ETP-Bachelor-KA/2025-06

Karlsruhe Institute of Technology

Studies on boosted $\tau\tau$ and bb pair identification in the NMSSM X $\rightarrow$ YH analysis

Bachelor Thesis

Jannik Ginter

At the Department of Physics
Institute of Experimental Particle Physics

Reviewer:	Prof. Dr. Ulrich Husemann
Second reviewer:	Dr. Thorsten Chwalek
Advisor:	Moritz Molch

Karlsruhe, 28. March 2025

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Karlsruhe, DATE

.....

(**Jannik Ginter**)

Contents

1 Introduction

The major ambition of physics is the mathematical explanation of phenomena appearing all over the universe. Particle physics in particular aims for an accurate description of the smallest objects and their interactions. Theories are developed which support the experimental observations and predict new physics, while at the same time, specialized experiments are conducted to exclude certain parameter ranges of these theories. The standard model (SM) of particle physics has been proven as the most successful theory to describe the interactions of elementary particles. With the detection of the Higgs boson in 2012 [1, 2], independently by both the Compact Muon Solenoid (CMS) [3] and the ATLAS [4] experiments at the Large Hadron Collider (LHC) at CERN [5], the last fundamental cornerstone in the picture of elementary particles has been found. Despite the years of challenging its statements and attempts to find flaws within the theory, the SM kept on being able to predict the interactions of the elementary particles with high precision.

However, it does not cover every phenomenon observed in nature. Among others, it fails to include the force of gravity, the mass of neutrinos [6], and the constituents of dark matter [7]. In an attempt to include these factors, extensions to the SM are proposed. Supersymmetry [8] has ever since been an attractive extension, as it introduces an additional symmetry linking fermions and bosons. Several of the shortcomings of the SM are embodied in supersymmetry and new particles are predicted that can be searched for in experiments. The next-to-minimal supersymmetric SM (NMSSM) [9, 10] is one of the extensions that is built upon supersymmetry. Out of the new particles it predicts, two additional Higgs bosons are relevant for this thesis.

This extended Higgs sector allows a di-Higgs process, in which one heavy Higgs boson X decays into a lighter Higgs boson Y and the SM Higgs boson H. In this thesis, the Y boson further decays into a τ lepton pair and the H boson decays into a bottom quark pair, resulting in a $\tau\tau$bb final state [11]. Different mass hypotheses for the X and Y bosons are simulated by the CMS experiment. The datasets generated by these simulations are studied in this thesis. For large differences in the mass of the X boson and its decay products, boosted decays occur that pose a challenge in correctly identifying them.

Previous studies on the boosted sector in the $X \rightarrow YH$ analysis [12] have used old identification algorithms like the one presented in [13]. A more efficient algorithm for the identification of boosted decays is desired. The aim of this thesis is to analyze the ParticleNet jet tagging algorithm [14] for the identification of boosted $\tau\tau$ and bb pair decays

in the X → YH analysis. The tagging efficiency of the algorithm will be studied in depth. The knowledge obtained about the behavior of the ParticleNet algorithm will provide important information for future boosted NMSSM di-Higgs analyses.

Following this introduction, Chapter 2 provides a breakdown of the SM and its supersymmetric extensions together with the X → YH process. Afterwards, the LHC and the CMS detector, which form the experimental setup for this analysis, are introduced in Chapter 3. Chapter 4 covers the dataset and object selection implemented to enrich boosted Higgs boson decays. A summary of the generator-level studies performed in this thesis is presented in Chapter 5. In Chapter 6 an introduction to the ParticleNet jet tagging algorithm and the studies on it are set out. Lastly, a conclusion of the results and an outlook is given in Chapter 7.

2 Theory

This chapter focuses on the theoretical background in the context of this thesis. The standard model, which forms the foundation of particle physics, is introduced in Section 2.1. This thesis studies a di-Higgs process within an extension of the standard model, both the extension and the process are presented in Section 2.2.

2.1 The standard model of particles physics

The information presented in this section is taken from [15, 16]. The standard model (SM) of particle physics is the most successful theory to describe the interactions of elementary particles. It stands today as the framework describing the interactions of elementary particles via the electromagnetic, weak, and strong force. The structure of the SM with its 17 composing elementary particles is depicted in Figure 2.1. It is divided into two groups of particles: fermions with spin $\frac{1}{2}$ and bosons with spin 0 or 1 in units of $\hbar$.

Fermions are divided into quarks and leptons with each containing six particles and their antiparticles. Fermions obey the Pauli exclusion principle [17], which states that two identical fermions in one system cannot occupy the same quantum state at the same time. This principle is responsible for fermions to form stable atoms making them the building blocks of all matter. Both groups of fermions can be further divided into three generations or families each containing two particles. Every generation shares a particle with the same electric charge with the other generations, they are positioned in the same row in Figure 2.1. Particles in higher generations get heavier and less stable. Because of this, everyday matter is solely made up of particles of the first generation.

The up (u), charm (c), and top (t) **quarks** have an electric charge of $\frac{2}{3}e$. The down (d), strange (s), and bottom (b) quarks have an electric charge of $-\frac{1}{3}e$. They all carry color charge, and because of color-confinement [18], quarks cannot exist as free particles. They only exist as hadrons, which are combinations of two (mesons) or three (baryons) quarks. Protons and neutrons, which together form an atomic nucleus, are both baryons consisting of combinations of u and d quarks from the first generation.

Leptons consist of the electron (e), muon (μ), and tau (τ) which share an electric charge of $-e$ as well as their corresponding neutrinos (ν_e,ν_μ,ν_τ) carrying no electric charge. Electrons orbit the nucleus of an atom making them the third building block of matter. In the SM, neutrinos are massless and they only interact via the weak force, which poses a challenge

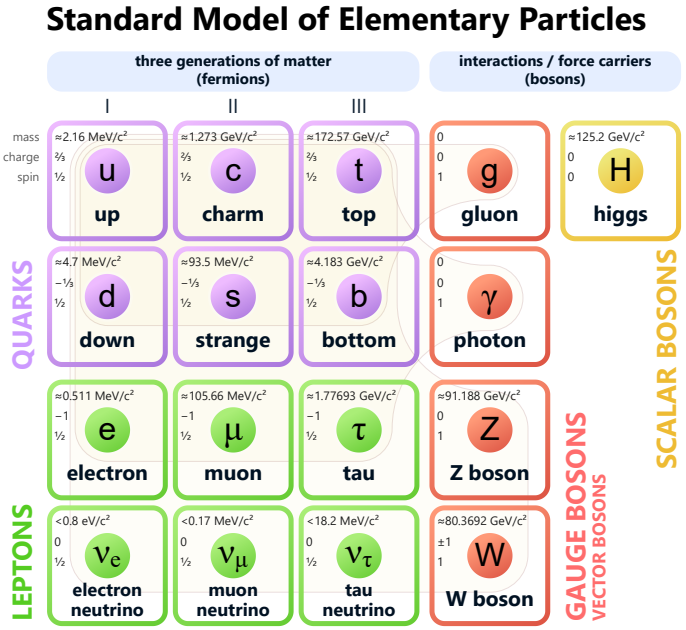


Figure 2.1: Elementary particles in the SM. Their rest mass, electric charge in units of e, and spin are specified in the top left corner of their box. Taken from [24].

to detect them. The observation of neutrino oscillations [6, 19] indicates that neutrinos do in fact have a small mass. This is one of many indications that there is physics beyond the SM.
An antiparticle has the same spin, mass, and mean lifetime as their corresponding particle but their electric charge and other charge-like properties are inverted [20]. This thesis uses the same notation for particles and their antiparticles, a τ can refer to a τ^- and a τ^+.

The standard model contains four vector (spin 1) **bosons** and one scalar (spin 0) **boson**. They do not obey the Pauli principle. Vector bosons are the mediators for the three fundamental interactions that are explained with the SM. The vector bosons consist of the massless and electrically neutral photon and gluons, the electrically neutral but massive Z boson and the electrically charged and massive W^+ and W^- bosons. Photons are exchanged in electromagnetic interactions, their range is infinite because photons are massless. Gluons are the mediators of the strong interaction. There are eight gluons in total and they carry color charge. That is why they can interact with each other limiting the range of the strong interaction to around 1 fm. Lastly, the Z and W bosons are exchanged in weak interactions. Because of their high mass the range of the weak force is limited to 10^{-3} fm.

The only scalar boson in the SM is the Higgs boson, it is the associated particle to the Higgs field. Massive fermions and bosons obtain their masses by coupling to the Higgs field. The Higgs boson was predicted to exist by the Brout-Englert-Higgs mechanism [21–23] in 1964 and was first observed at CERN in 2012 [1, 2], making it the last predicted particle of the SM being found.

2.2 Extensions to the standard model

While the SM can effectively predict the behavior of the elementary particles and their interactions, it also has many shortcomings. Besides the incorrect treatment of the neutrino mass in the SM, it fails to explain certain effects and observations like the fourth fundamental force, gravity, or dark matter [7] and dark energy [25]. In an effort to include such observations, physicists develop theories beyond the SM. One popular extension to the SM is supersymmetry (SUSY) [8] which postulates bosonic superpartners to every fermion and fermionic superpartners to every boson. The lightest superparticles are seen as dark matter candidates, but none of them have been detected so far. The minimal supersymmetric standard model (MSSM) [26] is the SM extension which implements SUSY while introducing the least amount of new fields and particles. To this day, no particles predicted by the MSSM have been observed experimentally constraining the theory severely. Physicist therefore lay focus on more complex theories like the next-to-minimal supersymmetric standard model (NMSSM) [9, 10]. It introduces one additional gauge-singlet superfield to the MSSM. The SUSY theories predict the existence of several Higgs bosons. In the MSSM there are a total of five predicted Higgs bosons. In the NMSSM two additional Higgs bosons are postulated for a total of seven Higgs bosons. There are three scalar bosons, the SM Higgs boson H, a light Higgs boson Y, and a heavy Higgs boson X. Additionally, there are two pseudoscalar Higgs bosons A_1 and A_2, and two charged Higgs bosons H^+ and H^-. The first three bosons mentioned are of importance for this thesis as all three participate in X $\rightarrow$ YH di-Higgs production.

