combining tracking and calorimetric information. An overview of how various particles interact with different detector elements can be seen in Figure 4.10. Charged particles, such as electrons, muons, and protons, leave tracks in the tracking detectors; electromagnetically-interacting particles deposit their energy in the electromagnetic calorimeter in the form of electromagnetic showers; strongly-interacting particles produce showers in the hadronic calorimeter. Neutrinos typically pass through without interacting with the detector material and are reconstructed as missing transverse energy.

Not all physics objects used in ATLAS analyses correspond to individual particles. In addition to missing transverse energy, which is the collective signature of neutrinos in an event, there are also jets - collective signatures of hadrons produced in the hadronization of a quark or gluon, as described in Chapter 2. These hadrons travel in a collimated cone and leave behind characteristic signatures with a large multiplicity of closely-grouped tracks and high-energy calorimeter clusters. The properties of a jet, such as its width and energy, can be used to infer the properties of the parton from which they originate. The jet reconstruction strategy can also be used to reconstruct and identify hadronically-decaying tau leptons. The reconstruction and identification strategies for some of the main physics objects used at ATLAS - electrons [156,157], muons [158,159], jets [160], hadronically-decaying tau leptons [161,162], and missing transverse energy [163] - are described in more detail below. It should also be noted that ATLAS uses a system of 'working points' (selection categories) - usually loose, medium, and tight - to score the quality of reconstructed and identified physics objects. These working points correspond to objects with increasingly higher selection purity and increasingly lower selection efficiency, defined as the number of correctly identified objects of a specific type divided by the total number of these objects that entered the detector.

4.4.1 Electrons

Electrons are reconstructed by matching tracks passing a certain p_T threshold to calorimeter clusters. Tracks reconstructed in the crack region are excluded. Calorimeter clusters are reconstructed using a sliding-window algorithm [165], which clusters calorimeter cells within a fixed-size window positioned so that the energy contained within is a local maximum. Tracks are reconstructed with a combinatorial Kalman filter [166] starting from seeds, which are built using combinations of silicon detector measurements. Tracks are then assigned scores based on criteria such as the track p_T and χ^2 of the track fit. An ambiguity solver uses track scores and other quality criteria to resolve
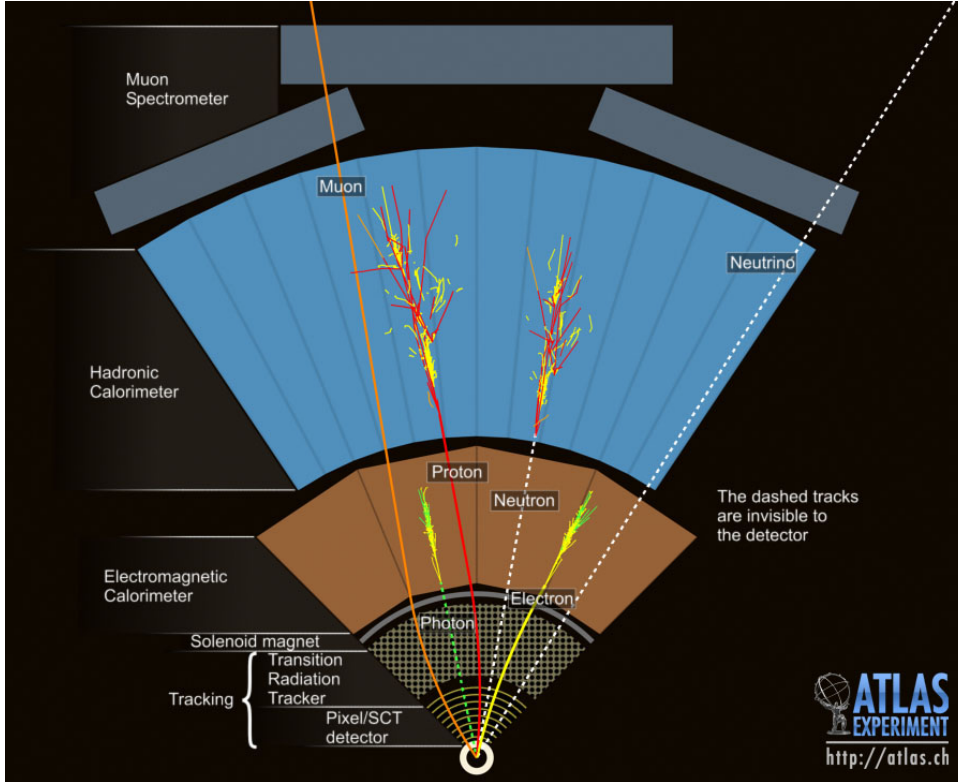
Figure 4.10: Interactions of various particles with the sub-components of the ATLAS detector. [164]

overlaps. Tracks are matched to calorimeter clusters if they pass $\Delta\phi$ criteria. If multiple tracks are matched to one cluster, distances in $\Delta\eta$ and the number of hits in different parts of the Inner Detector are taken into account to determine the most likely electron candidate.

Identification is based on a likelihood discriminant which uses variables related to the quality of the track and the match, the shape of the calorimetric cluster, hadronic leakage information, the suspected photo-conversion origin of electrons, and other criteria. Four working points - very loose, loose, medium, and tight - are defined, corresponding to fixed values of the likelihood discriminant score. The identification efficiencies vary as a function of electron E_T within each selection category. For electrons with $E_T = 20$ GeV the identification efficiencies are equal to 85% for the loose working point, 77% for the medium working point, and 67% for the tight working point; for electrons with $E_T = 100$ GeV the numbers are 96%, 94%, and 90%, respectively.

Additionally, isolation criteria can be applied to improve the differentiation between electrons from hard-scattering (signal) processes, which are typically expected to be well-isolated, and background processes like decays of heavy quarks and photo-conversion, where more detector activity near the electron is expected. Both calorimeter-based and track-based isolation algorithms are used. The implementation of isolation criteria depends on specific analysis needs. A variety of working points are defined in three categories: plain, where the isolation efficiency ϵ_{iso} is flat in electron p_T and η; gradient, where ϵ_{iso} is flat in η only; and fix, with fixed requirements on the value of the isolation variable.

4.4.2 Muons

Muon reconstruction begins with independent track reconstruction in the Inner Detector and the Muon Spectrometer. Inner Detector tracks are reconstructed in the same way as for electrons. In the Muon Detector, track segments are built individually for each of the four detector components, then combined using a segment-seeded combinatorial search. At least two segments must be matched to form a full track, except in the crack region, where one high-quality track is sufficient. Full muon candidates can be reconstructed in various ways, leading to four muon candidate types:

- **Combined muons:** Inner Detector and Muon Spectrometer tracks are matched geometrically, usually using outside-in pattern recognition.

- **Segment-tagged muons:** an Inner Detector track is geometrically matched to at least one Muon Spectrometer segment in the MDT or CSC chambers.

- **Calorimeter-tagged muons:** an Inner Detector track is geometrically matched to a calorimeter deposit consistent with a minimum-ionizing particle.

- **Extrapolated muons:** a Muon Spectrometer track passes a loose requirement on compatibility with originating from the interaction point; no Inner Detector track match is needed.

Overlaps between different muon types are resolved using hierarchy rules.

Muon identification is based on quality criteria such as the number of hits, differences between the charge and momentum of matched Inner Detector and Muon Spectrometer tracks, and the χ^2 of the combined track fit, where applicable. Five working points are defined: loose, medium, tight, low-p_T, and high-p_T. The loose working point is designed to maximize the identification efficiency while maintaining reasonable track quality; all muon types are allowed. The medium working point is the default at ATLAS and is designed to minimize the systematic uncertainties associated with muon reconstruction and calibration. Only combined and extrapolated muon types are used. The reconstruction efficiency of medium muons is over 95%. The tight working point is designed to maximize the selection purity and also uses only combined and extrapolated muon types. Lastly, the low-p_T working point maximizes the selection efficiency of low-p_T muons, while the high-p_T working point maximizes the momentum resolution of high-p_T muons, both using all muon types. Like in the case of electrons, additional isolation criteria are defined for muons in plain, gradient, and fix categories.

4.4.3 Jets

Jet reconstruction starts with clustering calorimeter cells using a topological algorithm [167], starting from a seed cell and iteratively adding neighboring cells with energies above a threshold defined as a function of the expected noise. These TopoClusters can be calibrated at either the electromagnetic scale, where the energies of constituent cells are simply summed, or at the local calibration (LC) scale, which accounts for the difference between electromagnetic and hadronic interactions in the ATLAS calorimeters. The second approach considers properties including the cluster energy density, isolation, and depth in the calorimeter, and corrects the average TopoCluster to the hadronic energy scale. TopoClusters then serve as inputs to the anti-k_T jet reconstruction algorithm [168, 169]. The typical size parameter for a single jet is $\Delta R = 0.4$. Sometimes the experimental signature includes very closely spaced jets, which cannot be separated given the calorimeter resolution; in this case, 'fat jets' with size parameter $\Delta R = 1.0$ can be reconstructed. Finally, jets are calibrated to suppress pile-up, and account for the energy that was not measured by the detector, either because it was deposited outside the active volume or because it was expended on non-ionizing nuclear processes. Jets can be matched to primary vertices with the Jet Vertex Tagger [170], a multivariate technique using tracking and vertexing information.

b-tagging

Jets resulting from the hadronization and decay of b quarks can be identified (tagged) due to some characteristic features [171–173]. Among these is the relatively long mean lifetime of the b quark, resulting in a displaced secondary (b-hadron decay) vertex. Since b quarks typically decay to c quarks, the b-hadron decay is usually followed by a c-hadron decay, forming a tertiary vertex. There are three types of basic algorithms used for b-tagging at ATLAS: impact-parameter-based algorithms, secondary vertex reconstruction algorithms, and decay chain multi-vertex reconstruction algorithms. Their outputs can be used as inputs for a multivariate classifier, which performs better than any of the basic algorithms do on their own.

The first basic algorithm type uses the impact parameter: the closest distance from the calculated primary collision vertex to a certain track. In b-hadron decays, the b-jet tracks originating from the displaced secondary vertex tend to have large impact parameters. An impact parameter

has two components: the transverse impact parameter d_0, which is the distance of closest approach in the $R - \phi$ plane, and the longitudinal impact parameter $z_0 \sin \phi$, which is the distance from the track to the primary vertex in the longitudinal plane at the point of closest approach in the $R - \phi$ plane. The two main impact-parameter-based algorithms at ATLAS are IP2D and IP3D. The input to IP2D is the significance of the transverse impact parameter, d_0/σ_{d_0}, where σ_{d_0} is the uncertainty. IP3D uses the significance of both the transverse and the longitudinal components, where the latter is $z_0 \sin \phi/\sigma_{z_0 \sin \phi}$ and $\sigma_{z_0 \sin \phi}$ is likewise the uncertainty. For each track in a jet, the IP2D and IPD3 algorithms calculate probabilities that the track is part of a b-jet, a c-jet, or a light jet. The final output for a jet can be a number of log-likelihood ratio discriminants comparing these probabilities. Each possible discriminant is calculated as the sum of contributions from all N tracks in a jet: $\sum_{i=1}^{N} \log \frac{p_X}{p_Y}$, where p_X and p_Y can be any of the three probabilities.

The secondary vertex reconstruction algorithm, SV, reconstructs a displaced secondary vertex by testing all track pairs within a jet to see if they originate from the same spot. The track pair and match must pass quality criteria, including minimum impact parameter thresholds and criteria designed to reject vertices from decays of various long-lived particles, interactions with detector material, and photon conversions. Once a set of high-quality matches is established, a single vertex is reconstructed by iteratively removing outlier tracks from the set. The main multi-vertex reconstruction algorithm at ATLAS is called JetFitter. This algorithm uses a Kalman filter to find primary, secondary, and tertiary vertices located in one line.

The main multivariate algorithm used during Run 2 was the MV2 algorithm. This is a boosted decision tree (BDT), which uses outputs from the IP2D, IP3D, SV, and JetFitter algorithms as inputs, in addition to the kinematic variables $p_T(\text{jet})$ and $\eta(\text{jet})$. Three variants of MV2 were developed, with different fractions of c-jets and light jets used in the background sample during model training. This was in order to obtain different levels of light versus c-jet rejection performance to suit different analysis group needs. The three variants are called MV2c00, MV2c10, and MV2c20, with the two digits at the end originally corresponding to the fraction of c-jets in the background sample (0%, 10%, and 20%). During later reoptimization, these fractions were changed to 0%, 7%, and 15% (respectively), without changing the variant names. The recommended variant is MV2c10. This variant has four working points (called 'operating points' in this instance) defined by fixed BDT score thresholds. The b-tagging efficiencies of these points are 60%, 70%, 77%, and 85%, based on $t\bar{t}$ sample measurements. The c-jet, light jet, and $\tau_{\text{had-vis}}$ rejection rates decrease as the b-tagging efficiency increases. Towards the end of Run 2, a deep learning algorithm called DL1 was also developed [174].

4.4.4 Hadronically-decaying tau leptons

The reconstruction and identification of tau leptons is challenging for several reasons. Firstly, the mean lifetime of tau leptons is only 2.9×10^{-13} s, translating to an average decay length that is much smaller than the radius of the LHC beampipe; this means that most tau leptons never enter the detector volume and their presence must be inferred from their decay products. Secondly, tau leptons have numerous possible decay modes (as listed in Section 1.1.7), resulting in a range of possible signatures. Furthermore, all decay modes have at least one neutrino in the final state.
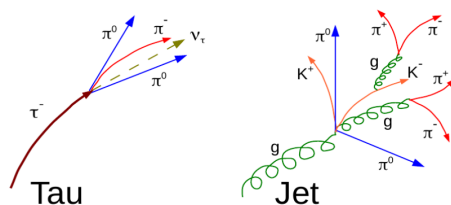

Figure 4.11: Examples of a typical tau jet and QCD jet.

Leptonically-decaying tau leptons, where only the electron or muon from the tau lepton decay interacts with the detector, are very difficult to differentiate from single light leptons from other sources. Because of this, leptonically-decaying tau leptons are not tagged as tau physics objects

47

at ATLAS. Hadronically-decaying tau leptons, on the other hand, produce collimated cones of low-multiplicity charged and neutral pions, kaons, and other mesons. These decay products leave jet-like signatures in the detector and can be reconstructed with the same algorithm. The hadron daughters of a hadronically-decaying tau object can be referred to collectively as a tau jet (even though they do not constitute a true jet as per the definition) or a visible tau, $\tau_{\text{had-vis}}$, since the neutrino is not accounted for in the jet reconstruction algorithm. QCD jets and tau jets can be differentiated by their multiplicity and width. Tau jets typically contain only one or three charged particles, which leave tracks in the Inner Detector; these are referred to as 1-prong and 3-prong decays, respectively. QCD jets tend to have a higher particle multiplicity. Tau jets also tend to be narrower than QCD jets, as illustrated in Figure 4.11. At ATLAS, tau jet identification is done with dedicated machine learning models, which use a number of input variables to exploit these differences. The default model was initially a boosted decision tree (BDT) [161], but in 2019 it was superseded by a recurrent neural network (RNN) [162]. Tau jet candidates entering the identification stage are jets reconstructed with the anti-k_T algorithm with p_T above 10 GeV and $|\eta| < 2.5$. After the tau lepton decay vertex is calculated, tracks with $p_T > 1$ GeV can be matched to the TopoCluster. The cone of $\Delta R < 0.2$ around the axis of the jet is called the core region, as all of the tau lepton decay products are expected to fit within in. The $0.2 < \Delta R < 0.4$ region is referred to as the isolation region. Working points correspond to BDT or RNN score thresholds.

After identification, the energies of tau jets are recalibrated from the calorimeter ('local cluster') scale to the so-called tau energy scale to account for energy not measured by the detector and suppress pile-up.

BDT tau lepton identification

BDTs were trained separately for 1-prong and 3-prong $\tau_{\text{had-vis}}$ candidates using the input variables below:

- **Central energy fraction (f_{cent}):** the fraction of the calorimeter transverse energy deposited in the region $\Delta R < 0.1$ with respect to all energy deposited in the region $\Delta R < 0.2$ around the $\tau_{\text{had-vis}}$ candidate.

- **Leading track momentum fraction ($f^{-1}_{\text{leadtrack}}$):** the transverse energy sum deposited in all cells belonging to TopoClusters in the core region of the $\tau_{\text{had-vis}}$ candidate, divided by the transverse momentum of the highest-p_T charged particle in the core region.

- **Track radius ($R^{0.2}_{\text{track}}$):** The p_T-weighted ΔR distance of the associated tracks to the $\tau_{\text{had-vis}}$ direction, using only the tracks in the core region.

- **Leading track impact parameter significance ($|S_{\text{leadtrack}}|$):** the absolute value of the transverse impact parameter (the closest distance from the calculated vertex to the track) of the highest-p_T track in the core region, calculated with respect to the $\tau_{\text{had-vis}}$ vertex, divided by its estimated uncertainty.

- **Fraction of tracks in p_T in the isolation region ($f^{\text{track}}_{\text{iso}}$):** the scalar sum of the p_T of tracks associated with the $\tau_{\text{had-vis}}$ candidate in the region $0.2 < \Delta R < 0.4$ divided by the sum of the p_T of all tracks associated with the $\tau_{\text{had-vis}}$ candidate.

- **Maximum ΔR (ΔR_{max}):** the maximum ΔR between a track associated with the $\tau_{\text{had-vis}}$ candidate and the $\tau_{\text{had-vis}}$ direction.

- **Transverse flight path significance (S^{flight}_T):** the decay length of the secondary vertex, reconstructed from the tracks associated with the core region of the $\tau_{\text{had-vis}}$ candidate, in the transverse plane, calculated with respect to the tau lepton vertex, divided by the estimated uncertainty.

- **Track mass (m_{track}):** the invariant mass calculated from the sum of the four-momenta of all tracks in the core region and isolation region, assuming a pion mass for each track.

- **Fraction of EM energy from charged pions ($f^{\text{track-HAD}}_{\text{EM}}$):** the ratio of energy deposited in the electromagnetic calorimeter to the energy of charged tracks in the core region. The

Table 4.1: Use of input variables in the 1-prong (1p) and 3-prong (3p) $\tau_\text{had-vis}$ identification BDTs.

Variable	1p BDT	3p BDT
f_cent	✔	✔
$f_\text{leadtrack}^{-1}$	✔	✔
$R_\text{track}^{0.2}$	✔	✔
$|S_\text{leadtrack}|$	✔	
$f_\text{iso}^\text{track}$	✔	
ΔR_max		✔
S_T^flight		✔
m_track		✔
$f_\text{EM}^\text{track-HAD}$	✔	✔
f_track^EM	✔	✔
$m_\text{EM+track}$	✔	✔
$p_T^\text{EM+track}/p_T$	✔	✔

denominator is the sum of the TopoCluster energy in the electromagnetic calorimeter associated with the $\tau_\text{had-vis}$ candidate. The numerator is the difference between the sum of the track momentum in the core region, and the sum of TopoCluster energies in the hadronic calorimeter.

- **Ratio of EM energy to track momentum ($f_\textbf{track}^\textbf{EM}$):** the fraction of the sum of energy in the electromagnetic calorimeter part of TopoCluster associated with the candidate and the sum of the track momentum in the core region.

- **Track-plus-EM-system mass ($m_\textbf{EM+track}$):** the invariant mass of the system composed of the tracks and up to two most energetic electromagnetic calorimeter clusters in the core region.

- **Ratio of track-plus-EM-system to p_T ($p_T^\textbf{EM+track}/p_\textbf{T}$):** the ratio of the $\tau_\text{had-vis}$ p_T estimated by the vector sum of track momentum and two highest-energetic electromagnetic calorimeter energy deposits in the core region to the calorimeter-measured $\tau_\text{had-vis}$ p_T.

A correction factor was applied to each input variable in order to obtain a flat distribution of the average number of interactions per bunch crossing μ (pile-up). Table 4.1 shows which variables were used for which of the two BDT sets. Three working points are defined: loose, medium, and tight. In each of these selection categories, the requirements on the BDT score vary in order to achieve a flat combined reconstruction and identification efficiency (signal efficiency) distribution as a function of $\tau_\text{had-vis}$ p_T. The signal efficiencies of the three working points for 1-prong and 3-prong $\tau_\text{had-vis}$ candidates are, respectively: 60% and 50% for loose selection, 55% and 40% for medium, 45% and 30% for tight.

RNN tau lepton identification

Tau lepton identification with a recurrent neural network was developed in order to use lower-level track and calorimeter cluster information that was not used by the BDT classifier, and thus improve performance. This method exploits the fact that tracks and clusters can be ordered in a physics-motivated way - for example in order of descending p_T - and thus form a sequence for use in an RNN.

The network uses three types of inputs: track information, cluster information, and higher-level variables (corresponding roughly to BDT input variables). These three information streams enter the network separately, as shown in Figure 4.12. Sequences of tracks and clusters, ordered by p_T and E_T, respectively, are passed though two fully-connected layers with shared weights followed by two LSTM layers. Higher-level variables do not form meaningful sequences, so they are passed through three fully-connected layers. All three information streams are then merged and passed through three additional dense layers. The output, a number between 0 and 1, gives a measure of the likelihood that a $\tau_\text{had-vis}$ candidate is a true $\tau_\text{had-vis}$ object. As in BDT $\tau_\text{had-vis}$ identification, two models are trained separately for 1-prong and 3-prong decays.

Table 4.2: Use of high-level input variables in the 1-prong (1p) and 3-prong (3p) $\tau_{\text{had-vis}}$ identification RNNs.

Variable	1p RNN	3p RNN
$p_T^{\text{uncalibrated}}$	✔	✔
f_{cent}	✔	✔
$f_{\text{leadtrack}}^{-1}$	✔	✔
$\lvert S_{\text{leadtrack}} \rvert$	✔	
$f_{\text{iso}}^{\text{track}}$	✔	✔
ΔR_{max}	✔	✔
S_T^{flight}		✔
m_{track}		✔
$f_{\text{track}}^{\text{EM}}$	✔	✔
$m_{\text{EM+track}}$	✔	✔
$p_T^{\text{EM+track}}/p_T$	✔	✔

For each track or cluster, the information that enters the network is a sequence of low-level variables. A schematic for the track information stream of the neural network is shown in Figure 4.13. For tracks, these variables are the transverse momentum of the track and the original seed jet, transverse and longitudinal impact parameters, angular distance to the $\tau_{\text{had-vis}}$ axis, and the number of hits on the track in the three tracker layers: the IBL, other pixel detector layers, and the SCT. For calorimeter clusters, the variables are the transverse momentum of the original seed jet, transverse energy of the cluster, the angular distance to the $\tau_{\text{had-vis}}$ axis, and three cluster moments: cluster depth (the distance of the cluster barycenter from the calorimeter front face, determined along the cluster axis), the longitudinal cluster extension (the second moment of the longitudinal distance of cluster cells from the cluster barycenter, along the cluster axis), and the radial cluster extension (the second moment of the radial distance of cluster cells from the cluster axis).

The higher-level variables were the same as those used in BDT $\tau_{\text{had-vis}}$ identification, with the omission of $R_{\text{track}}^{0.2}$ and $f_{\text{EM}}^{\text{track-HAD}}$ and the addition of one new variable: $p_T^{\text{uncalibrated}}$, the $\tau_{\text{had-vis}}$ p_T at the local cluster scale prior to tau energy scale calibration. This variable is calculated from the sum of cluster four-momenta within $\Delta R < 0.2$ around the $\tau_{\text{had-vis}}$ candidate axis. Table 4.2 shows which high-level variables were used for the 1-prong and 3-prong models; all track and cluster variables were used equally for both models.

Finally, the output score distribution is transformed to be flat as a function of $\tau_{\text{had-vis}}$ p_T and μ, decreasing the dependence of the combined reconstruction and identification efficiency (signal efficiency) on these two variables. Four working points are defined: very loose, loose, medium, and tight. Their efficiencies for 1-prong and 3-prong decays are as follows, in the same order: 95% and 95%, 85% and 75%, 75% and 60%, 60% and 45%.
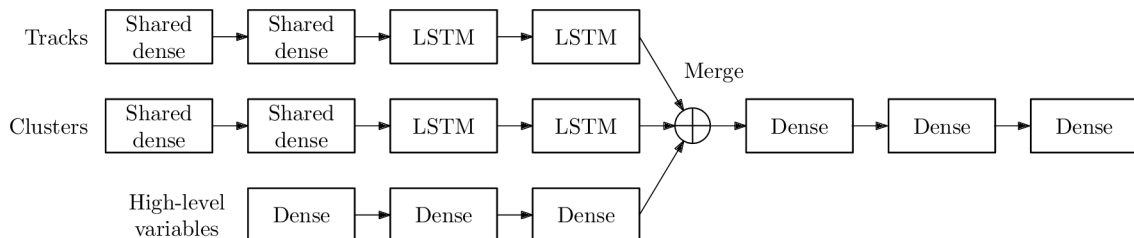

Figure 4.12: Schematic view of the tau lepton identification RNN model. [162]

Comparison of BDT and RNN tau lepton identification performance

The performance of BDT and RNN tau lepton identification models can be compared [162] by looking at the proportion of QCD jets misidentified as $\tau_{\text{had-vis}}$ that are successfully rejected for a given signal efficiency value. Figure 4.14 shows the rejection power - the reciprocal of the
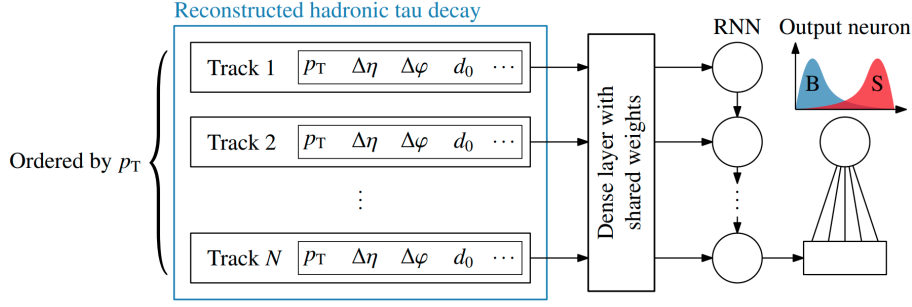
Figure 4.13: Schematic view of the track information stream in the tau lepton identification RNN model. The RNN label corresponds to the two LSTM layers. [162]

background selection efficiency - as a function of the true $\tau_{\text{had-vis}}$ efficiency for 1-prong and 3-prong BDT and RNN models. Markers correspond to the four working points defined for the RNN classifier. Across the entire signal efficiency range, the rejection power of the RNN classifier is approximately twice as high as the rejection power of the BDT classifier; this should lead to a $\sim 50\%$ reduction in QCD jets misidentified as $\tau_{\text{had-vis}}$ in physics analyses.
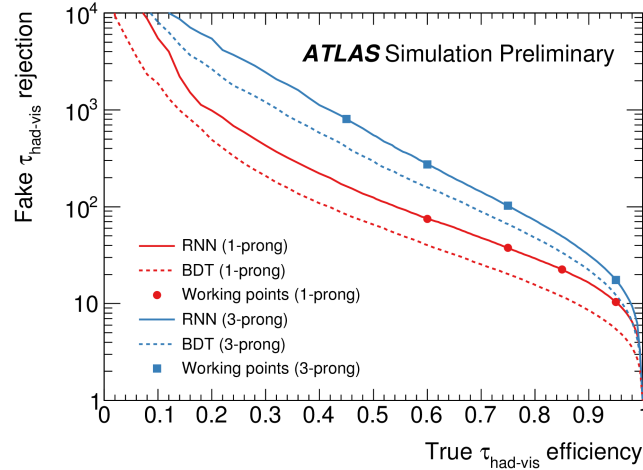

Figure 4.14: Rejection power as a function of true $\tau_{\text{had-vis}}$ efficiency for 1-prong and 3-prong BDT and RNN identification models. Circular and square markers correspond to 1-prong and 3-prong RNN working points, respectively: tight, medium, loose, and very loose, in order increasing signal efficiency. [162]

4.4.5 Missing transverse energy

The missing energy in an event refers to energy carried away by neutrinos or otherwise not recorded by the detector. Its magnitude and direction can be inferred from conservation of momentum. In hadron collisions the longitudinal component of the momentum of interacting partons is not known, while the transverse component should be zero. Therefore, the transverse momenta of reconstructed physics objects in an event can be used to reconstruct the total momentum of undetectable particles. Because the mass of neutrinos is close to zero, the missing transverse momentum is usually referred to as missing transverse energy, E_T^{miss}. The magnitude and direction (in ϕ) of E_T^{miss} is calculated using fully calibrated and reconstructed physics objects with the addition of a 'soft term' - an approximation of the p_T of low-energy particles that were not reconstructed. This soft term is determined from Inner Detector tracks associated with the hard scattering primary vertex, but not associated with any reconstructed physics objects, using the TST (Track-based Soft Term) algorithm [175].

4.4.6   Removal of overlapping physics objects

The presence of pile-up frequently causes ambiguities in the reconstruction of physics objects. If objects are reconstructed by different algorithms using the same detector hits, or otherwise in close proximity, overlap removal is performed. During the procedure, only one of the overlapping objects is kept and the others are discarded according to analysis-specific hierarchy rules.

4.5   The ATLAS data processing chain

Data processing at ATLAS requires several steps to obtain compatible real data and Monte Carlo samples which can be efficiently used by analysis groups. A schematic of the data processing chain defined by the ATLAS computing and analysis models [176, 177] is shown in Figure 4.15. Data processing up to and including the derivation stage is done with ATHENA [177], the ATLAS offline software based on the GAUDI event processing framework [178]. Physics analyses use a variety of specialized software frameworks, which are frequently based on ROOT [129], a CERN-born open-source data analysis framework.

Real data passing the trigger is collected in a RAW format [179], which encodes the detector response. Events (single records of bunch crossings that activated the trigger) are grouped in luminosity blocks, which correspond to approximately one minute of data taking. RAW files are passed through object reconstruction and identification steps, yielding the Analysis Object Data (AOD) format. During reconstruction, monitoring algorithms assess the data quality by checking whether all detector subsystems were functioning properly during a certain luminosity block and whether the performance of physics object reconstruction is acceptable. Luminosity blocks passing the quality criteria are included in 'good run lists' to be used for physics analyses. The proportion of good-quality data in the total data recorded by ATLAS during Run 2 can be seen in Figure 4.4.
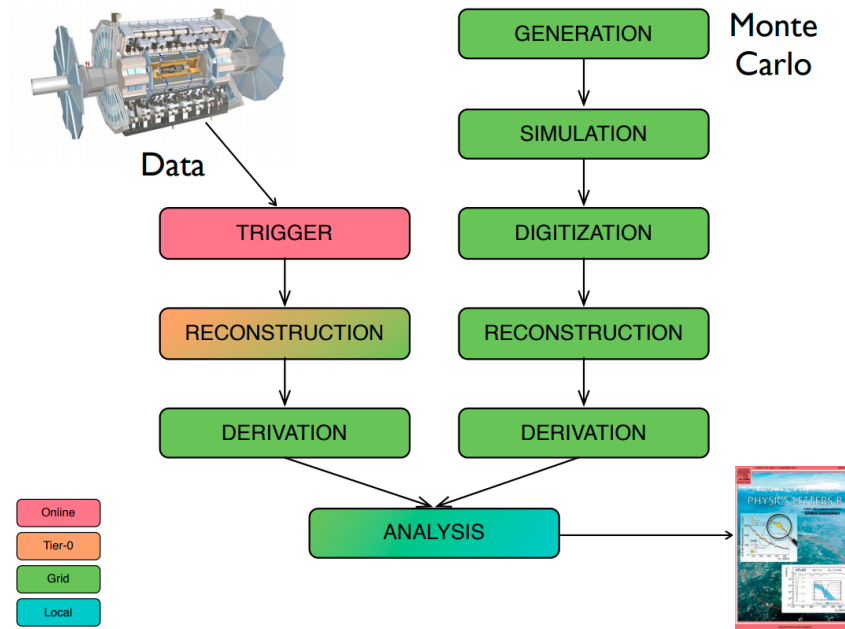

Figure 4.15: A flowchart showing the ATLAS data processing chain.

Official Monte Carlo production is done centrally for all ATLAS analyses. Samples of specific physics processes - from the simulation of interactions between quarks and gluons in the colliding protons to parton showering to hadronization and decays into stable particles - are generated by Monte Carlo generators in the EVNT format. These files contain the truth record in the HepMC format [180] wrapped in a form readable by ATHENA, but they are not ROOT-readable or compatible with the ATLAS Event Data Model, a model with a single set of objects representing

reconstructed physical objects, which is used throughout the data processing chain. In order to do generator-level truth studies of Monte Carlo samples, a ROOT-readable TRUTH format can be derived from EVNT files. The next step is detector simulation, done with the GEANT4 Detector Description and Simulation Tool [181] in ATHENA. During this step, GEANT4 calculates how particles in the EVNT samples interact with the ATLAS detector material, how they shower into secondaries, and how much energy they deposit in each sensitive element. The standard strategy is to use the full detector description; this gives the maximum accuracy but is also very CPU-intensive. A 'fast' strategy is also available, which uses pre-simulated particle showers loaded from a library. This reduces the CPU time by a factor of three. Alternatively, the ATLFASTII simulation package [182] can be used for applications where keeping the CPU time short is critical. This package uses the FASTCALOSIM calorimeter simulation [183], with energies deposited directly in calorimeter cells using parametrizations of the lateral and longitudinal particle shower energy profiles. This results in an order-of-magnitude improvement in CPU time over full GEANT4 simulation. The output of the detector response simulation step is in the HITS format, which in turn serves as the input for the digitization step. During thus step the simulated energy deposits are transformed into a detector response that looks like the real raw data. This simulated detector output, in the RDO format, is then passed through the reconstruction and identification steps, much like what is done for RAW files. The difference between Monte Carlo and real data AOD files is the additional presence of truth information (information about the true identities and properties of simulated particles, independent from reconstruction and identification) in the former, which is essential for various physics analyses.

Finally, both Monte Carlo and real data AODs are passed through the derivation stage, which involves data reduction and augmentation. During data reduction, samples might be skimmed (whole events removed based on some criteria), thinned (certain objects removed from all events), and slimmed (certain variables removed from all events). Data augmentation is done by adding new reconstructed objects (such as jets reconstructed with an alternative algorithm) and new, higher-level variables. The output of the derivation stage is in the Derived AOD (DAOD) format, which is much smaller than the AOD format, ROOT-readable, and can be used in physics analyses. Nevertheless, analysis groups often process DAODs further by producing ntuples - plain ROOT files with a flat structure, which can be read without ATLAS software and have the potential to be even smaller and faster.

Part II

Results

Chapter 5

Higgs Mass and Cosmological Constant Hierarchy Problems

As mentioned in Section 1.2, some of the open problems in physics are hierarchy puzzles, including the Higgs mass puzzle and the cosmological constant puzzle. The Higgs mass puzzle was discussed in some detail in Section 1.2.2. The cosmological constant puzzle is a conflict between cosmological observations and quantum field theoretical predictions. This constant, denoted here by C, appears in Einstein's general relativity field equations [57, 184], which relate the geometry of spacetime to the mass present within it:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = Cg_{\mu\nu} - \frac{8\pi G}{c^2}T_{\mu\nu},$$
(5.1)

where $R_{\mu\nu}$ is the Ricci curvature tensor, R is the scalar curvature (Ricci scalar), $g_{\mu\nu}$ is the metric tensor, G is Newton's constant, and $T_{\mu\nu}$ is the stress-energy tensor. The cosmological constant term, which was variously included in or excluded from Einstein's field equations as new advancements were made, must be larger than zero in order to account for the accelerating expansion of the Universe. Current measurements put the value of the cosmological constant at around $(0.002 \text{ eV})^4$ [185]. The cosmological constant is related to the vacuum energy density which is perceived by gravitation, ρ_{vac}:

$$C = \rho_{\text{vac}}8\pi G.$$
(5.2)

In quantum physics, one of the contributions to the energy density of empty space is quantum fluctuations (zero-point energies): the continuous creation and annihilation of virtual particles. Assuming that the Standard Model is valid up to the Planck scale, these fluctuations would be expected to give an energy density that is some 56 or 120 orders of magnitude larger (depending on the calculation method, as will be shown below) than the one determined from measuring the acceleration of the Universe [186–188]. This has been called the 'worst theoretical prediction in the history of physics' [184].

Since both the Higgs mass fine-tuning puzzle and the cosmological constant puzzle involve quantum fluctuations, they could be related. This relationship might involve supersymmetry, some other cancellation of bosonic and fermionic contributions, higher-dimensional operators of the Standard Model effective field theory, or some other mechanism. The following discussion is based on papers published by the author and Steven Bass [1, 2].

5.1 The Veltman condition and zero-point energies

The Higgs mass fine-tuning problem originates from the observation that the bare Higgs mass and the counterterm obtained during renormalization would have to be extremely fine-tuned to give the small Higgs mass seen experimentally. The form of this counterterm, at one-loop level, was first calculated by Veltman [189]:

$$\delta m_h^2 = \frac{\Lambda^2}{16\pi^2}\frac{6}{v^2}(m_h^2 + m_Z^2 + 2m_W^2 - 4m_t^2),$$
(5.3)

where Λ is the energy scale up to which the Standard Model remains valid (ex. the Planck scale), v is the Higgs vacuum expectation value, and m_i are masses of the heaviest Standard Model particles: the Higgs, Z, and $W^\pm$ bosons and the t quark. Contributions from lighter particles are neglected, as are NLO corrections, which are suppressed by a factor of $\frac{1}{(4\pi)^2}$. While this counterterm diverges quadratically with growing energy, the bare and renormalized masses can be made to coincide, eliminating fine-tuning, if the following condition is met:

$$m_h^2 + m_Z^2 + 2m_W^2 - 4m_t^2 = 0. \tag{5.4}$$

This is called the Veltman condition.

Pauli showed that a similar cancellation of fermionic and bosonic contributions can take place for zero-point energies [190], explaining the very small value of the cosmological constant. In flat spacetime, the total zero-point energy density of the vacuum would take the following form:

$$\rho_{\text{ZPE}} = \sum_{i \in \text{particles}} \frac{1}{2}\{\hbar\omega_i\} = \sum_i \frac{\hbar g_i}{2} \int_0^{k_{\max}} \frac{d^3 k}{(2\pi)^3} \sqrt{k^2 + m_i^2}, \tag{5.5}$$

where $\frac{1}{2}\{\hbar\omega_i\}$ are eigenvalues of the free Hamiltonian with $\omega_i = \sqrt{k^2 + m_i^2}$, k is the wavenumber, m_i is the mass of particle i, $g_i = (-1)^{2j}(2j+1)f$ is the degeneracy factor for particle i with spin j (with the minus sign following from the Pauli exclusion principle and the anticommutator relations for fermions), and the factor f equals 1 for bosons, 2 for each charged lepton (to account for antileptons), and 6 for each quark (to account for antiquarks and color charge). Here, using the Planck scale as the cutoff yields a value which is 10^{120} times larger than the experimentally-measured vacuum energy density, while using the electroweak scale (the Higgs vacuum expectation value) brings this factor down to 10^{56}. The pressure of the vacuum can be expressed with a similar formula [191]:

$$p_{\text{ZPE}} = \sum_{i \in \text{particles}} \frac{\hbar g_i}{6} \int_0^{k_{\max}} \frac{d^3 k}{(2\pi)^3} \frac{k^2}{\sqrt{k^2 + m_i^2}}. \tag{5.6}$$

Zero-point energies have ultraviolet divergences and must be regularized and renormalized. Using a (non-Lorentz-invariant) brute-force cutoff $k_{\max}$ yields the following cosmological equation of state:

$$p_{\text{ZPE}} = \frac{1}{3}\rho_{\text{ZPE}}. \tag{5.7}$$

This corresponds to the equation of state for radiation rather than for the cosmological constant (the vacuum energy) [192]. Switching to dimensional regularization in the modified minimal subtraction scheme $\overline{MS}$, a Lorentz-invariant approach, yields the correct equation of state [193]:

$$\rho_{\text{ZPE}} = -p_{\text{ZPE}} = \sum_{i \in \text{particles}} \frac{\hbar g_i m_i^4}{64\pi^2}\left(\frac{2}{\epsilon} + \frac{3}{2} - \gamma - 2\ln\left(\frac{m_i^2}{4\pi\mu^2}\right)\right) + \text{higher-order corrections}, \tag{5.8}$$

where $D = 4 - \epsilon$ is the number of dimensions, μ is the renormalization scale, and γ is the Euler–Mascheroni constant.

Equation 5.8 vanishes for massless particles such as the photon. For massive particles, a collective cancellation of zero-point energies from bosons and fermions takes place if the following conditions are met:

$$\sum_{i \in \text{particles}} g_i m_i^4 = 0, \tag{5.9}$$

$$\sum_{i \in \text{particles}} g_i m_i^4 \ln m_i^2 = 0. \tag{5.10}$$

Using Standard Model particles and neglecting light fermions, these take the form:

$$6m_W^4 + 3m_Z^4 + m_h^4 - 12m_t^4 = 0, \tag{5.11}$$

$$6m_W^4 \ln m_W^2 + 3m_Z^4 \ln m_Z^2 + m_h^4 \ln m_h^2 - 12m_t^4 \ln m_t^2 = 0. \tag{5.12}$$

Equations 5.4, 5.11, and 5.12 do not work with Standard Model particle masses. All three are fermion-heavy and would be (individually) satisfied if the Standard Model boson masses were higher - for instance, if the Standard Model Higgs boson had a mass of about 314, 319, and 311 GeV, respectively, rather than 125 GeV. However, these equations can be treated as constraints on possible undiscovered BSM particles.