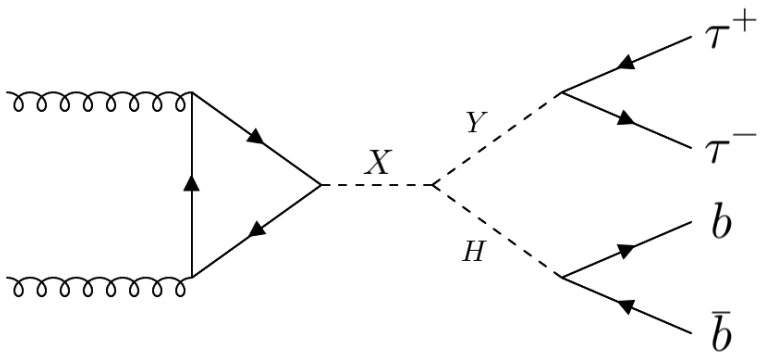


Figure 2.2: Example Feynman diagram of X $\rightarrow$ YH production and decay in proton-proton collisions. The heavy Higgs boson X is produced in gluon-gluon fusion. It decays into a light Higgs boson Y and the SM Higgs boson H. The Y boson further decays into a $\tau\tau$ pair and the H boson into a bb pair.

A Feynman diagram of the X $\rightarrow$ YH process is displayed in Figure 2.2. The X boson is produced in gluon-gluon fusion and decays into the H boson and the Y boson. The two bosons then further decay. In this thesis, the final state where the Y boson decays into a $\tau\tau$ pair and the H boson decays into a bb pair is studied. The bb pair produces collimated bundles of hadrons and the τ leptons decay according to Figure 2.3. The τ decays via weak interaction into a ν_τ and either a lepton-antineutrino pair of the first or second generation, or a quark-antiquark pair. The τ^+ decays in the same way but every particle has to be replaced by its antiparticle.
Since there is a $\tau\tau$ pair in this analysis, there are six possible combinations of decays. Either both τ leptons decay hadronically ($\tau_h\tau_h$), one decays hadronically and the other one leptonically ($\tau_h\mu$ and τ_he), or both decay leptonically (ee, $\mu\mu$, and eμ). The decay channels with two leptonic decays only have a branching fraction of 12% of all $\tau\tau$ decays [27] and are thus neglected in the scope of this thesis.

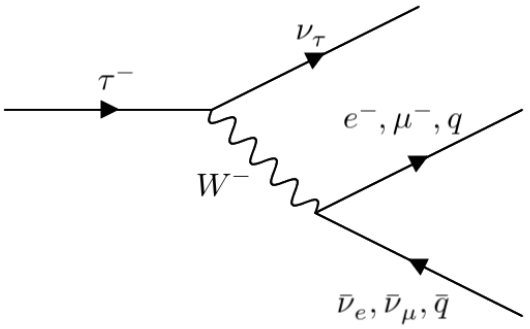


Figure 2.3: Feynman diagram of the τ decay. The τ^- decays into a ν_τ and either leptonically into an e or μ and their antineutrino, or hadronically into a quark-antiquark pair.

Particle detectors are needed to measure the particles appearing in these decays. In the following Chapter 3 the Compact Muon Solenoid detector and how it detects particles is described.

3 Experimental background

In order to test the predictions of theories like the NMSSM, which is presented in Chapter 2, particles are collided in scattering experiments. Large setups are needed that are capable of accelerating the particles to velocities close to the speed of light. In Section 3.1, the Large Hadron Collider (LHC) [28, 29], the particle accelerator providing the highest collision energy in the world located at the European Organization of Nuclear Research (CERN), is introduced. The activity in a collision is measured by different detectors located along the LHC. This thesis uses data samples from simulations of the Compact Muon Solenoid (CMS) detector, which is described in Section 3.2.

3.1 The LHC and the CERN accelerator complex

The information presented in this section is taken from [5, 28]. At CERN, there is a large complex of particle accelerators that are used jointly to reach high-energy particle beams. A sketch of the accelerator complex is displayed in Figure 3.1. Small accelerators serve as sources for particles, which then repeatedly get transmitted into bigger accelerators when they reach certain energies [30]. The LHC, which forms the final stage in the complex, is used to accelerate protons or lead ions.

In this thesis the products of proton-proton (pp) collisions are studied. In the following, the path of a proton through the complex is illustrate. Their journey starts in the Linear Accelerator 4 (Linac4), which accelerates negative hydrogen atoms H^-. While they get injected into the Proton Synchrotron Booster (PSB), the two electrons get stripped off the ions, leaving only the protons. The PSB accelerates the protons until they can get injected into the Proton Synchrotron (PS). The next station is the Super Proton Synchrotron (SPS). The final accelerator is the LHC, it accelerates two particle bunches up to an energy of 6.8 TeV each. With the LHC Run 3, which started in summer of 2022 [31], a record-breaking center-of-mass energy in pp collisions of $\sqrt{s} = 13.6\,\text{TeV}$ has been achieved. The LHC has a circumference of 26.7 km consisting of superconducting magnets and accelerating units. It has two separate beam pipes, in which the particle bunches travel in opposite directions. In total, 1232 dipole magnets are used to bend the beams in order for them to have a circular path, additionally, 392 quadrupole magnets are used to focus the beams. Right before collision, a quadrupole magnet focuses the beams to increase the luminosity of the LHC. The superconducting magnets are cooled down to -271.3 °C with superfluid helium so that they can function without resistance and energy loss.

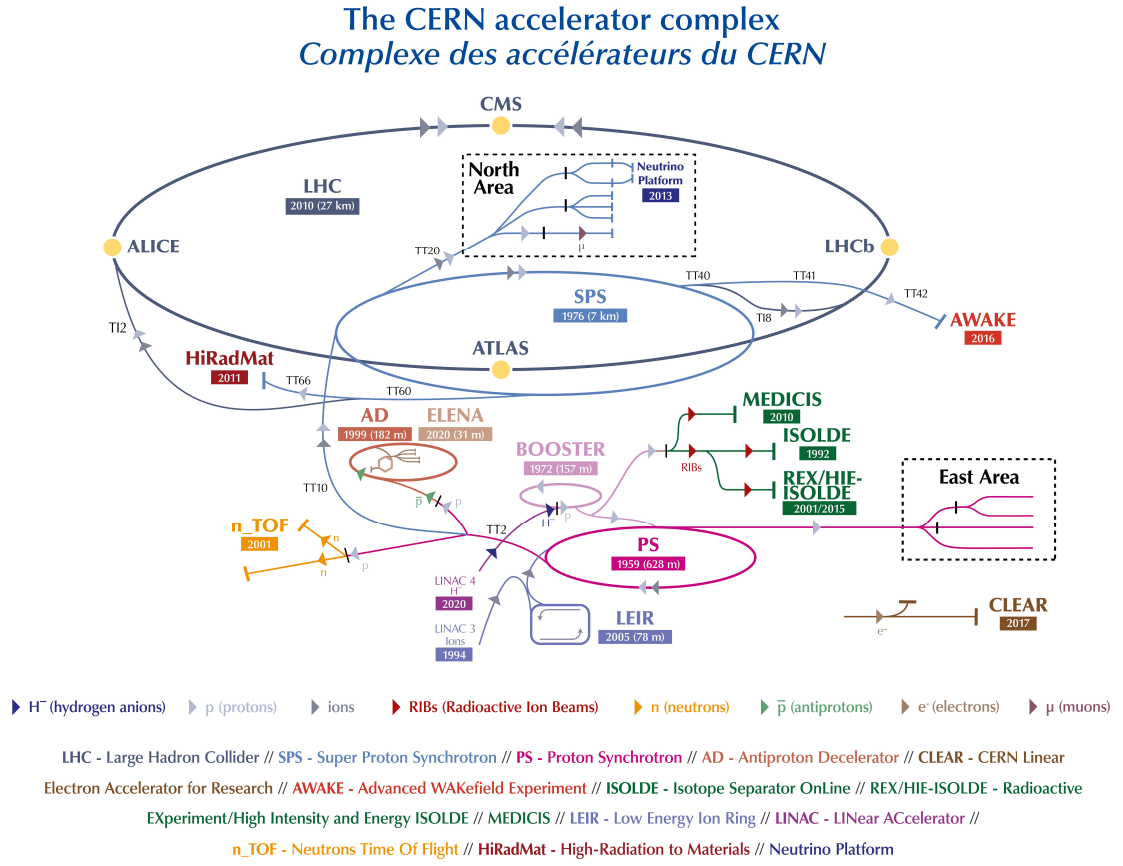


Figure 3.1: The CERN accelerator complex. Protons start in LINAC 4 as H^- atoms, the two electrons are stripped off on their way into the BOOSTER (PSB), they then enter the Proton Synchrotron (PS), after which they are transmitted to the Super Proton Synchrotron (SPS) and finally they get sent into the LHC. Taken from [34].

The collisions occur at four different points along the ring. There, the four large multipurpose particle detectors are located. The experiments at the LHC are ATLAS [4], CMS [3], LHCb [32], and ALICE [33].