Adding extra Higgs bosons, for example within the framework of 2HDMs (without the need for supersymmetry), could help balance out the equations, if the bosons have suitable masses. This could also help stabilize the electroweak vacuum, as the additional bosons would counteract the effect of the t quark loops. In practice, the low-mass 2HDM phase space is highly constrained by BSM Higgs boson searches and precision measurements of Higgs boson couplings, rare B meson decays, the muon anomalous magnetic moment, and electroweak parameters. For example, a global electroweak fit conducted by the Gfitter group [194] excluded m_H and m_A masses below 400 GeV for type-II and type-Y 2HDMs; Mahmoudi et al [195] set a limit of $m_H \geq 600$ GeV for the type-II 2HDM using flavor observables from Belle and BaBar experiments; while Misiak and Steinhauser [196] set the lower limit on m_H between 570 and 800 GeV for the type-II 2HDM using similar precision measurements. These numbers in the context of Pauli constraints imply that even if an extra type-II Higgs doublet were discovered, additional BSM fermions would be needed as well. However, certain 2HDMs are less constrained and could still satisfy the Pauli constraints. Extending the Veltman condition to accommodate extra Higgs bosons is more complicated. Some benchmark scenarios were developed by Darvishi and Krawczyk [197]; the scenario with the closest match to the ATLAS best-fit point [198] in the $\tan\beta$-$\cos(\beta - \alpha)$ plane for the type-II 2HDM assumes $m_A, m_{H^\pm} \sim 650$ GeV and $m_H \sim 830$ GeV.

5.2   Running masses

Veltman and Pauli conditions are evaluated from loop diagrams, causing the masses that appear within them to depend on the renormalization group scale. Since fermion and boson contributions evolve differently with the scale, there is a possibility that the conditions are met at some high energy scale without the need for extra particles.

This was tested by evolving Standard Model parameters with the `mr` C++ library [199], which uses the $\overline{MS}$ regularization scheme. The inputs were values of particle masses and the Higgs vacuum expectation value v published by the Particle Data Group (PDG) in 2019 [200]. The evolution of gauge coupling constants g, g', and g_s, the Higgs self-coupling λ, and the t quark Yukawa coupling y_t for energy scales up to 10^{20} GeV (an order of magnitude above the Planck scale) is shown in Figure 5.1. With 2019 PDG inputs, λ becomes negative at around 10^{10} GeV, and the vacuum loses stability. Stability all the way up to the Planck scale can be enforced by reducing the t quark mass from 173 GeV to 171 GeV. Figure 5.2 shows the evolution of particle masses and v, related to coupling constants via the following relations:

$$m_t = y_t \frac{v}{\sqrt{2}}, \tag{5.13}$$

$$m_W^2 = \frac{1}{4}g^2 v^2, \tag{5.14}$$

$$m_Z^2 = \frac{1}{4}(g^2 + g'^2)v^2, \tag{5.15}$$

$$m_h^2 = 2\lambda v^2. \tag{5.16}$$

Uncertainties on these parameters were calculated by varying inputs up and down within PDG uncertainties.

Figure 5.3 shows the evolution of the Veltman coefficient

$$C_{V1} = \frac{3}{v^2}(m_h^2 + m_Z^2 + 2m_W^2 - 4m_t) = \frac{9}{4}g^4 + \frac{3}{4}g'^4 + 6\lambda - 6y_t^2 \tag{5.17}$$

with the input value of m_t reduced to 171 GeV. The value of C_{V1} does not cross 0 below the Planck scale. To achieve crossing at the Planck scale, a Higgs boson pole mass of about 150 GeV is required. In alternative calculations, the crossing was achieved at about 10^{16} GeV with a metastable

vacuum in [201] and above 10^{20} GeV with a stable vacuum in [202–204]. If the crossing does occur below the Planck scale, it could point to an energy scale where a more fundamental theory takes over or where electroweak symmetry breaking is triggered, as suggested by Jegerlehner [201].

Figure 5.4 shows the evolution of the Pauli expressions given in Equations 5.11 and 5.12 for PDG inputs. The curve corresponding to the left side of Equation 5.11 crosses 0 at around 10^{16} GeV, at which point the bosonic contributions become larger than the fermionic contribution from the t quark. The evolution of the second Pauli expression terminates at 10^{10} GeV due to the negative value of λ, which causes the expression to develop an imaginary part, corresponding to the loss of vacuum stability. If the input value of m_t is reduced to 171 GeV, both Pauli constraints are satisfied between 10^{17} and 10^{18} GeV.

It should be noted that in the case of the Standard Model effective field theory, evolving parameter values above the energy cutoff would be unphysical.
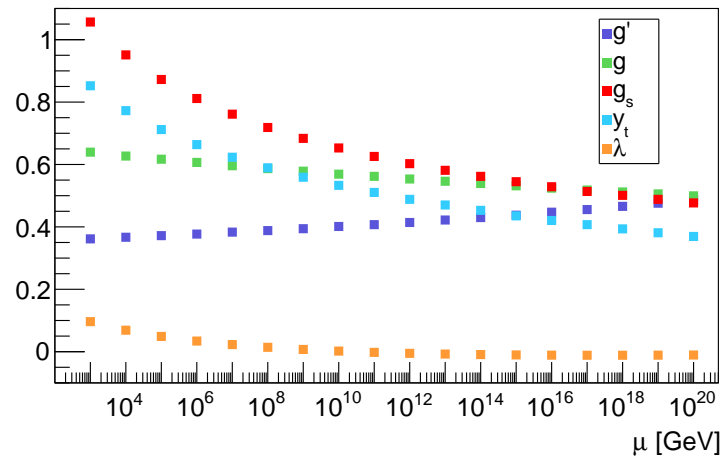

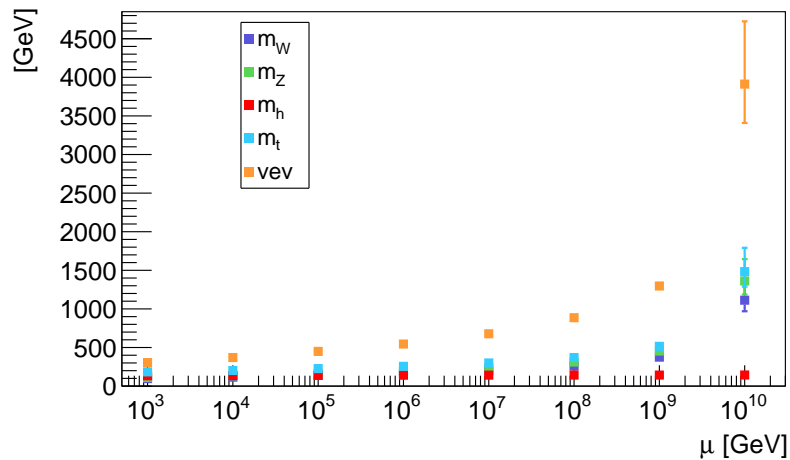Figure 5.1: Evolution of Standard Model coupling constants using the `mr` package.

Figure 5.2: Evolution of Standard Model particle masses and the Higgs vacuum expectation value using the `mr` package. The evolution terminates at the energy scale where λ becomes negative.
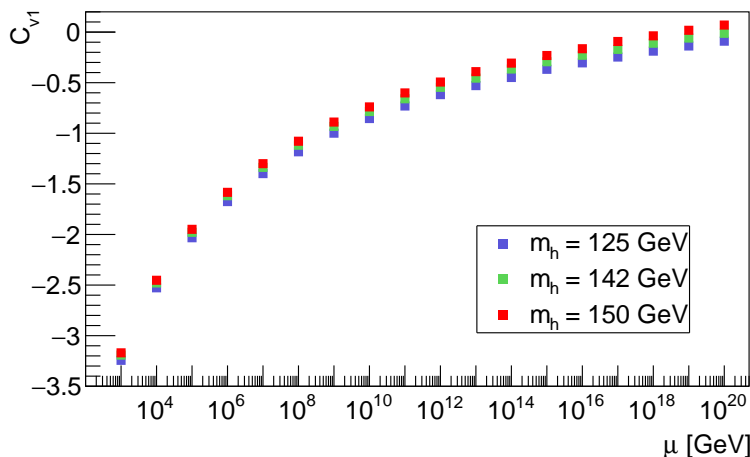
Figure 5.3: Evolution of the Veltman coefficient $C_{V1} = \frac{3}{v^2}(m_h^2 + m_Z^2 + 2m_W^2 - 4m_t) = \frac{9}{4}g^4 + \frac{3}{4}g'^4 + 6\lambda - 6y_t^2$ with $m_t = 171$ GeV and different Higgs boson pole masses using the `mr` package.

5.3 The Standard Model effective field theory and neutrino mass

In the absence of Veltman and Pauli crossing below the Planck energy scale, it is worthwhile to consider alternative approaches to solving the Higgs mass and cosmological constant puzzles. One such approach involves combining the idea of higher-dimensional operators in the Standard Model effective field theory with an observation about the components of the vacuum density.

The total vacuum energy density ρ_{vac} has several components. In addition to zero-point energies from quantum fluctuations, it receives contributions from the renormalized bare gravitational term ρ_{bare} [187] and dynamically-generated potential from Higgs and QCD condensates - vacuum mean values $\langle 0|\mathcal{O}_i|0\rangle$ of operators $\mathcal{O}_i$ arising due to non-perturbative effects. The full expression is then:

$$\rho_{\text{vac}} = \rho_{\text{ZPE}} + \rho_{\text{potential}} + \rho_{\text{bare}}. \tag{5.18}$$

All of these contributions depend on the renormalization scale μ. The zero-point energy contribution has both an explicit μ term (see Equation 5.5) and dependence on running masses. The Higgs potential depends on the running Higgs boson mass and the running Higgs self-coupling. However, ρ_{vac} is an observable - with a value of around $(0.002 \text{ eV})^4$, as mentioned above - and thus independent of the renormalization scale. This implies that the renormalization scale dependence of the components must cancel. A cancellation $\rho_{\text{vac}} = 0$ can be enforced (at dimension 4) by global space-time translational invariance [205]. In the Standard Model effective field theory, this leaves the non-renormalizable higher-dimensional operators, which should be taken into account. The leading contribution would be from the dimension-5 Weinberg operator (Equation 1.40), which is responsible for generating neutrino masses:

$$\rho_{\text{vac}} \sim \left(\frac{c^{(5)}v^2}{\Lambda}\right)^4 \tag{5.19}$$

with one factor of $\frac{c^{(5)}v^2}{\Lambda}$ for each dimension of spacetime. The Weinberg coefficient $c^{(5)}$ is assumed to be of the order of 1. As it happens, the cosmological constant scale of 0.002 eV is within experimentally-allowed values for neutrino masses and would correspond to new physics emerging at an energy scale of around 10^{16} GeV. Though this energy is very much beyond the reach of particle colliders, neutrino mass values and/or lepton number violation might eventually be discovered in next-generation neutrino experiments.

This scenario would point to an interesting connection between infrared and ultraviolet energy realms. While in the usual effective field theory approach it is assumed that the low-energy theory
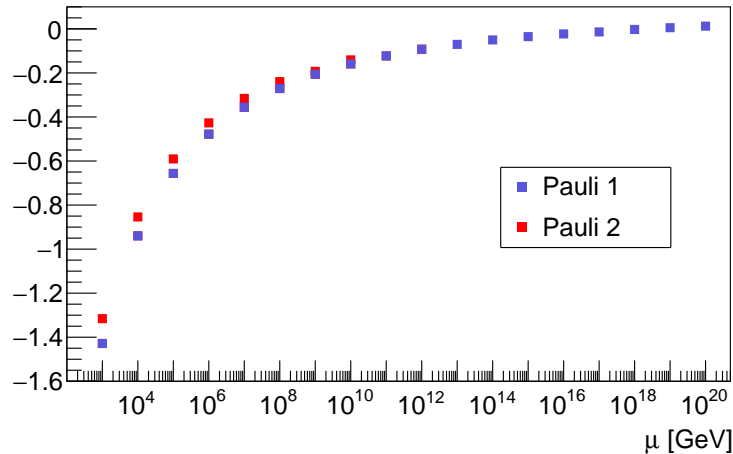
Figure 5.4: Evolution of Pauli expressions $P_1 = 6m_W^4 + 3m_Z^4 + m_h^4 - 12m_t^4 = v^4(\frac{9}{8}g^4 + \frac{3}{2}g^2g'^2 + \frac{3}{4}g'^4 + 4\lambda^2 - 3y_t^4)$ and $P_2 = 6m_W^4 \ln m_W^4 + 3m_Z^4 \ln m_Z^2 + m_h^4 \ln m_h^2 - 12m_t^4 \ln m_t^2 = v^4 \ln v^2(\frac{3}{8}g^4 \ln \frac{g^4}{4} + \frac{3}{16}(g^2 + g'^2)^2 \ln \frac{g^2+g'^2}{4} + 4\lambda^2 \ln 2\lambda - 3y_t \ln \frac{y_t^2}{2})$ with PDG pole masses using the `mr` package. P_1 was normalized to v^4 while P_2 was normalized to $v^4 \ln v^2$. The evolution of P_2 terminates at the energy scale where λ becomes negative.

is independent of what happens above the energy cutoff, in this case the new physics would be linked to what we observe at our energy scales. This connection could be responsible for the Standard Model Higgs boson and t quark masses being so close to the bounds needed for electroweak vacuum stability, and - more generally - for the values of Standard Model parameters, which seem fine-tuned to the existence of a life-supporting universe. In addition to the lower bound from vacuum stability requirements, an anthropic upper bound on the Standard Model Higgs boson mass can be derived from Equation 5.19. According to Weinberg [206], an order-of-magnitude increase in the value of $\rho_{\rm vac}$ would cause the Universe to begin its current period of accelerating expansion earlier, preventing the formation of galaxies. This anthropic upper bound on the value of $\rho_{\rm vac}$ corresponds to setting an upper bound on the value of v at 1.33 of its actual (measured) value, which leads to a corresponding bound on the Standard Model Higgs boson mass. This leaves quite a narrow window for possible mass values that could be observed by sentient beings.

5.4 Conclusion

The Higgs mass fine-tuning and cosmological constant puzzles are two open problems in physics, which might be related due to both involving quantum fluctuations. The cancellation of fermionic and bosonic contributions might be achieved with the Veltman condition and/or equations following from the Pauli conjecture. Unfortunately, this does not work with Standard Model particle masses. Here, the effect of adding extra BSM Higgs bosons predicted by 2HDMs was explored, taking into consideration existing experimental limits. The conclusion was that type-II 2HDM Higgs bosons with masses allowed by the existing limits would be insufficient to achieve cancellation using the Pauli conjecture, and additional fermions would be needed. Other 2HDM types are less constrained and could allow the cancellation to happen. The Veltman condition might still be satisfied for the type-II 2HDM and other 2HDM types. An alternative way of solving the cosmological constant puzzle within the framework of the Standard Model effective field theory was also explored, with the conclusion that the dimension-5 Weinberg operator could be simultaneously responsible for neutrino masses and the small value of the cosmological constant. This would also point to an interesting connection between the infrared and ultraviolet energy realms, which could explain the small value of the Standard Model Higgs boson mass. Future experiments in cosmology and neutrino physics, such as searches for neutrino-less double beta decay, a tell-tale sign that neutrinos are Majorana particles [71], should offer further clues and insights.

Chapter 6

Validation of Monte Carlo Samples with Tau Leptons

Final states with tau leptons are important in many high-energy physics searches and precision measurements. As the heaviest particles in the leptonic sector, tau leptons have a large Yukawa coupling to the Standard Model Higgs boson and possible MSSM Higgs bosons. The short lifetime and P-symmetry-violating decays of tau leptons allow for polarization measurements using decay product kinematics. These can differentiate between Z and H boson decays or help study the properties of new particles. Accurately modeling tau lepton kinematics, branching ratios, and polarization in Monte Carlo samples is key for analyses with tau leptons in the final state. Because Monte Carlo generators differ in how they model physics processes, it is necessary to validate each sample used in an experiment. This chapter will summarize a series of validation studies conducted mostly at 'truth' level - that is, using only information about real (truth) objects generated by Monte Carlo generators, without simulating interaction with the detector. More information about the physics of tau leptons and ATLAS physics sample formats can be found in Sections 1.1.7 and 4.5, respectively.

6.1 Processes and samples

The main physics processes considered in this set of studies were the Drell-Yan process and boson pair (diboson) production with tau leptons, and typically jets, in the final state. The former is a process where a quark and an antiquark (from the parton sea in colliding protons) annihilate into a virtual (massive) photon or Z boson, which in turn produces a lepton-antilepton pair. Feynman diagrams with virtual photons and Z bosons, shown collectively in Figure 6.1, interfere, and in practice it is impossible to separate the two process types; the Drell-Yan process with tau leptons in the final state is thus usually denoted by

$$q\bar{q} \to Z/\gamma^* \to \tau^+\tau^-. \tag{6.1}$$

The partons taking part in the hard-scattering interaction can radiate off other partons, which hadronize and form the final-state jets. The Drell-Yan process is especially important in hadron collision physics. It can be used for validating QCD predictions, examining the partonic structure of hadrons, and calibrating detectors, since the location of the Z peak on the lepton-antilepton invariant mass spectrum is well known. It also constitutes a large and irreducible background in Standard Model $H \to l^+l^-$ analyses and searches for new particles decaying to a lepton-antilepton pair.

The production of gauge boson pairs (any combination of $W^\pm$ bosons, Z bosons, and photons) occurs mostly via quark-antiquark annihilation, with a small contribution from gluon-gluon interaction. Some example Feynman diagrams are shown in Figure 6.2. Similarly to the Drell-Yan process, diboson events can be used to test QCD predictions and are an important background in some searches for new physics. In addition, they allow for precision measurements of triple and quartic gauge boson couplings, which is an important test of the electroweak sector of the Standard Model.
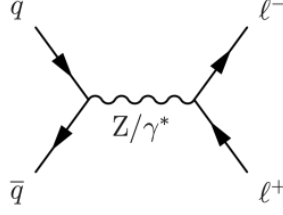
Figure 6.1: Tree-level Feynman diagram of the Drell-Yan process with a lepton-antilepton pair and no jets in the final state.
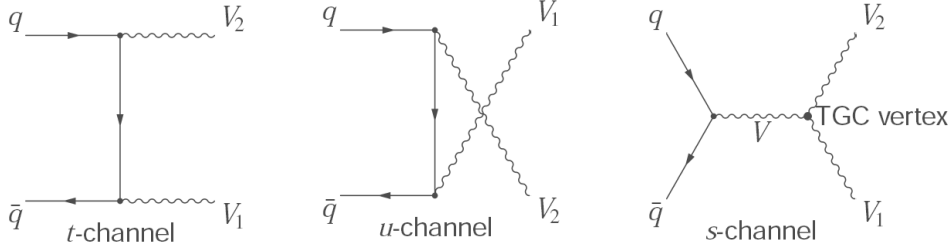

Figure 6.2: Leading-order Feynman diagram of three types of diboson production mechanisms: t-channel, u-channel, and s-channel (with labeled triple gauge coupling vertex).

The following Monte Carlo samples were used in validation studies described in this chapter:

- Drell-Yan $Z/\gamma^* \to \tau^+\tau^-$ samples generated with SHERPA 2.2.1, 2.2.7, 2.2.8, and 2.2.10 [207] using the NNPDF 3.0 NNLO PDF set [208],

- Drell-Yan $Z/\gamma^* \to e^+e^-/\mu^+\mu^-$ samples generated with SHERPA 2.2.10 using the NNPDF 3.0 NNLO PDF set,

- Drell-Yan $Z/\gamma^* \to \tau^+\tau^-$ samples generated with POWHEG-BOX v1 [209–212] interfaced to PYTHIA 8.1 [213] for the parton shower, with the CT10 PDF set [214] and the AZNLO tune [214] for the underlying event,

- samples with two vector bosons decaying via leptonic channels (combinations of Z/γ^* and $W^\pm$ bosons decaying to $l^+l^-l^+l^-$, $l^+l^-l^\pm\nu$, $l^\pm l^\pm\nu\nu$, $l^\pm l^\mp\nu\nu$ and $l^\pm\nu\nu\nu$ final states) generated with SHERPA 2.2.2 [207] using the NNPDF 3.0 NNLO PDF set.

The SHERPA 2.2.1 $Z/\gamma^* \to \tau^+\tau^-$ sample is a thoroughly validated 'legacy' sample used for many key analyses throughout Run 2. It is composed of individual invariant mass slices (sub-samples) weighted by their respective cross-sections, with a large number of events in each slice. This provides good statistics in the high-mass tail, which is important in analyses such as searches for neutral high-mass particles. The POWHEG-BOX+PYTHIA 8.1 sample, on the other hand, is an inclusive (non-sliced) sample with low statistics in the high-mass tail, which can be used in regions of phase space where the tail is not of importance. Drell-Yan samples generated with SHERPA 2.2.2+ were part of an ongoing effort to improve Monte Carlo dataset production for the demands of the High Luminosity LHC, which will see a large increase in luminosity but not in the CPU budget. Both theoretical and computational improvements were implemented, including improvements in parton shower simulation, the matching of parton showers to fixed-order perturbative calculations, efficiently increasing statistics in low cross-section regions, and introducing approximations that decrease the per-event computational time by a factor of three without affecting the modeling.

The fully-leptonic SHERPA 2.2.2 diboson samples are also legacy samples, and are used together with SHERPA 2.2.1 semi-leptonic diboson samples (samples where one of the bosons decays to leptons and the other to quarks) to model the diboson background in Run 2 analyses. Samples with two gauge bosons decaying exclusively to quarks are not used, as they are difficult to distinguish from multijet background. The reason for using different SHERPA setups for fully-leptonic and semi-leptonic samples is that a book-keeping bug-fix had to be applied to fully-leptonic SHERPA 2.2.1 samples.

6.2 Branching ratios and kinematic variables

Some of the most basic validation tests that can be conducted are checks of tau lepton branching ratios and distributions of kinematic variables like transverse momentum p_T, pseudorapidity η, and azimuthal angle ϕ. Reconstruction-level comparisons of kinematic variable distributions in samples generated with different Monte Carlo generators and in real data from the ATLAS detector were conducted in other validation studies. Here, only the generator-level comparison of SHERPA 2.2.10 $Z/\gamma^* \to e^+e^-/\mu^+\mu^-/\tau^+\tau^-$ will be discussed. Due to lepton universality, no differences are expected in kinematics prior to the decays of muons and tau leptons. Distributions of $\eta(e/\tau)$, $\phi(e/\tau)$, $p_T(e/\tau)$ and $p_T(e^+e^-/\tau^+\tau^-)$ (the scalar sum of lepton and antilepton transverse momenta), with no selection cuts applied, can be seen in Figures 6.3 and 6.4. The distribution of $p_T(e/\tau)$ exhibits a 'kink' structure at around 20 GeV, which is due to the generator-level requirement for the invariant mass of the lepton-antilepton pair, $m(l^+l^-) > 40$ GeV. Good agreement between electrons and tau leptons is seen in all of these distributions, with some statistical fluctuations. Muon distributions also exhibited good agreement, but are not pictured here (for brevity).
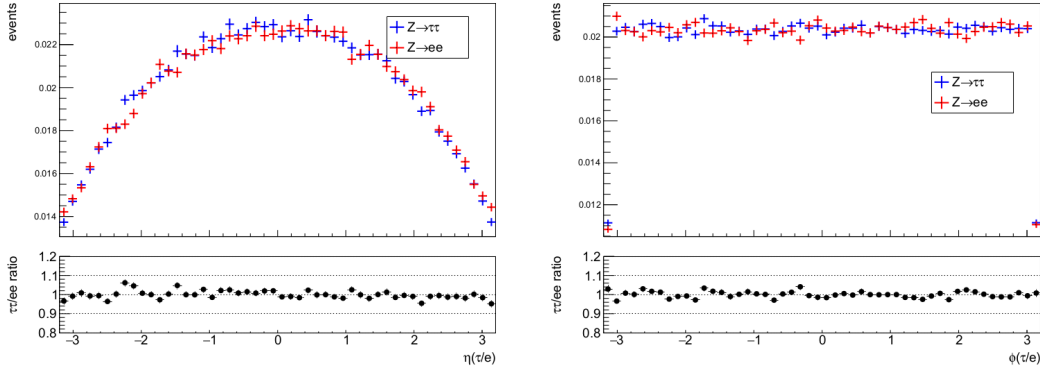

Figure 6.3: Comparison of $\eta(l)$ and $\phi(l)$ distributions for leptons from SHERPA 2.2.10 $Z/\gamma^* \to e^+e^-$ and $Z/\gamma^* \to \tau^+\tau^-$ samples.
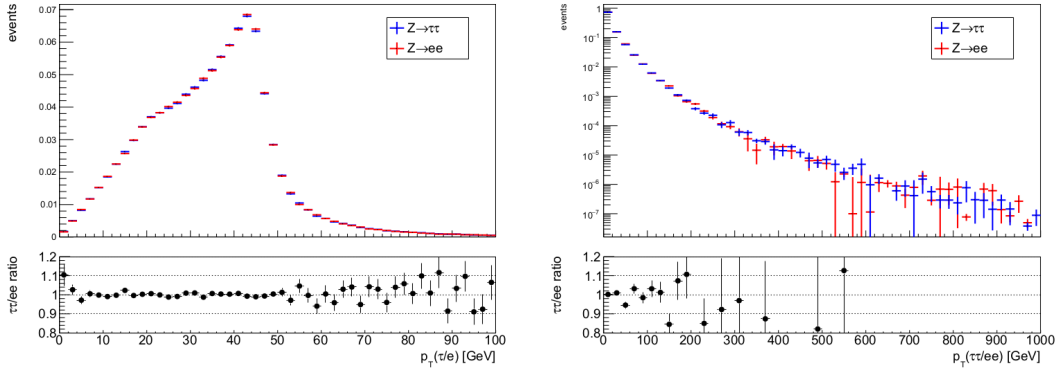

Figure 6.4: Comparison of $p_T(l)$ and $p_T(l^+l^-)$ distributions for leptons from SHERPA 2.2.10 $Z/\gamma^* \to e^+e^-$ and $Z/\gamma^* \to \tau^+\tau^-$ samples. The $p_T(l^+l^-)$ distribution is shown in logarithmic scale.

Branching ratios were found to agree reasonably well with PDG values across all samples. Table 6.1 shows the branching ratios of tau leptons from the SHERPA 2.2.2 $Z/\gamma^* \to \tau^+\tau^-$ sample, which can be compared to PDG values in Table 1.2 in Section 1.1.7. The branching ratios of 1-, 3-, and 5-prong decays, relative leptonic and hadronic decay fractions, and the two leptonic decay modes all agree within the quoted uncertainties. The hadronic decay modes show some disagreement, with SHERPA values being a bit higher than PDG values for $1h^\pm\ 0h^0$, $1h^\pm\ 1h^0$, $3h^\pm\ 0h^0$, and $3h^\pm\ 3h^0$ decay modes and a bit lower for $1h^\pm\ 2h^0$, $1h^\pm\ 3h^0$, and $3h^\pm\ 1h^0$ decay modes.

Table 6.1: Branching ratios of selected tau lepton decay modes measured in the SHERPA 2.2.2 $Z/\gamma^* \to \tau^+\tau^-$ sample. The decay modes are sorted by the number of prongs (charged decay products - leptons or hadrons), hadronic and leptonic modes, the two most common leptonic decay modes, and hadronic modes in $Nh^\pm$ Mh^0 format, where $h^\pm$ represents charged pions and/or kaons and h^0 represents neutral pions and/or kaons.

decay mode	branching ratio
1-prong	84.8±0.2%
3-prong	15.06±0.06%
5-prong	0.100±0.005%
Leptonic	35.2±0.1%
Hadronic	64.8±0.2%
$\tau^\pm \to e\nu_e\nu_\tau$	17.92±0.08%
$\tau^\pm \to \mu\nu_\mu\nu_\tau$	17.30±0.07%
$1h^\pm$ $0h^0$	12.76±0.06%
$1h^\pm$ $1h^0$	26.4±0.1%
$1h^\pm$ $2h^0$	9.31±0.05%
$1h^\pm$ $3h^0$	1.07±0.02%
$3h^\pm$ $0h^0$	9.78±0.05%
$3h^\pm$ $1h^0$	4.48±0.04%
$3h^\pm$ $2h^0$	0.52±0.01%
$3h^\pm$ $3h^0$	0.270±0.008%

6.3 Polarization measurement

Correctly modeling the polarization of tau leptons from vector boson decays is crucial in the case of analyses making use of spin correlation observables. In analyses that do not use such observables, correct spin effects are still important for modeling energy distributions and tau lepton acceptance in the detector. The polarization of fermions from such decays was discussed in Sections 1.1.4 and 1.1.7. In the case of $W^\pm \to \tau^\pm\nu$ decays, the tau lepton polarization is expected to be -100% due to $W^\pm$ bosons coupling exclusively to left-handed neutrinos and right-handed antineutrinos. The polarization of tau leptons in $Z \to \tau^+\tau^-$ boson decays, on the other hand, depends on the value of the Weinberg angle θ_W. The most accurate measurement of the polarization of tau leptons from Z boson decays was conducted at LEP, the Large Electron-Positron Collider [215] at CERN, in 2001 [216]. The obtained value was $(14.51\pm0.59)\%$, corresponding to an effective $\sin^2\theta_W$ parameter value of 0.23147±0.00057. More recent measurements of the effective value of $\sin^2\theta_W$ at other particle colliders came up with similar values: 0.23148±0.00033 from combined Tevatron results [217] and 0.23129±0.00033 from combined LHC results [218–221]. The value of this parameter can also be calculated from Z and $W^\pm$ boson masses using the tree-level formula in Equation 1.17, which gives a significantly smaller number of 0.22337±0.00010 [27]. A higher-order calculation using the $\overline{MS}$ renormalization scheme gives a value in line with experimental results, at 0.23121±0.00004 [27].

The polarization $P_\tau = \frac{\sigma_L - \sigma_R}{\sigma_L + \sigma_R}$ of hadronically-decaying tau leptons can be determined with a semi-analytical formula for the energy spectrum y of visible decay products in $\tau^\pm \to \pi^\pm\nu_\tau$ decays, derived by Jadach and Wąs [222]:

$$y = N(1 + P_\tau(2x_\pi - 1)), \tag{6.2}$$

where N is a normalization constant and x_π is the fraction of tau lepton energy carried by the pion in $\tau^\pm \to \pi^\pm\nu_\tau$ decays:

$$x_\pi = \frac{E(\pi^\pm)}{E(\tau^\pm)}. \tag{6.3}$$

This approach assumes that particle mass and QED bremsstrahlung can be neglected, which is a reasonable assumption for LHC proton-proton collisions. Formula 6.2 was fitted to x_π histograms using the `TH1::Fit` method in ROOT. Histograms for measuring the polarization of tau leptons from Z decays were constructed by selecting events with exactly two tau leptons with opposite electric charge and an invariant mass between 88 and 94 GeV. This invariant mass cut helped to select mostly on-shell Z peak events, with only a small contribution from virtual photons, which

couple equally to left-handed and right-handed fermions. Only $\tau^{\pm} \to \pi^{\pm} \nu_\tau$ decays were considered; these account for around 10.8% of all tau lepton decays [27].

Table 6.2 lists the polarizations of tau leptons from Z boson decays measured in SHERPA 2.2.2 diboson samples and SHERPA 2.2.7-10 Drell-Yan samples. In all of these samples, the polarization was simulated directly at generator level using specific electroweak input variable schemes. The standard $\{G_F,\ m_Z,\ m_W\}$ scheme mentioned in Section 1.1.8 was used in SHERPA 2.2.2 and 2.2.7. In this input scheme, the value of the $\sin^2 \theta_W$ parameter was calculated using the tree-level formula. This low value caused the measured polarization of tau leptons to be higher than expected from the LEP result. In SHERPA 2.2.8+, the input scheme was switched to $\{\sin^2 \theta_W,\ \alpha(m_Z),\ m_Z\}$ in order to use the effective value of the Weinberg angle and obtain the correct effective polarization. This worked as intended, with the tau lepton polarization matching the expected value within the statistical uncertainty in SHERPA 2.2.8 and 2.2.10 samples. Figure 6.5 shows the difference between fit lines obtained for samples with the two different electroweak input schemes. The first bin in the x_π histogram was excluded from the fit; its low value is caused by the requirement that the energy of the pion must be equal to or greater than its mass.

The polarization of tau leptons from $W^{\pm} \to \tau^{\pm} \nu$ decays in the SHERPA 2.2.2 diboson sample was also measured for completeness. In order to avoid selecting tau leptons from Z/γ^* boson decays, only events with one tau lepton in the final state were selected. No other cuts were applied. The measured polarization was -97%, less than the expected -100%. Since the polarization of tau leptons from $W^{\pm}$ boson decays does not depend on the electroweak input scheme, this was probably due to contamination from tau leptons originating from other (non-matrix-element) sources, like decays of B mesons.

Table 6.2: Polarization of tau leptons from $Z \to \tau^+ \tau^-$ decays in various SHERPA 2.2.X samples.

sample	polarization
diboson 2.2.2	$-19.0 \pm 1.2\%$
Drell-Yan 2.2.7	$-20.22 \pm 0.92\%$
Drell-Yan 2.2.8	$-15.6 \pm 1.1\%$
Drell-Yan 2.2.10	$-14.62 \pm 0.72\%$
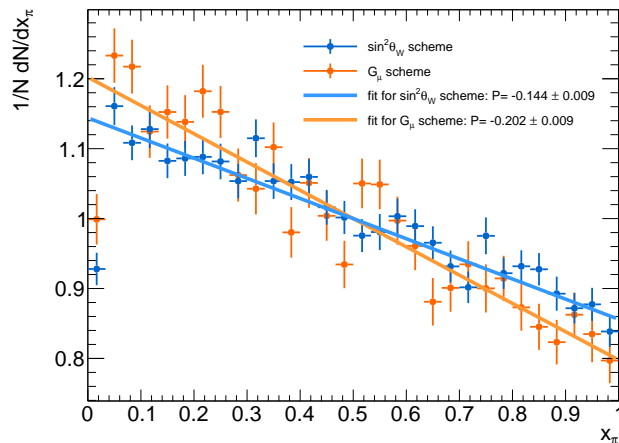
Figure 6.5: Comparison of tau lepton polarization measured in SHERPA $Z/\gamma^* \to \tau^+ \tau^-$ samples with the older (G_F) and newer ($\sin^2 \theta_W$) electroweak input scheme.

6.4   Comparison of input variables for the tauID BDT

Since the identification of hadronically-decaying tau leptons at ATLAS is done with machine learning techniques, as discussed in Section 4.4.4, it is important to make sure that input variables are

well modeled. At the time of conducting these validation studies, the machine learning model used for tau lepton identification was a BDT, later superseded by a recurrent neural network. In a study done by the author and Joseph Patton, Drell-Yan $Z/\gamma^* \to \tau^+\tau^-$ samples generated with SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 were compared at generator (truth) level and reconstruction level to check that there were no significant discrepancies between samples from the two Monte Carlo generators.

Distributions of all BDT input variables, listed with definitions in Section 4.4.4, were drawn and are shown in Figures 6.6-6.17. At reconstruction level, these variables were constructed straightforwardly from reconstructed physics objects in AOD-format samples. At generator level (in TRUTH-format samples), the detector interaction is not simulated, and only information about the properties of generated 'truth' particles is available. This meant that generator-level input variables had to be redefined to an extent, using only particle four-momenta and not detector response information such as calorimeter cluster energies. Estimated uncertainties used in the reconstruction-level calculation of $|S_{\text{leadtrack}}|$ and S_T^{flight} input variables were disregarded. Calorimeter cluster energies were replaced simply by the energies of electromagnetically-interacting particles. Only decays of tau leptons to pions were considered. Pions were associated with tau leptons if they were within $\Delta R < 0.2$ around the truth tau jet direction and passed a minimum p_T threshold of 1 GeV. The exception to this was the $f_{\text{iso}}^{\text{track}}$ input variable, where the value would have necessarily always been zero, as pions in the isolation cone would not be associated with the tau object - in this case, truth daughters were used instead. Tau lepton prongedness was determined from the number of associated pions rather than from truth information.

Good agreement is seen for most distributions, considering differences due to detector smearing (detector resolution) at reconstruction level. The smearing effect is seen most prominently in distributions of f_{cent} (Figure 6.6), $R_{\text{track}}^{0.2}$ (Figure 6.8), and $f_{\text{iso}}^{\text{track}}$ (Figure 6.10), where the truth-level distributions have more distinct peaks compared to the reconstruction-level distributions. Variables using electromagnetic calorimeter deposits ($p_T^{\text{EM+track}}/p_T$, $m_{\text{EM+track}}$, $f_{\text{EM}}^{\text{track-HAD}}$, $f_{\text{track}}^{\text{EM}}$) and estimated uncertainties (S_T^{flight}, $|S_{\text{leadtrack}}|$) differ necessarily between generator and reconstruction level due to a different construction. At truth level, significant differences between the two Monte Carlo generators are observed only in the ΔR_{max} distributions (which have a slope in the ratio plot, also seen in the reconstruction-level plot) and the tail of the m_{track} distribution. At reconstruction level, differences between generators appear in many other distributions, such as S_T^{flight}, where ratio of SHERPA and POWEHEGBOX+PYTHIA 8.1 distributions is consistently above 1.
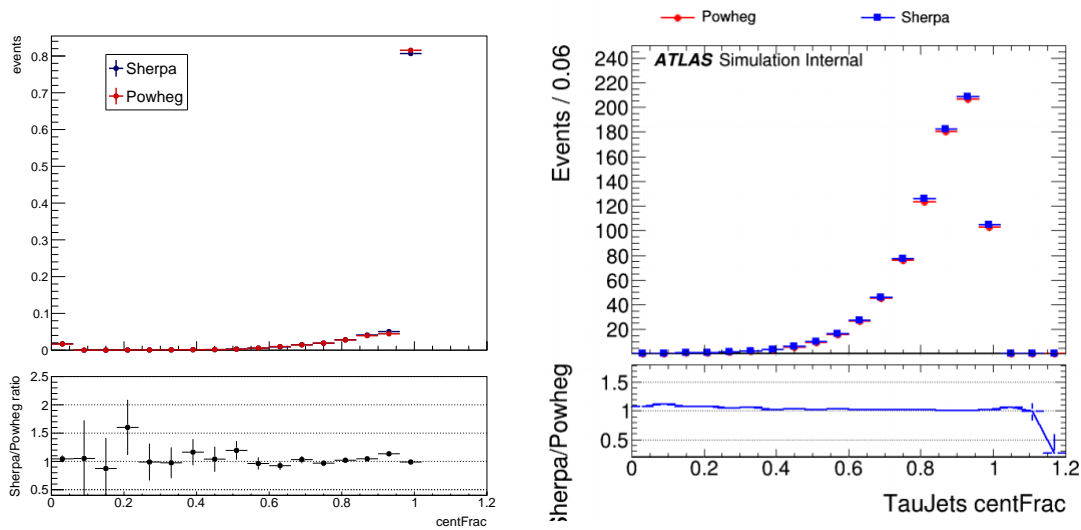

Figure 6.6: Generator-level (left) and reconstruction-level (right) comparison of the central energy fraction (f_{cent}) variable distribution using 1-prong and 3-prong tau lepton decays from samples generated by Sherpa 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.

Figure 6.7: Generator-level (left) and reconstruction-level (right) comparison of the leading track momentum fraction ($f_{\mathrm{leadtrack}}^{-1}$) variable distribution using 1-prong and 3-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.

Figure 6.8: Generator-level (left) and reconstruction-level (right) comparison of the track radius ($R_{\mathrm{track}}^{0.2}$) variable distribution using 1-prong and 3-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.

Figure 6.9: Generator-level (left) and reconstruction-level (right) comparison of the leading track impact parameter significance ($|S_{\text{leadtrack}}|$) variable distribution using 1-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.
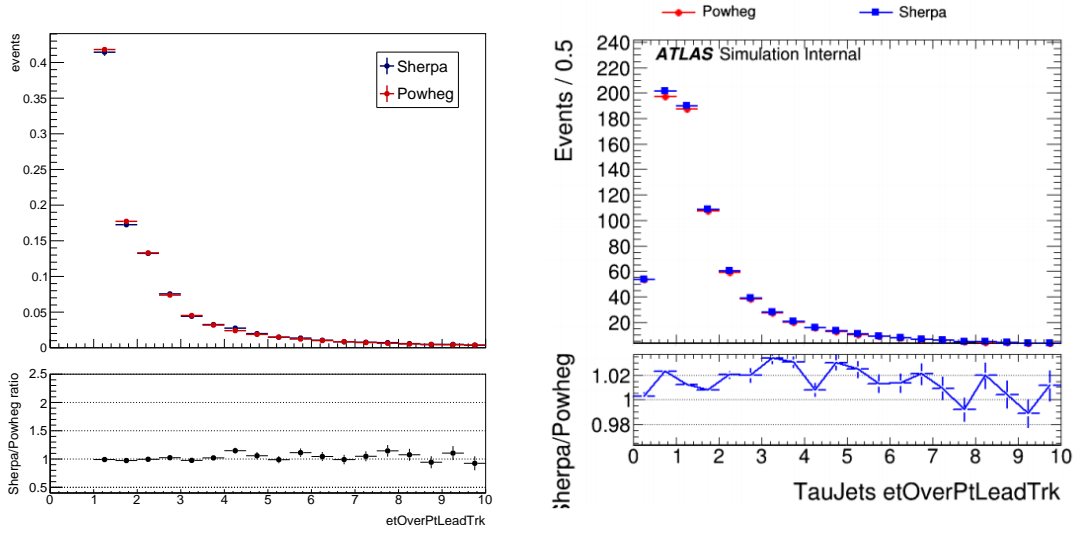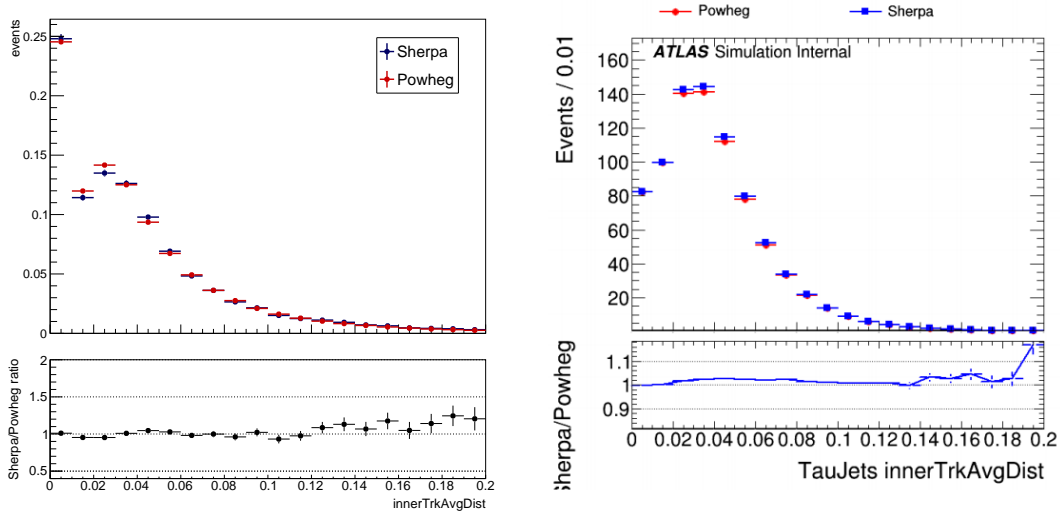

Figure 6.10: Generator-level (left) and reconstruction-level (right) comparison of the fraction of tracks in p_T in the isolation region ($f_{\text{iso}}^{\text{track}}$) variable distribution using 1-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.

Figure 6.11: Generator-level (left) and reconstruction-level (right) comparison of the maximum ΔR (ΔR_{max}) variable distribution using 3-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.
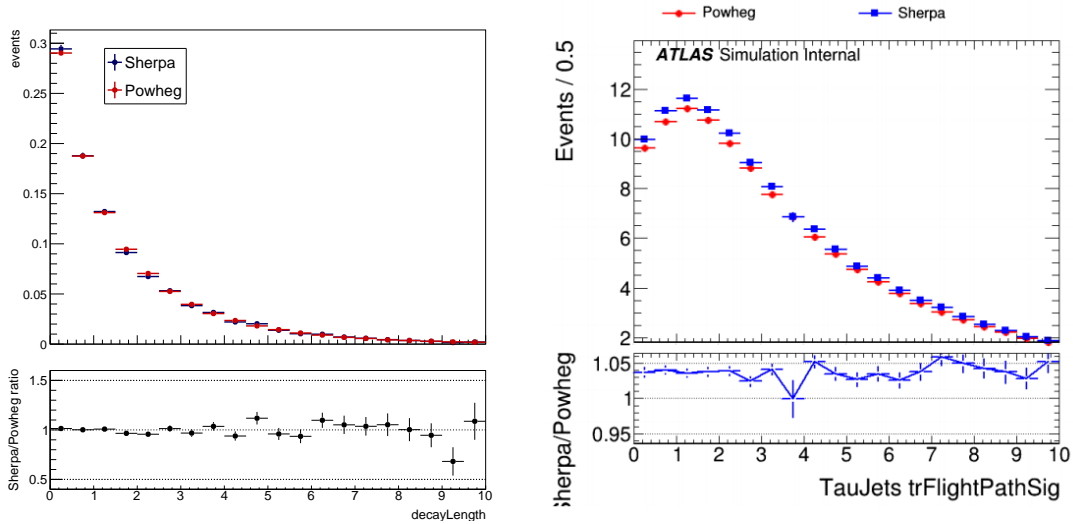
Figure 6.12: Generator-level (left) and reconstruction-level (right) comparison of the transverse flight path significance (S_T^{flight}) variable distribution using 3-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.
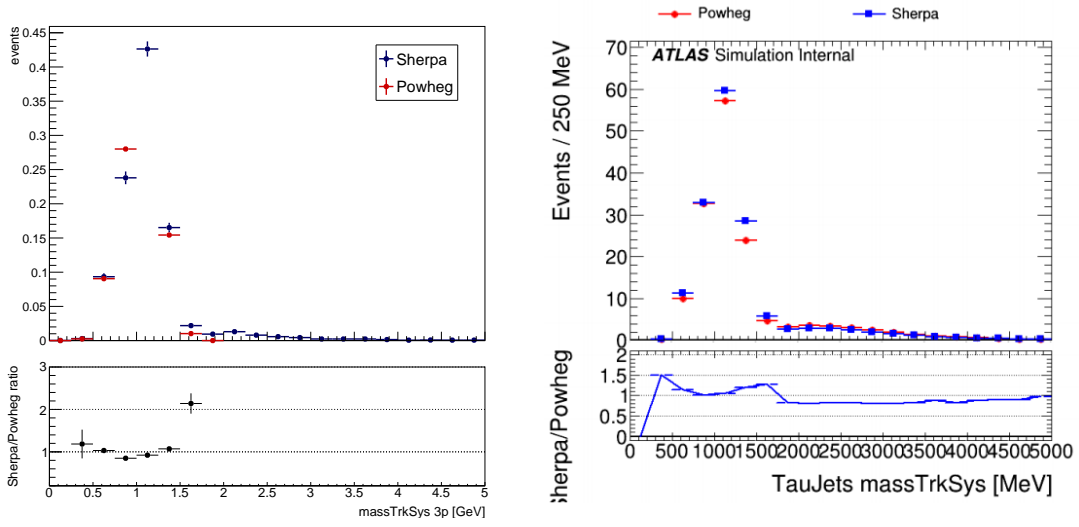
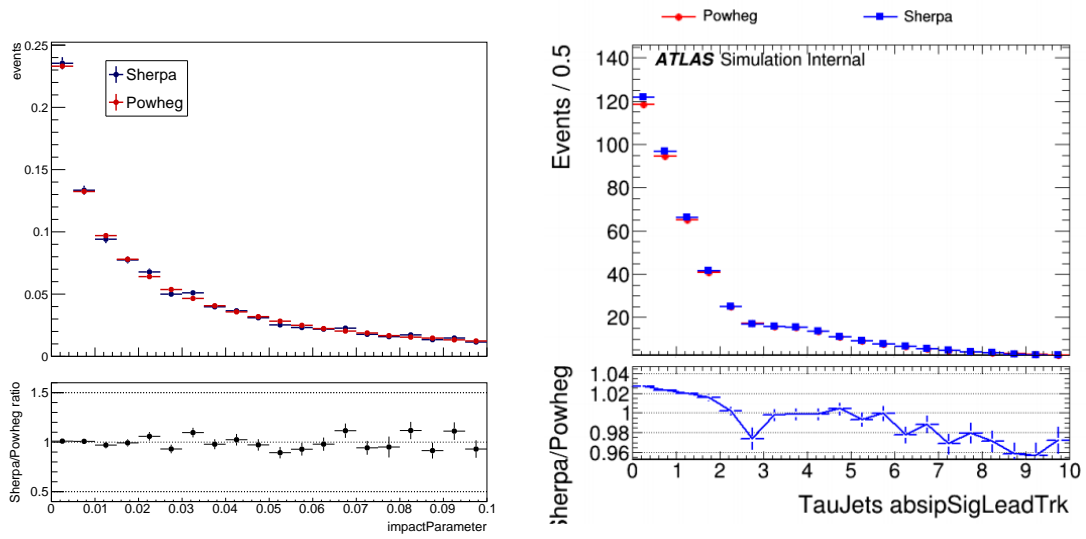Figure 6.13: Generator-level (left) and reconstruction-level (right) comparison of the track mass (m_track) variable distribution using 3-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG+PYTHIA 8.1 generators.
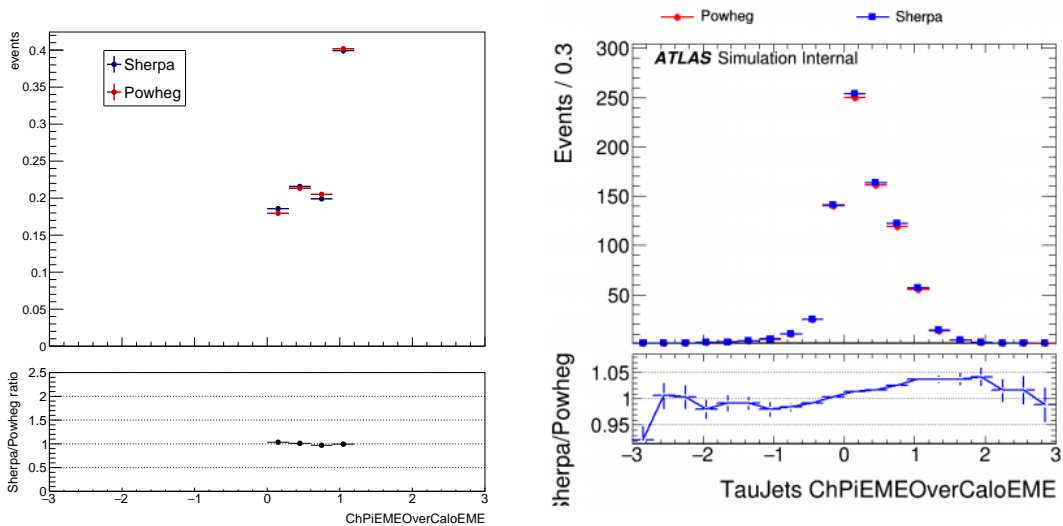
Figure 6.14: Generator-level (left) and reconstruction-level (right) comparison of the fraction of EM energy from charged pions ($f_\text{EM}^\text{track-HAD}$) variable distribution using 1-prong and 3-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.
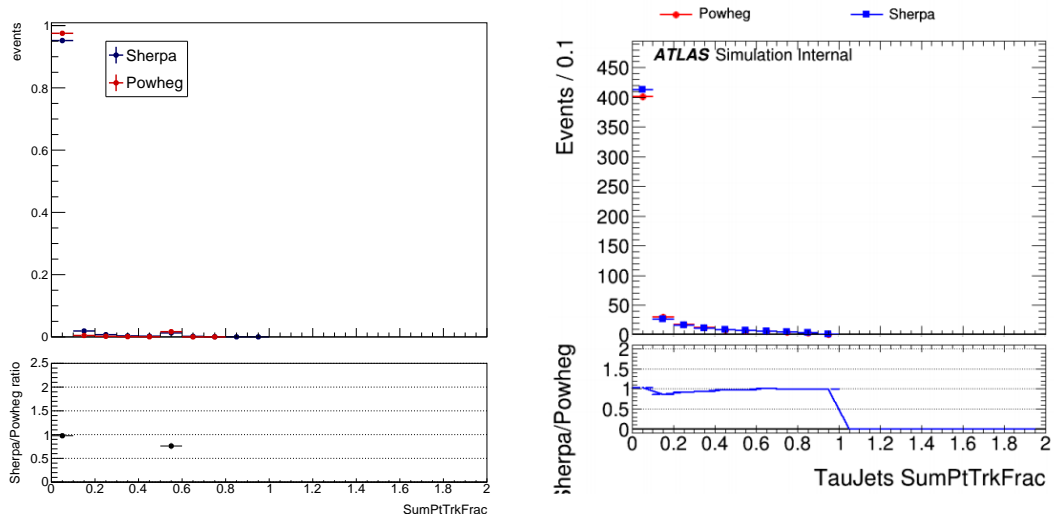
Figure 6.15: Generator-level (left) and reconstruction-level (right) comparison of the ratio of EM energy to track momentum ($f_{\text{track}}^{\text{EM}}$) variable distribution using 1-prong and 3-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.
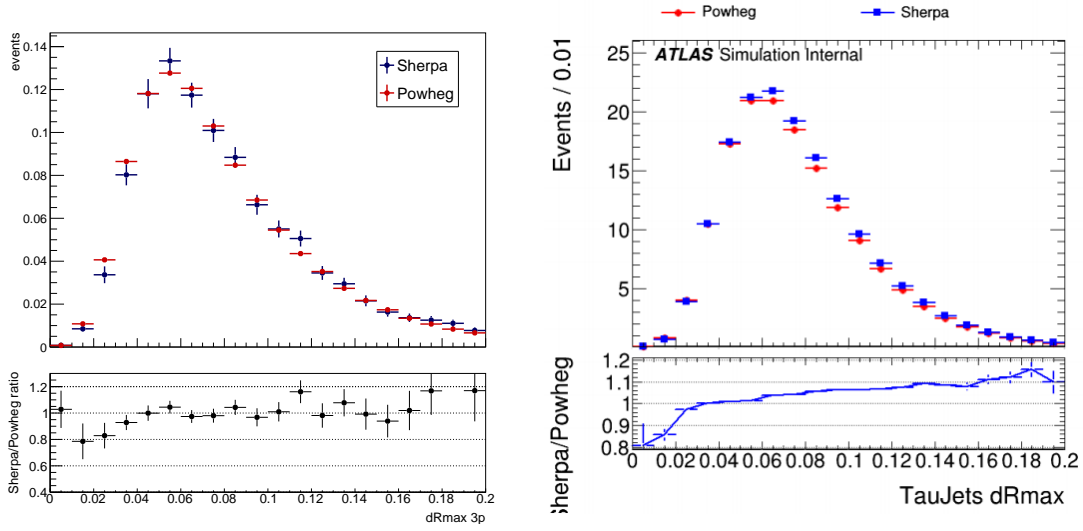

Figure 6.16: Generator-level (left) and reconstruction-level (right) comparison of the track-plus-EM-system mass ($m_{\text{EM+track}}$) variable distribution using 1-prong and 3-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.
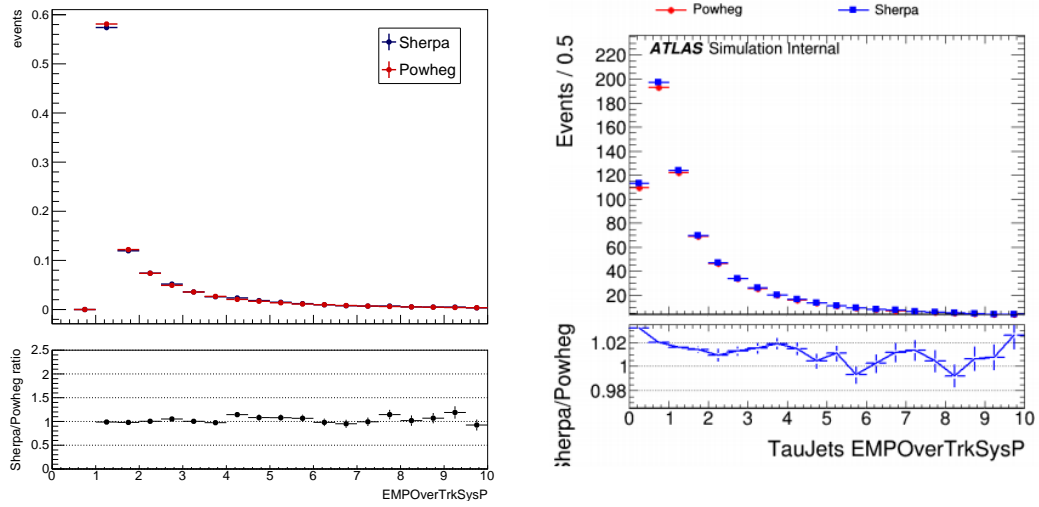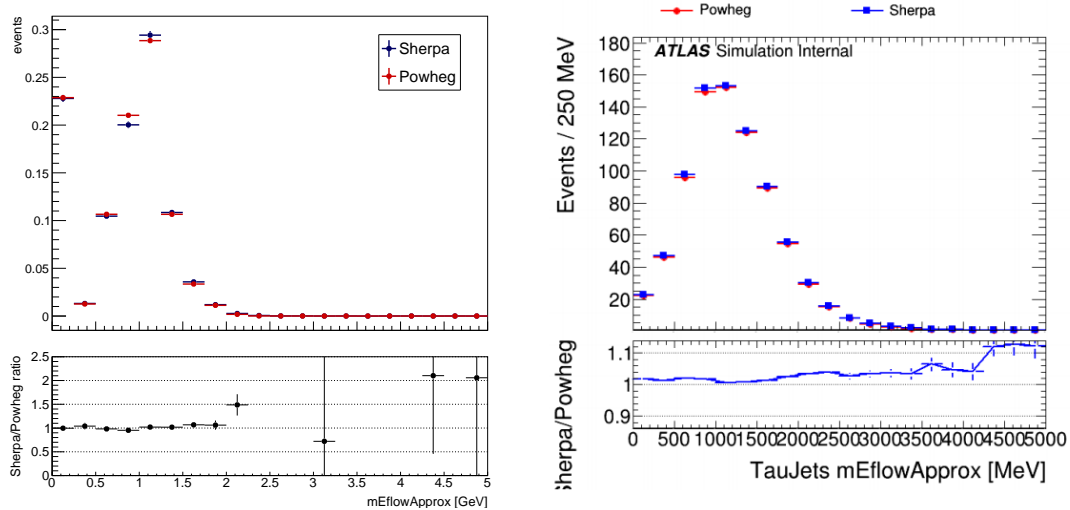
6.5 Comparison of tau lepton truth content in TRUTH and AOD formats

In order to bridge the gap between generator-level (TRUTH format) and reconstruction-level (AOD format) results, some additional checks were made. As described in Section 4.5, pile-up objects are added to Monte Carlo during the digitization step. Truth information about simulated pile-up objects is stored in AOD files along with truth information about the hard process propagated from EVNT files. Due to the presence of this information, it was possible to compare the numbers of truth tau leptons and truth pions in Powheg-Box+Pythia 8.1 $Z/\gamma^* \to \tau^+\tau^-$ samples before and after pile-up was applied. The results are shown in Table 6.3. The average number of truth tau jets per event remained the same, at 1.5. The number of truth tau lepton particles remained mostly the same, decreasing slightly below 2.0 in the AOD due to some cuts that were applied on the truth content in order to decrease the file sizes. The number of pions, on the other hand, increased by a large amount, from roughly 4 to over 245 per event. Of these, the number of pions passing the 1 GeV p_T cut required for pions to be matched to a tau jet increased 10-fold from roughly 3.5 to over 33 per event. The number of total truth particles increased 3-fold, and the number of total truth jets stayed constant at 6. Sources (parent particles listed in the truth record) of pions in the TRUTH and AOD samples are listed in Table 6.4. In the TRUTH sample, 84% of pions originated from tau leptons, with smaller percentages originating from resonances like ρ and ω mesons; nearly 7% had multiple parents listed in the truth record. In the AOD sample, only about 8.5% of pions came from tau lepton decays. The largest percentages originated from ρ mesons (nearly 35%) or had multiple parents (nearly 36%). Smaller percentages originated from other meson resonances and nucleons. The origin of these parent particles was not checked in this study.

The impact of the large number of pile-up pions on the reconstruction of hadronically-decaying tau lepton candidates was studied by comparing the numbers of pions within different ΔR distances from the truth tau jet: $\Delta R < 0.2$ (core region), $\Delta R < 0.4$ (core and isolation regions), and $0.2 < \Delta R < 0.4$ (isolation region). Distributions of these numbers are shown in Figures 6.18-6.19. As expected, the addition of pile-up increases the average number of pions within all three regions, which can lead to tau jets being incorrectly categorized as QCD jets, thus failing the hadronically-decaying tau lepton identification.

Table 6.3: The average numbers of truth particles and truth jets per event in AOD and TRUTH formats of a Powheg-Box+Pythia 8.1 $Z/\gamma^* \to \tau^+\tau^-$ sample.

type of physics object per event	AOD	TRUTH
truth particles / event	1631 ± 2	577.4 ± 0.8
truth pions / event	245.5 ± 0.4	4.321 ± 0.006
truth $p_T >$1 GeV pions / event	33.15 ± 0.04	3.473 ± 0.005
truth (particle) tau leptons / event	1.994 ± 0.004	2.019 ± 0.004
truth jets / event	6.00 ± 0.01	6.00 ± 0.01
truth tau jets / event	1.500 ± 0.003	1.500 ± 0.003

6.6 Conclusion

A number of validation studies were carried out on ATLAS Monte Carlo samples containing Dell-Yan and diboson production processes with tau leptons in the final state. No significant issues were found in tau lepton branching ratios and kinematic variable distributions. Measurements of the polarization of tau leptons from Z boson decays revealed values that differed significantly from values measured at the LEP collider. The cause was determined to be electroweak input scheme used in the Monte Carlo generator, which resulted in an incorrect value of the $\sin^2 \theta_W$ parameter. This issue was resolved by changing to an input scheme which used the effective value of this parameter as measured at the LEP collider. The distributions of input variables used for hadronically-decaying tau lepton identification were compared for two Monte Carlo event generators; no significant discrepancies were found. The effect of pile-up was also investigated
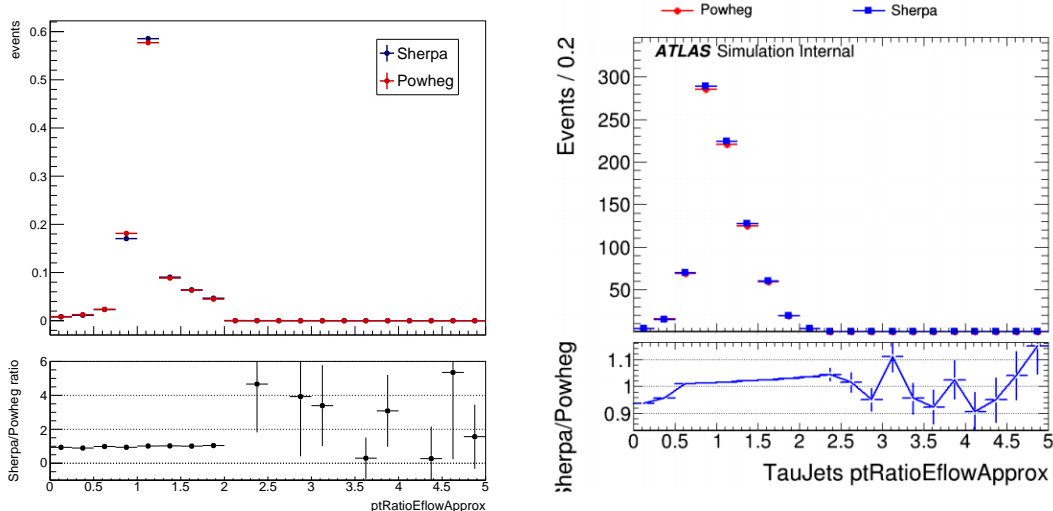
Figure 6.17: Generator-level (left) and reconstruction-level (right) comparison of the ratio of track-plus-EM-system to p_T ($p_T^{\mathrm{EM+track}}/p_T$) variable distribution using 1-prong and 3-prong tau lepton decays from samples generated by SHERPA 2.2.1 and POWHEG-BOX+PYTHIA 8.1 generators.

Table 6.4: Percentages of pion parents listed in the truth record of AOD and TRUTH formats of a POWHEG-BOX+PYTHIA 8.1 $Z/\gamma^* \to \tau^+\tau^-$ sample.

parent	AOD	TRUTH
tau lepton	$8.51 \pm 0.08\%$	$84.0 \pm 0.8\%$
kaon	$3.20 \pm 0.03\%$	$0.573 \pm 0.006\%$
ρ meson	$34.8 \pm 0.3\%$	$5.68 \pm 0.06\%$
ω meson	$6.53 \pm 0.06\%$	$1.42 \pm 0.01\%$
η meson	$2.00 \pm 0.02\%$	$0.911 \pm 0.009\%$
other meson	$6.29 \pm 0.06\%$	$0.497 \pm 0.005\%$
proton	$0.781 \pm 0.008\%$	$0.0 \pm 0.0\%$
neutron	$0.810 \pm 0.008\%$	$0.0 \pm 0.0\%$
multiple parents	$35.9 \pm 0.4\%$	$6.88 \pm 0.07\%$
other	$1.23 \pm 0.01\%$	$0.103 \pm 0.001\%$
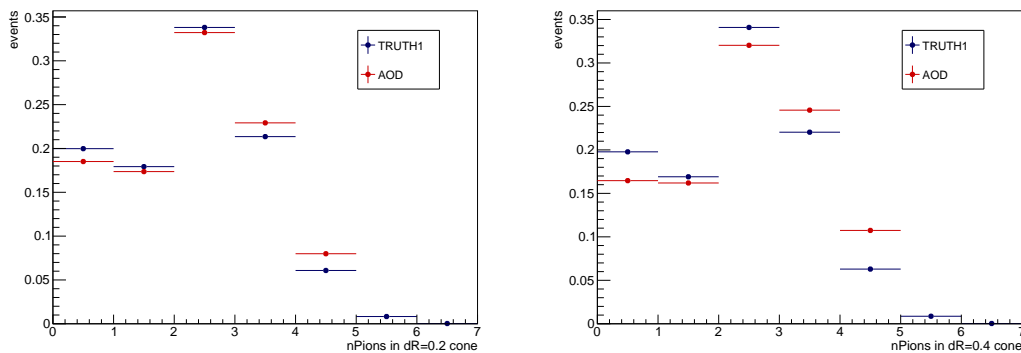
Figure 6.18: Pions within a $\Delta R < 0.2$ cone (left) and $\Delta R < 0.4$ cone (right) around the truth tau jet direction.

75
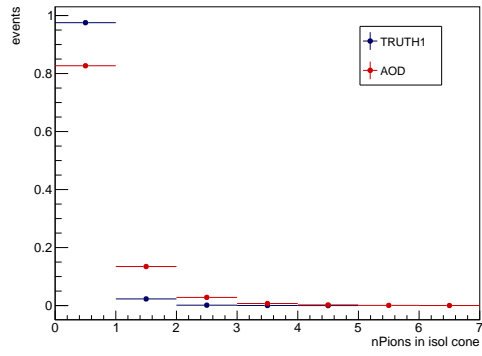
Figure 6.19: Pions within the isolation cone ($0.2 < \Delta R < 0.4$) around the truth tau jet direction.

to obtain a sense of how pions from pile-up processes can affect hadronically-decaying tau lepton reconstruction.

Chapter 7

BSM $H/A \to \tau^+\tau^-$ Search: Round 1

At ATLAS, there are multiple analyses searching for additional, heavy Higgs bosons, which are predicted by many extensions of the Standard Model. Of particular interest are BSM Higgs bosons predicted by the type-II Two-Higgs-Doublet Model (2HDM), which corresponds to the Higgs sector of the Minimal Supersymmetric Model (MSSM), as described in Section 1.2.4. These studies complement other searches for supersymmetric particles, such as (fermionic) higgsinos, squarks, and gluinos. BSM Higgs boson searches at ATLAS consider both neutral and charged Higgs boson production and various decay modes, such as $H^\pm \to \tau^\pm \nu$ [223], $H^\pm \to tb$ [224], and $H^0 \to b\bar{b}$ [225]. The analysis covered here is a search for a neutral, heavy BSM Higgs boson decaying to two oppositely-charged tau leptons. The BSM Higgs can be either CP-even (scalar, denoted H) or CP-odd (pseudoscalar, denoted A). This final state is especially useful for limit setting in the high-$\tan\beta$ phase space region, as H and A couplings to tau leptons - listed in Table 1.5 - are inversely proportional to $\cos\beta$. Additionally, this final state has a fairly clean experimental signature. The $H/A \to \tau^+\tau^-$ channel is responsible for much of the MSSM phase space exclusion achieved at ATLAS, as will be shown at the end of this chapter in the MSSM summary plot, Figure 7.17. The analysis covered in this chapter was a search for new resonances in the $[200, 2500]$ GeV mass range using the full Run 2 proton-proton collision dataset, corresponding to 139 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV in 2015-2018. A previous iteration of the analysis used 36 fb^{-1} of data collected during 2015 and 2016 [226]. The strategy for analyzing the full Run 2 dataset was a two-paper approach: publishing a quick paper following the same strategy as in the 36 fb^{-1} analysis, followed by a legacy paper with various analysis improvements. This chapter covers the first, quick analysis round - published in Physical Review Letters in 2019 [3] - while one of the improvements, the development of a multivariate classifier, is described in Chapter 8.

7.1 Heavy neutral Higgs boson production and decay

At the LHC, the dominant production mode of the Standard Model Higgs boson is gluon-gluon fusion [68]. Other production modes, such as b-associated production, $W^\pm$- and Z-associated production (Higgsstrahlung), and vector boson fusion, have cross-sections that are several orders of magnitude lower. However, in the MSSM the coupling between the Higgs bosons and b quarks is greatly enhanced for large $\tan\beta$ values and can cause b-associated production to become the dominant production mode. There are two viable approaches to calculating the b-associated production cross-section: the four-flavor scheme [227, 228] and the five-flavor scheme [229]. In the former, b quarks are considered to be massive and cannot appear in the initial state of the process. Only four quark flavors are available in the initial state: u, d, s, and c. In the latter, b quarks are considered to be massless and can appear in the initial state, raising the number of flavors to 5. The two types of b-associated production predictions can be combined via a number of matching procedures, ex. Santander [230] or NLO+NNLL [231] matching. Due to the relatively large predicted cross-sections, gluon-gluon fusion and b-associated production were the two production modes considered in this analysis. To target these processes, two selection categories were defined: a b-tag category necessitating the presence of at least one b-jet (originating from the associated b quark) and a b-veto category where no b-jets are present. Leading-order Feynman diagrams of the considered production modes are shown in Figure 7.1. Gluon-gluon fusion production will be denoted here by ggH/A, while b-associated production will be denoted by bbH/A.
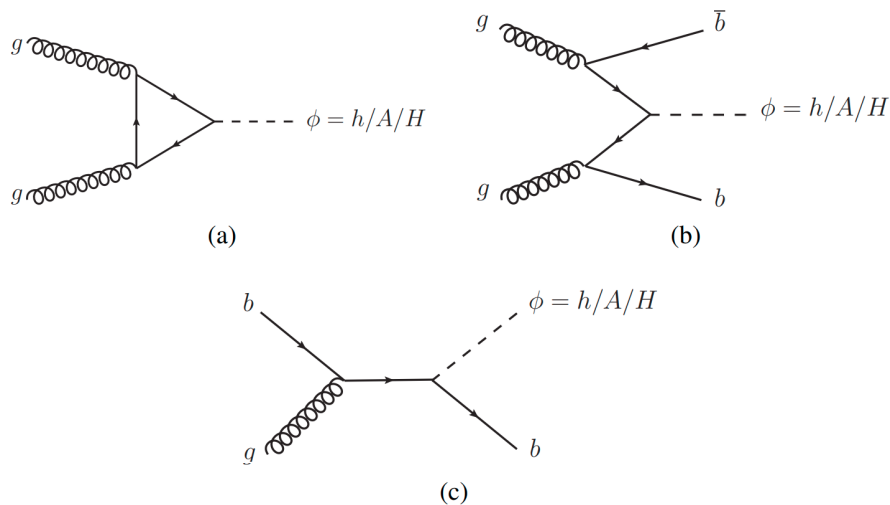
Figure 7.1: Leading-order Feynman diagrams for the following neutral BSM Higgs production modes: a) gluon-gluon fusion, b) b-associated production in the four-flavor scheme, c) b-associated production in the five-flavor scheme.

Since leptonically- and hadronically-decaying tau leptons leave very different signatures in the detector, the analysis has been further divided according to tau lepton decay modes into $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ and $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channels. The $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ channel was excluded, as it was responsible for a small percentage of events and the lack of a hadronically-decaying tau lepton made the Higgs signature less distinct and thus more difficult to separate from background. The two channels in use had different selection criteria, leading to different background distributions.

In keeping with the traditional bump hunt approach (described in Section 2.3, the discriminating variable used in this search was an invariant-mass-like variable defined in Section 7.3.4. Exclusion limits - both model-independent and an MSSM benchmark scenario interpretation - were calculated using a likelihood fit, described in Section 7.7.

7.2 Monte Carlo and data samples

The data samples used in this analysis correspond to 139 fb^{-1} of proton-proton collision data collected by the ATLAS detector with all subsystems operational during 2015-2018 at $\sqrt{s} = 13$ TeV. Luminosity blocks were filtered using good run lists, as described in Section 4.5. The Monte Carlo samples used in this analysis were produced as part of a centralized ATLAS campaign. Both data and Monte Carlo events were required to have at least one reconstructed primary vertex with at least two associated tracks.

Monte Carlo samples were used for the estimation of BSM Higgs boson signal produced via gluon-gluon fusion and b-associated production, and for the estimation of the following background processes: single Z and $W^{\pm}$ production with associated jets (Z+jets and W+jets background), single t quark and $t\bar{t}$ events, and the production of two vector bosons (diboson production). Table 7.1 lists the Monte Carlo generators, PDF sets for the hard process, models for the parton shower, hadronization, and underlying event simulation, and the order in which the cross-sections were calculated for each physics sample. All samples were produced with GEANT4 used for the full detector simulation, except for the b-associated Higgs production signal, which used the ATLFASTII fast simulation framework. The reason for this was that b-associated Higgs production samples were created with the MADGRAPH (MG5_AMC@NLO) generator, which produces many events with negative weights, necessitating the generation of much higher statistics compared to gluon-gluon fusion samples. The fast simulation was validated against the full simulation for one signal mass point.

Signal samples were generated using guidelines from the *Handbook of LHC Higgs Cross Sections* [68] for a number of mass hypotheses. The cross-sections of gluon-gluon fusion production events were calculated with the SuSHi (Supersymmetric Higgs) program [232]. These cross-sections included NLO supersymmetric-QCD corrections and NNLO QCD corrections for the t quark and light-quark electroweak effects. The cross-sections of b-associated production events were calculated in the four-flavor and five-flavors schemes and combined according to the prescription in [233–235]. The mixing and effective Yukawa couplings to tau leptons of the neutral Higgs bosons were calculated with the FeynHiggs program [236–243], while the branching ratios were calculated with a mixture of the FeynHiggs, HDecay [244, 245], and Prophecy4F [246] programs, with results combined according to the procedure described in the *Handbook of LHC Higgs Cross Sections* [247].

Table 7.1: List of physics processes modeled by Monte Carlo with respective generators, PDF sets for the hard process, models for the parton shower (PS), hadronization (H), and underlying event (UE), and orders for cross-section σ calculation. V refers to vector bosons $W^{\pm}$ and Z. Generator names are shortened to PB for Powheg-Box, Py for Pythia, MG for MG5_aMC@NLO, and Sh for Sherpa.

sample	generator	PDF	PS/H/UE	σ order
ggH/A	PB v2 [209–211, 248, 249]	CT10 [214]	Py 8.1 [213]	NLO/NNLO
bbH/A	MG 2.1.2 [250, 251]	CT10	Py 8.2 [252]	NLO/NNLO
W+jets	Sh 2.2.1 [207]	NNPDF 3.0 NNLO [208]	Sh 2.2.1 [253]	NNLO [254]
Z+jets	PB v1 [209–212]	CT10	Py 8.1	NNLO [254]
$VV/V\gamma^*$	Sh 2.2	NNPDF 3.0 NNLO	Sh 2.2	NLO
$t\bar{t}$	PB v2 [209–211, 255]	NNPDF 3.0 NLO	Py 8.2	NNLO+NNLL [256–262]
single t	PB v2 [209–211, 263–265]	NNPDF 3.0 NLO	Py 8.2	NNLO+NNLL [266, 267]

7.3 Event reconstruction and selection

The reconstruction and identification of physics objects at ATLAS is described in detail in Section 4.4. In this analysis, electrons had to pass the gradient isolation requirement, while muons had to pass the tight fixed radius cut isolation requirement. Jets were reconstructed with the anti-k_T algorithm with size parameter $\Delta R < 0.4$, and had to pass a 'medium' jet vertex tagger requirement in order to identify and select jets originating from the hard scatter interaction. This working point corresponded to a 92% efficiency for jets in the electromagnetic calorimeter. The b-tagging of jets was done using the MV2c10 algorithm with a multivariate score threshold corresponding to a 70% efficiency for b-jets in $t\bar{t}$ events.

Since collision events can have multiple identified hadronically-decaying tau leptons and light leptons (from pile-up and/or due to misidentification), it is important to find the most likely candidates for the visible parts of the tau lepton pair from the hard scattering. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, these candidates are the two highest-p_T hadronically-decaying tau leptons, called the leading and subleading tau lepton (in order of decreasing p_T). In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, they are the highest-p_T hadronically decaying tau lepton (leading tau lepton) and highest-p_T light lepton (leading lepton). Together with the missing transverse energy, these objects can be used to reconstruct various observables relating to the resonance production and decay.

7.3.1 Removal of geometrically-overlapping objects

The overlap removal procedure is largely the same in the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channel. The successive steps are as follows:

1. exclude jets within $\Delta R < 0.2$ of selected $\tau_{\text{had-vis}}$ candidates (the leading and subleading tau leptons in the $\tau_{\text{had}}\tau_{\text{had}}$ channel or the leading tau lepton in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel),

2. exclude jets within $\Delta R < 0.4$ of an electron or muon,

3. exclude $\tau_{\text{had-vis}}$ candidates within $\Delta R < 0.2$ of an electron or muon,

4. exclude electrons within $\Delta R < 0.2$ of a muon.

7.3.2 Event selection in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel

Events were preselected for the analysis with a single-lepton trigger, with different trigger requirements depending on luminosity conditions during different data-taking periods. Only one lepton passing the p_T threshold ($p_T > 15$ GeV for electrons and $p_T > 7$ GeV for muons) and the loose identification criterion can be present in the event, and it must be geometrically matched to the object that activated the trigger. This 'dilepton veto' suppresses $Z/\gamma^* \to \ell^+\ell^-$ background. The highest-p_T $\tau_{\text{had-vis}}$ candidate must have $p_T > 25$ GeV, $|\eta| < 2.3$, and pass the medium identification criterion. There is no limit on the number of $\tau_{\text{had-vis}}$ candidates in an event. The preselection is concluded by requiring that the lepton and leading $\tau_{\text{had-vis}}$ have opposite electric charge ('opposite sign').

The final selection for the signal region includes a few more cuts aimed at reducing background. The p_T of the lepton must be above 30 GeV. Since the heavy Higgs boson would usually be produced with a low p_T, the lepton and the leading $\tau_{\text{had-vis}}$ are required to be back-to-back, with $\Delta\phi(\ell, \tau_{\text{had-vis}}) > 2.4$. Background from W+jets production is suppressed with a cut on the transverse mass m_T of the lepton and E_T^{miss}:

$$m_T(\ell, E_T^{\text{miss}}) = \sqrt{2p_T(\ell)E_T^{\text{miss}}(1 - \cos\Delta\phi(\ell, E_T^{\text{miss}}))} < 40 \text{ GeV}. \tag{7.1}$$

Signal events are expected to have low values of $m_T(\ell, E_T^{\text{miss}})$ because the leptonically-decaying tau lepton produces one more neutrino compared to the hadronically-decaying tau lepton, causing E_T^{miss} to align more closely with the light lepton in the event. W+jets background, on the other hand, has a peak in $m_T(\ell, E_T^{\text{miss}})$ at around 80 GeV. Lastly, it was found that a large number of electrons from $Z/\gamma^* \to e^+e^-$ events were misidentified as $\tau_{\text{had-vis}}$ objects; to suppress this background, events with an invariant mass of the electron and leading $\tau_{\text{had-vis}}$ candidate between 80 and 110 GeV were vetoed. This veto is applied in the electron (e-had) sub-channel only. Finally, events are divided into b-veto and b-tag categories by requiring respectively 0 or ≥ 1 b-tagged jets with $p_T > 20$ GeV.