3.2 The Compact Muon Solenoid

Simulations based on the response of the CMS detector are studied in this thesis. The CMS is a multipurpose experiment designed to study different phenomena, like the search for the Higgs boson and its properties or searches for physics beyond the SM [31]. The CMS detector is positioned cylindrically around the beam pipes with the collision happening in its center; a cutaway sketch is shown in Figure 3.2. It has an overall length of 22 m and a diameter of 15 m with a total weight of 14,000 t. It consists of several layers, representing subdetectors, that are specialized for identifying different types of particles.

The **silicon tracker** is the innermost detector layer surrounding the collision point. It is of 5.8 m length and has a diameter of 2.5 m composed first of pixel detectors and in the outer region of strip detectors. It is used to determine the transverse momentum of charged particles. The track of the particles is reconstructed by measuring the particles' interactions with the detector layers. With the tracks the collision point can be calculated [36, 37].

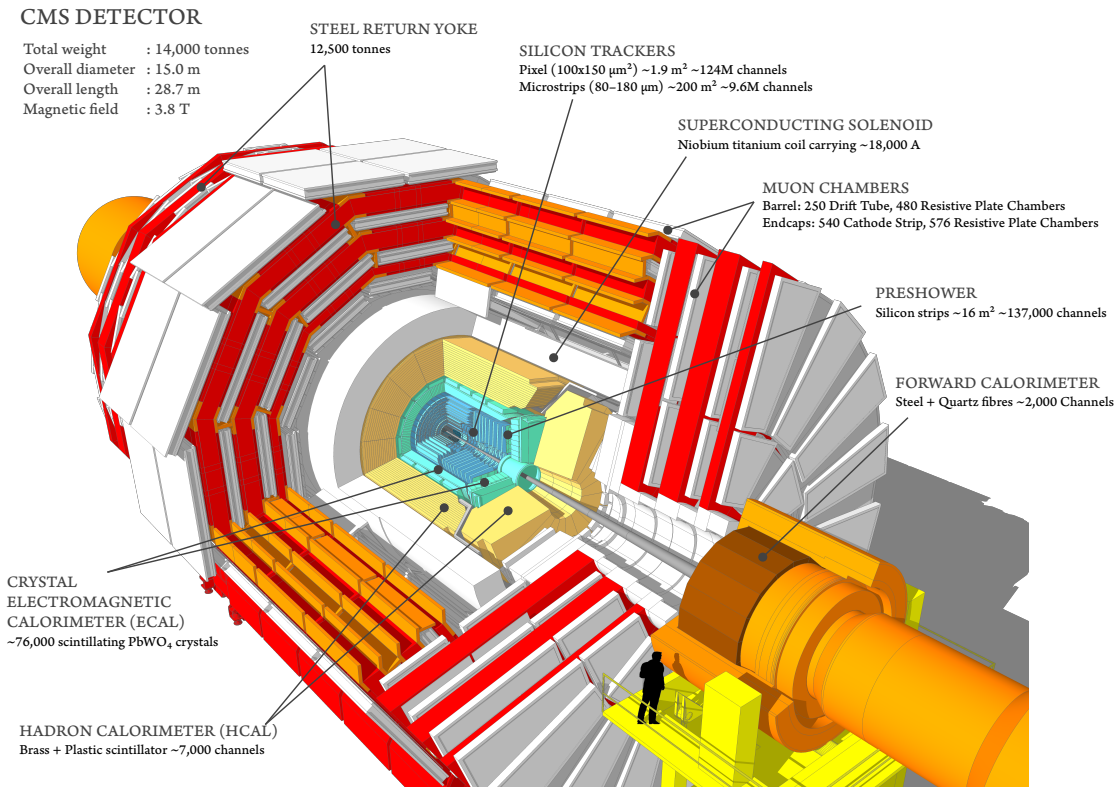


Figure 3.2: Cutaway diagram of the CMS detector. It cylindrically surrounds the beam pipes of the LHC with the pp collision happening in the center. Layers of subdetectors measure the resulting particles. Taken from [35].

The **electromagnetic calorimeter (ECAL)** [38] forms the second layer. It measures the energy deposition of electromagnetic showers produced by electrons and photons. It is made of lead tungstate ($PbWO_4$) crystals which have excellent characteristics for this application. They have a high density, short radiation length, and a small Molière radius.

The third layer is formed by the **hadron calorimeter (HCAL)** [39]. It measures the energy and direction of hadrons. The HCAL consists of altering layers of brass and steel plates as absorber material, and plastic scintillators. Hadrons flying through the absorber interact strongly with the nuclei of the plates and start hadronic showers.

The key feature of the CMS detector is the **superconducting solenoid** outside the tracker and calorimeters. It has a length of 12.5 m and an inner diameter of 6 m generating a 3.8 T magnetic field. The strong magnetic field is needed to bend the tracks of charged particles via the Lorentz force to determine the sign of their electric charge and their momentum through the radius of their trajectory.

Outside of the solenoid are **muon chambers** [40]. Their purpose is to detect and measure the momentum of muons, which are minimum ionizing particles and the only SM particles, besides neutrinos, that pass the previous detector layers. The magnetic field is significantly lower than inside the solenoid and points into the opposite direction, which changes the direction of the curvature of the trajectory. The chambers use three different technologies to detect the muons, drift tubes (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). The chambers are filled with a gas that ionizes when a muon travels through them. The produced free electrons drift towards positively charged wires, which induces a signal at the electrodes. This enables a precise measurement of the trajectory and momentum of a muon.

3.3 CMS coordinate system and kinematics

The Cartesian coordinate system (x, y, z) of the CMS detector has its origin at the nominal interaction point. The y axis points vertically upward and the x axis points towards the center of the LHC. Thus, the z axis points clockwise along the beam line [3]. The cylindrical shape of the detector favor the use of polar coordinates (r, θ, ϕ). The azimuthal angle ϕ is measured in the x-y plane starting at the x axis, the radial coordinate r denotes the distance to the collision point. The polar angle θ is measured from the z axis. The pseudorapidity η is normally used instead of θ, η is defined as

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \tag{3.1}$$

The transverse momentum p_{T} can be computed from the x and y components

$$\vec{p}_{\mathrm{T}} = \begin{pmatrix} p_{\mathrm{x}} \\ p_{\mathrm{y}} \end{pmatrix}. \tag{3.2}$$

The invariant mass m is equivalent to the mass of a particle in its rest frame. The four-momentum vector is expressed in terms of $(p_{\mathrm{T}}, \eta, \phi, m)$. With these coordinates, the dimensionless spatial distance ΔR of two objects can be calculated

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.3}$$

Here, $\Delta\eta$ denotes the difference in η and $\Delta\phi$ the azimuthal difference in ϕ of the two objects in radians.

This thesis works with natural units where $\hbar = c = k = 1$. Most notably this changes the units of mass and momentum to be equal to the unit of energy, eV.

4 Boosted Y and H decays

This thesis analyses boosted Y and H decays in the pp $\rightarrow$ X $\rightarrow$ YH process at the LHC by using the ParticleNet algorithm. A boosted decay occurs when the decaying particle has a high transverse momentum and therefore a high Lorentz boost. Requirements for the simulated events can be set that allow mainly boosted decays, these requirements are presented in this chapter. In Section 4.1, the data generation via simulation is explained and the samples that are chosen for this analysis are listed. To further enrich boosted decays an object selection within each sample is performed, Section 4.2 takes a closer look at this.

4.1 Simulation-based data generation

Since the fundamental interactions are probabilistic in nature, the testing of a theoretical prediction requires a strong statistical analysis only enabled by large datasets. Due to the complexity of real data, their interpretation requires simulations of interactions in event generators using the MC method. The NMSSM di-Higgs production process can be simulated with different mass hypotheses for both the X and the Y boson and with the desired final state decays. Simulated events are used in this study, because they contain generator-level information that can be used to verify how well the ParticleNet jet tagging works. The simulated data is produced with LHC Run 3 conditions in summer 2022, with a collision energy of 13.6 TeV. In this thesis, 19 different hypotheses of the X boson mass M_{X} are used. For all of them the mass of the Y boson M_{Y} is set to 125 GeV, which is equivalent to the mass of the SM Higgs boson M_{H}. Only the X $\rightarrow$ Y($\tau\tau$)H(bb) final state is considered, the particles enclosed in brackets represent the decay products of the particle preceding the respective bracket. The range of M_{X} goes from 300 GeV to 4000 GeV, the exact mass values are listed in Table 4.1. The reason behind fixing M_{Y} is to only analyze the dependence of M_{X} and not having to work with two variables. Because M_{Y} equals M_{H}, the X $\rightarrow$ Y(bb)H($\tau\tau$) and X $\rightarrow$ Y($\tau\tau$)H(bb) final states are identical, it does not make a difference which one is analyzed.

The distance ΔR can be used to get information about the boost of the decay. Boosted decays have a small $\Delta R \leq 0.8$ between the particles of a particle pair. Figure 4.1 visualizes the ΔR between the τ leptons that originate from the Y boson and the ΔR between the bottom quarks originating from the SM Higgs boson for four different M_{X}. For simulated events with high M_{X} the particle pairs are strongly boosted. The simulated event with $M_{\mathrm{X}} = 300$ GeV, the lowest M_{X} used in this thesis, barely contains any boosted decays.

Table 4.1: X boson masses in samples used for this analysis.

X boson mass [GeV]									
300	400	500	550	600	650	700	800	900	
1000	1200	1400	1600	1800	2000	2500	3000	3500	4000

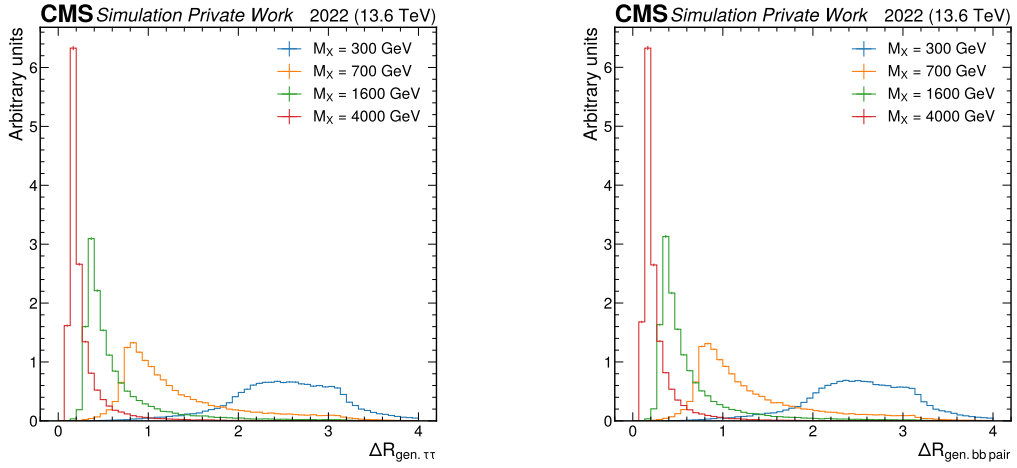


Figure 4.1: ΔR between both generator-level particles of a $\tau\tau$ (left) and bb pair (right). Four different M_X hypotheses are shown. The shapes are normalized to the same number of events.

4.2 Object selection

The main objects that are studied in this thesis are boosted jets. To identify particles in the CMS detector a particle-flow algorithm is used [41]. It reconstructs the particles with the combined information of the different detector layers [42]. The reconstructed particles are then clustered from their four-momentum into a particle jet. At CMS the anti-k_t jet clustering algorithm [43] is used for this application. In the boosted X $\rightarrow$ YH analysis the two jets produced by both τ leptons of the $\tau\tau$ pair or both b quarks of the bb pair overlap. Because of this, it is not possible to reconstruct both jets individually. The two jets are merged to one wide-cone AK8 jet. AK8 refers to the anti-k_t algorithm and the radius parameter used for the jet, in this case it is $R = 0.8$.
If the particles in a boosted decay have high transverse momenta, so does the AK8 jet they are reconstructed with. Therefore, only AK8 jets with a transverse momentum above a threshold of $p_T \geq 200\,\text{GeV}$ are considered. This threshold is suitable because the ParticleNet neural network was trained for AK8 jets above this value [44]. The reason behind their high momentum threshold is the difference in mass of the X boson and its decay products.
A second selection is applied to the AK8 jets regarding their pseudorapidity η, the CMS tracker system has a range of $|\eta| \leq 2.5$ [45], AK8 jets with η outside of this range therefore are not analyzed.
In Chapter 5 generator-level studies are performed, two sets of selection criteria are applied there, one for the τ leptons and one for the b quarks. For both their mother particle has to be either the Y or H boson. It is also important to use their kinematic properties right after the boson decay and not at a time where they might have lost momentum by emitting a photon or through other interactions.
In Chapter 6 information about electrons and muons near an AK8 jet is needed, for both of them, a minimum transverse momentum of $p_T \geq 10\,\text{GeV}$ is used. They both have their

own pseudorapidity range, $|\eta| \leq 2.5$ for electrons and $|\eta| \leq 2.4$ [45] for muons.
For all upcoming studies these selection criteria will be applied, starting with the studies on generator-level τ leptons and b quarks as well as reconstruction-level AK8 jets in Chapter 5.

5 Studies at generator level

A benefit of using simulated events over data measured at real collider experiments is the existing information about generator-level particles. Being able to use the properties of generator-level particles allows to validate the reconstruction-level results and test the performance of the algorithms.
This chapter focuses on generator-level studies performed in this thesis. In Section 5.1 the reconstruction of generator-level τ decays is discussed. Then, in Section 5.2 the process of matching the reconstruction-level AK8 jets to the generator-level particles is described.

5.1 Identification of the generator-level τ decay channel

Knowing which $\tau\tau$ pair decay occurs in each event is mandatory to be able to compare it to the reconstruction-level results. Therefore, a study is conducted that analyzes the generator-level $\tau\tau$ pair decay of every event. At first, the two generator-level τ leptons that directly originate from the Y boson decay are selected. A Y boson with a mass of 125 GeV is simulated, the four-vector sum of the two τ leptons should therefore have an invariant mass $m_{\tau\tau}$ equal to that of the Y boson. A histogram of the invariant masses of the combined τ leptons from the Y boson decay is shown in Figure 5.1, it centers around the expected value of $M_{\mathrm{Y}} = 125\,\mathrm{GeV}$. It can be concluded that the correct τ leptons are identified.
The next step is to search for electron (ν_{e}) or muon (ν_{μ}) neutrinos that are produced by a τ lepton decay. They serve as a starting point for an iteration that takes a particle, determines its mother particle and checks if it corresponds to one of the selected τ leptons from the Y boson decay. The iteration repeats itself with the mother particle until either one of the selected τ leptons or a generator-level particle that exists before the τ leptons is found. A neutrino counter $n_{\nu_{e/\mu}}$ counts the number of ν_{e} and ν_{μ} whose mother particle is one of the previously selected τ leptons. After all neutrinos have passed the iteration, the counter will be examined according to Table 5.1.

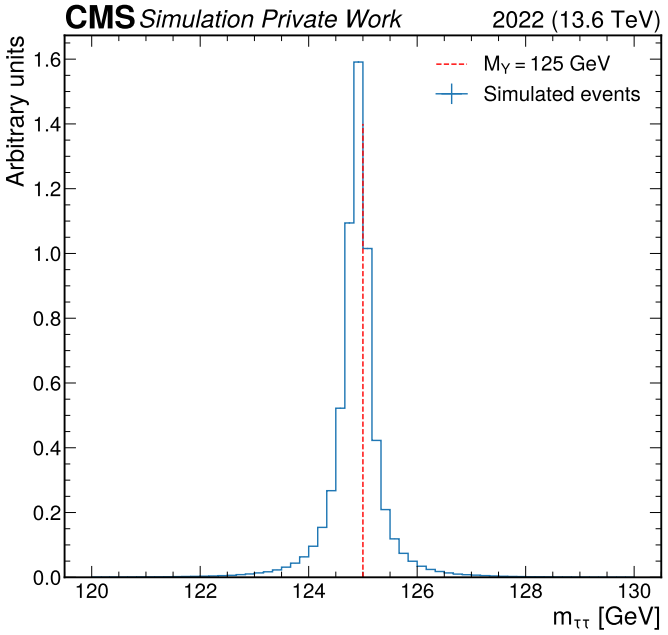


Figure 5.1: Reconstructed Y boson mass distribution by combination of τ leptons. The expected value of $M_{\mathrm{Y}} = 125\,\mathrm{GeV}$ is shown as a dashed line.

Table 5.1: Number of electron (muon) neutrinos n_{ν_e} (n_{ν_μ}) expected to appear in generator-level $\tau\tau$ pair decays for different decay channels. By counting the neutrinos from a generated $\tau\tau$ pair decay, the decay channel can be identified.

$\tau\tau$ pair decay channel	n_{ν_e}	n_{ν_μ}
$\tau_{\mathrm{h}}\tau_{\mathrm{h}}$	0	0
$\tau_{\mathrm{h}}\mathrm{e}$	1	0
$\tau_{\mathrm{h}}\mu$	0	1
$\mathrm{e}\mu$	1	1
ee	2	0
$\mu\mu$	0	2

The results of this study are displayed in Table 5.2. It contains the number of occurrences n of each $\tau\tau$ pair decay channel considered in this thesis. All mass hypotheses for the X boson are used and their respective $\tau\tau$ decay channel occurrences are added up. The total number N of all $\tau\tau$ pair decays is given, too. The total number includes the decay channels with two leptonic decays. The fractions p of events in a decay channel relative to the total number of all $\tau\tau$ pair decays and the expected fractions according to the Particle Data Group [27] are given as well. The occurrences of the $\tau\tau$ pair decays are binomially distributed, the variance σ^2 of binomial distributions

$$\sigma_n^2 = N \cdot p \cdot (1-p), \tag{5.1}$$

is used to calculate the statistical uncertainties σ of the measured fractions

$$\sigma_{fraction} = \frac{\sigma_n}{N} = \frac{\sqrt{N \cdot p \cdot (1-p)}}{N} = \sqrt{\frac{p \cdot (1-p)}{N}}. \tag{5.2}$$

The observed fractions in the simulated events comply with the expected fractions for the $\tau_{\mathrm{h}}\mathrm{e}$ and $\tau_{\mathrm{h}}\mu$ final states within the scope of measurement accuracy. This is not the

case for the $\tau_h\tau_h$ final state, which occurs 0.07% more often than expected. Systematical uncertainties are likely the cause for this and this generator-level study is deemed successful. The results will be used in Chapter 6.

Table 5.2: Number and fraction of events of every considered $\tau\tau$ pair decay channel for all simulated mass hypotheses.