7.3.3 Event selection in the $\tau_{\text{had}}\tau_{\text{had}}$ channel

Events were preselected for analysis with a single-tau-lepton trigger. During certain data-taking periods, multiple single-lepton triggers were active; in this case, events needed to pass at least one of the triggers. The leading-p_T $\tau_{\text{had-vis}}$ candidate was required to have a p_T of at least 5 GeV over the p_T threshold of the trigger, be geometrically matched to the object that activated that trigger, and pass the medium identification criterion. The sub-leading $\tau_{\text{had-vis}}$ candidate had to have $p_T > 65$ GeV and pass the loose identification criterion. Additionally, the leading and sub-leading $\tau_{\text{had-vis}}$ candidates had to be back-to-back, with $\Delta\phi(\tau_{\text{lead}}, \tau_{\text{sub-lead}}) > 2.7$, and have opposite electric charge.

The final selection for the signal region required only an additional cut on the number of b-tagged jets with $p_T > 20$ GeV (0 or ≥ 1) to separate events into b-veto and b-tag categories.

7.3.4 $\tau^+\tau^-$ mass reconstruction

The reconstruction of the invariant mass of the $\tau^+\tau^-$ system (and hence the mass of the resonance decaying to two tau leptons) is difficult due to the presence of multiple neutrinos from tau lepton decays. Tau leptons decaying hadronically have a single neutrino in the final state, while leptonic decays produce two neutrinos. Information about the relative directions of the neutrinos and the z component of the missing momentum is lost, complicating the invariant mass calculation. Multiple methods were developed, originally for $Z/\gamma^* \to \tau^+\tau^-$ measurements and Standard Model $H \to \tau^+\tau^-$ searches, to address this, including the Missing Mass Calculator [268] and MOSAIC [269] algorithms. However, in previous versions of this analysis it was found that the best-performing mass variable was the total transverse mass m_T^{tot}:

$$m_T^{\text{tot}} = \sqrt{m_T^2(\tau_1, E_T^{\text{miss}}) + m_T^2(\tau_2, E_T^{\text{miss}}) + m_T^2(\tau_1, \tau_2)}, \tag{7.2}$$

where τ_1 and τ_2 refer to the leading and subleading $\tau_{\text{had-vis}}$ candidates in the $\tau_{\text{had}}\tau_{\text{had}}$ channel or the $\tau_{\text{had-vis}}$ candidate and the light lepton in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, and transverse masses m_T are

defined in the following way:

$$m_T(\vec{a}, \vec{b}) = \sqrt{2p_T(\vec{a})p_T(\vec{b})(1 - \cos\Delta\phi(\vec{a}, \vec{b}))}. \qquad (7.3)$$

This variable is used as the final discriminant in this analysis.

7.4 Background estimation

Backgrounds in this analysis are predominantly due to irreducible $Z/\gamma^* \to \tau^+\tau^-$ events and events where jets are misidentified as hadronically-decaying tau leptons or light leptons. The second background category is difficult to estimate with Monte Carlo, for a few reasons. For one, the probabilities of jets being misidentified in this way are not modelled well in simulation. For another, multijet background has very large cross-sections and low acceptance, making it impractical or even impossible to simulate an adequate number of multijet events. Therefore, while backgrounds with real tau leptons and light leptons can be satisfactorily modeled with Monte Carlo, multijet events and other contributions with a large number of jets are estimated using data-driven techniques described below.

7.4.1 $\tau_{\text{lep}}\tau_{\text{had}}$ channel

In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, two background contributions are estimated with data-driven methods: background where both the lepton and the $\tau_{\text{had-vis}}$ are faked by jets, which corresponds mainly to multijet events, and background where only the $\tau_{\text{had-vis}}$ is faked by a jet, which corresponds mainly to W+jets events in the b-veto category and $t\bar{t}$ events (with $t \to W^+\bar{b}$ and $\bar{t} \to W^-b$) in the b-tag category. Both of these are estimated with the fake factor or template method, the premise of which is to construct a signal-less control region (CR) where jet background is enhanced. The method has two basic steps. First, Monte Carlo events with correctly identified ℓ and $\tau_{\text{had-vis}}$ objects are subtracted from real data to leave only the jet contribution. Next, this contribution is propagated to the signal region (SR) with a fake factor (transfer factor), an event weight which accounts for the difference in jet background yields between the CR and the SR.
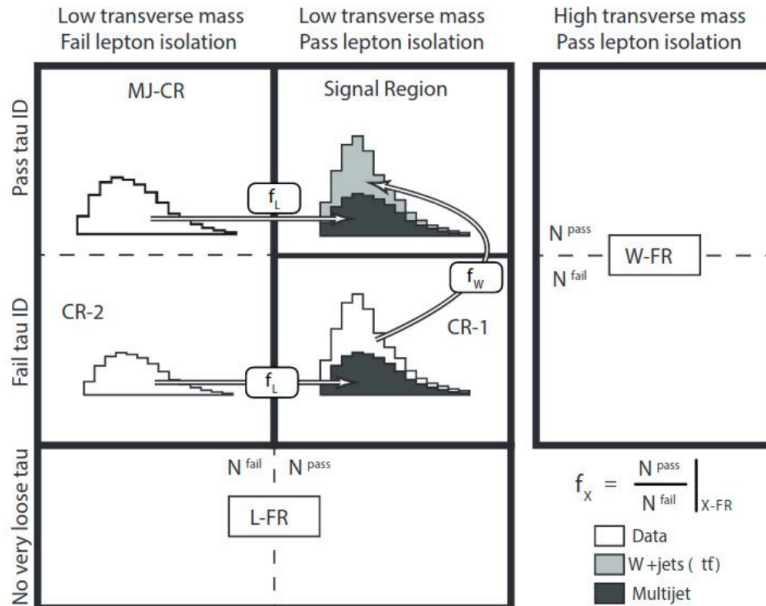
Figure 7.2: Schematic of the $\tau_{\text{lep}}\tau_{\text{had}}$ data-driven background estimation procedure.

This method is shown schematically in Figure 7.2. Three control regions were defined: the multijet control region (MJ-CR) for estimating the multijet background contribution in the signal region; CR-1 for estimating the W+jets/$t\bar{t}$ background contribution in the signal region; and CR-2 for estimating the multijet contribution in CR-1, which must be subtracted from data along

with Monte Carlo background in order to obtain the W+jets/$t\bar{t}$ contribution. MJ-CR differs from the signal region by inverting the ℓ isolation requirement, CR-1 inverts the requirement that the $\tau_{\text{had-vis}}$ candidate must pass medium tau lepton identification (though it must still pass very loose tau lepton identification), and CR-2 inverts both. Additionally, two fake regions are defined for calculating fake factors needed to propagate the jet background contributions. The multijet fake factor, also called the lepton fake factor (f_L), is calculated in the lepton fake region (L-FR), while the W+jets/$t\bar{t}$ fake factor (f_W) is calculated in the W fake region (W-FR). In the L-FR region, the $\tau_{\text{had-vis}}$ candidate must fail the very loose tau lepton identification requirement, ensuring a very high proportion of events with jets misidentified as taus. The fake factor, parametrized in $p_T(\ell)$, is calculated by comparing the number of events passing and failing the lepton isolation requirement in each $p_T(\ell)$ bin of the L-FR:

$$f_L(p_T(\ell)) = \frac{N_{\text{data}}^{\text{pass-iso}}(p_T(\ell)) - N_{MC}^{\text{pass-iso}}(p_T(\ell))}{N_{\text{data}}^{\text{fail-iso}}(p_T(\ell)) - N_{MC}^{\text{fail-iso}}(p_T(\ell))}. \tag{7.4}$$

Here and in the following equations, N_{data} refers to the number of data events, while N_{MC} refers to the number of Monte Carlo background events with real ℓ and $\tau_{\text{had-vis}}$ objects. The difference of these two numbers is considered to be the number of background events with jets misidentified as ℓ and/or $\tau_{\text{had-vis}}$ objects. The same f_L fake factor is used for propagating events from CR-2 to CR-1 and from MJ-CR to the signal region. In the W-FR region, the transverse mass cut is changed from $m_T(\ell, E_T^{\text{miss}}) < 40$ GeV to $m_T(\ell, E_T^{\text{miss}}) > 60$ GeV to increase the proportion of W+jets/$t\bar{t}$ events. The f_W fake factor, parametrized in $p_T(\tau_{\text{had-vis}})$ is calculated in a similar way to f_L, but comparing the number of events passing and failing the medium tau lepton identification requirement in the W-FR:

$$f_W(p_T(\tau_{\text{had-vis}})) = \frac{N_{\text{data}}^{\text{pass-ID}}(p_T(\tau_{\text{had-vis}})) - N_{MC}^{\text{pass-ID}}(p_T(\tau_{\text{had-vis}}))}{N_{\text{data}}^{\text{fail-ID}}(p_T(\tau_{\text{had-vis}})) - N_{MC}^{\text{fail-ID}}(p_T(\tau_{\text{had-vis}}))}. \tag{7.5}$$

With the fake factors calculated, the background-estimation procedure for the multijet contribution N_{MJ}^{SR} and the W+jets/$t\bar{t}$ contribution N_W^{SR} in the signal region can be summarized as follows (ignoring parametrization):

$$N_{MJ}^{SR} = f_L(N_{\text{data}}^{\text{MJ-CR}} - N_{\text{MC}}^{\text{MJ-CR}}), \tag{7.6}$$

$$N_W^{SR} = f_W(N_{\text{data}}^{\text{CR-1}} - N_{\text{MC}}^{\text{CR-1}} - f_L(N_{\text{data}}^{\text{CR-2}} - N_{\text{MC}}^{CR-2})). \tag{7.7}$$

It should be noted that the control and fake regions also differ from the SR by certain other changes in selection cuts, chosen in such a way as to purify the regions (increase the proportion of jet background) and/or improve statistics without significantly altering the shape of key distributions. The full definitions of the control and fake regions mentioned here are given in Table 7.2, along with the definition of a validation region (VR) for checking the full background modeling.

Table 7.2: Definitions of regions used in data-driven background estimation and background validation in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel.

region	selection
VR	pass SR except: $40 < m_T(\ell, E_T^{\text{miss}}) < 60$ GeV
MJ-CR	pass SR except: ℓ fails isolation
CR-1	pass SR except: $\tau_{\text{had-vis}}$ fails medium tau ID, passes very loose tau ID
CR-2	pass SR except: ℓ fails isolation and $\tau_{\text{had-vis}}$ fails medium tau ID, passes very loose tau ID
L-FR	single-ℓ trigger, exactly one ℓ, leading $\tau_{\text{had-vis}}$ fails very loose tau ID, $m_T(\ell, E_T^{\text{miss}}) < 30$ GeV, divided into fail-ℓ-isolation and pass-ℓ-isolation regions
W-FR	pass SR except: $60 < m_T(\ell, E_T^{\text{miss}}) < 150$ GeV for b-veto, $60 < m_T(\ell, E_T^{\text{miss}}) < 110$ GeV for b-tag, $\tau_{\text{had-vis}}$ passes very loose tau ID, divided into fail medium tau ID and pass medium tau ID regions

7.4.2 $\tau_{\text{had}}\tau_{\text{had}}$ channel

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, only the multijet background contribution is estimated with the fake factor method introduced in Section 7.4.1. This background contribution dominates due to the larger likelihood of QCD jets being misidentified as hadronically-decaying tau leptons rather than light leptons. The W+jets($t\bar{t}$) contribution is less prominent than in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel and can be successfully estimated with Monte Carlo, corrected with a data-driven fake rate approach. Only one control region is defined: CR-1, where the tau lepton identification cut on the subleading $\tau_{\text{had-vis}}$ is inverted. The multijet fake factor f_{MJ} is calculated in the dijet fake region DJ-FR, where the selection criteria require that events pass the single-jet (rather than single-tau) trigger, and the p_T of the subleading $\tau_{\text{had-vis}}$ must be at least 30% of the p_T of the leading tau lepton. The fake factor, parametrized in subleading $\tau_{\text{had-vis}}$ p_T and the number of tau lepton prongs, is calculated by comparing the number of events passing and failing the subleading tau loose ID requirement:

$$f_{MJ}(p_T(\tau_{\text{sublead}})) = \frac{N_{\text{data}}^{\text{pass-ID}}(p_T(\tau_{\text{sublead}})) - N_{\text{MC}}^{\text{pass-ID}}(p_T(\tau_{\text{sublead}}))}{N_{\text{data}}^{\text{fail-ID}}(p_T(\tau_{\text{sublead}})) - N_{\text{MC}}^{\text{fail-ID}}(p_T(\tau_{\text{sublead}}))}. \tag{7.8}$$

The number of multijet events propagated to the signal region N_{MJ}^{SR} is then:

$$N_{MJ}^{SR} = f_{MJ}(N_{\text{data}}^{\text{CR-1}} - N_{\text{MC}}^{\text{CR-1}}). \tag{7.9}$$

Monte Carlo events with one jet misidentified as a tau lepton are weighted by fake rates FR calculated from data in the W+jets/$t\bar{t}$ fake regions W-FR/T-FR, defined in Table 7.3. These fake regions are populated by events were the $W^\pm$ boson (from W+jets events or $t \to Wq$ decays) decays to $\mu^\pm\nu$ rather than $\tau^\pm\nu$. Due to lepton universality, the kinematics of these events are the same as the kinematics of events with $W^\pm \to \tau^\pm\nu$ decays. It can then be assumed that $\tau^{\text{had-vis}}$ candidates in the events are misidentified jets, not true $\tau_{\text{had-vis}}$ objects. Fake rates are calculated to determine the probability of a jet being misidentified as a $\tau_{\text{had-vis}}$ object. This is done in a similar manner to the calculation of fake factors, comparing the number of events where the leading-p_T $\tau_{\text{had-vis}}$ candidate passes the identification criterion to the total number of events, $N^{\text{tot}} = N^{\text{pass-ID}} + N^{\text{fail-ID}}$:

$$FR(p_T(\tau_{\text{had-vis}})) = \frac{N_{\text{data}}^{\text{pass-ID}}(p_T(\tau_{\text{had-vis}})) - N_{\text{MC}}^{\text{pass-ID}}(p_T(\tau_{\text{had-vis}}))}{N_{\text{data}}^{\text{tot}}(p_T(\tau_{\text{had-vis}})) - N_{\text{MC}}^{\text{tot}}(p_T(\tau_{\text{had-vis}}))}. \tag{7.10}$$

Fake rates are calculated separately for $\tau_{\text{had-vis}}$ candidates passing the loose and medium tau lepton identification criteria, in order to apply them to both leading and subleading $\tau_{\text{had-vis}}$ candidates in the analysis.

Table 7.3: Definitions of control and fake regions used in data-driven background estimation in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

region	selection
DJ-CR	pass SR except: single-jet trigger, $\frac{p_T(\tau_{\text{sublead}})}{p_T(\tau_{\text{lead}})} > 0.3$, no tau ID requirement
CR-1	pass SR except: subleading $\tau_{\text{had-vis}}$ fails loose ID
W-FR (T-FR)	isolated μ passing trigger, τ_{lead} with no tau ID requirement, $|\Delta\phi(\mu, \tau_{\text{lead}})| > 2.4$ $m_T(\mu, E_T^{\text{miss}}) > 40$ GeV, b-veto (b-tag)

7.5   Validation of background modeling

To make sure that background estimates describe reality with reasonable accuracy, the agreement between real data and background estimated with Monte Carlo and data-driven techniques is checked in key observables. However, it is vital to avoid experimenter's bias by not looking at the signal region until the analysis is finalized; for this reason, checks are done in validation regions, (mostly) signal-less regions where the background composition is similar to the signal region (SR). VRs are defined by changing some selection criteria to exclude signal events. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, the VR differs from the SR only by the $m_T(\ell, E_T^{\text{miss}})$ cut, which is changed from

$m_T(\ell, E_T^{\mathrm{miss}}) < 40$ GeV to $40 < m_T(\ell, E_T^{\mathrm{miss}}) < 60$ GeV - it thus bridges the gap between the SR and the W-FR. In the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel, the VR is constructed by changing the opposite-sign (OS) requirement for the $\tau^+\tau^-$ pair to a same-sign (SS) requirement.

While many validation checks were carried out by the analysis team, this section will focus on the author's contribution: validation studies of the data-driven W+jets background estimation in the b-veto $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel.

7.5.1 Validation of W+jets background estimation in the b-veto $\tau_{\mathbf{lep}}\tau_{\mathbf{had}}$ channel

Two versions of f_W parametrization, both available in the previous analysis round, were compared in the initial stage of W+jets background validation: f_W parametrized in $p_T(\tau_{\mathrm{had\text{-}vis}})$ only and f_W parametrized in $p_T(\tau_{\mathrm{had\text{-}vis}})$ and $\Delta\phi(\tau_{\mathrm{had\text{-}vis}}, E_T^{\mathrm{miss}})$ simultaneously. Distributions of $p_T(\ell)$, $p_T(\tau_{\mathrm{had\text{-}vis}})$, E_T^{miss}, $\Delta\phi(\ell, E_T^{\mathrm{miss}})$, and $\Delta\phi(\tau_{\mathrm{had\text{-}vis}}, E_T^{\mathrm{miss}})$ in the b-veto pass-tau-ID W-FR, using these two p_W parametrizations, are shown in Figure 7.3. It was found that both parametrizations resulted in the mismodeling of $p_T(\ell)$ and E_T^{miss} distributions, and, as a result, in the mismodeling of the final discriminant m_T^{tot}. The p_T distribution of the $\tau_{\mathrm{had\text{-}vis}}$ object, the third physics object used for calculating m_T^{tot}, was - trivially - not mismodeled, because the f_W fake factor was constructed from fail-tau-ID and pass-tau-ID $p_T(\tau)$ distributions. A similar mismodeling of variable distributions was also seen in the VR (Figures 7.5-7.6), which contains a large number of W+jets events.

The shape of the data-to-estimated-background ratio was similar in $p_T(\ell)$, E_T^{miss}, and m_T^{tot}: there was an excess of data compared to estimated background in the low-momentum region (up to roughly 60 GeV in $p_T(\ell)$ and roughly 40 GeV in E_T^{miss}) and a deficit of data compared to estimated background in the high-momentum region. Since the low-momentum region corresponds largely to the Z peak, this led to the initial suspicion that the Z background might be mismodeled, for example by having an incorrect normalization. The low-momentum region could also contain data events from decays of the 125 GeV Higgs boson, not included in the estimated background; however, the expected contribution was calculated to be negligible. The following studies were done to investigate the source of the mismodeling and mitigate it:

- removing the low-$p_T(\ell)$ region,

- investigating the Z peak by inverting the $Z/\gamma^* \to e^+e^-$ veto,

- changing the normalization of the $Z/\gamma^* \to \tau^+\tau^-$ and/or $Z/\gamma^* \to \ell^+\ell^-$ background,

- comparing the performance of fake factors parametrized in: $p_T(\tau_{\mathrm{had\text{-}vis}})$ only, $p_T(\tau_{\mathrm{had\text{-}vis}})$ and $\Delta\phi(\tau_{\mathrm{had\text{-}vis}}, E_T^{\mathrm{miss}})$ (with and without a $p_T(\ell)$ correction factor), $p_T(\tau_{\mathrm{had\text{-}vis}})$ and $p_T(\ell)$, $p_T(\tau_{\mathrm{had\text{-}vis}})$ and E_T^{miss}.

They will be described in detail below.

Removing the low-$p_T(\ell)$ region

Events in the Z peak region were removed by changing the $p_T(\ell)$ selection cut from 30 GeV to 75 GeV. New $f_W(p_T(\tau_{\mathrm{had\text{-}vis}}))$ fake factors were calculated using this cut. Figure 7.7 shows the modeling of $p_T(\ell)$, E_T^{miss}, and m_T^{tot} distributions in the b-veto pass-tau-ID W-FR with the new $p_T(\ell)$ cut. Removing the Z peak regions improves the modeling of both E_T^{miss} and m_T^{tot}, though E_T^{miss} still shows significant mismodeling in the high-momentum region. The distribution of $p_T(\ell)$ remains mismodeled even after the cut is applied. This provides mixed evidence as to $Z/\gamma^* \to \tau^+\tau^-/\ell^+\ell^-$ events causing the mismodeling.

Investigating the Z peak by inverting the $Z/\gamma^* \to e^+e^-$ veto

The nominal event selection in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel includes a cut to remove most of the Z peak in the $e\tau_{\mathrm{had\text{-}vis}}$ sub-channel, by vetoing events with $80 < m(e\tau_{\mathrm{had\text{-}vis}}) < 110$ GeV. In this study, this cut was inverted to create a W-FR enriched in $Z/\gamma^* \to \ell^+\ell^-/\tau^+\tau^-$ events.
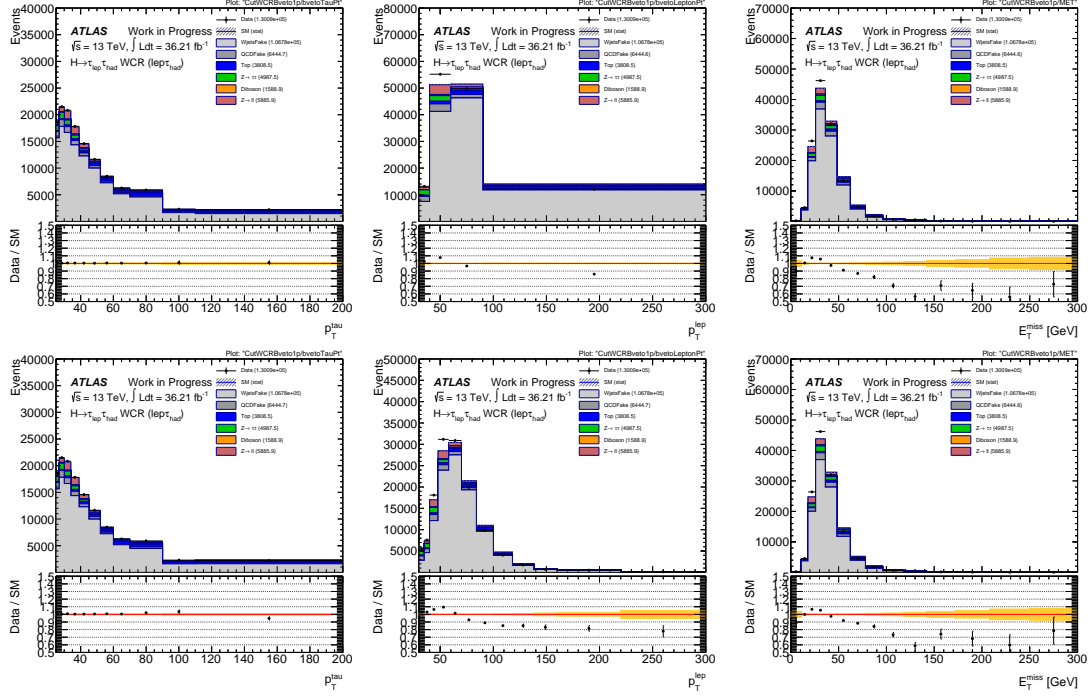
Figure 7.3: Distributions of $p_T(\tau_{\text{had-vis}})$ (left), $p_T(\ell)$ (center), and E_T^{miss} (right) in the b-veto 1-prong $\ell\tau_{\text{had-vis}}$ W-FR using $W+$jets background events calculated with $f_W(p_T(\tau_{\text{had-vis}}))$ (top) and $f_W(p_T(\tau_{\text{had-vis}}), \Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}}))$ (bottom).

Due to the $\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}}) > 60$ GeV cut used in the W-FR, the dominant background was still jets misidentified as tau leptons; however, moving to a low-$\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})$ region ($\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}}) < 40$ GeV) for this comparison was not possible, as that region would be expected to have signal events. The fake factor used was $f_W(p_T(\tau_{\text{had-vis}}))$ calculated using the nominal selection and full $p_T(\tau_{\text{had-vis}})$ range in the W-FR. Distributions of $p_T(\ell)$, E_T^{miss}, and m_T^{tot} in the $e\tau_{\text{had-vis}}$ W-FR with an inverted $m(e\tau_{\text{had-vis}})$ cut are shown in Figure 7.8. In this region, the mismodeling was less pronounced despite the higher percentage of $Z/\gamma^* \to e^+e^-/\tau^+\tau^-$ events, suggesting that an incorrect normalization of Monte Carlo Z events is not the cause.

Changing the normalization of $Z/\gamma^* \to \ell^+\ell^-/\tau^+\tau^-$ background

In the last step of the Z peak investigation, the normalization of the $Z/\gamma^* \to \ell^+\ell^-$ and $Z/\gamma^* \to \tau^+\tau^-$ background was altered, both simultaneously and one at a time, and $f_W(p_T(\tau_{\text{had-vis}}))$ fake factors were recalculated with the modified Monte Carlo event content. For ease of implementation in the analysis framework, the first approach was to double the normalization of $Z/\gamma^* \to \ell^+\ell^-/\tau^+\tau^-$ events. Such a large normalization error would be unlikely in centrally-produced Monte Carlo samples, but could be due to a bug in the analysis code. Additionally, if this very coarse normalization change produced a drastic change in the shape of mismodeling, the normalization change could then be fine-tuned to find a value that minimized the mismodeling. Distributions of $p_T(\ell)$, E_T^{miss}, and m_T^{tot} in the b-veto pass-tau-ID W-FR with a doubling of the $Z/\gamma^* \to \ell^+\ell^-$ and/or $Z/\gamma^* \to \tau^+\tau^-$ normalization are shown in Figure 7.9. The modeling of m_T^{tot} improved significantly and the modeling of $p_T(\ell)$ improved slightly with a doubling of only $Z/\gamma^* \to \ell^+\ell^-$. The doubling of $Z/\gamma^* \to \tau^+\tau^-$ background (with or without doubling $Z/\gamma^* \to \ell^+\ell^-$ background) worsened the modeling in the very low-m_T^{tot} region.

The improvement obtained with the change in $Z/\gamma^* \to \ell^+\ell^-$ normalization was not considered large enough to continue this study with other normalization values. However, later in the analysis, this issue was revisited by 'floating' the normalization of the $Z/\gamma^* \to \ell^+\ell^-$ background in the fit - that is, treating the normalization as a free parameter, much like the normalization of the t quark background. However, this did not result in a significant change compared to using the nominal normalization.
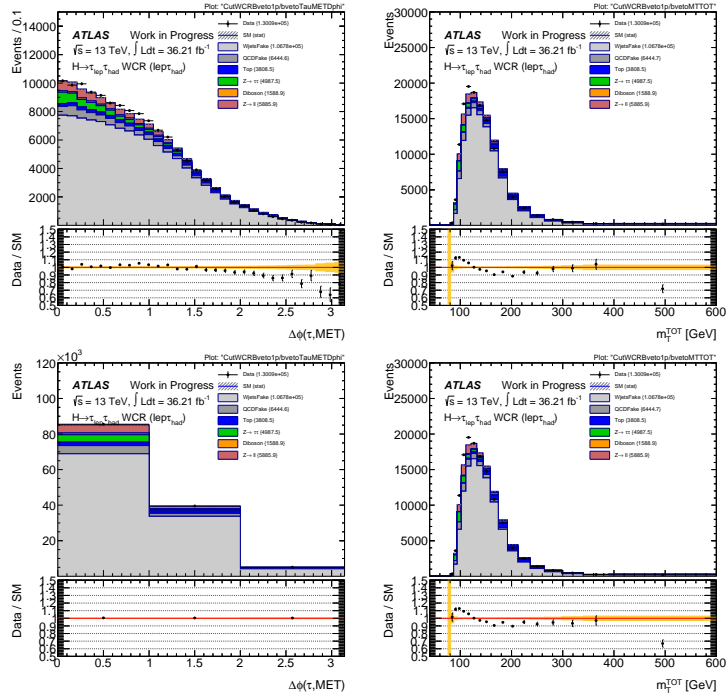
Figure 7.4: Distributions of $\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})$ (left), and m_T^{tot} (right) in the b-veto 1-prong $\ell\tau_{\text{had-vis}}$ W-FR using W+jets background events calculated with $f_W(p_T(\tau_{\text{had-vis}}))$ (top) and $f_W(p_T(\tau_{\text{had-vis}}), \Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}}))$ (bottom).

Figure 7.5: Distributions of $p_T(\tau_{\text{had-vis}})$ (left), $p_T(\ell)$ (center), and m_T^{tot} (right) in the b-veto 1-prong $e\tau_{\text{had-vis}}$ VR with W+jets background estimated using $f_W(p_T(\tau_{\text{had-vis}}))$ (top) and $f_W(p_T(\tau_{\text{had-vis}}), \Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}}))$ (bottom).
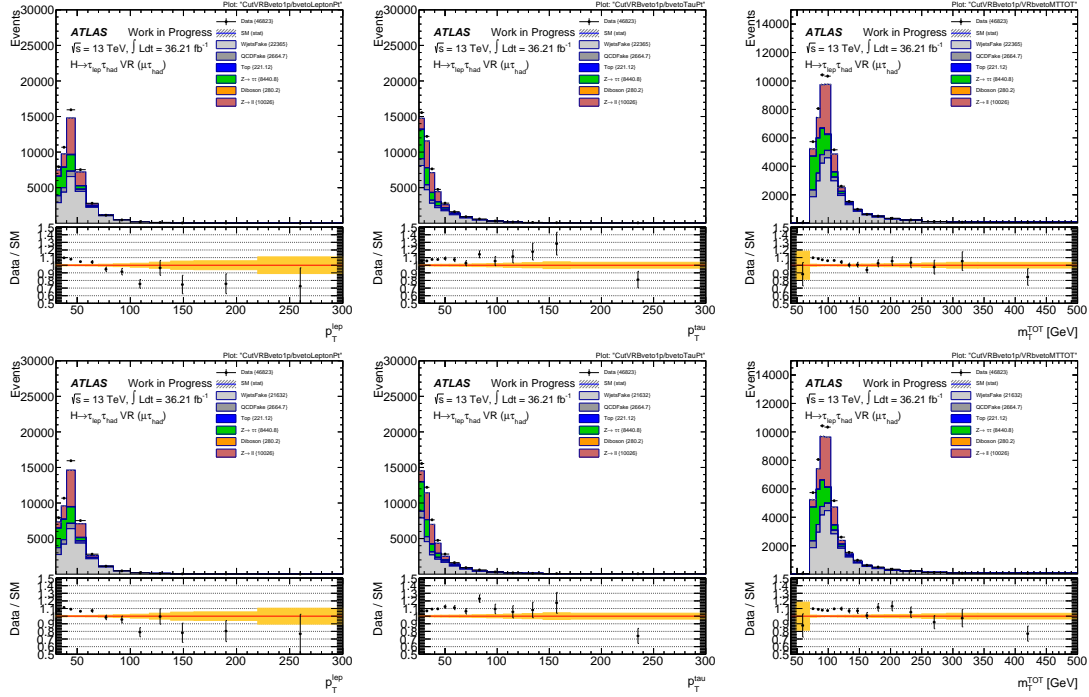
Figure 7.6: Distributions of $p_T(\tau_{\text{had-vis}})$ (left), $p_T(\ell)$ (center), and m_T^{tot} (right) in the b-veto 1-prong $\mu\tau_{\text{had-vis}}$ VR with $W+$jets background estimated using $f_W(p_T(\tau_{\text{had-vis}}))$ (top) and $f_W(p_T(\tau_{\text{had-vis}}), \Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}}))$ (bottom).
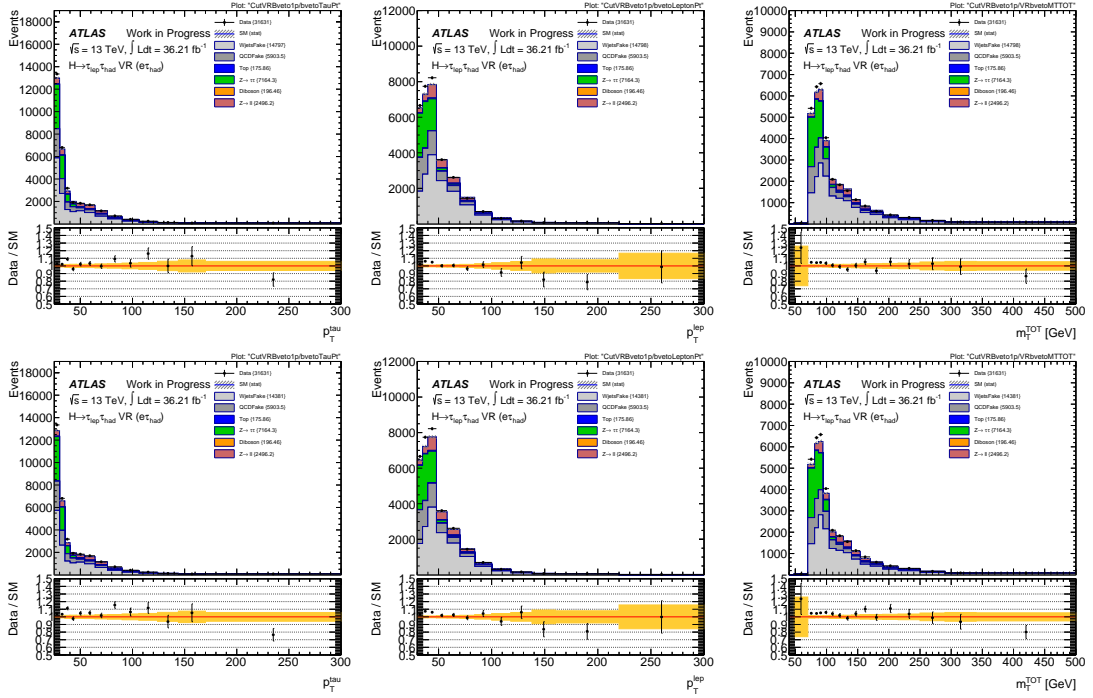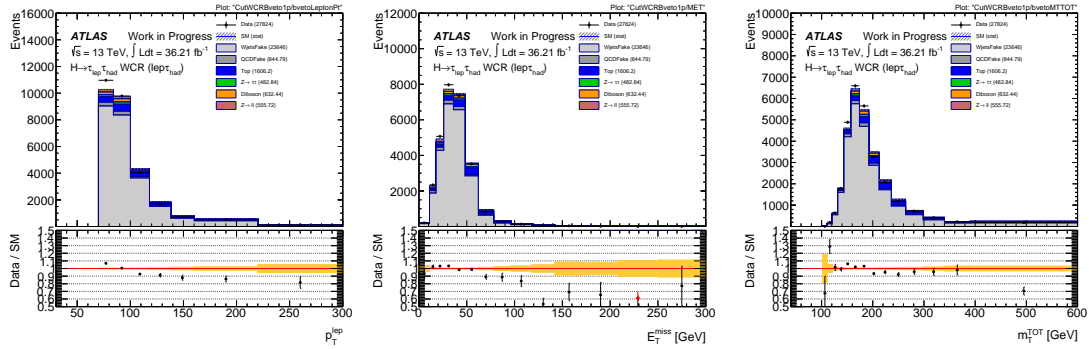

Figure 7.7: Distributions of $p_T(\ell)$ (left), E_T^{miss} (center), and m_T^{tot} in the b-veto 1-prong $\ell\tau_{\text{had-vis}}$ W-FR with a $p_T(\ell) > 75$ GeV cut. The f_W fake factor was calculated using the same $p_T(\ell)$ cut.
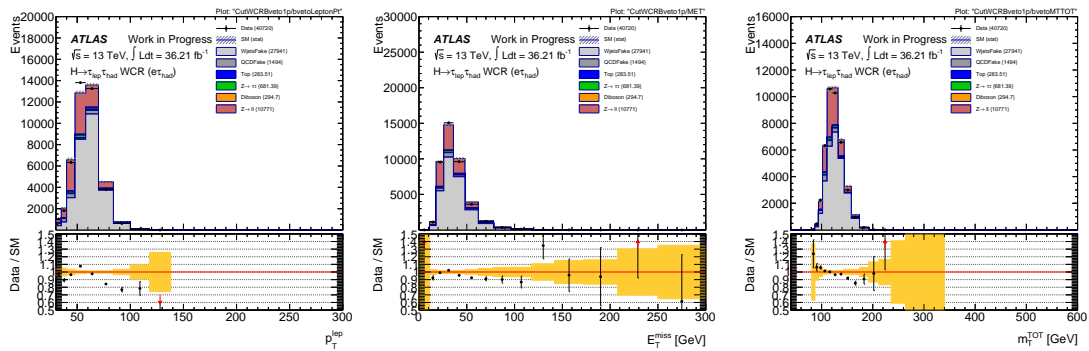
Figure 7.8: Distributions of $p_T(\ell)$ (left), E_T^{miss} (center), and m_T^{tot} in the b-veto 1-prong $e\tau_{\text{had-vis}}$ W-FR with a $80 < m(e\tau_{\text{had-vis}}) < 110$ GeV cut. The nominal $f_W(p_T(\tau_{\text{had-vis}}))$ fake factor was used for $W+$jets estimation.
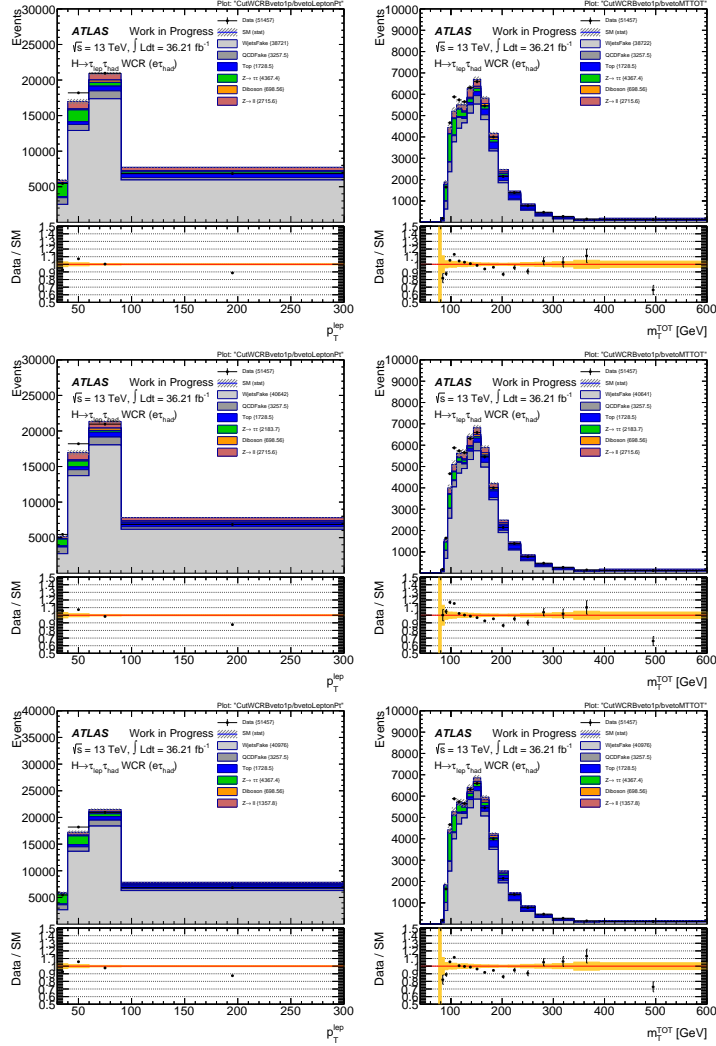
Figure 7.9: Distributions of $p_T(\ell)$ (left) and m_T^{tot} (right) in the b-veto 1-prong $\ell\tau_{\text{had-vis}}$ W-FR with different normalizations of the $Z/\gamma^* \rightarrow \ell^+\ell^-/\tau^+\tau^-$ background contribution: double the nominal normalization of both background contributions (top row), double the normalization of the $Z/\gamma^* \rightarrow \ell^+\ell^-$ contribution (center), and double the normalization of the $Z/\gamma^* \rightarrow \tau^+\tau^-$ contribution (bottom).

Investigating different parametrizations of the W+jets fake factor

Several different parametrizations of the f_W fake factor were considered, to determine which one resulted in the best overall modeling:

- parametrization in $p_T(\tau_{\text{had-vis}})$ with $\{25, 27, 30, 34, 39, 45, 52, 60, 70, 90, 110, 200, 300\}$ GeV (nominal) binning,

- simultaneous parametrization in $p_T(\tau_{\text{had-vis}})$ with nominal binning and $\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})$ with $\{0, 1, 2, 2.5, \pi\}$ GeV (nominal) binning,

- simultaneous parametrization in $p_T(\tau_{\text{had-vis}})$ with nominal binning and $p_T(\ell)$ with $\{30, 40, 60, 90, 300\}$ GeV binning,

- simultaneous parametrization in $p_T(\tau_{\text{had-vis}})$ with nominal binning and E_T^{miss} with $\{0, 20, 40, 60, 90, 300\}$ GeV binning,

- parametrization in $p_T(\tau_{\text{had-vis}})$ with nominal binning with a sequential $p_T(\ell)$ correction, with $\{30, 40, 60, 90, 300\}$ GeV binning, applied on top,

- simultaneous parametrization in $p_T(\tau_{\text{had-vis}})$ and $\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})$, both with nominal binning, with a sequential $p_T(\ell)$ correction, with $\{30, 40, 60, 90, 300\}$ GeV binning, applied on top.

All versions of f_W were additionally parametrized in the number of charged tracks of $\tau_{\text{had-vis}}$ (either one or thee). While the binning in $p_T(\tau_{\text{had-vis}})$ was kept rather fine to capture the differences between different $p_T(\tau_{\text{had-vis}})$ regions, binning in any second variable was coarser, in order to keep the number of events in each 2-dimensional bin reasonably high and keep the statistical uncertainty reasonably low. Adding a correction factor sequentially on top of the fake factor was another method of avoiding a small number of events in individual bins. Correction factors f_{corr} were calculated by comparing distributions of data and background estimated with Monte Carlo and data-driven methods in the pass-identification W-FR in each bin of the variable to be corrected (in this case always $p_T(\ell)$):

$$f_{\text{corr}}(p_T(\ell)) = \frac{N_{\text{data}}(p_T(\ell))}{N_{\text{MC}}(p_T(\ell)) + N_{\text{data-driven}}(p_T(\ell))}. \tag{7.11}$$

The distributions of $p_T(\tau_{\text{had-vis}})$, $p_T(\ell)$, E_T^{miss}, $\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})$, $\Delta\phi(\ell, E_T^{\text{miss}})$, and m_T^{tot} in the W-FR for each of these f_W parametrizations are shown in Figure 7.3 and 7.10. Where f_W was parametrized in one or two variables without a correction factor, the distributions of these variables show a perfect agreement between data and background estimated with Monte Carlo and data-driven methods, as long as the binning used for calculating the fake factor and visualizing the distributions was the same. Where the $p_T(\ell)$ correction was used, only the $p_T(\ell)$ distribution shows perfect agreement. Correcting the $p_T(\ell)$ distribution causes mismodeling (with a slope in the data-to-estimated-background ratio) in the $p_T(\tau_{\text{had-vis}})$ distribution, as the two variables are closely correlated. Parametrizing the f_W in $p_T(\ell)$ or adding a sequential $p_T(\ell)$ correction improves the m_T^{tot} modeling compared to the two nominal f_W parametrizations, while parametrizing the f_W in E_T^{miss} does not improve the modeling.

Final version of the W+jets fake factor

In a separate study conducted within the analysis group, it was found that the tau-ID BDT score distributions of jets misidentified as hadronically-decaying taus are different in the SR, VR, and W-FR, especially for high values of $\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})$. These jets had the highest average BDT score in the SR, followed by VR, and - finally W-FR. This is most likely due to different fractions of quark- and gluon-initiated jets, which have different probabilities of being misidentified as hadronically-decaying tau leptons, in regions that all share the back-to-back requirement for ℓ and $\tau_{\text{had-vis}}$, but have different $m_T(\tau_{\text{had-vis}}, E_T^{\text{miss}})$ requirements. Thus, it was determined that $\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})$ was an unsuitable variable for parametrizing the W+jets fake factor. Parametrization in $p_T(\tau_{\text{had-vis}})$ with a sequential $p_T(\ell)$ correction was chosen as the final version,
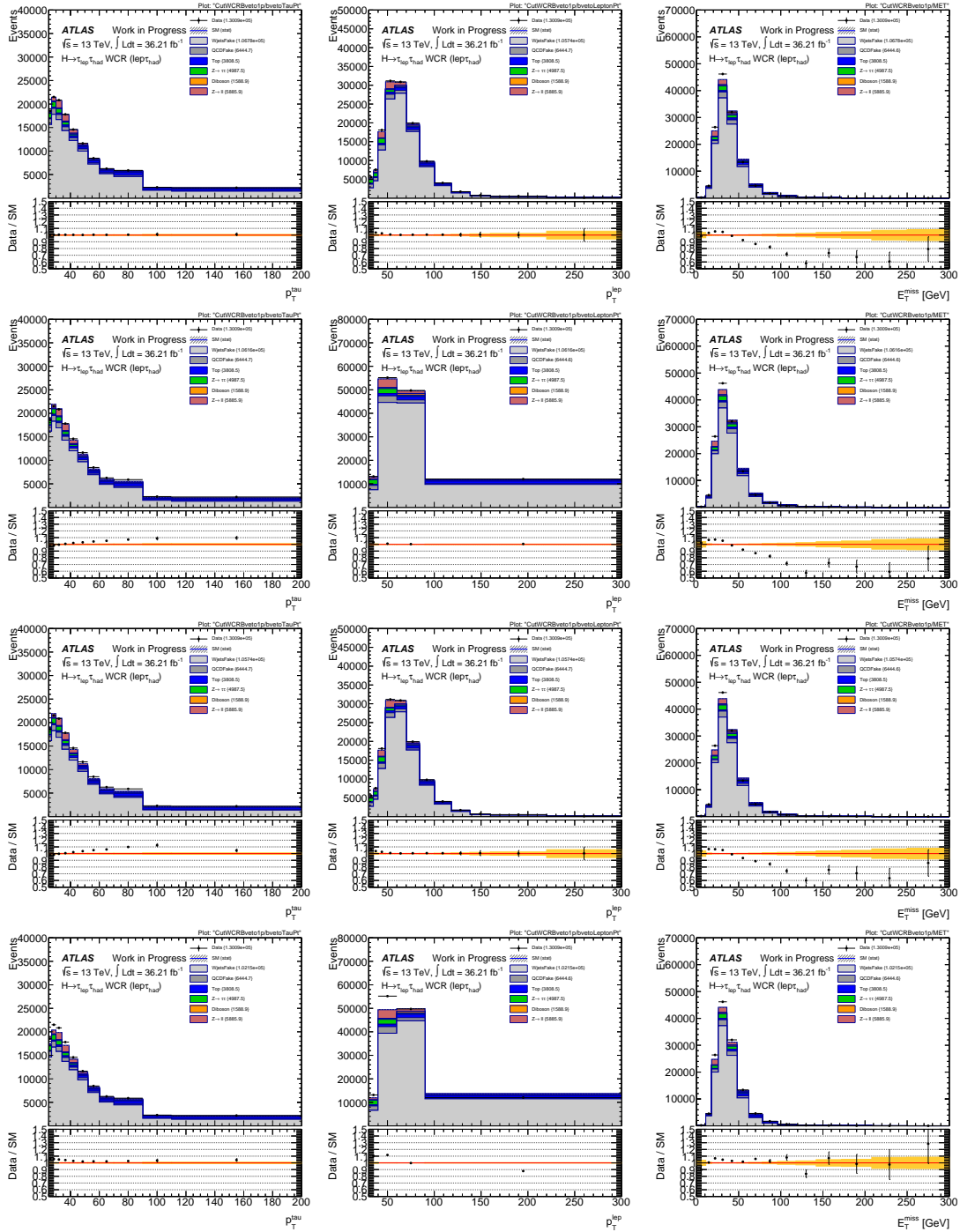
Figure 7.10: Distributions of $p_T(\tau_{\text{had-vis}})$ (left), $p_T(\ell)$ (center), and E_T^{miss} (right) in the b-veto 1-prong $\ell\tau_{\text{had-vis}}$ W-FR using different f_W parametrizations: $f_W(p_T(\tau_{\text{had-vis}}), p_T(\ell))$ (top row), $f_W(p_T(\tau_{\text{had-vis}})) + f_{\text{corr}}(p_T(\ell))$ (second from top), $f_W(p_T(\tau_{\text{had-vis}}), \Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})) + f_{\text{corr}}(p_T(\ell))$ (third from top), $f_W(p_T(\tau_{\text{had-vis}}), E_T^{\text{miss}})$ (bottom).
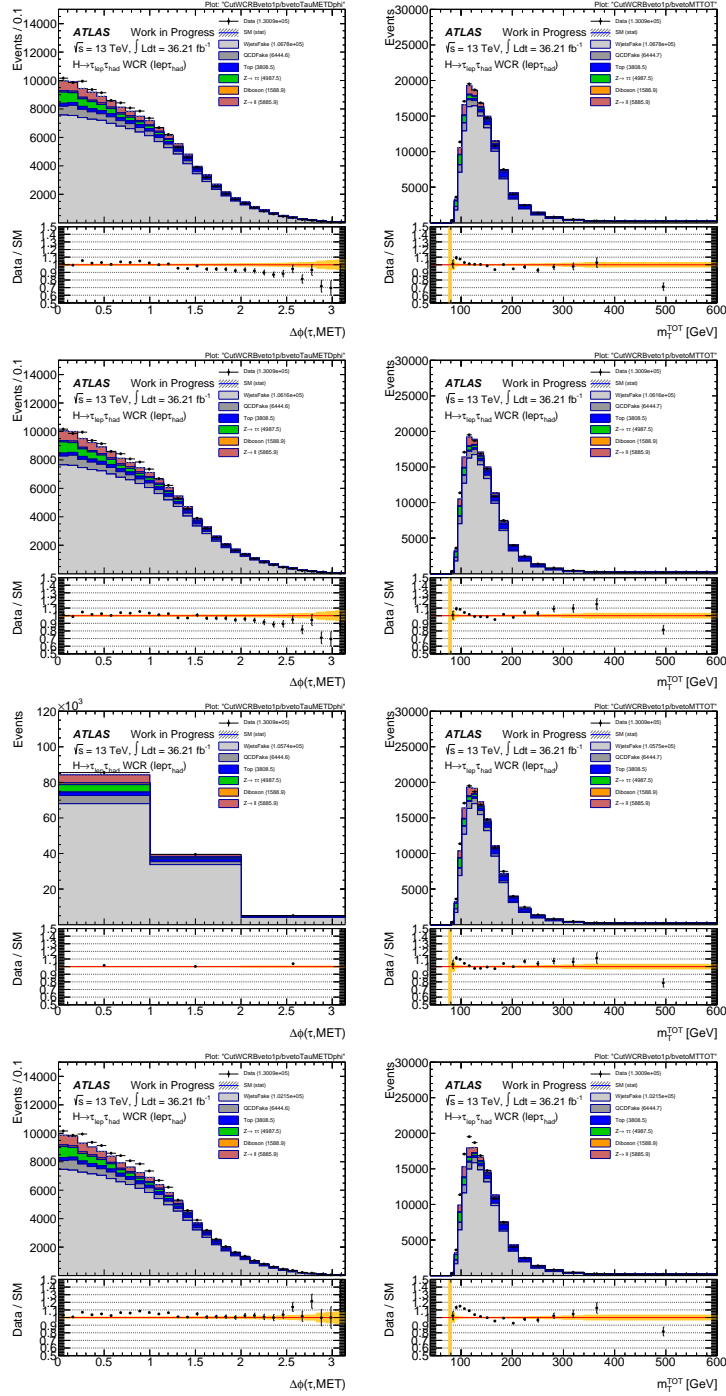
Figure 7.11: Distributions of $\Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})$ (left) and m_T^{tot} (right) in the b-veto 1-prong $\ell\tau_{\text{had-vis}}$ W-FR using different f_W parametrizations: $f_W(p_T(\tau_{\text{had-vis}}), p_T(\ell))$ (top row), $f_W(p_T(\tau_{\text{had-vis}})) + f_{\text{corr}}(p_T(\ell))$ (second from top), $f_W(p_T(\tau_{\text{had-vis}}), \Delta\phi(\tau_{\text{had-vis}}, E_T^{\text{miss}})) + f_{\text{corr}}(p_T(\ell))$ (third from top), $f_W(p_T(\tau_{\text{had-vis}}), E_T^{\text{miss}})$ (bottom).

as it resulted in only a mild mismodeling in $p_T(\tau_{\text{had-vis}})$ and improved m_T^{tot} modeling considerably compared to the previous nominal fake factors. The correction was treated as a systematic uncertainty in later stages of the analysis. Plots 7.12-7.13 show distributions of $p_T(\tau_{\text{had-vis}})$, $p_T(\ell)$, and m_T^{tot} in the VR with full statistical and systematic uncertainty (described in Section 7.6) using $f_W(p_T(\tau_{\text{had-vis}}))$ with and without a sequential $p_T(\ell)$ correction. When the correction is applied, the data and estimated background distributions of all three variables agree within the uncertainty; without the correction, $p_T(\ell)$ shows mismodeling outside the uncertainty, but $p_T(\tau_{\text{had-vis}})$ and m_T^{tot} remain in agreement. Since the SR contains less W+jets background compared to the VR, it is expected that distributions of these variables will also agree within the total uncertainty in the SR.

7.6 Systematic uncertainties

In this analysis, experimental sources of systematic uncertainty included uncertainties on the integrated luminosity measurement, pile-up, the efficiency of reconstruction, identification, isolation, triggering, and tagging algorithms, and the energy scale and resolution of reconstructed objects. Theoretical sources included the uncertainties of cross-section calculations, the QCD scale (the energy scale above which QCD can be described by perturbative theory), and the modeling of initial- and final-state radiation and multi-parton interactions. The use of fake factors and fake rates, which are affected by limited statistics in control regions and fake regions and by subtraction uncertainties, resulted in systematic uncertainties on data-driven background components.

Most systematic uncertainties are calculated centrally at ATLAS and updated periodically. They are applied only to signal and background contributions estimated with Monte Carlo, usually by re-creating Monte Carlo samples with parameters varied up and down within the uncertainty. Table 7.4 lists all categories of centrally-calculated systematic uncertainties used in this analysis, along with the number of individually-considered sources of uncertainty in each category. The total number of distinct sources of uncertainty (distinct contributions) was 111.

Table 7.4: A list of centrally-calculated sources of systematic uncertainty used in the $H/A \to \tau^+\tau^-$ analysis.

category	description	no. contributions
luminosity	integrated luminosity measurement	1
$\tau_{\text{had-vis}}$ objects	reconstruction, ID, electron veto, energy scale	25
muon objects	trigger, reconstruction, ID, isolation, energy scale	15
electron objects	trigger, reconstruction, ID, isolation, energy scale	12
E_T^{miss} objects	uncertainty related to E_T^{miss} soft terms	3
jet objects	jet energy scale and energy resolution	17
b-tagging	tagging efficiency	14
pile-up	pile-up measurement	1
diboson cross-section	cross-section calculation	1
$t/t\bar{t}$ cross-section	cross-section calculation	1
Z+jets modeling	PDFs, energy scale, QCD coupling constant	12
$t\bar{t}$ modeling	uncertainty on variable distribution shapes	4
signal acceptance	factorization and renormalization scale, PDFs, QCD scale, Monte Carlo tune (initial- and final-state radiation)	5
total	-	111

Systematic uncertainties related to data-driven background estimation are analysis-specific and were calculated by the analysis team. There are four basic sources of uncertainty and error when using the fake-factor method:

1. the statistical uncertainties of data and subtracted Monte Carlo backgrounds in relevant control regions,

2. the modeling uncertainty of subtracted Monte Carlo backgrounds in relevant control regions

Figure 7.12: Distributions of $p_T(\tau_{\text{had-vis}})$ (left), $p_T(\ell)$ (center), and m_T^{tot} (right) in the b-veto 1-prong+3-prong $e\tau_{\text{had-vis}}$ VR with W+jets background estimated using $f_W(p_T(\tau_{\text{had-vis}}))$ with (top) and without (bottom) the $f_{\text{corr}}(p_T(\ell))$ correction factor.
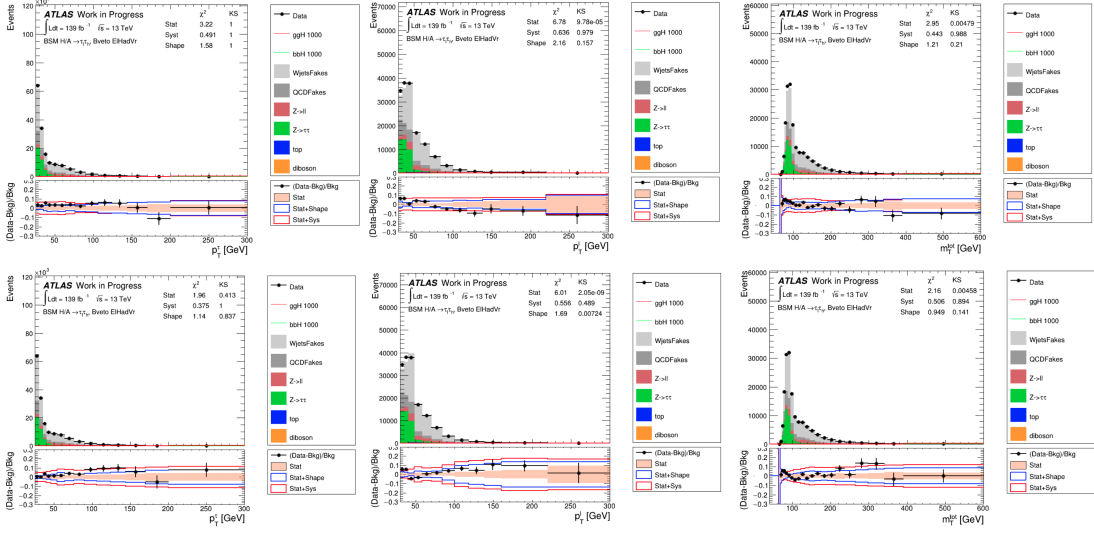

Figure 7.13: Distributions of $p_T(\tau_{\text{had-vis}})$ (left), $p_T(\ell)$ (center), and m_T^{tot} (right) in the b-veto 1-prong+3-prong $\mu\tau_{\text{had-vis}}$ VR with W+jets background estimated using $f_W(p_T(\tau_{\text{had-vis}}))$ with (top) and without (bottom) the $f_{\text{corr}}(p_T(\ell))$ correction factor.
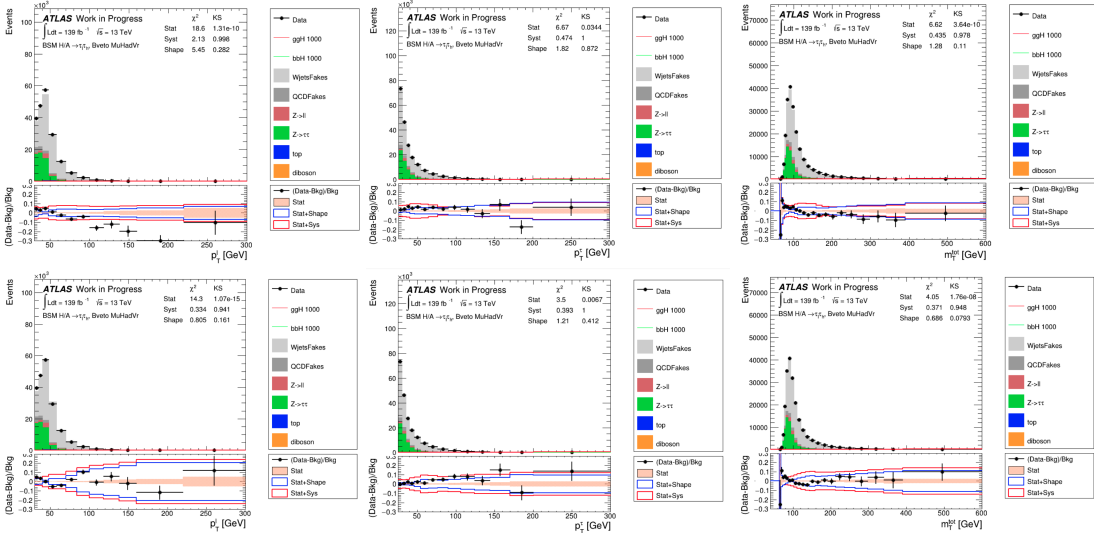
3. uncertainty due to extrapolation of the fake factors from the control regions to the signal region, and

4. error due to mismodeling of variable distributions not used for parametrizing the fake factors.

The statistical uncertainties were determined straightforwardly from histogram subtraction. The modeling uncertainty of subtracted Monte Carlo background was accounted for by varying the background normalization by a certain value and recalculating the fake factors. This value was 20% for the lepton fake factor in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel and 10% for the W+jets fake factor in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. The 10% value was determined from conservative estimates of the precision of theoretical cross-section calculation and measurements on electron, muon and $\tau_{\text{had-vis}}$ performance at ATLAS. This percentage was doubled for the lepton fake factor, since the lepton fake control region does not use lepton isolation requirements, and non-isolated leptons had not been thoroughly studied and validated. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the normalization of Monte Carlo backgrounds used for calculating the QCD fake factor was varied by 1σ of the statistical uncertainty and the total uncertainty from sources given in Table 7.4, added in quadrature. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, the full correction from data-driven correction factors is assigned as an additional source of uncertainty. Additionally, uncertainty due to extrapolating the W+jets fake factor from the W+jets fake factor region (high $m_T(\ell, E_T^{\text{miss}})$) to the signal region (low $m_T(\ell, E_T^{\text{miss}})$) was accounted for by checking the modeling in the validation region (intermediate $m_T(\ell, E_T^{\text{miss}})$) and taking 100% of the disagreement between data and estimated background in the $p_T(\tau_{\text{had-vis}})$ distribution as the systematic uncertainty. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel no significant mismodeling was observed in control and validation regions, and thus uncertainty due to fake factor extrapolation and mismodeling was not taken into account. Systematic uncertainties on fake rates used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel are determined from statistical uncertainties only; this is due to a very small amount of subtracted Monte Carlo in the W-FR and T-FR regions used for fake rate estimation.

7.7 Statistical model and results

The parameter of interest in the analysis is the signal strength μ, defined as the ratio of the observed value of the resonance production cross-section times the resonance branching ratio to the predicted value:

$$\mu = \frac{(\sigma \times \mathcal{B})_{\text{observed}}}{(\sigma \times \mathcal{B})_{\text{predicted}}}. \tag{7.12}$$

Predicted values are calculated for specific theoretical models, called benchmark scenarios. A model-independent case is also evaluated, with $(\sigma \times \mathcal{B})_{\text{predicted}}$ set to 1.0 pb. The parameter of interest is estimated using a complex profile likelihood which parametrizes all available knowledge and tests it against data [270–272]. The likelihood function is constructed as the product of Poisson probability terms with one term derived from each bin of discriminating variable distributions in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels and b-veto and b-tag categories:

$$L(n|\mu S + B) = \prod_{b \in \text{bins}}^{N} \frac{(\mu S_b + B_b)^{n_b}}{n_b!} e^{-(\mu S_b + B_b)}. \tag{7.13}$$

Here, n is the number of observed (data) events, S is the number of expected signal events, and B is the expected number of background events. The discriminating variable is the total transverse mass m_T^{tot}, defined in Eq. 7.2. Systematic uncertainties are treated as nuisance parameters θ - parameters which are not of interest, but must be accounted for in the statistical model - and constrained using Gaussian functions. The expected number of signal and background events are functions of these nuisance parameters. Since the values of experimental nuisance parameters are typically determined from measurements done with auxiliary datasets, an additional term is added to reflect this knowledge:

$$L(m|u) = \prod_{b \in \text{bins}}^{N} \frac{(u_b)^{m_b}}{m_b!} e^{-u_b}, \tag{7.14}$$

where m and u are the measured and expected number of events used to calculate the central value of a nuisance parameter. This term acts as a penalty, increasing if a nuisance parameter is shifted

away from its central value. It should be noted that some nuisance parameters, like floating normalizations, do not have assigned priors and are determined exclusively from the analysis dataset.

The full statistical model, a function of μ and θ, is then defined as follows:

$$\mathcal{L} = \prod_{b \in \text{bins}}^{N} \frac{(\mu S_b(\theta) + B_b(\theta))^{n_b}}{n_b!} e^{-(\mu S_b(\theta) + B_b(\theta))} L(m|u). \qquad (7.15)$$

Next, a fit is conducted to maximize $\mathcal{L}(\mu, \theta)$ with respect to all parameters. The following test statistic q_μ is constructed to quantify the level of agreement between the observed data and a hypothesis with a specific value of μ:

$$q_\mu = 2 \ln \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \qquad (7.16)$$

where $\hat{\mu}$ and $\hat{\theta}$ are parameter values that maximize the likelihood globally and $\hat{\hat{\theta}}$ are nuisance parameter values that maximize the likelihood for a given μ. To test the background-only hypothesis, μ is set to 0. The background-only hypothesis is used to set exclusion limits. Two types of limits are typically calculated: expected and observed. The expected limit corresponds to what would be obtained if real data from the detector matched the estimated background exactly. This is a way of determining the sensitivity of the analysis before it is finalized, while keeping it blinded. The observed limit is calculated using real data.

The post-fit distributions of the discriminating variable m_T^{tot} in the signal region are shown in Figure 7.14. The binning in these plots is the same as the binning used in the fit. The superimposed signal histograms correspond to a few signal hypotheses. The fit was conducted simultaneously in the signal region and the unbinned top control region (T-CR) in order to fix the normalization of the $t/t\bar{t}$ background.

The impact of various systematic uncertainties on the 95% confidence level upper exclusion limits for two signal mass hypotheses is given in Table 7.5. The four most significant sources of uncertainty are tau lepton identification efficiency, tau lepton energy scale, the modeling of Z+jets background, and the modeling of background with jets misidentified as taus. The combined effect of all other sources of uncertainty is small by comparison. The impact of systematic uncertainties is much larger for the lower mass hypothesis (400 GeV) compared to the higher one (1 TeV). It is also larger for ggH/A signal compared to bbH/A signal.

Table 7.5: The impact of systematic uncertainties on the 95% confidence level upper exclusion limit for the production cross-section times branching ratio, expressed as the relative increase over the expected limit. Values are given for two mass hypotheses (400 GeV and 1 TeV) and separately for the two considered Higgs boson production modes.

source	ggH/A 400 GeV	ggH/A 1 TeV	bbH/A 400 GeV	bbH/A 1 TeV
$\tau_{\text{had-vis}}$ ID efficiency	14%	16%	12%	8%
$\tau_{\text{had-vis}}$ energy scale	33%	9%	22%	3%
Z+jets modeling	27%	19%	8%	4%
data-driven bkg modeling	22%	1%	14%	3%
other	9%	4%	11%	2%
total	54%	28%	45%	13%

7.7.1 Model-independent exclusion limits

In the model-independent approach, exclusion limits are set on the BSM Higgs boson production cross-section times the branching ratio of Higgs boson decays to two tau leptons for each available mass hypothesis. No assumptions are made about the relative fractions of Higgs bosons produced via gluon-gluon fusion and b-associated production. Exclusion limits are set separately for each production mode, as shown in Figure 7.15. Additionally, limits are calculated as a function of both
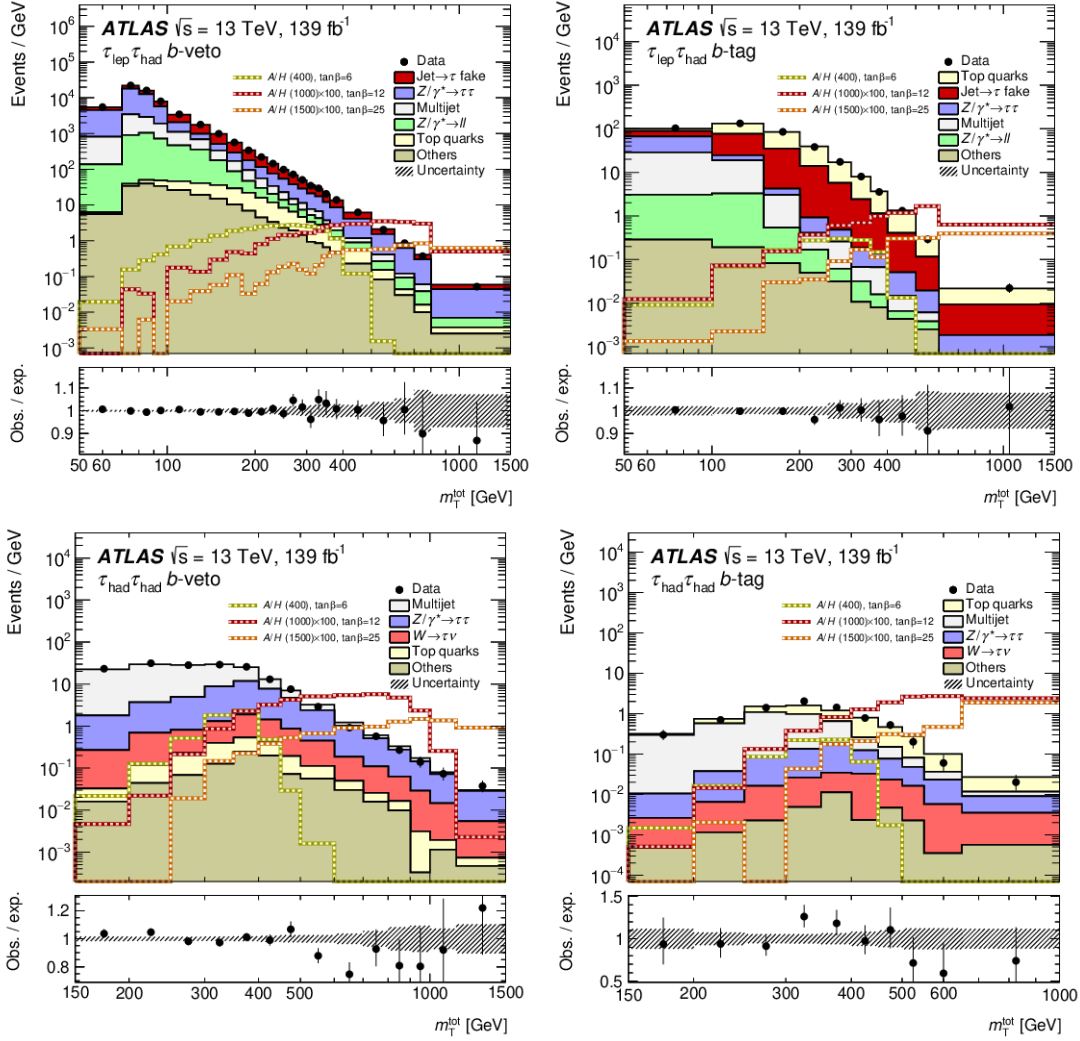
Figure 7.14: Post-fit plots showing m_T^{tot} distributions in the b-veto (left) and b-tag (right) categories of the $\tau_{lep}\tau_{had}$ (top) and $\tau_{had}\tau_{had}$ (bottom) channels.

the Higgs boson mass and the relative production mode fractions by scanning over the fraction of b-associated Higgs production in the range from 0 (gluon-gluon fusion only) to 1 (b-associated production only) in increments of 0.05. These two-dimensional model-independent exclusion limits were then used to derive model-dependent limits.

7.7.2 Model-dependent exclusion limits

In this analysis, a model-dependent exclusion limit was first set for the m_h^{125} benchmark scenario, described in Section 1.2.4. The limit, shown in Figure 7.16, constrains the $m_A - \tan\beta$ phase space of the model. The lowest considered value of $\tan\beta$ was 0.5. Mass degeneracy of H and A Higgs bosons (decoupling limit) was assumed; the blue area in the bottom left corner of the plot corresponds to a region of phase space where the mass splitting would be above 50% of the mass resolution, causing the limit to lose validity. Additionally, this region provides predictions that are incompatible with the measured mass of the 125 GeV Higgs boson by over 3σ. The shape of the exclusion limit in the low-m_A region, near 350 GeV, reflects the behavior of the $H/A \rightarrow$ branching ratio close to the $H/A \rightarrow t\bar{t}$ kinematic threshold for low $\tan\beta$ values. The limit excludes values of $\tan\beta > 8$ and $\tan\beta > 21$ for $m_A = 500.0$ GeV and $m_A = 1.5$ TeV, respectively, at the 95% confidence level.

This result was later reinterpreted for the hMSSM scenario, also described in Section 1.2.4, and compared with exclusion limits obtained from other ATLAS searches and precision measurements. A summary plot is shown in Figure 7.17. The area in grey corresponds to the exclusion limit set by this analysis; as expected, the $H/A \rightarrow \tau^+\tau^-$ channel is especially useful for setting limits for high $\tan\beta$ values. For $\tan\beta = 60$, the limit on allowed m_A masses is twice as stringent as the next-best limit set by the $H^\pm \rightarrow \tau^\pm\nu$ analysis (2 TeV vs 1 TeV). For $\tan\beta$ values between approximately 6 and 25, no other ATLAS search has set limits on m_A. The only other limits in this range come from precision measurements of 125 GeV Higgs boson couplings, and for most $\tan\beta$ values they are less stringent than the limits from the $H/A \rightarrow \tau^+\tau^-$ analysis.

7.8 Conclusion

The BSM $H/A \rightarrow \tau^+\tau^-$ analysis is one of several searches for BSM Higgs bosons at ATLAS. So far no particles of this type have been found, but new results continue to constrain the phase space of various BSM models. In this analysis round, the full Run 2 dataset was analyzed to set improved exclusion limits using the analysis strategy from a previous paper. Limits were set on BSM Higgs boson production cross-section times the branching ratio to two tau leptons as well as on the $m_A - \tan\beta$ phase space of the m_h^{125} and hMSSM benchmark scenarios.
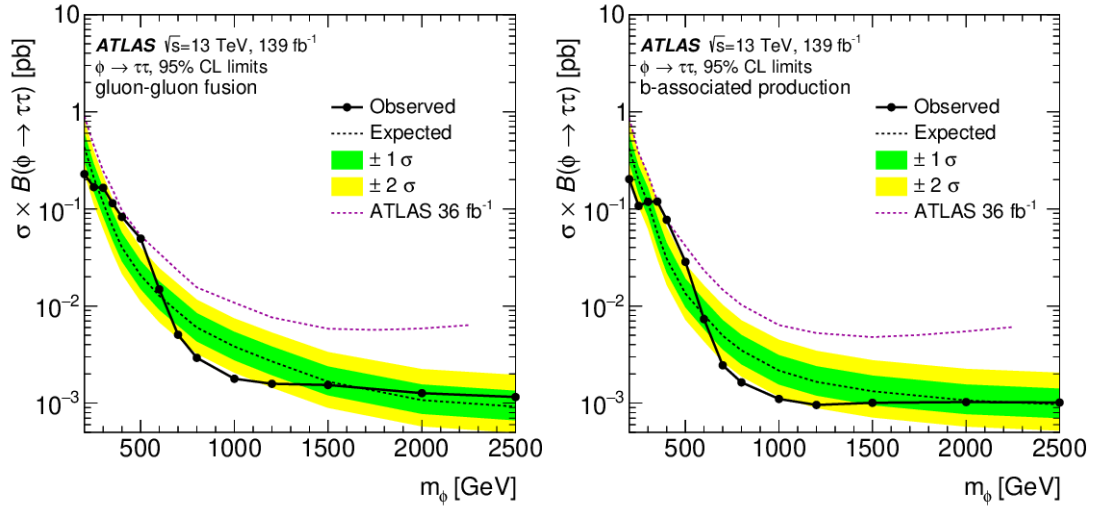
Figure 7.15: 95% confidence level exclusion limits on the BSM Higgs boson production cross-section times $H/A \to \tau^+\tau^-$ branching ratio for gluon-gluon fusion (left) and b-associated (right) Higgs boson production modes. Exclusion limits from the previous search, which analyzed 36 fb^{-1} of data from ATLAS, are shown in red.
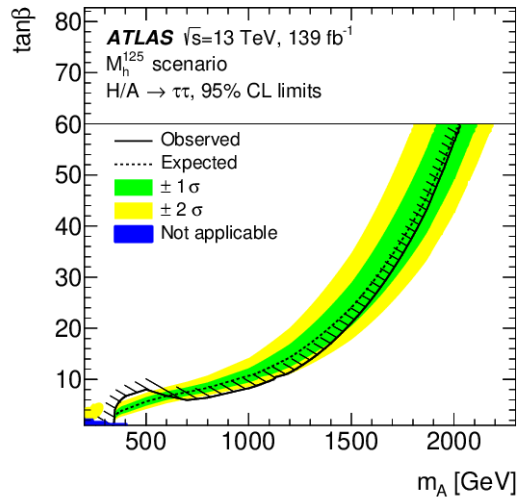
Figure 7.16: 95% confidence level exclusion limits on $\tan\beta$ as a function of Higgs boson mass in the m_h^{125} benchmark scenario. The hatched area shows which side of the curve is excluded by the analysis.
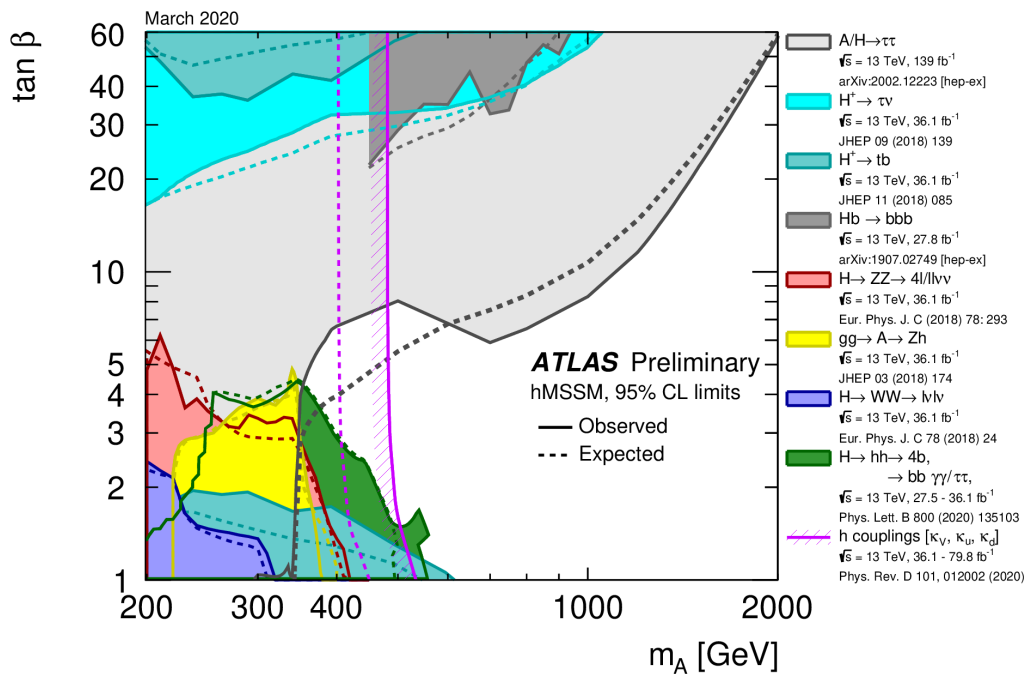
Figure 7.17: A summary plot showing 95% confidence level exclusion limits on the $\tan\beta$-m_A phase space of the hMSSM from various ATLAS searches and precision measurements. [273]

Chapter 8

BSM $H/A \to \tau^+\tau^-$ Search: Multivariate Classifier for Round 2

In the second round of the $H/A \to \tau^+\tau^-$ analysis, which was geared towards producing a 'legacy' paper, various improvements to the analysis strategy were implemented. These included switching to a multivariate final discriminant, adding additional signal interpretations (decays of the BSM Z' boson and third-generation scalar leptoquark to two tau leptons), improving the estimation of background with jets misidentified as tau leptons, and extending the phase space. This chapter will focus on the author's contribution: the development of a multivariate final discriminant in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel.

The purpose of developing a multivariate classifier to replace the m_T^{tot} final discriminating variable was to improve the sensitivity of the analysis. This classifier, which could combine information from various input variables carrying information about the kinematics and mass of the $\tau^+\tau^-$ system, would be a supervised machine learning model trained on labeled background and signal events. The ultimate goal was to train a mass-parametrized neural network, which could be trained on all signal mass points simultaneously and then used for calculating the limit at each mass point, eliminating the need for training a different model for each mass point.

8.1 Methods

The ROOT-based TMVA toolkit (described in Section 3.4) was used to train and validate models. Events were given weights for training and validation. These were the same weights as the ones used in the standard analysis, taking into account Monte Carlo weights, cross-sections, luminosities, and fake/scale factors for background events estimated with data-driven methods. Events with negative weights were dealt with by taking the absolute value of the weight during training. For most studies involving ROC AUC calculation, a 50:50 train:test split was done in the interest of time. For limit calculation, k-folding with five folds was utilized. Where the training and validation datasets differed (ex. when tuning mass-parametrized neural networks to a certain mass) or a very precise calculation of the ROC AUC was needed, models were evaluated with a separate python script, employing stand-alone ROC AUC calculation, rather than using the built-in TMVA method.

At each step, some manual hyperparameter optimization was done. Quoted ROC AUC values come from the best-performing, non-overfitted classifier found during optimization. For BDTs, optimization involved varying the number of trees, minimum node size, maximum tree depth, boost type (AdaBoost or gradient boost), and number of cuts. For neural networks, the optimization involved changing the architecture, the activation function (tanh, sigmoid, ReLU), and the variable transformation method (normalization, decorrelation, Gaussianization).

The final figures of merit in this series of studies were 95% confidence level exclusion limits on the BSM Higgs boson production cross-section times the branching ratio to two tau leptons. These limits were calculated using the same approach as the one detailed in Section 7.7, but using the classifier response distributions rather than m_T^{tot}. Histograms were binned using the following

algorithm:

- start with a very fine binning: 50 000 bins per histogram

- starting from the signal-like end, merge bins one by one until the following criteria are satisfied:

  - at least N background events in the merged bin (with N usually set to 5 or 10)
  - at least $M\%$ of the signal in the merged bin (with M usually set to 2.5% or 5%)

  then move on, starting another merged bin

- the final bin can have less than $M\%$ of the signal

In the case of m_T^{tot}, the rebinning procedure started from the high-mass end of the distribution. Limits will be quoted here in the limit$_{-1\sigma}^{+1\sigma}$ format.