$\tau\tau$ pair decay channel	Number of occurrence	Fraction in simulation [%]	Expected fraction [%] [27]
τ_he	337,264	23.03 ± 0.03	23.09 ± 0.06
$\tau_h\mu$	329,768	22.51 ± 0.03	22.53 ± 0.06
$\tau_h\tau_h$	617,639	42.17 ± 0.04	41.98 ± 0.08
1,464,676 total $\tau\tau$ pair decays			

5.2 Matching of AK8 jets to generator-level particles

This thesis is interested in the efficiency of the ParticleNet jet tagging algorithm for the NMSSM di-Higgs analysis. Every AK8 jet gets scores given by the algorithm that evaluate how likely a jet was initiated by a certain particle. To calculate the tagging efficiency, the information about which generator-level particle decays initiate the reconstructed AK8 jets is needed. Therefore, a second generator-level study is performed to match reconstruction-level AK8 jets to the generator-level τ lepton or b quark pairs. This matching is done by calculating the ΔR between the AK8 jets and the generator-level particles. Because of the conservation of momentum the sum of all decay products will travel in the same direction as their parent particles did. Since AK8 jets are a good approximation for the sum of all decay products, as they generally contain most of the decay products, the ΔR between an AK8 jet and the original generator-level particle is supposed to be small. As the radius of an AK8 jet cone roughly corresponds to $R = 0.8$ in the η-ϕ plane, it is sufficient to match two objects if $\Delta R(\text{gen. particle}, \text{AK8 jet}) < 0.8$ applies. An AK8 jet will only be matched to a pair if it meets the minimum ΔR criteria for both constituents of the generated particle pair. The properties of the four-vector sum of the generated particle pair cannot be used in this case because it is possible for the particles to have a large ΔR between them but their four-vector sum is close to an AK8 jet originating from another source. If there are two AK8 jets that match to a particle pair, the AK8 jet with the smaller ΔR to the four-vector sum of the two generated particles is chosen as the matched jet.
The ΔR between AK8 jets and their matched particles for three representative values of M_X is displayed in Figure 5.2. The histograms have a sharp cutoff at a spatial distance of $\Delta R = 0.8$, showing that the matching process works. The distributions show big differences between the values of M_X. The higher the X boson mass, the smaller the ΔR between the AK8 jet and the generated particle pair. This can be explained by the stronger boost of the particles caused by the large difference in mass of the X boson and its decay products. There is no difference for the ΔR between the AK8 jets and τ leptons and the ΔR between the AK8 jets and b quarks besides a small peak at $\Delta R_{\text{gen.matched }\tau,\text{AK8 jet}} = 0$ for low M_X masses in the left plot of Figure 5.2. The origin of this peak must be investigate further.
Additionally, the ΔR between the AK8 jets and $\tau\tau$ and bb pairs is displayed in Figure 5.3. This time, there is a difference in the distributions for the τ leptons and the b quarks, the curves for the b quarks are slimmer. This can be explained by the neutrinos produced in the τ lepton decay, which cannot be detected by the CMS detector. They change the momentum of the τ leptons slightly, which leads to a bigger ΔR.

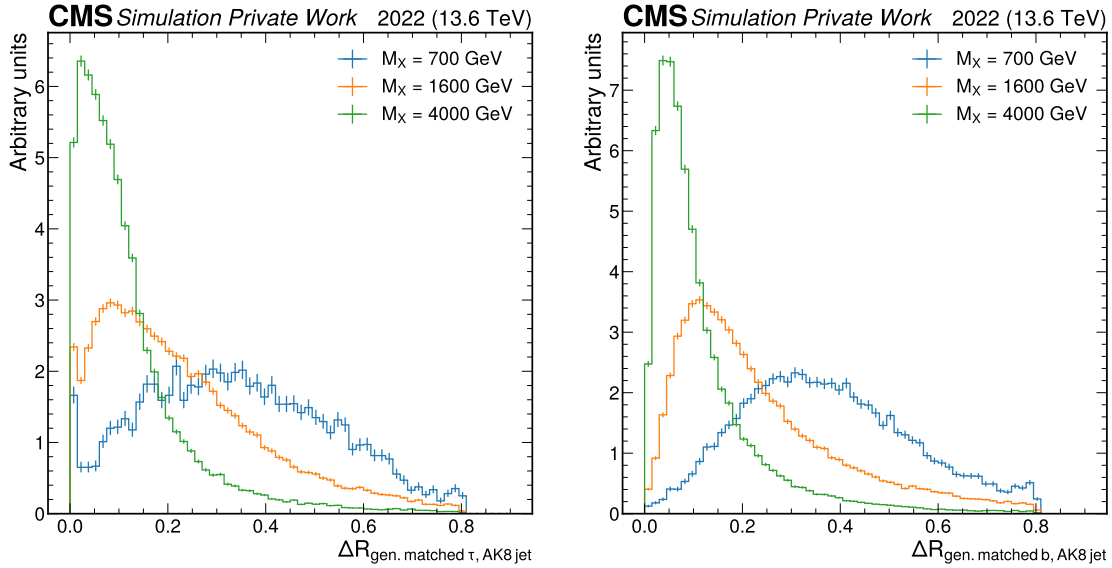


Figure 5.2: ΔR between AK8 jets and τ leptons (left) or b quarks (right) if they are matched to each other for different mass hypotheses. The shapes are normalized to the same number of events.

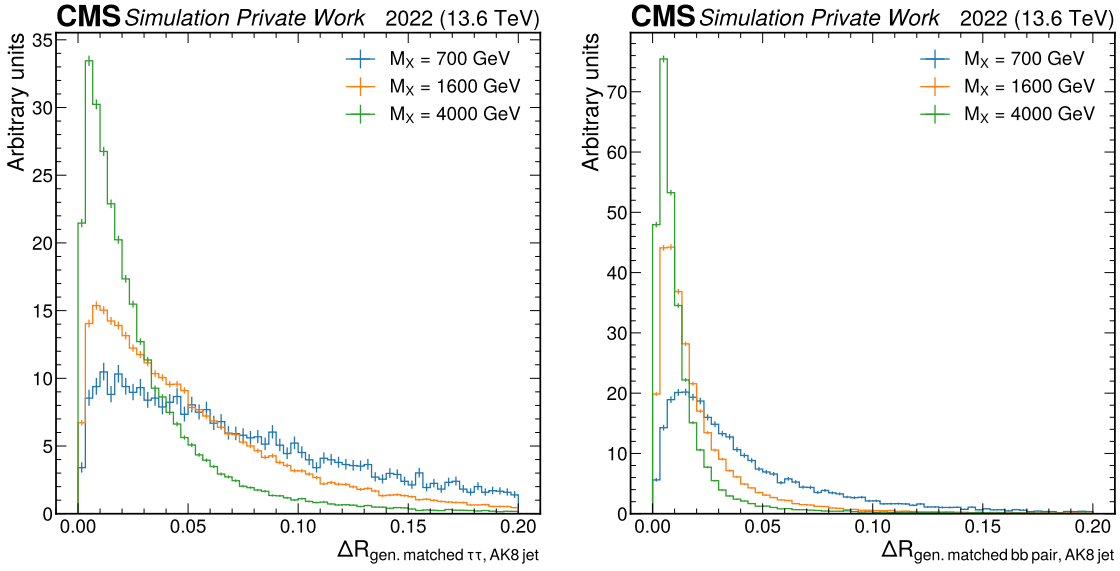


Figure 5.3: ΔR between AK8 jets and $\tau\tau$ (left) or bb (right) pair if they are matched to each other for different X boson mass hypotheses. The shapes are normalized to the same number of events.

6 Studies on ParticleNet

With the methods presented in Chapter 5, the ParticleNet jet tagging efficiency can be explored. Section 6.1 starts with a small introduction to the ParticleNet neural network. The results of the studies and a discussion can be found in Section 6.2.

6.1 Introduction to ParticleNet

This section provides an overview of the essentials of the ParticleNet architecture, a more detailed explanation can be found in [14]. ParticleNet is a neural network architecture for jet tagging problems. It provides scores for jets that express the probability of the jet originating from a certain particle decay. In this thesis, ParticleNet is used to tag decay products of boosted particles that are clustered into one AK8 jet. These boosted particles are the $\tau\tau$ and bb pairs. For this analysis five of the ParticleNet raw scores are important: Three raw scores for each $\tau\tau$ pair decay channel, p_{te} for $\tau_{\mathrm{h}}\mathrm{e}$, p_{tm} for $\tau_{\mathrm{h}}\mu$, and p_{tt} for $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. Additionally, there is a raw score for the bb pair decay which will be called p_{bb} and lastly a raw QCD-score called p_{QCD} for jets produced in pure QCD processes. The first four raw scores can be put in relation to p_{QCD} to get ratios

$$r_{\mathrm{te}} = \frac{p_{\mathrm{te}}}{p_{\mathrm{te}} + p_{\mathrm{QCD}}}, \tag{6.1}$$

where p_{te} is taken as an example, the calculation is the same for the other raw scores. For the remainder of this thesis these ratios will be referred to as scores. If a jet is similar to jets produced in pure QCD processes, then their four scores are low. The score of a jet will rise if the jet has more similarity to a jet of the respective target decay than to a jet produced in pure QCD processes. To get the raw scores the equation can be transformed

$$p_{\mathrm{te}} = \frac{r_{\mathrm{te}} \cdot p_{\mathrm{QCD}}}{1 - r_{\mathrm{te}}}. \tag{6.2}$$

One problem with the raw scores is the denominator, as it can be zero, so the raw scores cannot be calculated in all cases. In Figure 6.1 the four $r_{\mathrm{te,tm,tt,bb}}$ score distributions with every AK8 jet considered are shown. The r_{bb} score is evenly distributed with small peaks at zero and one. The scores for the $\tau\tau$ pair decays on the other hand are a lot more concentrated on these edge values, the neural network confidently tags them.