8.2 Choosing input variables in the b-veto category

When choosing input variables for machine learning methods, it is important to consider how signal differs from background and which variables might potentially be biased in Monte Carlo and data-driven fake events compared to real data. In the b-veto category, the main source of background is from $Z/\gamma^* \to \tau^+\tau^-$ events. This background is irreducible with respect to the signal, since the composition of final state particles is the same. Differences in topologies between Higgs boson and Z boson decays are primarily due to two factors: the mass and polarization of the boson.

8.2.1 Mass variables

Since it is difficult to reconstruct the invariant mass of a $\tau^+\tau^-$ pair (due to the presence of multiple neutrinos in the final state, which are not seen by the detector), distributions of mass variables like m_T^{tot} are necessarily wide, causing significant overlap for decays of Z bosons and low-mass BSM Higgs boson decays. A comparison of m_T^{tot} and true invariant mass distributions for $Z/\gamma^* \to \tau^+\tau^-$ and 200 GeV ggH/A and bbH/A Monte Carlo samples, after all selection cuts, is shown in Figure 8.1. Since multiple methods of approximating the invariant mass of $\tau^+\tau^-$ pairs exist, combining different mass variables in a multivariate classifier can mitigate this effect, improving the mass resolution. In addition to m_T^{tot}, the following mass variables might be useful:

- **visible mass**. This is the invariant mass of the lepton and visible tau. This variable is used in event selection, to remove the $Z \to e^+e^-$ peak in the $e\tau_{\text{had-vis}}$ channel.

- **Missing Mass Calculator mass** [268]. The Missing Mass Calculator (MMC) is an algorithm for estimating the full invariant mass of the $\tau^+\tau^-$ pair by minimizing a likelihood function in the kinematically-allowed phase space region. The underlying requirement is that mutual orientations of the decay products, including neutrinos, are consistent with the mass and decay kinematics of a tau lepton. A scan is performed over the possible angles between the neutrinos and the visible decay products, with each solution weighted according to probability density functions derived from simulated tau lepton decays. The MMC comes in two versions: the original (MAXW) and a refactored version (MLM) with improvements for making the tool more comprehensible, faster, and easier to re-tune. For comparison purposes, MMC masses were calculated using both MMC versions in the xTauFramework.

- **collinear approximation mass** [274]. This mass calculation method relies on two assumptions: that neutrinos from tau lepton decays are collinear with the visible tau lepton decay products and that all of the missing energy in the event is due to neutrinos from tau lepton decays. These are reasonable assumptions to make if the $\tau^+\tau^-$ system is boosted. What follows from these assumptions is that the azimuthal angle ϕ of the vector sum of neutrino momenta and the angle of the missing energy is the same. With ν_{lep} referring to the electron or muon neutrino in the leptonic tau lepton decay, ν_{had} referring to a combination of the two neutrinos originating from the hadronic tau lepton decay (a tau neutrino and an electron or muon antineutrino, in the case of τ^- decays), and $\vec{p}_T = (p_x, p_y)$, the invariant mass of the

$\tau^+\tau^-$ pair can then be calculated as follows.

Starting from the assumption that missing energy is due to tau-decay neutrinos:

$$\vec{p}_T(\nu_{\text{lep}}) + \vec{p}_T(\nu_{\text{had}}) = \vec{E}_T^{\text{miss}}, \tag{8.1}$$

add the collinearity requirement:

$$a \cdot \vec{p}_T(\tau_{\text{lep}}) + b \cdot \vec{p}_T(\tau_{\text{had}}) = \vec{E}_T^{\text{miss}}. \tag{8.2}$$

This can be broken up into x and y components:

$$a \cdot p_x(\tau_{\text{lep}}) + b \cdot p_x(\tau_{\text{had}}) = E_x^{\text{miss}}, \tag{8.3}$$

$$a \cdot p_y(\tau_{\text{lep}}) + b \cdot p_y(\tau_{\text{had}}) = E_y^{\text{miss}}. \tag{8.4}$$

Solving for a and b, the components of neutrino three-momenta can be calculated:

$$\vec{p}(\nu_{\text{lep}}) = a \cdot \vec{p}(\tau_{\text{lep}}), \tag{8.5}$$

$$\vec{p}(\nu_{\text{had}}) = b \cdot \vec{p}(\tau_{\text{had}}). \tag{8.6}$$

Finally, the invariant mass of the full system can be calculated in the usual way:

$$m(\tau\tau) = \sqrt{[E(\tau_{\text{lep}}) + E(\tau_{\text{had}}) + |\vec{p}(\nu_{\text{lep}})| + |\vec{p}(\nu_{\text{had}})|]^2 - (\Sigma p_x)^2 - (\Sigma p_y)^2 - (\Sigma p_z)^2}, \tag{8.7}$$

where the sums are over τ_{lep}, τ_{had}, ν_{lep}, and ν_{had}. There is no solution in cases where visible tau leptons are exactly back-to-back.
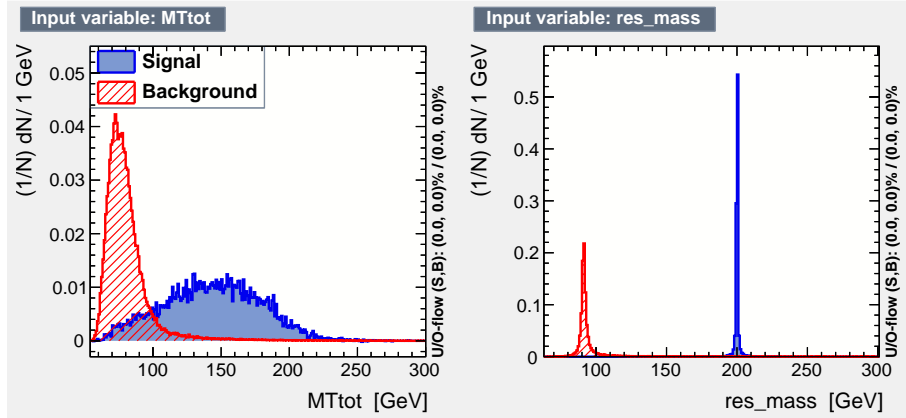

Figure 8.1: Distributions of m_T^{tot} (left) and the true resonance mass (right) for $Z/\gamma^* \to \tau^+\tau^-$ background and 200 GeV ggH/A and bbH/A signal.

Distributions of these mass variables for $Z/\gamma^* \to \tau^+\tau^-$, 200 GeV ggH/A, and 2500 GeV ggH/A samples are shown in Figures 8.2-8.3. No selection cuts were applied other than 'truth-matching' the tau leptons to make sure they came from Z and H/A boson decays rather than from other sources, like B meson decays. Events that did not have a solution for the collinear approximation mass calculation were assigned a value of 0 for this variable. For all samples, the different mass variables had quite different distributions and resolutions. Visible mass was consistently the mass variable with the lowest average value (due to not accounting for neutrinos) and the best resolution. The m_T^{tot} variable had a slightly higher average value - though still below the true resonance mass - and a slightly worse resolution. Collinear approximation mass and MMC MAXW mass both had long tails for $Z/\gamma^* \to \tau^+\tau^-$ and 200 GeV ggH/A samples. For some events, no solution was found for the MMC methods; This was especially true for 2500 GeV ggH/A signal. Table 8.1 lists the percentages of events with zero-valued MMC mass variables. The correlations between all five mass variables in $Z/\gamma^* \to \tau^+\tau^-$ and 200 GeV ggH/A events are shown in Figure 8.4. Events with zero-valued MMC were not included in the correlation matrix calculation. In both samples, the collinear approximation mass shows the least correlation with other mass variables, in the 29%-64% range. Other variables are correlated with each other at a 69%-77% level.

103
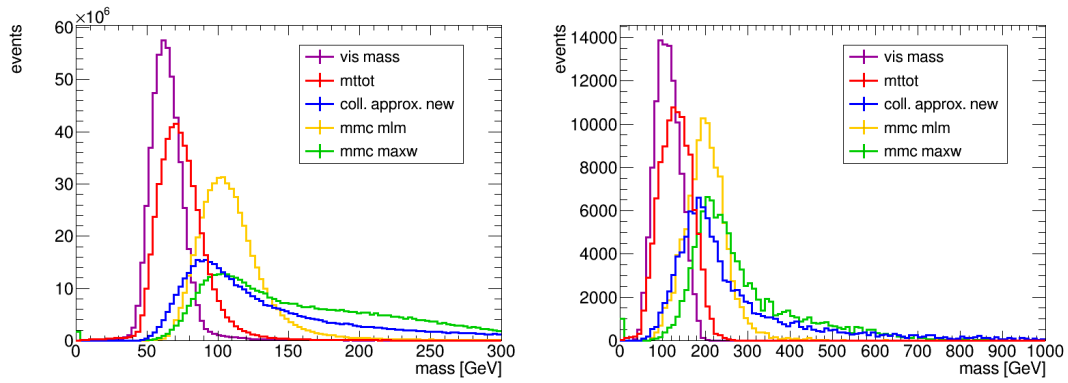
Figure 8.2: Distributions of various $\tau^+\tau^-$ mass variables in $Z/\gamma^* \to \tau^+\tau^-$ (left) and 200 GeV ggH/A (right) samples. No selection cuts have been applied other than truth-matching the tau leptons. Note that the x-axis ranges are different.

Figure 8.3: Distributions of various $\tau^+\tau^-$ mass variables in the 2500 GeV ggH/A sample, with (left) and without (right) the first bin containing zero-value events. No selection cuts have been applied other than truth-matching tau leptons.

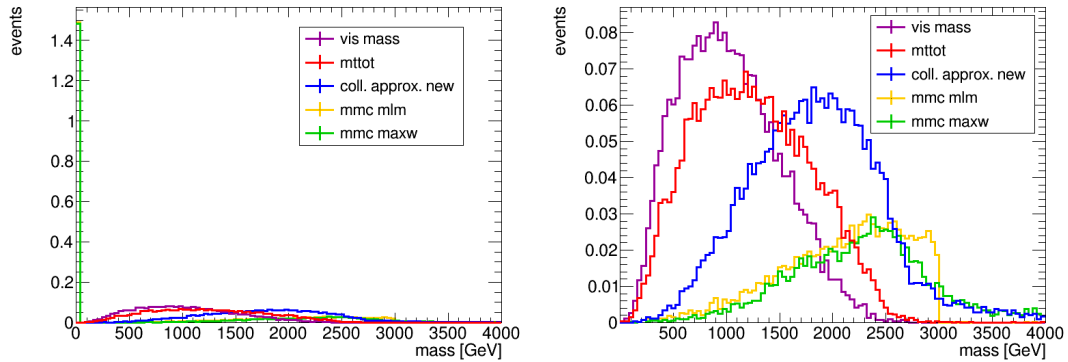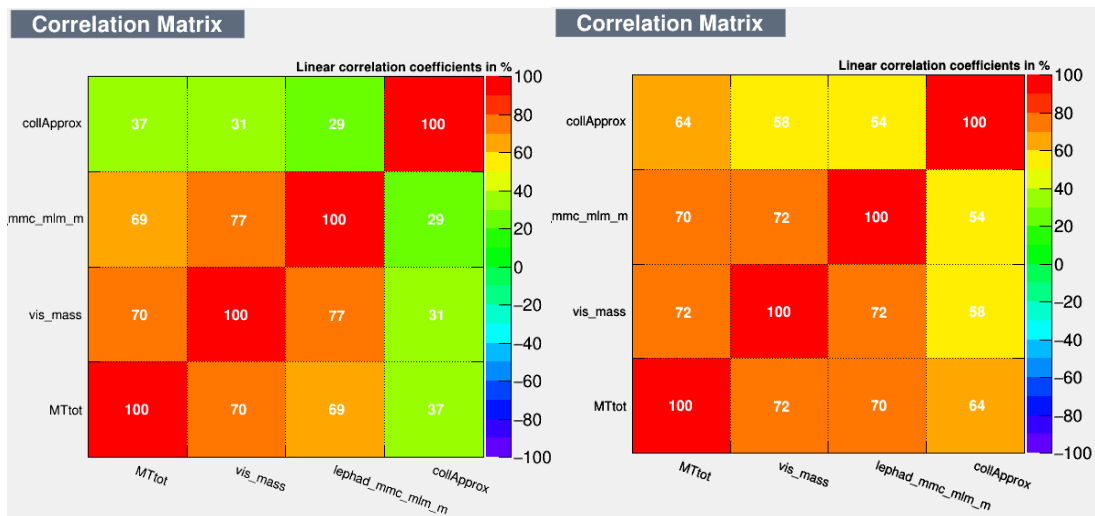Figure 8.4: Correlation matrices for the considered mass variables, calculated for $Z/\gamma^* \to \tau^+\tau^-$ (left) and 200 GeV ggH/A (right) samples.

Table 8.1: Percentages of events with zero-valued mass variables in the studied samples.

Sample	MMC MAXW = 0	MMC MLM = 0
$Z/\gamma^* \to \tau^+\tau^-$	0.4%	0.4%
200 GeV ggH/A	0.9%	0.9%
2500 GeV ggH/A	61.8%	62.0%

Figure 8.5: ROC AUC values of BDTs trained on $Z/\gamma^* \to \tau^+\tau^-$ background and 200 GeV $ggH/A+bbH/A$ signal in the b-veto category, using different mass variables.

8.2.2 Polarization variable

The other difference between Z and Higgs boson decays is due to different polarizations of the bosons. The Z boson is a vector boson with a polarization of -1, causing tau leptons from $Z \to \tau^+\tau^-$ decays to have a polarization of approximately -14.5%, as discussed in Section 6.3. The Higgs boson under consideration in this analysis is a scalar, lacking polarization. In this scenario, tau leptons from $H \to \tau^+\tau^-$ decays would also be unpolarized. One observable which can be used for determining the polarization of tau leptons is the visible momentum fraction of $\tau^\pm \to \pi^\pm\nu$ decays x_π, which was used in Monte Carlo sample validation in Section 6.3. However, as this observable can only be calculated for one tau lepton decay channel, this is not useful for event-wise model training and application. Another observable which can be more broadly used is the charged energy asymmetry Υ for 1-prong tau lepton decays [275]:

$$\Upsilon = \frac{E(\pi^\pm) - E(\pi^0)}{\pi^\pm + \pi^0} \approx 2\frac{p_T(\text{track})}{p_T(\tau_{\text{had-vis}})} - 1, \tag{8.8}$$

though this variable is still limited in that in can only be calculated (in a meaningful way) for 1-prong decays. In order to use this variable in a multivariate model, two approaches can be taken: either train separate models for 1-prong and 3-prong events (with Υ used as an input variable only in the former) or train one model which uses Υ for both types of events, with the understanding that it does not contribute meaningful information in the case of 3-prong events. The distribution of Υ for 1-prong events, after all selection cuts, is shown in Figure 8.6.

The usefulness of this variable was tested using only 1-prong $Z/\gamma^* \to \tau^+\tau^-$ and 200 GeV ggH/A events. By itself, Υ is a weakly discriminating variable, yielding a ROC AUC value of 0.573, compared to 0.939 for m_T^{tot}. A model trained on m_T^{tot} and Υ performed only marginally better than a model trained only on m_T^{tot}, with a ROC AUC value of 0.940.

8.2.3 Other variables and final variable list

The b-veto category also has small but significant contributions from reducible backgrounds: $Z/\gamma^* \to l^+l^-$, multijet background, and W+jets background. Contributions from diboson and $t/t\bar{t}$ backgrounds are very small. These backgrounds differ from signal more significantly. For
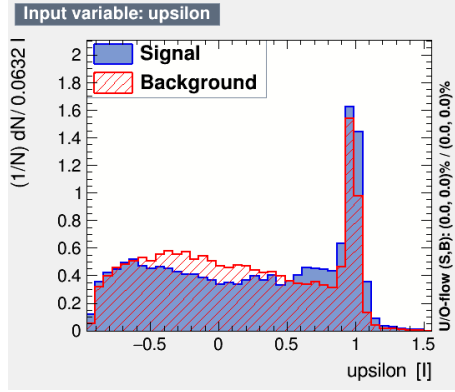
Figure 8.6: Distribution of the Υ variable for 200 GeV ggH/A signal and $Z/\gamma^* \to \tau^+\tau^-$ background 1-prong events after selection.

backgrounds with jets or light leptons misidentified as tau leptons ('fake background'), an improvement in performance can be gained from kinematic variables like missing transverse energy, $\tau_{\text{had-vis}}$ and light lepton three-momenta, and angles between the $\tau_{\text{had-vis}}$ object, the light lepton, and missing transverse energy. The use of the tau lepton identification score should in principle give a large performance boost, discriminating between true tau leptons and jets misidentified as tau leptons, but its use is difficult for two major reasons. Firstly, the data-driven fake estimation method used in the first analysis round uses a fail-tau-ID region, with events from this region propagated to the signal region with fake factor weights. Thus, the tau lepton identification score of jets misidentified as tau leptons is by design mismatched from what would be seen in real data in the signal region, as tau leptons in real data events would have to pass the medium tau lepton identification selection requirement. Likewise, variables related to the tau lepton identification score, like the number of tracks in core and isolation regions of $\tau_{\text{had-vis}}$ candidates, are biased because of the fake estimation method. A few approaches to get around these issues were investigated when using ntuples with BDT tau lepton identification; these studies are documented in Section 8.4. After the move to RNN tau lepton identification, which resulted in the number of jets misidentified as tau leptons falling by about 50%, these attempts were abandoned, as the improvement gained by using the tau lepton identification score would be marginal.

In principle, the same classifier performance can be obtained with different combinations of input variables carrying the same information, though model architectures might need to be fine-tuned (for example by using deeper neural networks when training with lower-level variables). For efficiency, the final input variable lists were designed in such a way as to maximize the ROC AUC (without overfitting) with the minimum number of variables. Variable ranking lists calculated by TMVA were used as a rough guide, though they fluctuated significantly with changes in model hyperparameters. Variables considered in the rankings included all mass variables mentioned above, Υ, three-vectors of the lepton and tau, the 2-vector of missing transverse energy, distances in η and ϕ between objects, and differences of p_T between the objects. The shortest input variable list which maximized the ROC AUC for each background contribution is as follows: m_T^{tot}, visible mass, MMC mass (it is sufficient to choose either MAXW or MLM), $p_T(\tau_{\text{had-vis}})$, $\eta(\tau_{\text{had-vis}})$, $p_T(\ell)$, and $\Delta\phi(\ell, \tau_{\text{had-vis}})$. Table 8.2 shows a breakdown of useful variables by background contribution. Figure 8.7 shows the changes in ROC AUC for each background contribution as these variables are added in order to model training; this figure also shows the ROC AUC for two methods of combining background contributions in one classifier, which will be discussed below. Based on both combinations, the final input variable list can be truncated to only include m_T^{tot}, visible mass, MMC mass, and $p_T(\tau_{\text{had-vis}})$.

The same input list can also be used for all signal mass points above 200 GeV, since the Higgs production process remains the same. However, for the highest masses m_T^{tot} is already an excellent discriminating variable, and it is unlikely that any non-negligible improvement can be obtained with a multivariate classifier. The ROC AUC obtained using only m_T^{tot} compared to using all b-veto input variables is the same: 0.999, as calculated by TMVA. To say conclusively whether any

difference in performance exists, the limit must be calculated.

	$Z/\gamma^* \to \tau^+\tau^-$	$Z/\gamma^* \to \ell^+\ell^-$	W+jets	multijet	combined
m_T^{tot}	✔	✔	✔	✔	✔
visible mass	✔	✔	✔	✔	✔
MMC mass	✔	✔	✔	✔	✔
$p_T(\tau_{\text{had-vis}})$	✔	✔	✔	✔	✔
$\eta(\tau_{\text{had-vis}})$		✔			
lep pt				✔	
$\Delta\phi(\ell, \tau_{\text{had-vis}})$		✔	✔	✔	

Table 8.2: Lists of variables which increased the ROC AUC when added in sequence for different background contributions in the b-veto category.
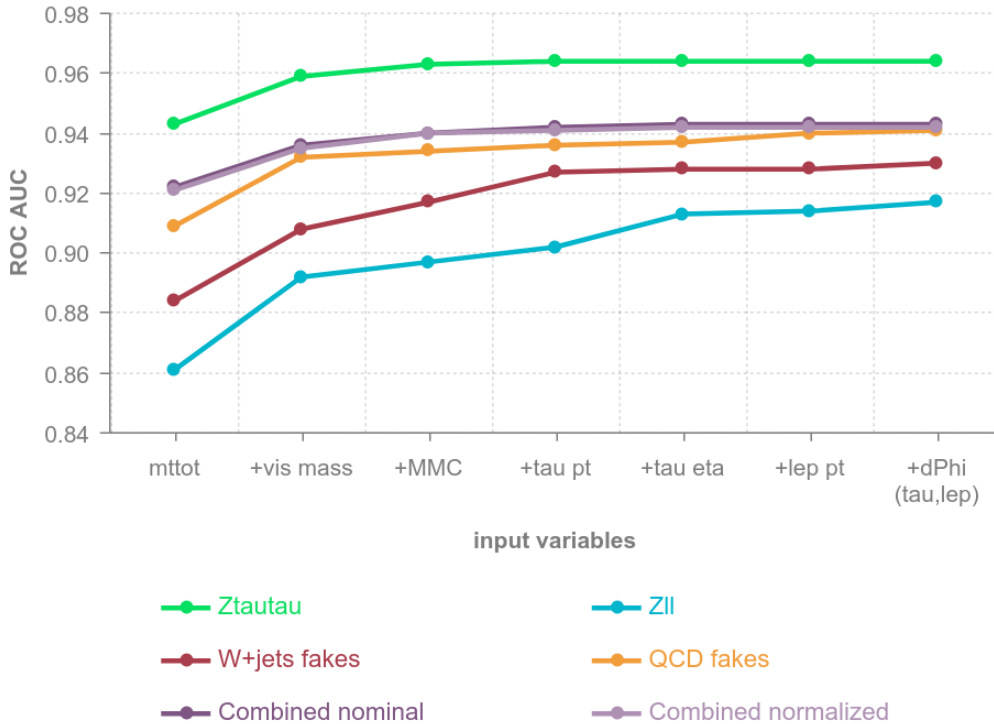

Figure 8.7: ROC AUC values of BDTs trained on different backgrounds and input variable lists in the b-veto category. The signal in each case was 200 GeV $bbH/A+ggH/A$. 'Combined nominal' refers to training and testing with all four backgrounds contributions, using nominal weights for all events. 'Combined normalized' refers to training with background where each of the four contributions was normalized to 1 using the sum of weights, and testing on background with nominal weights.

8.2.4   Combining background contributions

Ultimately, information about all the background contributions must be combined in one classifier. There are a few ways of going about this, including:

1. **training a binary classifier on nominally-scaled background contributions**; that is, use nominal weights (including fake factors) for all background events, so that the background composition in the training is the same as the background composition in the signal region. All background contributions form one background event class.

2. **training a binary classifier on normalized background contributions**; that is, reweight background so that each contribution is normalized to the same number of events in train-

ing. This helps the model to learn all the backgrounds, but it might not result in optimal performance on standard background composition.

3. **training a multi-class classifier with multiple labels**; ex. Z decays, fakes (jets misidentified as tau leptons), and signal. Instead of outputting one 'model score', the model would output a probability for each class. The distribution of one of these, the probability of events belonging to the signal class, can be used for limit setting. This can allow the model to learn different topologies for different background classes instead of 'averaging' over all background events. However, it is much slower to train than a binary classifier and might not result in an improved performance.

First, options 1 and 2 were compared. In the first case, BDT and NN models with a 50:50 train-test split were trained on b-veto 200 GeV ggH/A and bbH/A signal and $Z/\gamma^* \rightarrow \tau^+\tau^-$, $Z/\gamma^* \rightarrow \ell^+\ell^-$, multijet, and W+jets background, all with nominal weights. In the second case, no changes were made to signal weighting, but the four background contributions were all scaled to the same number of events. Final input variable lists described above were used in both cases. The models were checked for overfitting, and then applied to c16d signal and background with nominal weights. The ROC AUC values were very similar: 0.943 for option 1 and 0.942 for option 2. Therefore, option 1 was chosen for simplicity.

To check the viability of using option 3 - which is supported by TMVA but in a more limited capacity compared to binary classification - a simplified study was conducted. Two background classes were used: $Z/\gamma^* \rightarrow \tau^+\tau^-$ background and W+jets background. 200 GeV bbH/A and ggH/A signal samples were grouped in one signal category. The performance of the multiclass classifier was judged based on the signal-vs-rest ROC AUC calculated on validation events. The best achieved value when training a BDT with default background weights was 0.899, compared to 0.949 for a binary BDT classifier using the same input parameters and samples. In the case of a BDT trained and tested on background samples reweighted to the same number of events, the best ROC AUC was 0.916, compared to 0.934 for a comparable binary classifier. A multiclass MLP classifier performed even more poorly, achieving a maximum ROC AUC of 0.908 when testing and training on reweighted background samples. Additionally, multiclass classifiers almost always overfitted, even when using a simple architecture and few training epochs. For example, the multiclass BDT trained on events with nominal weights achieved a ROC AUC of 0.915 on training events - 0.016 higher than on testing events. In comparison, the corresponding binary BDT achieved the same ROC AUC for both training and testing events. The difference in performance between binary and multiclass classifiers is understandable, as different metrics are optimized during training. A binary classifier is trained to maximize the separation between signal and background (by maximizing the ROC AUC), but a multiclass classifier is trained to maximize the accuracy of identifying events from different classes - which, in general, will not lead to an optimal signal-vs-rest ROC AUC.

8.3 Choosing input variables in the b-tag category

No irreducible background is present in the b-tag category; therefore, a larger improvement is expected when using a multivariate classifier. Additionally, the presence of a b-jet in the final state provides a set of additional useful input variables, such as the b-jet three-momentum, the invariant mass of the $\tau^+\tau^-$-jet triple, and $\Delta\phi$, $\Delta\eta$, and Δp_T between the b-jet and other final-state objects. All of these variables, in addition to variables listed in Section 8.2, were used to find the optimum input variable list.

Similarly to the b-veto category, the impact of the input variables on the classifier performance were studied for the main background contributions: single t quark and $t\bar{t}$ background, W+jets, multijet, $Z/\gamma^* \rightarrow \tau^+\tau^-$, and $Z/\gamma^* \rightarrow \ell^+\ell^-$ background. The minimum list of input variables maximizing the ROC AUC for each background contribution - and the combined background - is as follows: m_T^{tot}, visible mass, MMC mass, $p_T(\tau_{\text{had-vis}})$, $\Delta\phi(\ell, \tau_{\text{had-vis}})$, $p_T(b$-jet$)$, $\Delta\eta(b$-jet$, \ell)$, and $\Delta\eta(b$-jet$, \tau_{\text{had-vis}})$. The breakdown of useful input variables by background contribution is shown in Table 8.3, and the ROC AUC values for different sets of input variables are plotted in Figure 8.8.
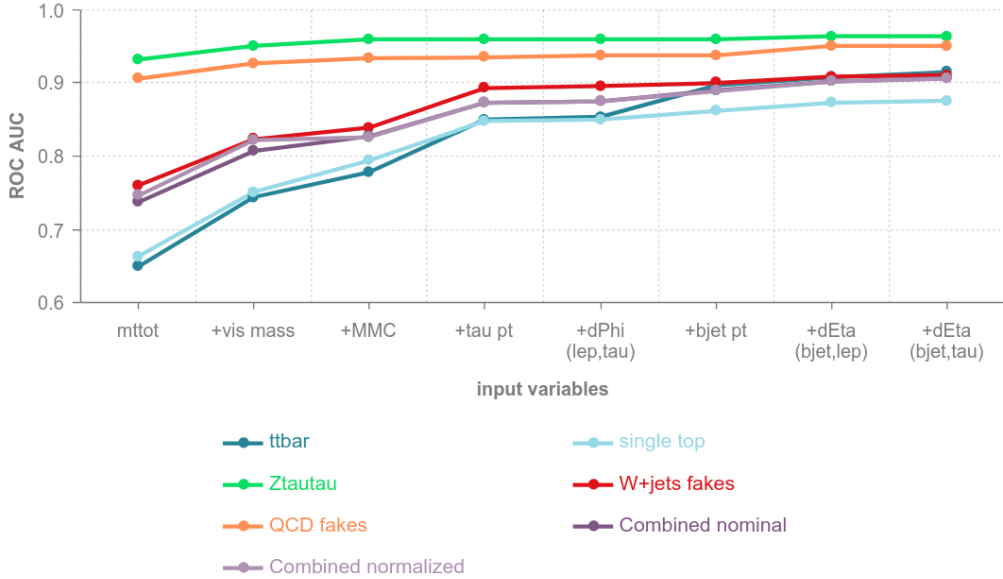
Figure 8.8: ROC AUC values of BDT models trained on different backgrounds and input variable lists in the *b*-tag category. The signal in each case was 200 GeV $bbH/A+ggH/A$. 'Combined nominal' refers to training and testing with all four backgrounds contributions, using nominal weights for all events. 'Combined normalized' refers to training with background where each of the four contributions was normalized to 1 using the sum of weights, and testing on background with nominal weights.

	$t\bar{t}$	single t	$Z/\gamma^* \to \tau^+\tau^-$	W+jets	multijet	combined
m_T^{tot}	✔	✔	✔	✔	✔	✔
visible mass	✔	✔	✔	✔	✔	✔
MMC mass	✔	✔	✔	✔	✔	✔
$p_T(\tau_{\mathrm{had\text{-}vis}})$	✔	✔		✔		✔
$\Delta\phi(\ell, \tau_{\mathrm{had\text{-}vis}})$	✔	✔		✔		✔
$p_T(b\text{-jet})$	✔	✔		✔		✔
$\Delta\eta(b\text{-jet}, \ell)$	✔	✔	✔	✔	✔	✔
$\Delta\eta(b\text{-jet}, \tau_{\mathrm{had\text{-}vis}})$	✔	✔		✔		✔

Table 8.3: Lists of variables which increased the ROC AUC when added in sequence for different background contributions in the *b*-tag category.

For completeness, the three approaches to background composition used in training were studied also in the *b*-tag category. The results were largely the same as in the *b*-veto category. The binary classifier outperformed the multiclass classifier regardless of whether the background contributions were normalized or not.

While at the high end of the signal mass spectrum m_T^{tot} is still an excellent discriminating variable (as in the *b*-veto category), some more improvement might be obtained here with a multivariate discriminant. This is due to two reasons. For one, the $t/t\bar{t}$ background is broader than $Z/\gamma^* \to \tau^+\tau^-$ background, so there is greater overlap between signal and background than in the *b*-veto category. For another, in the *b*-tag category we can go beyond improving mass resolution by using kinematic variables, so even events that overlap in the mass distribution can be discriminated.

8.4   Pseudo-continuous tau lepton identification study

As discussed in Section 8.2, using the tau lepton identification score would be useful for discriminating between true tau leptons and jets misidentified as tau leptons, but is technically challenging. The simplest way to obtain a correctly-modeled tau lepton identification score distribution is to

use the existing working points and define a few disjoint categories: medium-not-tight, tight, and - if the selection cuts are loosened - loose-not-medium. Fake estimation can then be done using the standard method, but with a separate fake factor calculated for each category. Fake events would still be propagated from fail-ID regions, and the tau lepton identification value would simply be reassigned to match the pass-ID region: loose for the loose-not-medium background component, medium for the medium-not-tight background component, and tight for the tight background component.
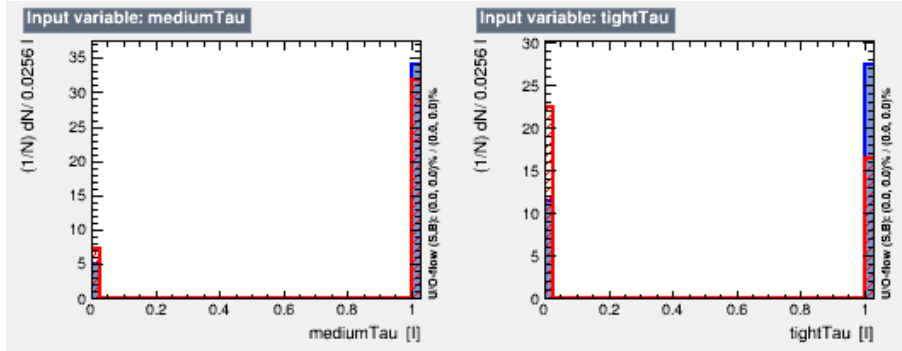

Figure 8.9: The proportions of W+jets background (red) and ggH/A 200 GeV signal (blue) events passing at least medium (left) and at least tight (right) tau lepton identification in the signal region, where the minimum tau lepton identification requirement has been changed from medium to loose. The distributions have been normalized to 1.

Figure 8.9 shows the distributions of the number of W+jets fake and ggH/A 200 GeV events passing and failing the medium and tight tau lepton identification requirements in the b-veto category after estimating background with jets misidentified as tau leptons using the method described above. Selection cuts are nominal with the exception of the tau lepton identification cut, which was changed from medium to loose. As expected, signal events tend to have higher tau lepton identification scores than events with jets misidentified as tau leptons, which is particularly visible for the tight tau lepton identification distribution. The ROC AUC obtained after adding medium and tight tau lepton identification variables to the training was 0.884, compared to 0.867 when training without these variables. This is a significant improvement for the background contribution with jets misidentified as tau leptons; however, when all the backgrounds are combined, the change in the ROC AUC value is negligible.

A better improvement could be obtained when using the continuous tau lepton identification score distribution rather than discrete working points. The challenge with this, however, is that tau lepton identification scores cannot be easily 'reassigned' in the same way that working points can. Instead, a different fake estimation method is required - one that does not use fail-ID region. One possibility is using the same-sign region. To test the feasibility of this, SS→OS fake factors were calculated in the W-FR (with a medium tau lepton identification cut). As for standard fail-tauID→pass-tauID fake factors, the SS→OS fake factors were binned in $p_T(\tau_{\text{had-vis}})$, and a scale factor was derived to fix the mismodeling seen in the $p_T(\ell)$ distribution. The modeling of a few key variables was checked in the W-FR after applying the scale factor. Unfortunately, significant mismodeling was observed in many variable distributions, likely due to the different compositions of events with jets misidentified as tau leptons in the OS and SS regions. In true W+jets events, the $W^{\pm}$ boson and the leading quark-initiated jet tend to have opposite signs. In the SS region, quark-initiated jets can originate from gluon splitting ($g \to q\bar{q}$), produced independently from the W boson.

Due to the relatively small size of the background contribution from events with jets misidentified as tau leptons, especially after the switch to RNN tau lepton identification, studies on pseudo-continuous and continuous tau lepton identification use were discontinued at this stage.

8.5 Training a mass-parametrized neural network

An alternative approach to training one machine learning model per signal mass point was developed in [276]. In this approach, one neural network is trained on all signal samples simultaneously, with the signal mass (m_{signal}) included as an additional input parameter. Signal mass values must also be assigned to background events in such a way as to not bias the framework - in other words, the m_{signal} distributions in signal and background events used in training must be the same. This can be achieved in two ways:

- assigning a randomly-selected m_{signal} value to each background event

- using N copies of each background event in the training, where N is the number of signal mass points, with each copy assigned a different m_{signal} value.

In both approaches, the m_{signal} distribution can be either flat (by normalizing each mass point to the same value) or non-flat (typically by using nominal weight values). With a non-flat m_{signal} distribution, care must be taken to use the correct probability distribution for choosing random m_{signal} values or to reweight background copies by appropriate values. The second approach, using N copies of background, is generally preferable to the first, as the background distribution remains the same for each signal hypothesis. This is especially important in the high-mass tail, where there are fewer background events. A drawback is that using such a large number of training events can max out available computer memory. When applying the model to a specific signal hypothesis, the user must specify one m_{signal} value for the signal and background; this is called tuning the network.

An advantage of parametrized neural networks (PNNs) is that they are able to interpolate between mass points that were used in training; that is, they can be tuned to m_{signal} values that were not previously seen by the network. Models trained on single mass points, on the other hand, typically only have good sensitivity in a small region around the mass point used in training. This can be verified by dropping an available mass point from training and then checking the performance of the trained network on this mass point. Another advantage is the convenience of having a single model rather than a set of individually-trained BDTs or neural networks. On the other hand, PNNs take a long time to train, typically orders of magnitude longer than a BDT, making them difficult to optimize.

In this study, PNNs were trained using the entire mass range in the b-veto and b-tag categories, using the same input parameters as the ones listed in Sections 8.2 and 8.3, with the addition of the m_{signal} variable. The following mass points were used in training: {200, 250, 300, 350, 400, 500, 600, 700, 800, 1200, 1500, 2000, 2500} GeV. The large gap between the 800 GeV and 1200 GeV mass points is due to a missing sample: the 1000 GeV ggH/A signal sample was unavailable (at xAOD level) when these studies were conducted. All mass points were normalized to the same number of events, resulting in a flat m_{signal} distribution.

Both approaches to assigning m_{signal} values to background events were compared in the b-veto category with 200 GeV ggH/A signal. The background used for this check was $Z/\gamma^* \to \tau^+\tau^-$ and W+jets background, and all background contributions (including diboson and t quark events) were used in evaluation. Performance was compared by calculating the expected 95% confidence level exclusion limit, as described in Section 8.1, to get a more accurate sense of the difference between the two approaches than would be possible with ROC AUC values only. The PNN with N copies of the background used in training performed better, with an upper limit of $0.119^{+0.110}_{-0.031}$ pb compared to $0.141^{+0.131}_{-0.036}$ pb for a PNN with a single copy of the background. Therefore, training with N copies of background was used for all further studies.

To check the interpolation between mass points used in training, two PNNs were trained in the b-veto category: the first one was trained on the full mass range, and the second one on the full mass range without the 500 GeV mass point. Both PNNs were subsequently tuned to 500 GeV and ROC AUC values were calculated: 0.978 for the first PNN and 0.974 for the second PNN. While these values are similar, there is a certain drop in performance when the PNN is asked to interpolate.

The goal with the PNN approach is obtaining a performance as least as good as that obtained for a set of BDTs. Optimizing PNN hyperparameters to reach this goal proved to be difficult due to the time needed to train a single PNN and to a certain instability - small changes in hyperparameters sometimes produced large changes in model performance - in the TMVA toolkit. The b-veto category proved to be especially troublesome, with poor PNN performance typically seen for the higher mass points. In this mass range, little to no improvement over the standard m_T^{tot} final discriminating variable was expected, but including mass points from this range in training seemed to produce some unwelcome effects. A number of attempts and checks were conducted to fix the performance, including changing the input variable list, dropping some lower mass points from training to obtain a more even mass point spacing, and using various non-flat m_{signal} distributions (ex. ones with higher mass points normalized to higher values). Nevertheless, some drop in performance was always observed. The 'final' b-veto PNN model discussed in the next section is the best-performing model trained using the nominal input variable list with a flat m_{signal} distribution with all points included in training.

8.6   Model performance comparison

The performance of the PNNs was compared to the performance of BDTs and the standard m_T^{tot} method by calculating expected 95% exclusion limits, with statistical uncertainties only, for each mass point. The final PNN models for the b-veto and b-tag categories were trained using the full 200-2500 GeV mass range. Both PNNs had the same architecture: one input layer with N neurons, where N was the number of input variables (5 for b-veto and 9 for b-tag), one hidden layer with $N + 1$ neurons using the tanh activation function, and one output layer with 1 neuron using the softmax activation function. Variables were transformed using TMVA normalization (`VarTransform=N`). The number of training epochs was set at 200 for all PNNs. The final set of BDTs was created by training two gradient-boosted BDTs for each mass point: one for the b-veto category and one for the b-tag category. The BDTs were typically composed of 80 trees with a minimum node size of 2.5% and maximum depth of 4. Each model was checked for overfitting before calculating the limits. Histogram binning for limit calculation was done using the algorithm described in Section 8.1 requiring $N = 10$ background events and $M = 5\%$ of signal events per bin.

In the b-tag category, Figures 8.10-8.12, PNN and BDT models outperform m_T^{tot} across the entire mass range, with the largest difference for the lowest mass points. For the 200 GeV mass hypothesis, the limit obtained with the PNN is 66.3% lower than the limit obtained with the m_T^{tot} final discriminant for ggH/A signal and 56.2% lower for the bbH/A signal. The corresponding percentages for the 200 GeV BDT compared to m_T^{tot} are 67.5% and 62.4%. For the 2500 GeV mass hypothesis, the improvement obtained with the PNN compared to m_T^{tot} is 2.8% for ggH/A signal and 3.1% for bbH/A signal. The corresponding percentages for the 2500 GeV BDT compared to m_T^{tot} are 12.5% and 10.7%. PNN and BDT performance is compared in Figure 8.12. BDTs outperform the PNN across the entire mass range for both ggH/A and bbH/A signal, with differences between scores reaching up to 15%. These differences are small compared to the $\pm 1\sigma$ uncertainty band.