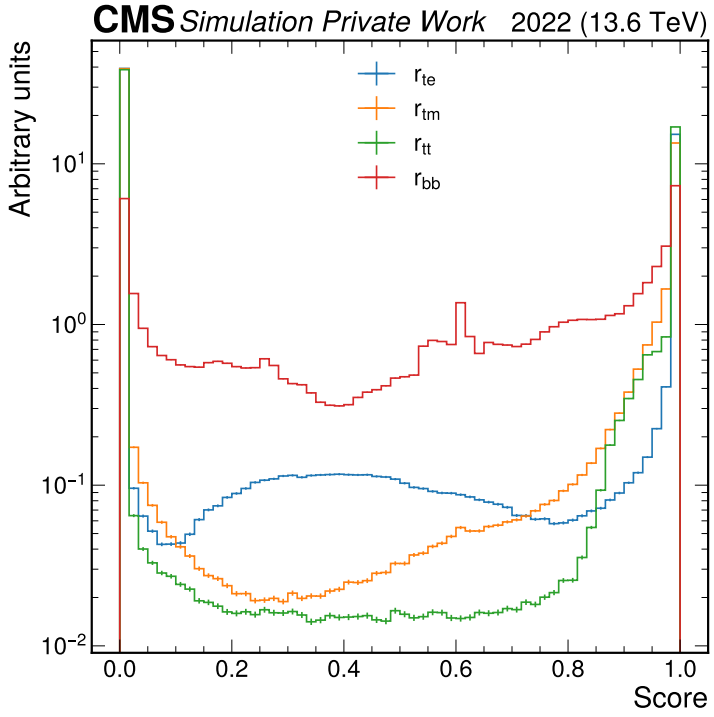


Figure 6.1: ParticleNet score distributions for every AK8 jet. The distributions are normalized to the same total number of events.

6.2 Analyzing ParticleNet jet tagging scores

The highest ParticleNet score of an AK8 jet is used to determine the decay that initiated it. For example, if the r_{te} score is the highest score for an AK8 jet, the AK8 jet is tagged as a $\tau_{\mathrm{h}}\mathrm{e}$ decay. By comparing the tagging results with the generator-level information, the efficiency of the algorithm can be calculated. Various tests with the ParticleNet scores are performed in this thesis. The final results with the best agreement to the generator-level studies gathered in Chapter 5 will be presented in this section. At first the efficiency and purity of the ParticleNet jet tagging is discussed in Section 6.2.1. In Section 6.2.2 a deeper look at misidentified AK8 jets is taken. Lastly, the $\tau\tau$ pair tagging efficiency distribution for the $\tau\tau$ pair and AK8 jet p_{T} as well as the AK8 jet η is studied in Section 6.2.3.

6.2.1 Efficiency and purity

For every AK8 jet, information about the generator-level decay that initiated it is compared to the decay they are tagged as. By counting the number of AK8 jets matched to a generator-level decay A and being tagged as a decay B, the tagging efficiency can be calculated. With the five types of decays, $\tau_{\mathrm{h}}\mathrm{e}$, $\tau_{\mathrm{h}}\mu$, $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, bb, and QCD events, a total of 25 values c are gathered. For example, the number of AK8 jets that are matched to a generator-level $\tau_{\mathrm{h}}\mathrm{e}$ decay that get tagged as a $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay is $c_{\tau_{\mathrm{h}}\mathrm{e},\tau_{\mathrm{h}}\tau_{\mathrm{h}}}$. A good visual representation of the tagging efficiency and purity can be achieved by constructing confusion matrices. The efficiency matrix shows the fractions p of a generator-level decay getting tagged by the individual scores. Every count for the same generator-level decay is added to get the total number of AK8 jets that stem from this generator-level decay. By dividing

the single counts by the total number, the fractions are calculated as

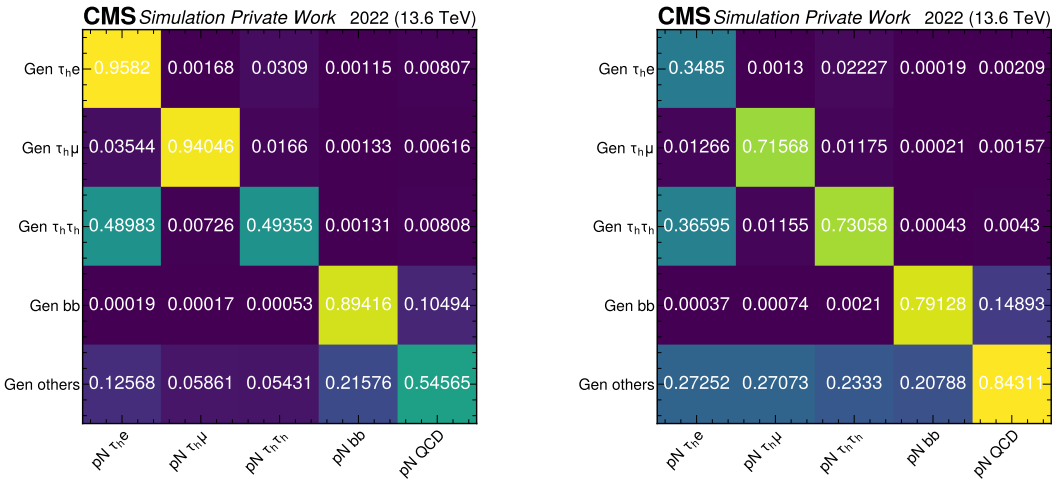
$$p_{\tau_\mathrm{h}\mathrm{e},\tau_\mathrm{h}\tau_\mathrm{h}} = \frac{c_{\tau_\mathrm{h}\mathrm{e},\tau_\mathrm{h}\tau_\mathrm{h}}}{c_{\tau_\mathrm{h}\mathrm{e},\tau_\mathrm{h}\mathrm{e}} + c_{\tau_\mathrm{h}\mathrm{e},\tau_\mathrm{h}\mu} + c_{\tau_\mathrm{h}\mathrm{e},\tau_\mathrm{h}\tau_\mathrm{h}} + c_{\tau_\mathrm{h}\mathrm{e},\mathrm{bb}} + c_{\tau_\mathrm{h}\mathrm{e},\mathrm{QCD}}}. \tag{6.3}$$

The purity matrix shows the fractions of a tagging score distributed over the generator-level decays. Here, every count with the same tagging score is added to a total number of AK8 jets that are tagged as a certain decay. The fractions are again calculated by dividing the single counts by the total number of AK8 jets

$$p_{\tau_\mathrm{h}\mathrm{e},\tau_\mathrm{h}\tau_\mathrm{h}} = \frac{c_{\tau_\mathrm{h}\mathrm{e},\tau_\mathrm{h}\tau_\mathrm{h}}}{c_{\tau_\mathrm{h}\mathrm{e},\tau_\mathrm{h}\tau_\mathrm{h}} + c_{\tau_\mathrm{h}\mu,\tau_\mathrm{h}\tau_\mathrm{h}} + c_{\tau_\mathrm{h}\tau_\mathrm{h},\tau_\mathrm{h}\tau_\mathrm{h}} + c_{\mathrm{bb},\tau_\mathrm{h}\tau_\mathrm{h}} + c_{\mathrm{QCD},\tau_\mathrm{h}\tau_\mathrm{h}}}. \tag{6.4}$$

With these calculations, the contents of the efficiency matrix add up to one for every row, while the contents of the purity matrix add up to one for every column.

A first tagging efficiency without further distinctions is calculated, the efficiency matrix is displayed in Figure 6.2a and the purity matrix in Figure 6.2b. The rows correspond to the generator-level decay a AK8 jet is matched to, see Chapter 5, with "Gen others" referring to every AK8 jet that is not matched to any $\tau\tau$ or bb pair decay. The columns correspond to the class to which the reconstructed AK8 jet is assigned to. While this works well for the $\tau_\mathrm{h}\mathrm{e}$, $\tau_\mathrm{h}\mu$, and bb decay, there emerges an unexpected behavior for the $\tau_\mathrm{h}\tau_\mathrm{h}$ decay. Around half of the generator-level $\tau_\mathrm{h}\tau_\mathrm{h}$ decays are tagged as $\tau_\mathrm{h}\mathrm{e}$ decays. The cause of this behavior is that for a lot of AK8 jets the r_te and r_tt scores are equal. This can be seen in Figure 6.3. It shows the difference of the two scores for AK8 jets matched to either one of the decays. The differences are dominated by discrete values, most likely due to the compressed data format of the simulated events. A discrete distribution of the scores only appears for high scores, as can be seen in Figure 6.4 for the r_te scores.



(a) Efficiency (b) Purity

Figure 6.2: Confusion matrices for efficiency (a) and purity (b) of ParticleNet jet tagging scores. The x axis corresponds to which decay the reconstructed AK8 jets are assigned to. The y axis corresponds to the generator-level decay to which the reconstructed AK8 jets are matched to.