In the b-veto category, Figures 8.13-8.15, the improvement obtained with machine learning methods compared to m_T^{tot} was not as dramatic; this was due to a larger percentage of irreducible background. The PNN outperformed m_T^{tot} only in the 200-500 GeV range, with improvements of 22.9% for 200 GeV ggH/A signal and 22.5% for 200 GeV bbH/A signal. Above 500 GeV the PNN underperformed compared to m_T^{tot}, with discrepancies up to 42% for the 1200 GeV mass point. The BDTs yield an improvement up to around 800 GeV. For the 200 GeV mass point, the improvement over m_T^{tot} was 28.6% for ggH/A signal and 28.5% for bbH/A signal. Past 800 GeV, the BDTs performed on par with or worse than m_T^{tot}. This issue was not picked up during validation with TMVA (prior to calculating limits), as the ROC AUC values for high mass points were very close to 1. While it should be possible to find BDT hyperparameters to match m_T^{tot} performance for all mass points, it seems unlikely that any significant improvement could be obtained. In this case, using m_T^{tot} rather than a machine learning model for the highest mass points would be more efficient. In the case of the PNN, it could be better to retrain the model using only the 200-500 GeV range, or attempt to match BDT performance using the 200-800 GeV range.
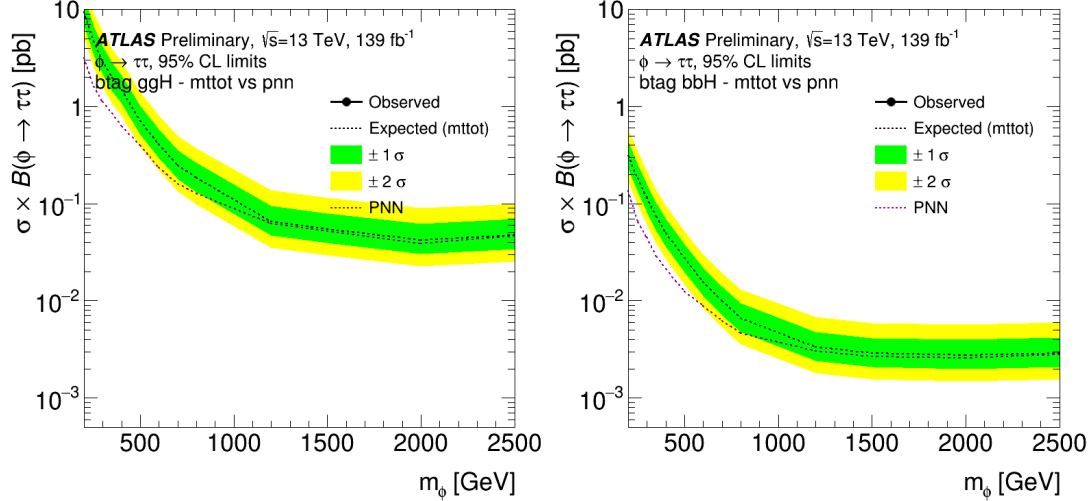
Figure 8.10: Expected 95% confidence level exclusion limit comparison for m_T^{tot} and PNN discriminants in the b-tag category. The limits are on the BSM Higgs boson production cross-section times $H/A \to \tau^+\tau^-$ branching ratio for gluon-gluon fusion (left) and b-associated production (right) Higgs boson production modes, using only $\tau_{\text{lep}}\tau_{\text{had}}$ events. The black dashed line and the green and yellow bands correspond respectively to the expected limit and $\pm 1\sigma$, $\pm 2\sigma$ uncertainty calculated using the m_T^{tot} final discriminant. The red dashed line corresponds to the expected limit calculated using the b-tag PNN tuned to each mass point; the uncertainty bands for the PNN are not shown. Only statistical uncertainty is included.

8.7 Modeling of variable distributions

While machine learning models do not introduce additional systematic uncertainties, it is necessary to check that all input variable distributions are correctly modeled in order to avoid bias when applying the models to real data. Validation-region plots of all input variables for the b-tag and b-veto models are shown in Figures 8.16-8.17 and 8.18, respectively. Only statistical uncertainties are included, as systematic variations were not yet available in the analysis workflow at the time of writing this thesis. In both the b-tag and the b-veto categories the number of data events is a few percentage points higher than the number of estimated background events. This points towards a possible problem with the data-driven background component. This issue was not investigated, as a new, improved data-driven background estimation method was being developed in parallel (but had not been finished by the time of writing this thesis).

In both the b-tag and the b-veto categories, a slope is seen in the ratio of data to estimated background for low values of mass variables. This is similar to what was observed during the last analysis round. In the b-tag category, this type of slope is also seen in $p_T(b\text{-jet})$ and $p_T(\tau_{\text{had-vis}})$ distributions, while $\Delta\eta(b\text{-jet}, \tau_{\text{had-vis}}/\ell)$ variables show some mismodeling at values around -2. As in the previous analysis round, this amount of mismodeling should be covered by systematic uncertainties; alternatively, it might be resolved with the improved data-driven background estimation method. In addition to comparing the distributions of input variables, the machine learning model response distributions were also compared in the validation region (Figures 8.19-8.20). Interestingly, the ratio of data to estimated background is quite flat for both PNN and BDT classifiers in both the b-tag and b-veto categories, and discrepancies should be easily covered by systematic uncertainties.

8.8 Conclusion

Both machine learning approaches - training a tunable PNN on all Higgs boson mass hypotheses and training a BDT for each mass hypothesis - resulted in an increase in analysis sensitivity compared to the standard m_T^{tot} final discriminant for the entire mass range of the b-tag category and the low mass range of the b-veto category. The improvement in sensitivity, measured by calculating the expected
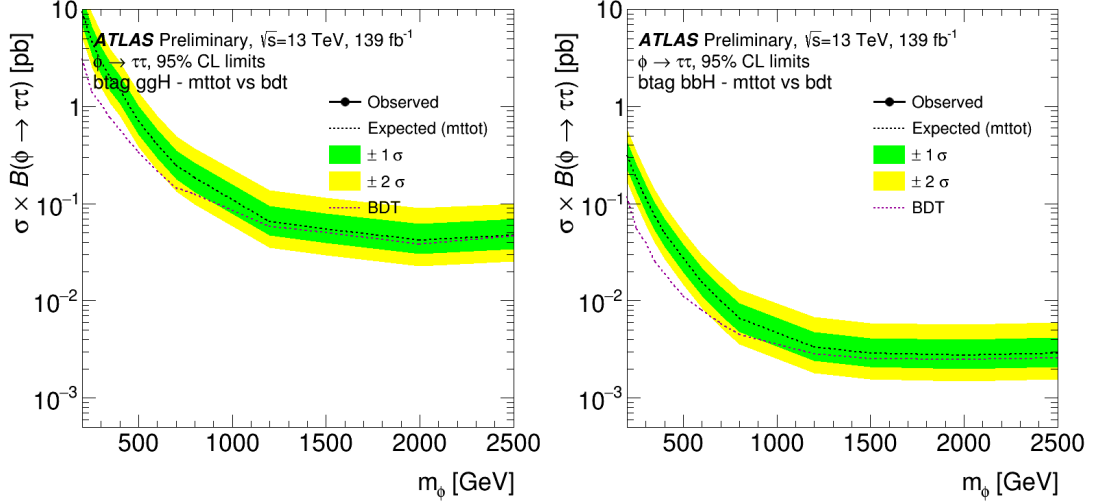
Figure 8.11: Expected 95% confidence level exclusion limit comparison for m_T^{tot} and BDT discriminants in the b-tag category. The limits are on the BSM Higgs boson production cross-section times $H/A \rightarrow \tau^+\tau^-$ branching ratio for gluon-gluon fusion (left) and b-associated production (right) Higgs boson production modes, using only $\tau_{lep}\tau_{had}$ events. The black dashed line and the green and yellow bands correspond respectively to the expected limit and $\pm1\sigma$, $\pm2\sigma$ uncertainty calculated using the m_T^{tot} final discriminant. The red dashed line corresponds to the expected limit calculated using the b-tag BDT set; the uncertainty bands for the BDTs are not shown. Only statistical uncertainty is included.

95% confidence level exclusion limits, was as high as 67.5% for the 200 GeV mass hypothesis in the b-tag category and 28.6% for the 200 GeV mass hypothesis in the b-veto category. The best performance was obtained with BDTs, which were much faster to train and easier to optimize than a PNN. Improved PNN performance might be obtained with more advanced machine learning frameworks instead of the TMVA toolkit, though a cost-benefit analysis should be carried out to determine whether the time investment would be worth the (possibly quite modest) increase in performance.
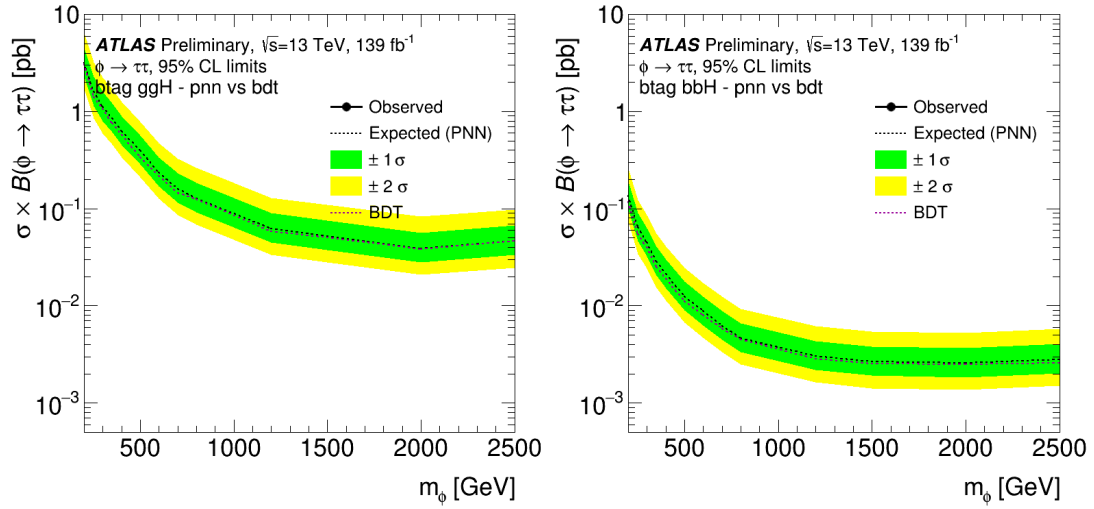
Figure 8.12: Expected 95% confidence level exclusion limit comparison for PNN and BDT discriminants in the b-tag category. The limits are on the BSM Higgs boson production cross-section times $H/A \to \tau^+\tau^-$ branching ratio for gluon-gluon fusion (left) and b-associated production (right) Higgs boson production modes, using only $\tau_{\text{lep}}\tau_{\text{had}}$ events. The black dashed line and the green and yellow bands correspond respectively to the expected limit and $\pm 1\sigma$, $\pm 2\sigma$ uncertainty calculated using the PNN final discriminant tuned to each mass point. The red dashed line corresponds to the expected limit calculated using the b-tag BDT set; the uncertainty bands for the BDTs are not shown. Only statistical uncertainty is included.
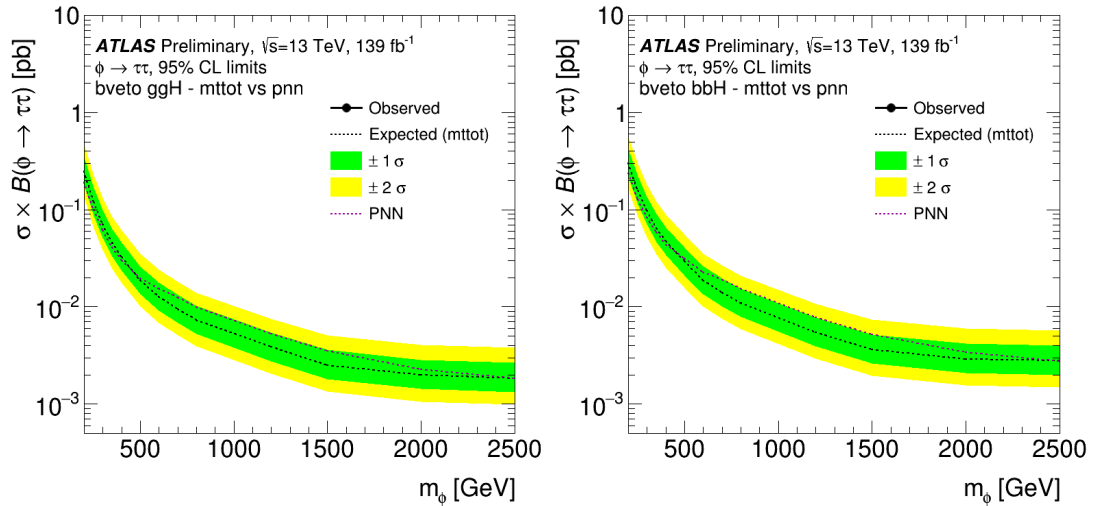

Figure 8.13: Expected 95% confidence level exclusion limit comparison for m_T^{tot} and PNN discriminants in the b-veto category. The limits are on the BSM Higgs boson production cross-section times $H/A \to \tau^+\tau^-$ branching ratio for gluon-gluon fusion (left) and b-associated production (right) Higgs boson production modes, using only $\tau_{\text{lep}}\tau_{\text{had}}$ events. The black dashed line and the green and yellow bands correspond respectively to the expected limit and $\pm 1\sigma$, $\pm 2\sigma$ uncertainty calculated using the m_T^{tot} final discriminant. The red dashed line corresponds to the expected limit calculated using the b-veto PNN tuned to each mass point; the uncertainty bands for the PNN are not shown. Only statistical uncertainty is included.
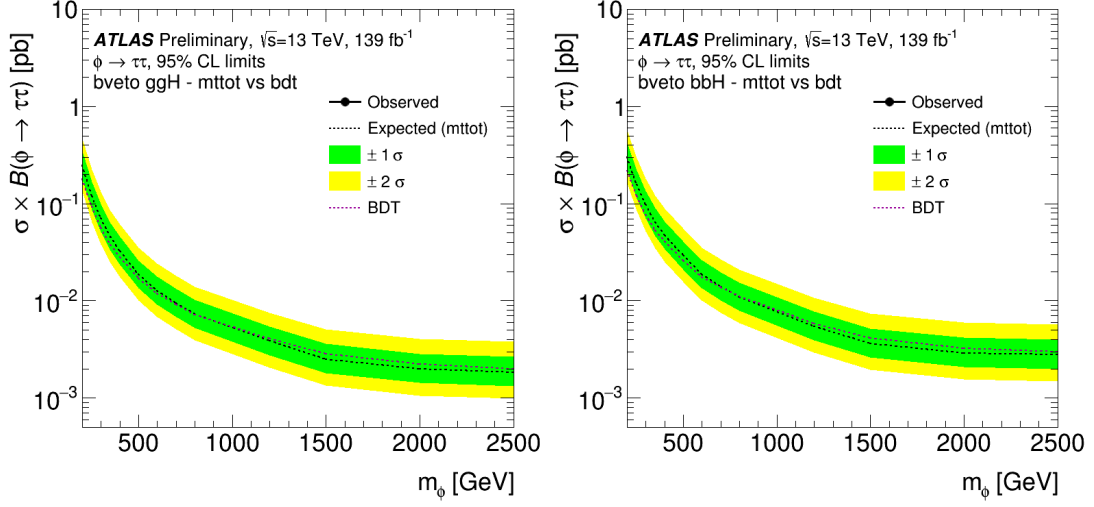
Figure 8.14: Expected 95% confidence level exclusion limit comparison for m_T^{tot} and BDT discriminants in the b-veto category. The limits are on the BSM Higgs boson production cross-section times $H/A \to \tau^+\tau^-$ branching ratio for gluon-gluon fusion (left) and b-associated production (right) Higgs boson production modes, using only $\tau_{\text{lep}}\tau_{\text{had}}$ events. The black dashed line and the green and yellow bands correspond respectively to the expected limit and $\pm 1\sigma$, $\pm 2\sigma$ uncertainty calculated using the m_T^{tot} final discriminant. The red dashed line corresponds to the expected limit calculated using the b-veto BDT set; the uncertainty bands for the BDTs are not shown. Only statistical uncertainty is included.

Figure 8.15: Expected 95% confidence level exclusion limit comparison for PNN and BDT discriminants in the b-veto category. The limits are on the BSM Higgs boson production cross-section times $H/A \to \tau^+\tau^-$ branching ratio for gluon-gluon fusion (left) and b-associated production (right) Higgs boson production modes, using only $\tau_{\text{lep}}\tau_{\text{had}}$ events. The black dashed line and the green and yellow bands correspond respectively to the expected limit and $\pm 1\sigma$, $\pm 2\sigma$ uncertainty calculated using the PNN final discriminant. The red dashed line corresponds to the expected limit calculated using the b-veto BDT set; the uncertainty bands for the BDTs are not shown. Only statistical uncertainty is included.
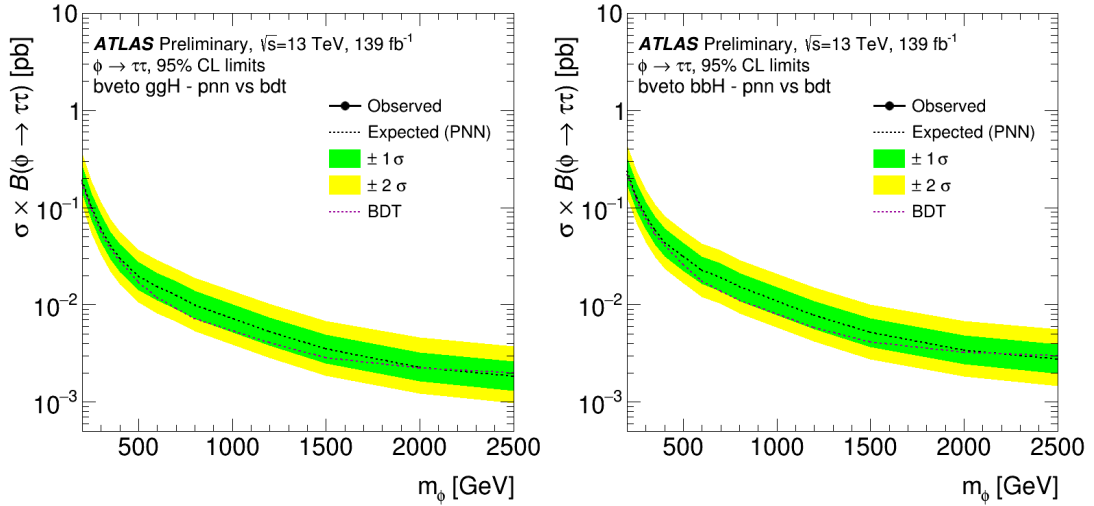
Figure 8.16: *b*-tag validation region distributions of the following input variables: m_T^{tot} (top left), m_{vis} (top right), MMC mass (center left), $p_T(\tau_{\text{had-vis}})$ (center right), $\Delta\phi(\ell, \tau_{\text{had-vis}})$ (bottom left), $p_T(b\text{-jet})$ (bottom right). Only statistical uncertainty is shown.
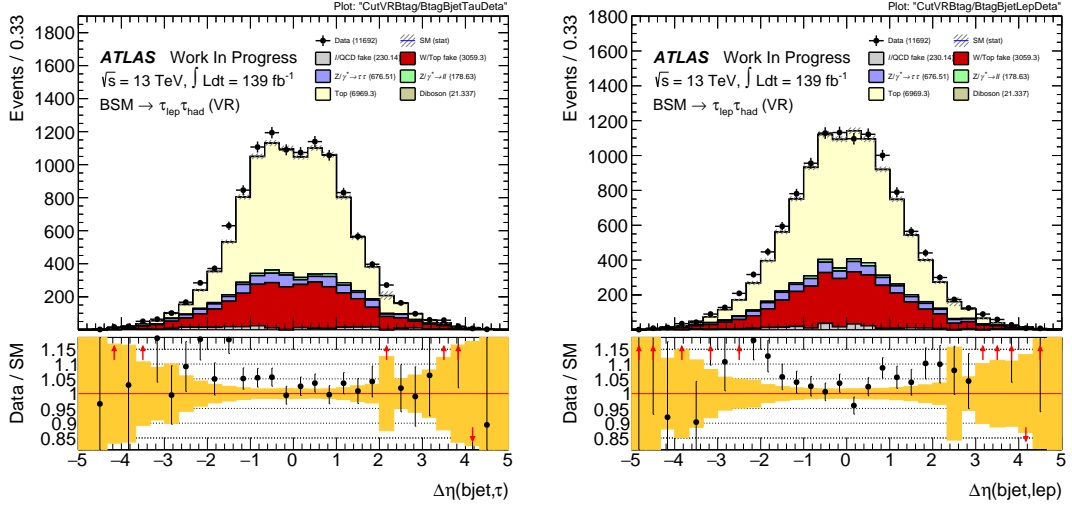
Figure 8.17: *b*-tag validation region distributions of the following input variables: $\Delta\eta(b\text{-jet}, \ell)$ (left) and $\Delta\eta(b\text{-jet}, \tau_{\text{had-vis}})$ (right). Only statistical uncertainty is shown.
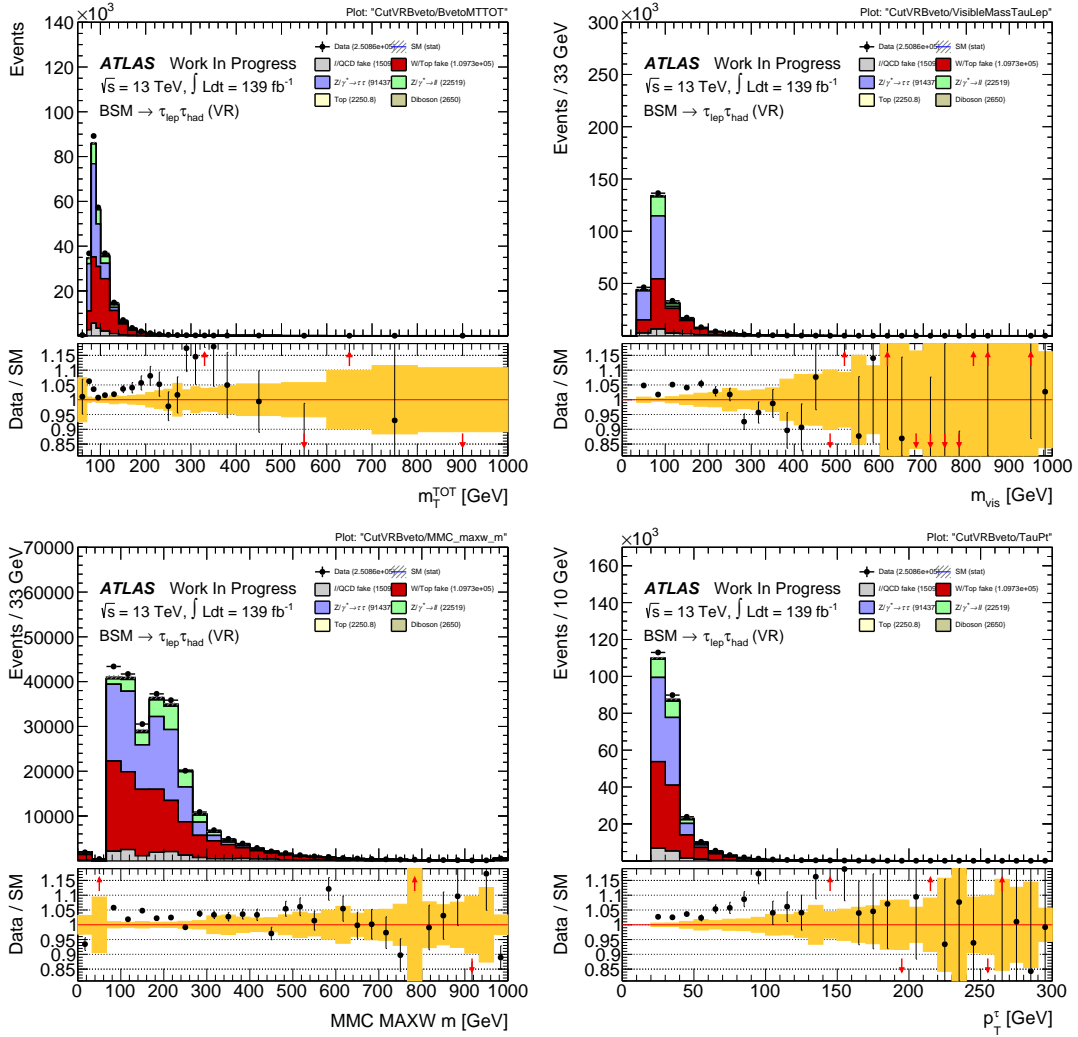
Figure 8.18: *b*-veto validation region distributions of the following input variables: m_T^{tot} (top left), m_{vis} (top right), MMC mass (bottom left), $p_T(\tau_{\text{had-vis}})$ (bottom right). Only statistical uncertainty is shown.
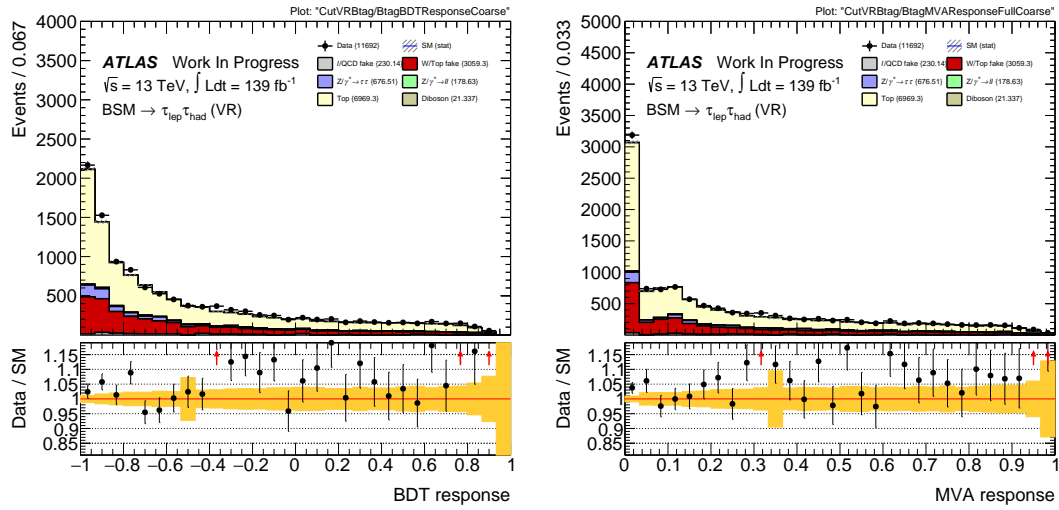
Figure 8.19: *b*-tag validation region distributions of 200 GeV *b*-tag BDT (left) and PNN (right) model scores; the BDT was trained on 200 GeV signal only, while the PNN was trained on all signal hypotheses and tuned to 200 GeV. Only statistical uncertainty is shown.
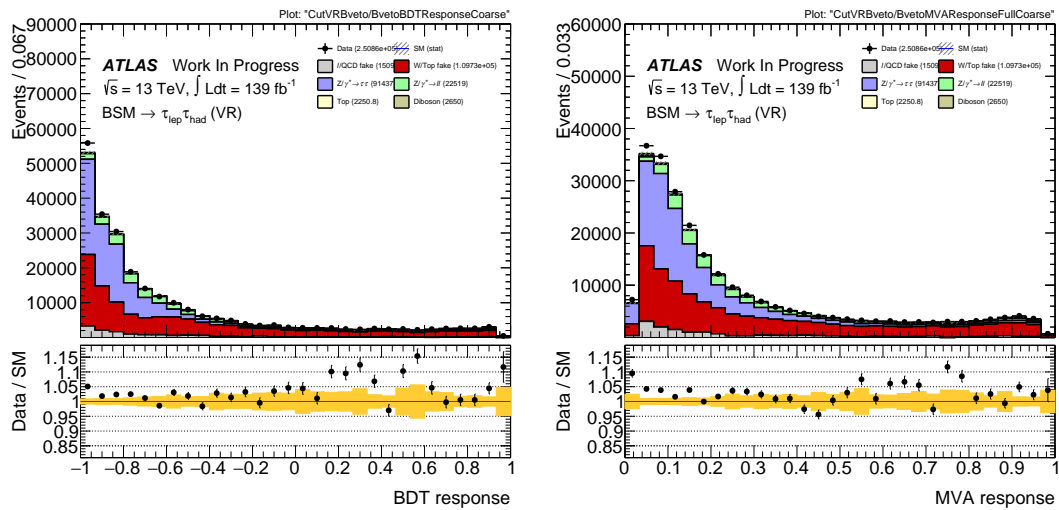
Figure 8.20: *b*-veto validation region distributions of 200 GeV *b*-veto BDT (left) and PNN (right) model scores; the BDT was trained on 200 GeV signal only, while the PNN was trained on all signal hypotheses and tuned to 200 GeV. Only statistical uncertainty is shown.

Chapter 9

Summary

The focus of this thesis was on additional Higgs bosons predicted by Two-Higgs-Doublet Models (2HDMs), a family of simple extensions of the Standard Model. Of particular interest is the type-II 2HDM, which corresponds to the Higgs sector of the Minimal Supersymmetric Standard Model (MSSM). The discovery of supersymmetry would be a big breakthrough in particle physics, cleanly solving the Higgs mass fine-tuning problem and providing a dark matter candidate, among other benefits. Even without supersymmetry, the presence of additional Higgs bosons could help cancel out bosonic and fermionic quantum fluctuations responsible for the Higgs fine-tuning problem and the cosmological constant problem, as was explored in Chapter 5. Two chapters of this thesis, Chapters 7 and 8, were dedicated to the search for additional heavy neutral Higgs bosons decaying to two tau leptons at the ATLAS detector, part of the LHC complex at CERN. The results were interpreted in two MSSM benchmark scenarios. While no trace of a new boson was found, the analysis was able to set the most stringent constraints on the MSSM phase space to date.

Tau leptons, featuring in the final state of the ATLAS search described here, are interesting particles in their own right, so much so that they warranted their own chapter. Chapter 6 was dedicated exclusively to validation studies of ATLAS Monte Carlo samples with these complex leptons. With a relatively short lifetime and large mass, their myriad decay modes allow for glimpses into the nature of the electroweak sector, by allowing measurements of tau lepton polarization and, consequently, of the weak mixing angle. At the same time, the abundance of tau lepton decay modes makes their reconstruction and identification in the ATLAS detector quite a challenge. This challenge, like many others in experimental particle physics and beyond, is being tackled with machine learning methods, with a large improvement seen in the recent upgrade from a boosted decision tree classifier to a recurrent neural network (deep learning) model, described in the Background part in Chapter 4. The potential of machine learning was also exploited by the author in the hopes of constructing a classifier capable of differentiating between signal and background events in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel of the ATLAS $H/A \to \tau^+\tau^-$ search and of outperforming the nominal final discriminating variable. This endeavor was successful, yielding improvements of up to 67.5% and 28.6% in the exclusion limits calculated for low mass points in the two analysis categories under consideration: final states with and without b-jets, respectively.

As the phase space of Two-Higgs-Doublet Models continues to be constrained by searches and precision measurements, the realization of supersymmetry in Nature seems increasingly unlikely. However, is the author's hope that exciting new discoveries and insights into physics beyond the Standard Model are just around the corner, and the field will continue to flourish for many years to come.

Bibliography

[1] S. Bass and J. Krzysiak, "Vacuum energy with mass generation and Higgs bosons," *Physics Letters B*, vol. 803, p. 135351, 2020. doi: 10.1016/j.physletb.2020.135351

[2] S. Bass and J. Krzysiak, "The cosmological constant and Higgs mass with emergent gauge symmetries," *Acta Physica Polonica B*, vol. 51, p. 1251, 2020. doi: 10.5506/APhysPolB.51.1251

[3] ATLAS Collaboration, "Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at $\sqrt{s}$ =13 TeV," *Physical Review Letters*, vol. 125, 2020. doi: 10.1103/PhysRevLett.125.051801

[4] J. Krzysiak on behalf of the ATLAS Collaboration, "Searching for BSM Higgs and gauge bosons decaying to two tau leptons using 36 fb^{-1} of data collected at $\sqrt{s}$ = 13 TeV with the ATLAS detector," *Acta Physica Polonica B*, vol. 51, pp. 1405–1410, 2020. doi: 10.5506/APhysPolB.51.1405

[5] C. Yang and R. Mills, "Conservation of isotopic spin and isotopic gauge invariance," *Physical Review*, vol. 96, pp. 191–195, 1954. doi: 10.1103/PhysRev.96.191

[6] A. Salam and J. Ward, "Weak and electromagnetic interactions," *Nuovo Cimento*, vol. 11, pp. 568–577, 1959. doi: 10.1007/BF02726525

[7] S. Glashow, "Partial symmetries of weak interactions," *Nuclear Physics*, vol. 22, pp. 579–588, 1961. doi: 10.1016/0029-5582(61)90469-2

[8] J. Goldstone, A. Salam, and S. Weinberg, "Broken symmetries," *Physical Review*, vol. 127, pp. 965–970, 1962. doi: 10.1103/PhysRev.127.965

[9] A. Salam and J. Ward, "Gauge theory of elementary interactions," *Physical Review*, vol. 136, pp. B763–B768, 1964. doi: 10.1103/PhysRev.136.B763

[10] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Physical Review Letters*, vol. 13, pp. 321–323, 1964. doi: 10.1103/PhysRevLett.13.321

[11] P. Higgs, "Broken symmetries, massless particles and gauge fields," *Physics Letters*, vol. 12, pp. 132–133, 1964. doi: 10.1016/0031-9163(64)91136-9

[12] P. Higgs, "Broken symmetries and the masses of gauge bosons," *Physical Review Letters*, vol. 13, pp. 508–509, 1964. doi: 10.1103/PhysRevLett.13.508

[13] G. Guralnik, C. Hagen, and T. Kibble, "Global conservation laws and massless particles," *Physical Review Letters*, vol. 13, pp. 585–587, 1964. doi: 10.1103/PhysRevLett.13.585

[14] S. Weinberg, "A model of leptons," *Physical Review Letters*, vol. 19, pp. 1264–1266, 1967. doi: 10.1103/PhysRevLett.19.1264

[15] A. Salam, "Elementary particle physics: Relativistic groups and analyticity," in *Eighth Nobel Symposium. Stockholm: Almquvist and Wiksell*, 1968, p. 367.

[16] G. 't Hooft, "Renormalization of massless Yang-Mills fields," *Nuclear Physics B*, vol. 33, pp. 173–199, 1971. doi: 10.1016/0550-3213(71)90395-6

[17] G. 't Hooft, "Renormalizable Lagrangians for massive Yang-Mills fields," *Nuclear Physics B*, vol. 35, pp. 167–188, 1971. doi: 10.1016/0550-3213(71)90139-8

[18] G. 't Hooft and M. Veltman, "Regularization and renormalization of gauge fields," *Nuclear Physics B*, vol. 44, pp. 189–213. doi: 10.1016/0550-3213(72)90279-9

[19] D. Gross and F. Wilczek, "Ultraviolet behavior of non-Abelian gauge theories," *Physical Review Letters*, vol. 30, pp. 1343–1346, 1973. doi: 10.1103/PhysRevLett.30.1343

[20] H. Politzer, "Reliable perturbative results for strong interactions," *Physical Review Letters*, vol. 30, pp. 1346–1349, 1973. doi: 10.1103/PhysRevLett.30.1346

[21] R. Oerter, *The Theory of Almost Everything : The Standard Model, The Unsung Triumph of Modern Physics*. Plume, 2006. ISBN 0452287863

[22] Image by MissMJ, licensed under CC BY-SA 3.0.

[23] M. Veltman, *Diagrammatica: the Path to Feynman Rules*. Cambridge University Press, 1994. ISBN 9780521456920

[24] P. Dirac, "Relativity quantum mechanics with an application to Compton scattering," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 111, pp. 405–423, 1926. doi: 10.1098/rspa.1926.0074

[25] W. Heisenberg, "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," *Zeitschrift für Physik*, vol. 43, no. 3-4, pp. 172–198, 1927. doi: 10.1007/BF01397280

[26] L. Mandelshtam and I. Tamm, "The uncertainty relation between energy and time in non-relativistic quantum mechanics," *Journal of Physics (USSR)*, vol. 9, pp. 249–254, 1945.

[27] Particle Data Group, "Review of particle physics," *Progress of Theoretical and Experimental Physics*, vol. 2020, no. 8, 2020 and 2021 update. doi: 10.1093/ptep/ptaa104

[28] G. 't Hooft, "Dimensional regularization and the renormalization group," *Nuclear Physics B*, vol. 61, pp. 455–468, 1973. doi: 10.1016/0550-3213(73)90376-3

[29] S. Weinberg, "New approach to the renormalization group," *Physical Review D*, vol. 8, no. 10, pp. 3497–3509, 1973. doi: 10.1103/PhysRevD.8.3497

[30] R. Feynman, *QED: The Strange Theory of Light and Matter*. Princeton University Press, 1985. ISBN 978-0-691-12575-6

[31] J. Greensite, *An Introduction to the Confinement Problem*. Springer, 2011. ISBN 978-3-642-14381-6

[32] V. Mathieu, N. Kochelev, and V. Vento, "The physics of glueballs," *International Journal of Modern Physics E*, vol. 18, no. 1, pp. 1–49, 2009. doi: 10.1142/S0218301309012124

[33] N. Cabibbo, "Unitary symmetry and leptonic decays," *Physical Review Letters*, vol. 10, no. 12, pp. 531–533, 1963. doi: 10.1103/PhysRevLett.10.531

[34] M. Kobayashi and T. Maskawa, "CP-violation in the renormalizable theory of weak interaction," *Progress of Theoretical Physics*, vol. 49, no. 2, pp. 652–657, 1973. doi: 10.1143/PTP.49.652

[35] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge," *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*, vol. 34, p. 247, 1957.

[36] Super-Kamiokande Collaboration, "The Super-Kamiokande detector," *Nuclear Instruments and Methods in Physics Research A*, vol. 501, no. 2-3, pp. 418–462, 2003. doi: 10.1016/S0168-9002(03)00425-X

[37] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos," *Physical Review Letters*, vol. 81, no. 8, pp. 1562–1567, 1998. doi: 10.1103/PhysRevLett.81.1562

[38] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles," *Progress of Theoretical Physics*, vol. 28, no. 5, p. 870, 1962. doi: 10.1143/PTP.28.870

[39] I. Esteban et al., "The fate of hints: updated global analysis of three-flavor neutrino oscillations," *Journal of High Energy Physics*, vol. 2020, no. 9, 2020. doi: 10.1007/jhep09(2020)178

[40] C. Wu et al., "Experimental test of parity conservation in beta decay," *Physical Review*, vol. 105, no. 4, pp. 1413–1415, 1957. doi: 10.1103/PhysRev.105.1413

[41] F. Klinkhamer, "Neutrino mass and the Standard Model," *Modern Physics Letters A*, vol. 28, no. 5, p. 1350010, 2013. doi: 10.1142/S0217732313500107

[42] J. Christenson et al., "Evidence for the 2π decay of the K_2^0 meson system," *Physical Review Letters*, vol. 13, no. 4, p. 138, 1964. doi: 10.1103/PhysRevLett.13.138

[43] D. Perkins, *Introduction to High Energy Physics*. Cambridge University Press, 2015. ISBN 978-0521621960

[44] MuLan Collaboration, "Improved measurement of the positive-muon lifetime and determination of the Fermi constant," *Physical Review Letters*, vol. 99, no. 3, p. 032001, 2007. doi: 10.1103/PhysRevLett.99.032001

[45] 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*, vol. 716, pp. 1–29, 2012. doi: 10.1016/j.physletb.2012.08.020

[46] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," vol. 716, pp. 30–61, 2012. doi: 10.1016/j.physletb.2012.08.021

[47] H. Yukawa, "On the interaction of elementary particles. I," *Progress of Theoretical Physics Supplement*, vol. 1, pp. 1–10, 1955. doi: 10.1143/PTPS.1.1

[48] Z.-Z. Xing, "Neutrino physics," in *1st Asia-Europe-Pacific School of High-Energy Physics*, 2014. doi: 10.5170/CERN-2014-001.177 pp. 177–217.

[49] T. Markkanen, A. Rajantie, and S. Stopyra, "Cosmological aspects of Higgs vacuum metastability," *Frontiers in Astronomy and Space Sciences*, vol. 5, 2018. doi: 10.3389/fspas.2018.00040

[50] N. Cabibbo et al., "Bounds on the fermions and Higgs boson masses in grand unified theories," *Nuclear Physics B*, vol. 158, no. 2-3, pp. 295–305, 1979. doi: 10.1016/0550-3213(79)90167-6

[51] S. Alekhin, A. Djouadi, and S. Moch, "The top quark and Higgs boson masses and the stability of the electroweak vacuum," *Physics Letters B*, vol. 716, no. 1, pp. 214–219, 2012. doi: 10.1016/j.physletb.2012.08.024

[52] C. Callan and S. Coleman, "Fate of the false vacuum. II. First quantum corrections," *Physical Review D*, vol. 16, no. 6, pp. 1762–1768, 1977. doi: 10.1103/physrevd.16.1762

[53] Y.-S. Tsai, "Decay correlations of heavy leptons in $e^+e^- \to l^+l^-$," *Physical Review D*, vol. 4, no. 9, pp. 2821–2837, 1971. doi: 10.1103/physrevd.4.2821

[54] L. Canetti, M. Drewes, and M. Shaposhnikov, "Matter and antimatter in the universe," *New Journal of Physics*, vol. 14, no. 9, p. 095012, 2012. doi: 10.1088/1367-2630/14/9/095012

[55] A. Sakharov, "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe," *Zhurnal Eksperimentalnoi Teoreticheskoi Fiziki*, vol. 5, pp. 32–35, 1967. doi: 10.1070/pu1991v034n05abeh002497

[56] M. Gavela et al., "Standard Model CP-violation and baryon asymmetry," *Modern Physics Letters A*, vol. 9, pp. 795–809, 1994. doi: 10.1142/s0217732394000629

[57] A. Einstein, "Die Feldgleichungen der Gravitation," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, pp. 844–847, 1915.