The next important step is to define an additional selection criterion to distinguish the $\tau\tau$ pair decays if two scores have the same value. This can be achieved by searching for particles within the AK8 jet that only appear in one of the decays, for example, an electron

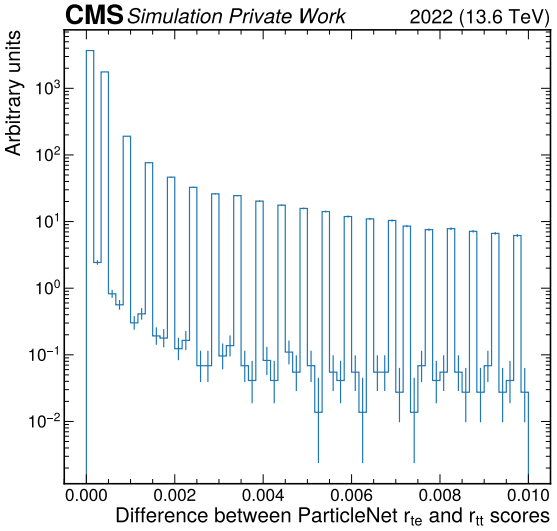
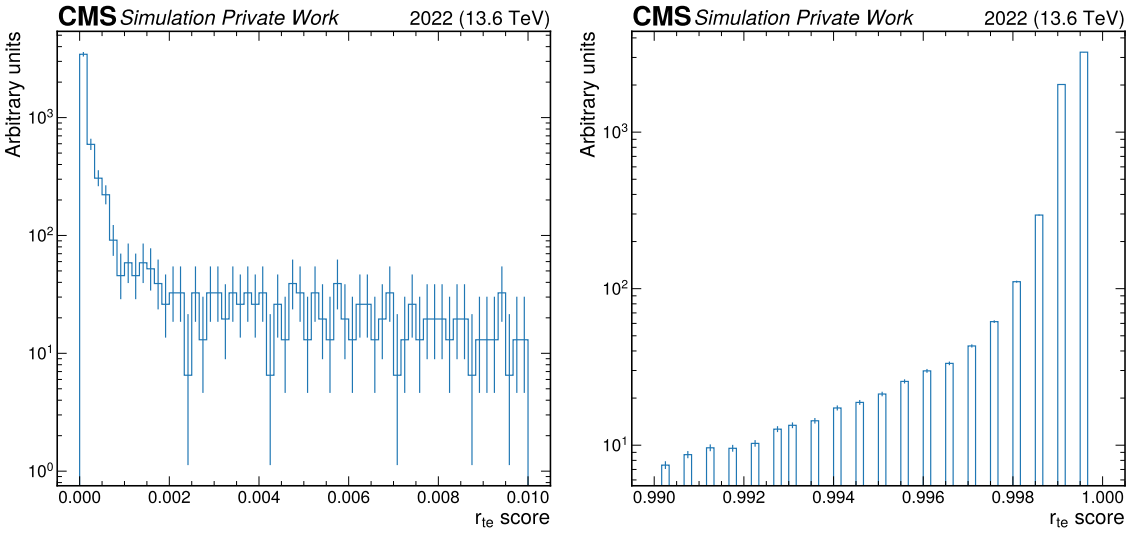


Figure 6.3: Difference of r_{te} and r_{tt} scores with logarithmic y scale. A discrete behavior of the values is visible.

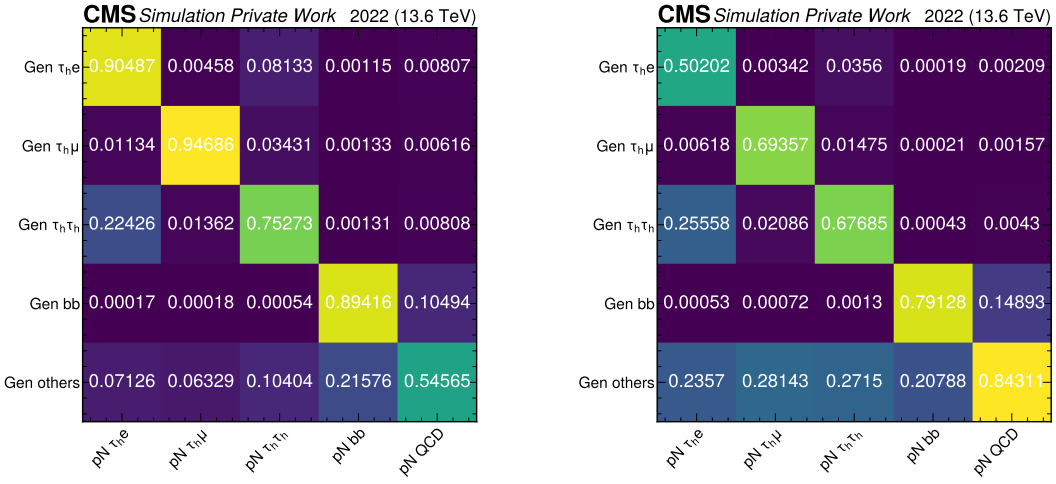
in the τ_{h}e decay, and a muon in the $\tau_{\mathrm{h}}\mu$ decay. If an electron is measured in a range of $\Delta R \leq 0.8$ to the AK8 jet where r_{te} equals r_{tt}, the jet is tagged as a τ_{h}e decay. Otherwise, if no electron is present, the AK8 jet is tagged as a $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay. The same procedure is done for muons.
The confusion matrices obtained with this extra selection criterion of an electron or a muon detected in the AK8 jets are displayed in Figure 6.5. This increases the efficiency for the r_{tt} score by around 50% while only decreasing it by a small amount for the r_{te} and r_{tm} scores. Both matrices have the highest values along their diagonal, especially the $\tau_{\mathrm{h}}\mu$ decay has a very high efficiency. About 22% of generator-level $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decays are still classified as τ_{h}e decays by ParticleNet. There is a confusion between the bb and QCD AK8 jets, too. While about 10% of the AK8 jets matched to a generator-level bb pair are tagged as a QCD event, about 22% of the jets not originating from a $\tau\tau$ or bb pair are tagged as bb jets. This is an expected behavior as a bb decay is similar to a QCD event since both most of the times are expressed by a hadronization. There also are AK8 jets that contain only one b jet together with another quark jet. They contribute to the 22%, if they are tagged as a bb decay. With the same reasoning the 10% of AK8 jets not matched to a $\tau\tau$ or bb pair tagged as a $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay can be explained. In the purity matrix, for each jet tagging class the fraction of jets that are neither matched to a bb or a $\tau\tau$ pair is greater than 20%. For the τ_{h}e and $\tau_{\mathrm{h}}\mu$ decays these high fractions partly come from the leptonic $\tau\tau$ pair decays. Again it shows that a lot of AK8 jets tagged as a τ_{h}e decay are actually a $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay. In the future it would be beneficial to look at ways to better distinguish between those two decays. Generally, the tagging efficiency is good but might still need improvement for the different $\tau\tau$ pair decays. An additional approach to increase the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ tagging efficiency is to demand a high quality in the identification of the leptons. But this leads to a large decrease of the τ_{h}e tagging efficiency and is therefore dismissed.



(a) r_{te} scores close to zero.

(b) r_{te} scores close to one.

Figure 6.4: Distribution of r_{te} score close to zero (a) and close to one (b). Continuous values for low scores and discrete values for high scores are visible.

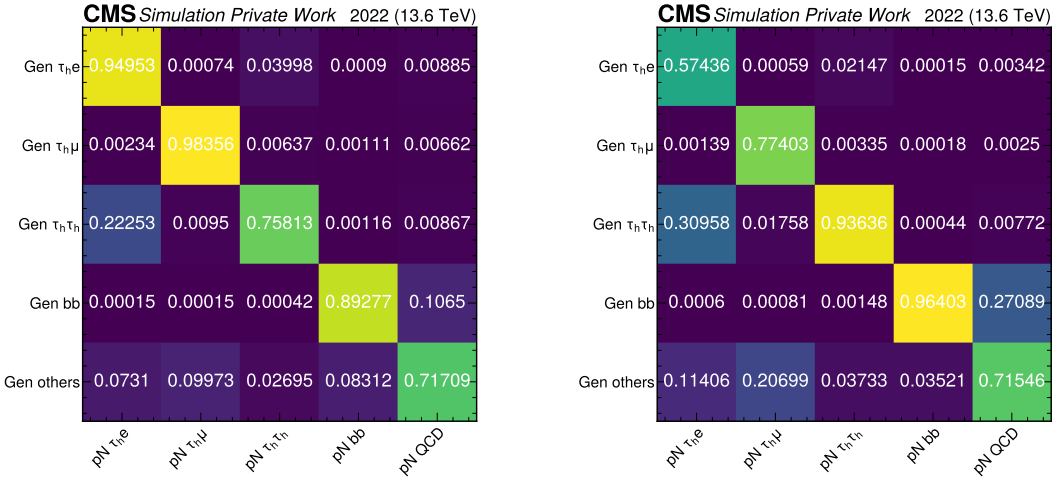


(a) Efficiency

(b) Purity

Figure 6.5: Confusion matrices for efficiency (a) and purity (b) of ParticleNet jet tagging scores together with information about electrons and muons in the jet. The x axis corresponds to which decay the reconstructed AK8 jets are assigned to. The y axis corresponds to the generator-level decay to which the reconstructed AK8 jets are matched to.

Instead of using every AK8 jet, only AK8 jets that meet certain criteria can be considered in order to improve the tagging efficiency. Cutting away AK8 jets significantly reduces the number of analyzed objects in exchange of high tagging efficiency. One criterion to look at is a threshold to the softdrop mass m_{softdrop} [46] of the AK8 jets. With soft drop declustering, soft wide-angle radiation of an AK8 jet is removed to improve jet reconstruction. Only AK8 jets with $m_{\text{softdrop}} \geq 30\,\text{GeV}$ are now considered. This threshold is chosen because the ParticleNet algorithm is trained in this m_{softdrop} region. With such AK8 jets the efficiency and purity matrices change, the new matrices are displayed in Figure 6.6. This increases the efficiency of the τ_he, $\tau_\text{h}\mu$ and QCD tagging without lowering it for the $\tau_\text{h}\tau_\text{h}$ and bb tagging. This also improves the purity, though more bb decays are now tagged as QCD events.