[58] A. Einstein, "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, p. 142, 1917.

[59] LIGO Collaboration, "Observation of gravitational waves from a binary black hole merger," *Physical Review Letters*, vol. 116, no. 6, 2016. doi: 10.1103/physrevlett.116.061102

[60] A. Zee, *Quantum Field Theory in a Nutshell*. Princeton University Press, 2010. ISBN 978-0-691-14034-6

[61] M. Planck, *Sitzungsbericht Deutsche Akademie der Wissenschaften, Mathematisch-Physikalische Technische Klasse*, vol. 5, pp. 440–480, 1899.

[62] J. de Swart, G. Bertone, and J. van Dongen, "How dark matter came to matter," *Nature Astronomy*, vol. 1, no. 3, 2017. doi: 10.1038/s41550-017-0059

[63] L. Baudis, "The search for dark matter," *European Review*, vol. 26, no. 1, pp. 70–81, 2017. doi: 10.1017/s1062798717000783

[64] K. Freese, "Status of dark matter in the universe," *International Journal of Modern Physics D*, vol. 26, no. 06, p. 1730012, 2017. doi: 10.1142/s0218271817300129

[65] M. Milgrom, "A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis," *The Astrophysical Journal*, vol. 270, p. 365, 1983. doi: 10.1086/161130

[66] T. Clifton, P. Ferreira, A., and C. Skordis, "Modified gravity and cosmology," *Physics Reports*, vol. 513, no. 1-3, pp. 1–189, 2012. doi: 10.1016/j.physrep.2012.01.001

[67] F. Riva, "Effective (Field) Theories," lecture given at the 2018 Electroweak Symmetry Breaking Summer School in Maratea, Italy.

[68] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN, 2017.

[69] S. Weinberg, "Baryon and lepton nonconserving processes," *Physical Review Letters*, vol. 43, pp. 1566–1570, 1979. doi: 10.1103/PhysRevLett.43.1566

[70] E. Majorana, "Teoria simmetrica dell'elettrone e del positrone," *Il Nuovo Cimento*, vol. 14, no. 4, pp. 171–184, 1937. doi: 10.1007/bf02961314

[71] F. Deppisch, *A Modern Introduction to Neutrino Physics*. Morgan & Claypool Publishers, 2019. ISBN 978-1-64327-679-3

[72] D. Kaplan, "Five lectures on effective field theory," lectures delivered at the 2015 Invisibles15 School in Madrid, Spain.

[73] B. Grzadkowski et al., "Dimension-six terms in the Standard Model Lagrangian," *Journal of High Energy Physics*, vol. 10, no. 85, 2010. doi: 10.1007/JHEP10(2010)085

[74] L. Lehman, "Extending the Standard Model effective field theory with the complete set of dimension-7 operators," *Physical Review D*, vol. 90, no. 12, p. 125023, 2014. doi: 10.1103/PhysRevD.90.125023

[75] B. Henning et al., "2, 84, 30, 993, 560, 15456, 11962, 261485, ...: higher dimension operators in the SM EFT," *Journal of High Energy Physics*, vol. 2017, no. 16, 2017. doi: 10.1007/jhep08(2017)016

[76] H. Primakoff and S. Rosen, "Baryon number and lepton number conservation laws," *Annual Review of Nuclear and Particle Science*, vol. 31, no. 1, pp. 145–192, 1981. doi: 10.1146/annurev.ns.31.120181.001045

[77] A. Manohar, "Introduction to effective field theories," proceedings of the CVIII Session of the Ecole de Physique des Houches, July 2017.

[78] I. Bah and F. Bonetti, "Anomaly inflow, accidental symmetry, and spontaneous symmetry breaking," *Journal of High Energy Physics*, vol. 2020, no. 1, 2020. doi: 10.1007/jhep01(2020)117

[79] A. Cherchiglia et al., "Guises and disguises of quadratic divergences," *Annals of Physics*, vol. 351, pp. 751–772, 2014. doi: 10.1016/j.aop.2014.10.002

[80] J. Wells, "Lectures on Higgs boson physics in the Standard Model and beyond," in *38th British Universities Summer School in Theoretical Elementary Particle Physics*.

[81] H. Miyazawa, "Baryon number changing currents," *Progress of Theoretical Physics*, vol. 36, no. 6, pp. 1266–1276, 1966. doi: 10.1143/PTP.36.1266

[82] J. Gervais and B. Sakita, "Field theory interpretation of supergauges in dual models," *Nuclear Physics B*, vol. 34, no. 2, pp. 632–639, 1971. doi: 10.1016/0550-3213(71)90351-8

[83] D. Volkov and V. Akulov, "Possible universal neutrino interaction," *Zhurnal Eksperimentalnoi Teoreticheskoi Fiziki*, vol. 16, pp. 621–624, 1972.

[84] J. Wess and B. Zumino, "Supergauge transformations in four dimensions," *Nuclear Physics B*, vol. 70, no. 1, pp. 39–50, 1974. doi: 10.1016/0550-3213(74)90355-1

[85] C. Csaki, "The Minimal Supersymmetric Standard Model (MSSM)," *Modern Physics Letters A*, vol. 11, no. 08, pp. 599–613, 1996. doi: 10.1142/s021773239600062x

[86] U. Ellwanger, C. Hugonie, and A. Teixeira, "The Next-to-Minimal Supersymmetric Standard Model," *Physics Reports*, vol. 496, no. 1-2, pp. 1–77, 2010. doi: 10.1016/j.physrep.2010.07.001

[87] M. Dugan, H. Georgi, and D. Kaplan, "Anatomy of a composite Higgs model," *Nuclear Physics B*, vol. 254, pp. 299–326, 1985. doi: 10.1016/0550-3213(85)90221-4

[88] K. Agashe, R. Contino, and A. Pomarol, "The minimal composite Higgs model," *Nuclear Physics B*, vol. 719, no. 1-2, pp. 165–187, 2005. doi: 10.1016/j.nuclphysb.2005.04.035

[89] Z. Chacko, H.-S. Goh, and R. Harnik, "Natural electroweak breaking from a mirror symmetry," *Physical Review Letters*, vol. 96, no. 23, 2006. doi: 10.1103/physrevlett.96.231802

[90] G. Arcadi, A. Djouadi, and M. Raidal, "Dark matter through the Higgs portal," *Physics Reports*, vol. 842, pp. 1–180, 2020. doi: 10.1016/j.physrep.2019.11.003

[91] I. Ivanov, "Building and testing models with extended Higgs sectors," *Progress in Particle and Nuclear Physics*, vol. 95, pp. 160–208, 2017. doi: 10.1016/j.ppnp.2017.03.001

[92] G. Branco et al., "Theory and phenomenology of two-Higgs-doublet models," *Physics Reports*, vol. 516, no. 1-2, pp. 1–102, 2012. doi: 10.1016/j.physrep.2012.02.002

[93] S. Glashow and S. Weinberg, "Natural conservation laws for neutral currents," *Physical Review D*, vol. 15, no. 7, pp. 1958–1965, 1977. doi: 10.1103/physrevd.15.1958

[94] E. Paschos, "Diagonal neutral currents," *Physical Review D*, vol. 15, no. 7, pp. 1966–1972, 1977. doi: 10.1103/physrevd.15.1966

[95] A. Djouadi et al., "The post-Higgs MSSM scenario: habemus MSSM?" *The European Physical Journal C*, vol. 73, no. 12, 2013. doi: 10.1140/epjc/s10052-013-2650-0

[96] E. Bagnaschi et al., "Benchmark scenarios for low $\tan\beta$ in the MSSM," Tech. Rep., 2015. [Online]. Available: url:https://cds.cern.ch/record/2039911

[97] E. Bagnaschi et al., "MSSM Higgs boson searches at the LHC: benchmark scenarios for Run 2 and beyond," *The European Physical Journal C*, vol. 79, no. 7, 2019. doi: 10.1140/epjc/s10052-019-7114-8

[98] LEP Collaboration, "Search for neutral MSSM Higgs bosons at LEP," *The European Physical Journal C*, vol. 47, no. 3, 2006. doi: 10.1140/epjc/s2006-02569-7

[99] R. Walker, "Synchrotron radiation," 1994. doi: 10.5170/CERN-1994-001.437

[100] T. Gleisberg et al., "Event generation with SHERPA 1.1," *Journal of High Energy Physics*, vol. 2009, no. 02, pp. 007–007, 2009. doi: 10.1088/1126-6708/2009/02/007

[101] D. Brandt, "CAS - CERN Accelerator School: Intermediate course on accelerator physics," 2006. doi: 10.5170/CERN-2006-002

[102] M. Mangano and T. Stelzer, "Tools for the simulation of hard hadronic collisions," *Annual Review of Nuclear and Particle Science*, vol. 55, no. 1, pp. 555–588, 2005. doi: 10.1146/annurev.nucl.55.090704.151505

[103] K. Albertsson et al., "Machine learning in high energy physics community white paper," *Journal of Physics: Conference Series*, vol. 1085, p. 022008, 2018. doi: 10.1088/1742-6596/1085/2/022008

[104] T. Hastie, R. Tibshirani, and J. Friedman, *Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 2016. ISBN 0387848576

[105] L. Kaelbling, M. Littman, and A. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996. doi: 10.1613/jair.301

[106] Image by Chabacano, licensed under CC BY-SA 4.0.

[107] Image by kakau, licensed under CC BY-SA 3.0.

[108] C. Ferri, J. Hernández-Orallo, and M. Salido, "Volume under the ROC surface for multi-class problems," in *Machine Learning: European Conference on Machine Learning 2003.* Springer Berlin Heidelberg, 2003, pp. 108–120.

[109] D. Hand and R. Till, *Machine Learning*, vol. 45, no. 2, pp. 171–186, 2001. doi: 10.1023/a:1010920819831

[110] Image by Gufosowa, licensed under CC BY-SA 4.0.

[111] L. Breiman et al., *Classification and Regression Trees.* Chapman and Hall, 1984. ISBN 9780412048418

[112] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, 1999. doi: 10.1613/jair.614

[113] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1-2, pp. 1–39, 2009. doi: 10.1007/s10462-009-9124-7

[114] M. Kearns, "Thoughts on hypothesis boosting," 1988, unpublished manuscript (Machine Learning class project, December 1988).

[115] M. Kearns and L. Valiant, "Crytographic limitations on learning Boolean formulae and finite automata," in *Proceedings of the twenty-first annual ACM symposium on Theory of computing - STOC 89*, 1989. doi: 10.1145/73007.73049

[116] R. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, 1990. doi: 10.1007/bf00116037

[117] L. Breiman, "Arcing classifier (with discussion and a rejoinder by the author)," *The Annals of Statistics*, vol. 26, no. 3, 1998. doi: 10.1214/aos/1024691079

[118] R. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999. doi: 10.1023/a:1007614523901

[119] L. Mason et al., "Boosting algorithms as gradient descent," in *Proceedings of the 12th International Conference on Neural Information Processing Systems*, ser. NIPS'99, 1999. doi: 10.5555/3009657.3009730 pp. 512–518.

[120] K. Gurney, *An Introduction to Neural Networks.* CRC Press, 1997. ISBN 1857285034

[121] Image by Glosser.ca, licensed under CC BY-SA 3.0.

[122] V. Nair and G. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning*, 2010, pp. 807–814.

[123] Y. LeCun, F. Huang, and L. Bottou, "Learning methods for generic object recognition with in-variance to pose and lighting," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2004. doi: 10.1109/cvpr.2004.1315150

[124] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986. doi: 10.1038/323533a0

[125] Image by fdeloche, licensed under CC BY-SA 4.0.

[126] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735

[127] Image by MingxianLin, licensed under CC BY-SA 4.0.

[128] H. Voss et al., "TMVA, the Toolkit for Multivariate Data Analysis with ROOT," in *Proceedings of XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research - PoS(ACAT)*, 2009. doi: 10.22323/1.050.0040

[129] R. Brun and F. Rademakers, "ROOT - an object oriented data analysis framework," in *Proceedings AIHENP'96 Workshop*, vol. 389, no. 1-2, 1997. doi: 10.1016/s0168-9002(97)00048-x pp. 81–86.

[130] The CERN accelerator complex by Esma Mobs, CERN-GRAPHICS-2019-002-1.

[131] ATLAS Collaboration, "The ATLAS experiment at the CERN Large Hadron Collider," *Journal of Instrumentation*, vol. 3, no. 08, p. S08003, 2008. doi: 10.1088/1748-0221/3/08/s08003

[132] CMS Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, p. S08004, 2008. doi: 10.1088/1748-0221/3/08/s08004

[133] ALICE Collaboration, "The ALICE experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, p. S08002, 2008. doi: 10.1088/1748-0221/3/08/s08002

[134] LHCb Collaboration, "The LHCb detector at the LHC," *Journal of Instrumentation*, vol. 3, no. 08, p. S08005, 2008. doi: 10.1088/1748-0221/3/08/s08005

[135] CMS-TOTEM Collaboration, "CMS-TOTEM precision proton spectrometer," Tech. Rep., 2014. [Online]. Available: https://cds.cern.ch/record/1753795

[136] MoEDAL Collaboration, "Technical design report of the MoEDAL experiment," Tech. Rep., 2009. [Online]. Available: cds.cern.ch/record/1181486

[137] LHCf Collaboration, "LHCf experiment: Technical design report," Tech. Rep., 2006. [Online]. Available: https://cds.cern.ch/record/926196

[138] J. Feng at al., "ForwArd Search ExpeRiment at the LHC," *Physical Review D*, vol. 97, no. 3, 2018. doi: 10.1103/physrevd.97.035001

[139] G. Apollinari et al., "High Luminosity Large Hadron Collider HL-LHC," Tech. Rep., 2015.

[140] Photo by Saintfevrier, licensed under CC BY-SA 4.0.

[141] ATLAS Collaboration, "ATLAS experiment - public results." [Online]. Available: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2

[142] "Worldwide LHC computing grid." [Online]. Available: http://cern.ch/lcg

[143] ATLAS Collaboration, "ATLAS inner detector: Technical design report, vol. 1." Tech. Rep., 1997. [Online]. Available: https://cds.cern.ch/record/331063

[144] ATLAS Collaboration, "ATLAS inner detector: Technical design report, vol. 2." Tech. Rep., 1997. [Online]. Available: https://cds.cern.ch/record/331063

[145] M. Capeans et al., "ATLAS Insertable B-Layer Technical Design Report," Tech. Rep., 2010. [Online]. Available: https://cds.cern.ch/record/1291633

[146] N. Wermes and G. Hallewel, "ATLAS pixel detector: Technical design report," Tech. Rep., 1998. [Online]. Available: https://cds.cern.ch/record/381263

[147] B. Mandelli, "The pixel detector of the ATLAS experiment for the Run 2 at the Large Hadron Collider," *Nuclear and Particle Physics Proceedings*, vol. 273-275, pp. 1166–1172, 2016. doi: 10.1016/j.nuclphysbps.2015.09.183

[148] ATLAS Collaboration, "The silicon microstrip sensors of the ATLAS semiconductor tracker," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 578, no. 1, pp. 98–118, 2007. doi: 10.1016/j.nima.2007.04.157

[149] ATLAS Collaboration, "Operation and performance of the ATLAS semiconductor tracker," *Journal of Instrumentation*, vol. 9, no. 08, pp. P08 009–P08 009, 2014. doi: 10.1088/1748-0221/9/08/p08009

[150] ATLAS Collaboration, "The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance," *Journal of Instrumentation*, vol. 3, no. 02, pp. P02 013–P02 013, 2008. doi: 10.1088/1748-0221/3/02/p02013

[151] B. Mindur, "ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 845, pp. 257–261, 2017. doi: 10.1016/j.nima.2016.04.026

[152] K. Potamianos, "The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider," 2015, p. 261.

[153] ATLAS Collaboration, "ATLAS liquid-argon calorimeter: Technical design report," Tech. Rep., 1996. [Online]. Available: https://cds.cern.ch/record/331061

[154] ATLAS Collaboration, "ATLAS tile calorimeter: Technical design report," Tech. Rep., 1996. [Online]. Available: https://cds.cern.ch/record/331062

[155] ATLAS Collaboration, "ATLAS muon spectrometer: Technical design report," Tech. Rep., 1997. [Online]. Available: https://cds.cern.ch/record/331068

[156] ATLAS Collaboration, "Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data," Tech. Rep. 7, 2014.

[157] 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," Tech. Rep. 8, 2019.

[158] ATLAS Collaboration, "Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV," Tech. Rep. 5, 2016.

[159] S. Rettie, "Muon identification and performance in the ATLAS experiment," Tech. Rep., 2018. [Online]. Available: https://cds.cern.ch/record/2626330

[160] ATLAS Collaboration, "Jet reconstruction and performance using particle flow with the ATLAS detector," Tech. Rep. 7, 2017.

[161] ATLAS Collaboration, "Reconstruction, energy calibration, and identification of hadronically decaying tau leptons in the ATLAS experiment for Run-2 of the LHC," Tech. Rep., 2015. [Online]. Available: https://cds.cern.ch/record/2064383

[162] ATLAS Collaboration, "Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment," Tech. Rep., 2019. [Online]. Available: https://cds.cern.ch/record/2688062

[163] 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*, vol. 78, no. 11, 2018. doi: 10.1140/epjc/s10052-018-6288-9

[164] J. Pequenao and P. Schaffner, "How ATLAS detects particles: diagram of particle paths in the detector," 2013. [Online]. Available: https://cds.cern.ch/record/1505342

[165] W. Lampl et al., "Calorimeter clustering algorithms: Description and performance," 2008. [Online]. Available: https://cds.cern.ch/record/1099735

[166] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2-3, pp. 444–450, 1987. doi: 10.1016/0168-9002(87)90887-4

[167] ATLAS Collaboration, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1," *The European Physical Journal C*, vol. 77, no. 7, 2017. doi: 10.1140/epjc/s10052-017-5004-5

[168] M. Cacciari, G. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, no. 04, pp. 063–063, 2008. doi: 10.1088/1126-6708/2008/04/063

[169] M. Cacciari and G. Salam, "Dispelling the N^3 myth for the k_t jet-finder," *Physics Letters B*, vol. 641, no. 1, pp. 57–61, 2006. doi: 10.1016/j.physletb.2006.08.037

[170] ATLAS Collaboration, "Tagging and suppression of pileup jets with the ATLAS detector," Tech. Rep., 2014. [Online]. Available: https://cds.cern.ch/record/1700870

[171] M. Lehmacher, "b-tagging algorithms and their performance at ATLAS," Tech. Rep., 2008. [Online]. Available: http://cds.cern.ch/record/1128662

[172] ATLAS Collaboration, "Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV," *Journal of High Energy Physics*, vol. 2018, no. 8, 2018. doi: 10.1007/jhep08(2018)089

[173] ATLAS Collaboration, "Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run," Tech. Rep., 2017. [Online]. Available: https://cds.cern.ch/record/2273281

[174] M. Lanfermann on behalf of the ATLAS Collaboration, "Deep learning in flavour tagging at the ATLAS experiment," in *Proceedings of The European Physical Society Conference on High Energy Physics - PoS(EPS-HEP2017)*, 2018. doi: 10.22323/1.314.0764

[175] D. Cavalli et al., "Reconstruction of soft missing transverse momentum with inner detector tracks," Tech. Rep., 2015. [Online]. Available: https://cds.cern.ch/record/2002888

[176] D. Adams et al., "The ATLAS computing model," 2004.

[177] P. Calafiura et al., "The Athena control framework in production, new developments and lessons learned," in *Computing in High Energy Physics and Nuclear Physics 2004, Interlaken*, 2005. doi: 10.5170/CERN-2005-002.456 pp. 456–458.

[178] M. Cattaneo, "Status of the GAUDI event-processing framework," 2001, presented at the 2000 International Conference on Computing in High Energy and Nuclear Physics in Padua, Italy. [Online]. Available: https://cds.cern.ch/record/1745145

[179] C. Bee et al., "The raw event format in the ATLAS trigger & DAQ," Tech. Rep., 2016. [Online]. Available: https://cds.cern.ch/record/683741

[180] M. Dobbs and J. Hansen, "The HepMC C++ Monte Carlo event record for high energy physics," *Computer Physics Communications*, vol. 134, no. 1, pp. 41–46, 2001. doi: 10.1016/s0010-4655(00)00189-2

[181] S. Agostinelli et al., "Geant4 - a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003. doi: 10.1016/s0168-9002(03)01368-8

[182] ATLAS Collaboration, "The ATLAS simulation infrastructure," *The European Physical Journal C*, vol. 70, no. 3, pp. 823–874, 2010. doi: 10.1140/epjc/s10052-010-1429-9

[183] T. Yamanaka on behalf of the ATLAS Collaboration, "The ATLAS calorimeter simulation FastCaloSim," *Journal of Physics: Conference Series*, vol. 331, no. 3, p. 032053, 2011. doi: 10.1088/1742-6596/331/3/032053

[184] M. Hobson, G. Efstathiou, and A. Lasenby, *General Relativity: An Introduction for Physicists.* Cambridge University Press, 2006. ISBN 978-0-521-82951-9

[185] Planck Collaboration, "Planck 2018 results," *Astronomy & Astrophysics*, vol. 641, p. A6, 2020. doi: 10.1051/0004-6361/201833910

[186] R. Adler, B. Casey, and O. Jacob, "Vacuum catastrophe: An elementary exposition of the cosmological constant problem," *American Journal of Physics*, vol. 63, no. 7, pp. 620–626, 1995. doi: 10.1119/1.17850

[187] J. Solà, "Cosmological constant and vacuum energy: old and new ideas," *Journal of Physics: Conference Series*, vol. 453, p. 012015, 2013. doi: 10.1088/1742-6596/453/1/012015

[188] I. Buchbinder and I. Shapiro, *Introduction to Quantum Field Theory with Applications to Quantum Gravity.* Oxford University Press, 2021.

[189] M. Veltman, *Acta Physica Polonica B*, vol. 12, p. 437, 1981.

[190] W. Pauli, *Pauli Lectures on Physics: Vol. 6, Selected Topics in Field Quantization.* MIT Press, 1971. ISBN 978-0486414591

[191] M. Visser, "Lorentz invariance and the zero-point stress-energy tensor," *Particles*, vol. 1, no. 1, p. 10, 2018. doi: 10.3390/particles1010010

[192] F. Melia, "The cosmic equation of state," *Astrophysics and Space Science*, vol. 356, no. 2, pp. 393–398, 2014. doi: 10.1007/s10509-014-2211-5

[193] J. Martin, "Everything you always wanted to know about the cosmological constant problem (but were afraid to ask)," *Comptes Rendus Physique*, vol. 13, no. 6-7, pp. 566–665, 2012. doi: 10.1016/j.crhy.2012.04.008

[194] J. Haller et al., "Update of the global electroweak fit and constraints on two-Higgs-doublet models," *The European Physical Journal C*, vol. 78, no. 8, 2018. doi: 10.1140/epjc/s10052-018-6131-3

[195] A. Arbey et al., "Status of the charged Higgs boson in two Higgs doublet models," *The European Physical Journal C*, vol. 78, no. 3, 2018. doi: 10.1140/epjc/s10052-018-5651-1

[196] M. Misiak and M. Steinhauser, "Weak radiative decays of the B meson and bounds on $M_{H^\pm}$ in the two-Higgs-doublet model," *The European Physical Journal C*, vol. 77, no. 3, 2017. doi: 10.1140/epjc/s10052-017-4776-y

[197] N. Darvishi and M. Krawczyk, "Implication of quadratic divergences cancellation in the two Higgs doublet model," *Nuclear Physics B*, vol. 926, pp. 167–178, 2018. doi: 10.1016/j.nuclphysb.2017.10.027

[198] ATLAS Collaboration, "Constraints on new phenomena via Higgs boson couplings and invisible decays with the ATLAS detector," *Journal of High Energy Physics*, vol. 2015, no. 11, 2015. doi: 10.1007/jhep11(2015)206

[199] B. Kniehl, A. Pikelner, and O. Veretin, "mr : A C++ library for the matching and running of the Standard Model parameters," *Computer Physics Communications*, vol. 206, pp. 84–96, 2016. doi: 10.1016/j.cpc.2016.04.017

[200] Particle Data Group, "Review of particle physics," *Physical Review D*, vol. 98, p. 030001, 2018 and 2019 update. doi: 10.1103/PhysRevD.98.030001

[201] F. Jegerlehner, "The Standard Model as a low-energy effective theory: What is triggering the Higgs mechanism?" *Acta Physica Polonica B*, vol. 45, no. 6, p. 1167, 2014. doi: 10.5506/aphyspolb.45.1167

[202] I. Masina and M. Quiros, "On the Veltman condition, the hierarchy problem, and high-scale supersymmetry," *Physical Review D*, vol. 88, no. 9, 2013. doi: 10.1103/physrevd.88.093003

[203] G. Degrassi et al., "Higgs mass and vacuum stability in the Standard Model at NNLO," *Journal of High Energy Physics*, vol. 2012, no. 8, 2012. doi: 10.1007/jhep08(2012)098

[204] Y. Hamada, H. Kawai, and K.-Y. Oda, "Bare Higgs mass at Planck scale," *Physical Review D*, vol. 87, no. 5, 2013. doi: 10.1103/physrevd.87.053009

[205] S. Weinberg, "The cosmological constant problem," *Reviews of Modern Physics*, vol. 61, no. 1, pp. 1–23, 1989. doi: 10.1103/revmodphys.61.1

[206] S. Weinberg, "Anthropic bound on the cosmological constant," *Physical Review Letters*, vol. 59, no. 22, pp. 2607–2610, 1987. doi: 10.1103/physrevlett.59.2607

[207] E. Bothmann et al., "Event generation with Sherpa 2.2," *SciPost Physics*, vol. 7, no. 3, 2019. doi: 10.21468/scipostphys.7.3.034

[208] R. Ball et al., "Parton distributions for the LHC Run II," *Journal of High Energy Physics*, vol. 2015, no. 4, 2015. doi: 10.1007/jhep04(2015)040

[209] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms," *Journal of High Energy Physics*, vol. 2004, no. 11, p. 40, 2004. doi: 10.1088/1126-6708/2004/11/040

[210] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method," *Journal of High Energy Physics*, vol. 2007, no. 11, p. 70, 2007. doi: 10.1088/1126-6708/2007/11/070

[211] S. Alioli et al., "NLO Higgs boson production via gluon fusion matched with shower in POWHEG," *Journal of High Energy Physics*, vol. 2009, no. 04, p. 2, 2009. doi: 10.1088/1126-6708/2009/04/002

[212] S. Alioli et al., "NLO vector-boson production matched with shower in POWHEG," *Journal of High Energy Physics*, vol. 2008, no. 07, p. 60, 2008. doi: 10.1088/1126-6708/2008/07/060

[213] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1," *Computer Physics Communications*, vol. 178, no. 11, pp. 852–867, 2008. doi: 10.1016/j.cpc.2008.01.036

[214] H.-L. Lai et al., "New parton distributions for collider physics," *Physical Review D*, vol. 82, no. 7, 2010. doi: 10.1103/physrevd.82.074024

[215] S. Myers and E. Picasso, "The design, construction and commissioning of the CERN Large Electron-Positron Collider," *Contemporary Physics*, vol. 31, no. 6, pp. 387–403, 1990. doi: 10.1080/00107519008213789

[216] ALEPH Collaboration, "Measurement of the tau polarisation at LEP," *The European Physical Journal C*, vol. 20, no. 3, pp. 401–430, 2001. doi: 10.1007/s100520100689

[217] Tevatron Collaboration, "Tevatron Run II combination of the effective leptonic electroweak mixing angle," *Physical Review D*, vol. 97, no. 11, 2018. doi: 10.1103/physrevd.97.112007

[218] ATLAS Collaboration, "Measurement of the forward-backward asymmetry of electron and muon pair-production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," *Journal of High Energy Physics*, vol. 2015, no. 9, 2015. doi: 10.1007/jhep09(2015)049

[219] ATLAS Collaboration, "Measurement of the effective leptonic weak mixing angle using electron and muon pairs from Z-boson decay in the ATLAS experiment at $\sqrt{s} = 8$ TeV," Tech. Rep., 2018. [Online]. Available: https://cds.cern.ch/record/2630340

[220] CMS Collaboration, "Measurement of the weak mixing angle using the forward–backward asymmetry of Drell–Yan events in pp collisions at 8 TeV," *The European Physical Journal C*, vol. 78, no. 9, 2018. doi: 10.1140/epjc/s10052-018-6148-7

[221] LHCb Collaboration, "Measurement of the forward-backward asymmetry in $Z/\gamma^{\to \mu^+\mu^-}$ decays and determination of the effective weak mixing angle," *Journal of High Energy Physics*, vol. 2015, no. 11, 2015. doi: 10.1007/jhep11(2015)190

[222] S. Jadach and Z. Was, "QED $O(\alpha^3)$ radiative corrections to the reaction $e^+e^- \to \tau^+\tau^-$ including spin and mass effects," *Acta Physica Polonica B*, vol. 15, p. 1151, 1984.

[223] ATLAS Collaboration, "Search for charged Higgs bosons decaying via $H^\pm \to \tau^\pm \nu_\tau$ in the τ+jets and τ+lepton final states with 36 fb^{-1} of pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS experiment," *Journal of High Energy Physics*, vol. 2018, no. 9, 2018. doi: 10.1007/jhep09(2018)139

[224] ATLAS Collaboration, "Search for charged Higgs bosons decaying into top and bottom quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Journal of High Energy Physics*, vol. 2018, no. 11, 2018. doi: 10.1007/jhep11(2018)085

[225] ATLAS Collaboration, "Search for heavy neutral Higgs bosons produced in association with b-quarks and decaying into b-quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Physical Review D*, vol. 102, no. 3, 2020. doi: 10.1103/physrevd.102.032004

[226] ATLAS Collaboration, "Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Journal of High Energy Physics*, vol. 2018, no. 1, 2018. doi: 10.1007/jhep01(2018)055

[227] S. Dittmaier, M. Krämer, and M. Spira, "Higgs radiation off bottom quarks at the Fermilab Tevatron and the CERN LHC," *Physical Review D*, vol. 70, no. 7, 2004. doi: 10.1103/physrevd.70.074010

[228] S. Dawson et al., "Exclusive Higgs boson production with bottom quarks at hadron colliders," *Physical Review D*, vol. 69, no. 7, 2004. doi: 10.1103/physrevd.69.074027

[229] C. Duhr et al., "Higgs production in bottom quark fusion: matching the 4- and 5-flavour schemes to third order in the strong coupling," *Journal of High Energy Physics*, vol. 2020, no. 8, 2020. doi: 10.1007/jhep08(2020)017

[230] R. Harlander, M. Kramer, and M. Schumacher, "Bottom-quark associated Higgs-boson production: reconciling the four- and five-flavour scheme approach," Tech. Rep., 2011. [Online]. Available: https://cds.cern.ch/record/1407669

[231] A. Kulesza et al., "Associated $t\bar{t}h$ production at the LHC: Theoretical predictions at NLO+NNLL accuracy," *Physical Review D*, vol. 97, no. 11, 2018. doi: 10.1103/physrevd.97.114007

[232] R. Harlander, S. Liebler, and H. Mantler, "SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM," *Computer Physics Communications*, vol. 184, no. 6, pp. 1605–1617, 2013. doi: 10.1016/j.cpc.2013.02.006

[233] M. Bonvini, A. Papanastasiou, and F. Tackmann, "Resummation and matching of b-quark mass effects in $b\bar{b}H$ production," *Journal of High Energy Physics*, vol. 2015, no. 11, 2015. doi: 10.1007/jhep11(2015)196

[234] M. Bonvini, A. Papanastasiou, and F. Tackmann, "Matched predictions for the $b\bar{b}H$ cross section at the 13 TeV LHC," *Journal of High Energy Physics*, vol. 2016, no. 10, 2016. doi: 10.1007/jhep10(2016)053

[235] S. Forte, D. Napoletano, and M. Ubiali, "Higgs production in bottom-quark fusion in a matched scheme," *Physics Letters B*, vol. 751, pp. 331–337, 2015. doi: 10.1016/j.physletb.2015.10.051

[236] S. Heinemeyer, W. Hollik, and G. Weiglein, "FeynHiggs: a program for the calculation of the masses of the neutral CP-even Higgs bosons in the MSSM," *Computer Physics Communications*, vol. 124, no. 1, pp. 76–89, 2000. doi: 10.1016/s0010-4655(99)00364-1

[237] S. Heinemeyer, W. Hollik, and G. Weiglein, "The masses of the neutral CP-even Higgs bosons in the MSSM: Accurate analysis at the two-loop level," *The European Physical Journal C*, vol. 9, no. 2, pp. 343–366, 1999. doi: 10.1007/s100529900006

[238] G. Degrassi et al., "Towards high-precision predictions for the MSSM Higgs sector," *The European Physical Journal C*, vol. 28, no. 1, pp. 133–143, 2003. doi: 10.1140/epjc/s2003-01152-2

[239] M. Frank et al., "The Higgs boson masses and mixings of the complex MSSM in the Feynman-diagrammatic approach," *Journal of High Energy Physics*, vol. 2007, no. 02, pp. 047–047, 2007. doi: 10.1088/1126-6708/2007/02/047

[240] T. Hahn et al., "High-precision predictions for the light CP-even Higgs boson mass of the Minimal Supersymmetric Standard Model," *Physical Review Letters*, vol. 112, no. 14, 2014. doi: 10.1103/physrevlett.112.141801

[241] K. Williams, H. Rzehak, and G. Weiglein, "Higher-order corrections to Higgs boson decays in the MSSM with complex parameters," *The European Physical Journal C*, vol. 71, no. 6, 2011. doi: 10.1140/epjc/s10052-011-1669-3

[242] H. Bahl and W. Hollik, "Precise prediction for the light MSSM Higgs-boson mass combining effective field theory and fixed-order calculations," *The European Physical Journal C*, vol. 76, no. 9, 2016. doi: 10.1140/epjc/s10052-016-4354-8

[243] H. Bahl et al., "Reconciling EFT and hybrid calculations of the light MSSM Higgs-boson mass," *The European Physical Journal C*, vol. 78, no. 1, 2018. doi: 10.1140/epjc/s10052-018-5544-3

[244] A. Djouadi, J. Kalinowski, and M. Spira, "HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension," *Computer Physics Communications*, vol. 108, no. 1, pp. 56–74, 1998. doi: 10.1016/s0010-4655(97)00123-9

[245] A. Djouadi et al., "HDECAY: Twenty++ years after," *Computer Physics Communications*, vol. 238, pp. 214–231, 2019. doi: 10.1016/j.cpc.2018.12.010

[246] A. Bredenstein at al., "Precise predictions for the Higgs-boson decay $H \rightarrow WW/ZZ \rightarrow 4$ leptons," *Physical Review D*, vol. 74, no. 1, 2006. doi: 10.1103/physrevd.74.013004

[247] S. Dawson et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. CERN, 2011.

[248] S. Alioli et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *Journal of High Energy Physics*, vol. 2010, no. 6, 2010. doi: 10.1007/jhep06(2010)043

[249] E. Bagnaschi et al., "Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM," *Journal of High Energy Physics*, vol. 2012, no. 2, 2012. doi: 10.1007/jhep02(2012)088

[250] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *Journal of High Energy Physics*, vol. 2014, no. 7, 2014. doi: 10.1007/jhep07(2014)079

[251] M. Wiesemann et al., "Higgs production in association with bottom quarks," *Journal of High Energy Physics*, vol. 2015, no. 2, 2015. doi: 10.1007/jhep02(2015)132

[252] T. Sjöstrand et al., "An introduction to PYTHIA 8.2," *Computer Physics Communications*, vol. 191, pp. 159–177, 2015. doi: 10.1016/j.cpc.2015.01.024

[253] S. Schumann and F. Krauss, "A parton shower algorithm based on Catani-Seymour dipole factorisation," *Journal of High Energy Physics*, vol. 2008, no. 03, pp. 038–038, 2008. doi: 10.1088/1126-6708/2008/03/038

[254] C. Anastasiou et al., "High-precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order," *Physical Review D*, vol. 69, no. 9, 2004. doi: 10.1103/physrevd.69.094008

[255] S. Frixione, G. Ridolfi, and P. Nason, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction," *Journal of High Energy Physics*, vol. 2007, no. 09, p. 126, 2007. doi: 10.1088/1126-6708/2007/09/126

[256] M. Beneke et al., "Hadronic top-quark pair production with NNLL threshold resummation," *Nuclear Physics B*, vol. 855, no. 3, pp. 695–741, 2012. doi: 10.1016/j.nuclphysb.2011.10.021

[257] M. Cacciari et al., "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation," *Physics Letters B*, vol. 710, no. 4-5, pp. 612–622, 2012. doi: 10.1016/j.physletb.2012.03.013

[258] P. Bärnreuther, M. Czakon, and A. Mitov, "Percent-level-precision physics at the Tevatron: Next-to-next-to-leading order QCD corrections to $q\bar{q} \to t\bar{t} + X$," *Physical Review Letters*, vol. 109, no. 13, 2012. doi: 10.1103/physrevlett.109.132001

[259] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels," *Journal of High Energy Physics*, vol. 2012, no. 12, 2012. doi: 10.1007/jhep12(2012)054

[260] M. Czakon and A. Mitov, "NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction," *Journal of High Energy Physics*, vol. 2013, no. 1, 2013. doi: 10.1007/jhep01(2013)080

[261] M. Czakon, P. Fiedler, and A. Mitov, "Total top-quark pair-production cross section at hadron colliders though $\mathcal{O}(\alpha_s^4)$," *Physical Review Letters*, vol. 110, no. 25, 2013. doi: 10.1103/physrevlett.110.252004

[262] M. Czakon and A. Mitov, "Top++: A program for the calculation of the top-pair cross-section at hadron colliders," *Computer Physics Communications*, vol. 185, no. 11, pp. 2930–2938, 2014. doi: 10.1016/j.cpc.2014.06.021

[263] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method," *The European Physical Journal C*, vol. 71, no. 2, 2011. doi: 10.1140/epjc/s10052-011-1547-z

[264] R. Frederix, E. Re, and P. Torrielli, "Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMCNLO," *Journal of High Energy Physics*, vol. 2012, no. 9, 2012. doi: 10.1007/jhep09(2012)130

[265] S. Alioli et al., "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions," *Journal of High Energy Physics*, vol. 2009, no. 09, p. 111, 2009. doi: 10.1088/1126-6708/2009/09/111

[266] M. Aliev et al., "HATHOR - HAdronic top and heavy quarks crOss section calculatoR," *Computer Physics Communications*, vol. 182, no. 4, pp. 1034–1046, 2011. doi: 10.1016/j.cpc.2010.12.040

[267] P. Kant et al., "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions," *Computer Physics Communications*, vol. 191, pp. 74–89, 2015. doi: 10.1016/j.cpc.2015.02.001

[268] A. Elagin et al., "A new mass reconstruction technique for resonances decaying to," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 654, no. 1, pp. 481–489, 2011. doi: 10.1016/j.nima.2011.07.009

[269] A. Piqueras et al., "Search for neutral MSSM Higgs bosons $H/A \to \tau_{lep}\tau_{had}$ and $Z' \to \tau_{lep}\tau_{had}$ produced in 13 TeV collisions with the ATLAS detector," 2016. [Online]. Available: https://cds.cern.ch/record/2131232

[270] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C*, vol. 71, no. 2, 2011. doi: 10.1140/epjc/s10052-011-1554-0

[271] K. Cranmer, "Practical statistics for the LHC," 2015. doi: 10.5170/CERN-2015-001.247. [Online]. Available: https://cds.cern.ch/record/2004587

[272] J. Ocariz, "Probability and statistics for particle physicists," 2014. doi: 10.5170/CERN-2014-001.253. [Online]. Available: http://cds.cern.ch/record/1701936

[273] ATLAS Collaboration, "hMSSM summary plots from direct and indirect searches," 2020. [Online]. Available: https://cds.cern.ch/record/2713580

[274] A. Elagin et al., "A new mass reconstruction technique for resonances decaying to," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 654, no. 1, pp. 481–489, 2011. doi: 10.1016/j.nima.2011.07.009

[275] ATLAS Collaboration, "Measurement of τ polarisation in $Z/\gamma^* \to \tau\tau$ decays in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," *The European Physical Journal C*, vol. 78, no. 2, 2018. doi: 10.1140/epjc/s10052-018-5619-1

[276] P. Baldi et al., "Parameterized neural networks for high-energy physics," *The European Physical Journal C*, vol. 76, no. 5, 2016. doi: 10.1140/epjc/s10052-016-4099-4