(a) Efficiency matrix

(b) Purity matrix

Figure 6.6: Confusion matrices for efficiency and purity of ParticleNet jet tagging scores for AK8 jets with softdrop mass above 30 GeV. The x axis corresponds to which decay the reconstructed AK8 jets are assigned to. The y axis corresponds to the generator-level decay to which the reconstructed AK8 jets are matched to.

6.2.2 Distribution of softdrop mass

After showing the efficiency of the tagging process, it is interesting to look at why the misidentification happens. As an example, the softdrop mass distribution of selected jets is considered and split into the different true origins of the jets. The histograms can be viewed in Figure 6.7. For this analysis only the sample with a X boson mass of $m_{\mathrm{X}} = 4000\,\mathrm{GeV}$ is used.

The purity matrix, Figure 6.5b, shows that a large fraction of the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decays are tagged as a τ_{h}e decay. This can also be observed in Figure 6.7a, which shows the distribution of the softdrop mass for AK8 jets that are tagged as a τ_{h}e decay. The $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decays make up a big fraction especially for higher softdrop masses. Other $\tau\tau$ pair decay channels like the ee or μe final states make up a small fraction, too. The signature of the ee decay is more similar to the τ_{h}e decay than to the other decays and since the ee decay does not have its own ParticleNet score, it is tagged as a τ_{h}e decay. The μe decays will also partly be tagged as a τ_{h}e decay for this reason, but also partly as a $\tau_{\mathrm{h}}\mu$ decay.

Switching to the $\tau_{\mathrm{h}}\mu$ decay in Figure 6.7b, with the same reasoning as before, the occurrence of the leptonic $\tau\tau$ pair decays can be explained. The $\mu\mu$ decays are more similar to the $\tau_{\mathrm{h}}\mu$ decay than any of the other decays and the μe decay is partly tagged as a $\tau_{\mathrm{h}}\mu$ decay as well. Apart from the leptonic $\tau\tau$ pair decays, the softdrop mass is almost exclusively represented by the $\tau_{\mathrm{h}}\mu$ decay in agreement with the matrices.

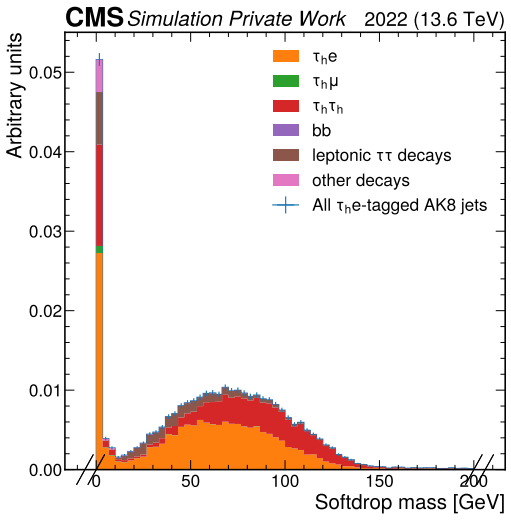
For softdrop masses above 20 GeV the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ distribution in Figure 6.7c mainly consists of generator-level $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decays. In the lower softdrop mass region, a few QCD processes and τ_{h}e decays are misidentified as a $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay, but still generator-level $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decays are mainly tagged. This is unexpected because according to the purity matrix 27.13% of the AK8 jets tagged as a $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay are supposed to be other decays. This fraction decreases for AK8 jets with $m_{softdrop} \geq 30\,\mathrm{GeV}$ to 3.7 %, according to Figure 6.6b, so the misidentified AK8 jets must have a low softdrop mass. The misidentified jets need to appear in simulated events with a different X boson mass hypotheses. To test this theory, the same histogram can be created for a simulated event with a lower X boson mass. The softdrop mass distribution for the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay channel for $M_{\mathrm{X}} = 4000\,\mathrm{GeV}$ and $M_{\mathrm{X}} = 550\,\mathrm{GeV}$ is displayed in Figure 6.8. A logarithmic y scale is used because only a few jets are tagged for high softdrop masses for $M_{\mathrm{X}} = 550\,\mathrm{GeV}$. A lot of QCD jets get tagged as $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ at low softdrop masses for $M_{\mathrm{X}} = 550\,\mathrm{GeV}$. Meanwhile, there are only a few QCD events tagged as a $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay for $M_{\mathrm{X}} = 4000\,\mathrm{GeV}$. For simulated events with a lower X boson mass, ParticleNet misidentifies more AK8 jets as a $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay. This is not an intuitive behavior as stronger boosted decays are expected to be more difficult to identify. A possible reason for this behavior is the occurrence of more resolved Y boson decays in simulated events with a low X boson mass. The products of a single hadronic τ lepton decay might be reconstructed with an AK8 jet and incorrectly identified by ParticleNet as a $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ decay.

What can be observed for all $\tau\tau$ pair decays is that the maximum of the softdrop mass distribution is significantly below 125 GeV, which would be the mass of the Y boson. This is due to the neutrinos from the τ decays missing in the AK8 jets.

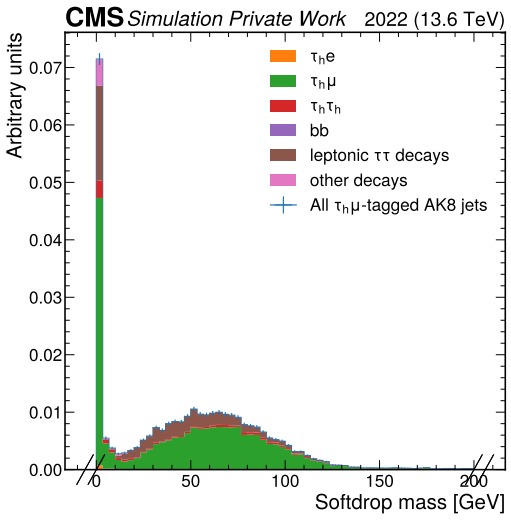
The tagging for the bb pair decay has a relatively high purity, although QCD decays are tagged too. This can also be observed in Figure 6.7d. The bb tagging works very well for high softdrop masses. Below a softdrop mass of 30 GeV the bb score is also high for QCD jets, which is partly due to the ParticleNet algorithm training, and partly due to QCD jets having a low softdrop mass generally. Here, a peak at 125 GeV can be observed because most of the decay particles are identified in the AK8 jet.

Figure 6.7e shows the distribution of the softdrop mass of AK8 jets tagged as a QCD decay. The ParticleNet has difficulties with tagging bb jets correctly for higher softdrop masses.

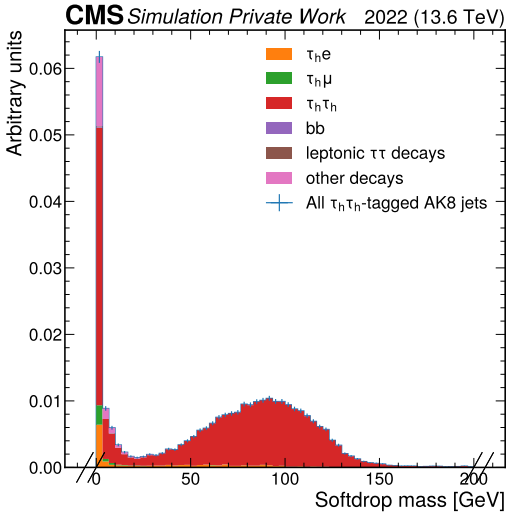
Around the 125 GeV mark several bb jets are identified as a QCD decay. Like mentioned before in Section 6.2.1, this might happen due to both decays being similar. Jets tagged as a QCD decay with smaller softdrop masses are mainly QCD decays with a few exceptions where bb decays are tagged.



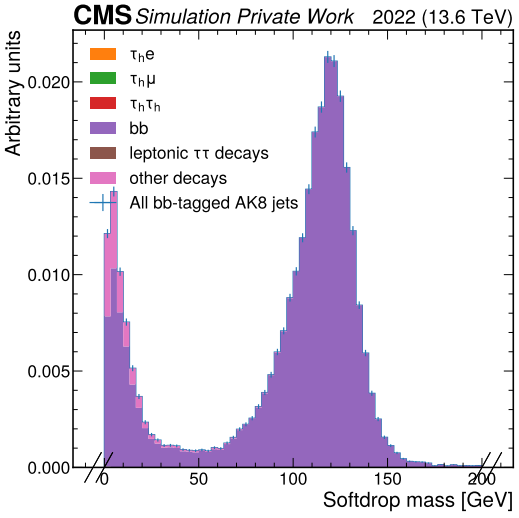
(a) Tagged as $\tau_h e$ decay



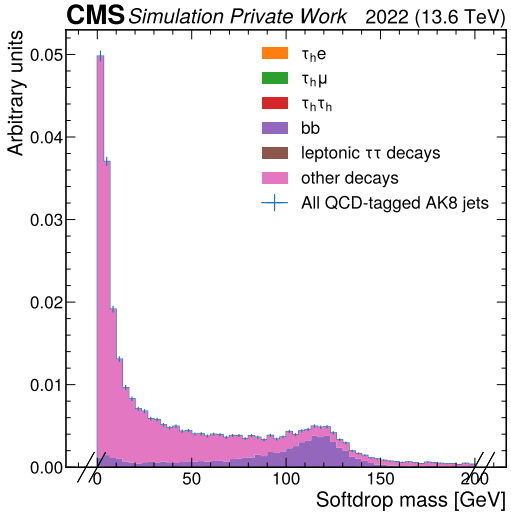
(b) Tagged as $\tau_h \mu$ decay



(c) Tagged as $\tau_h \tau_h$ decay

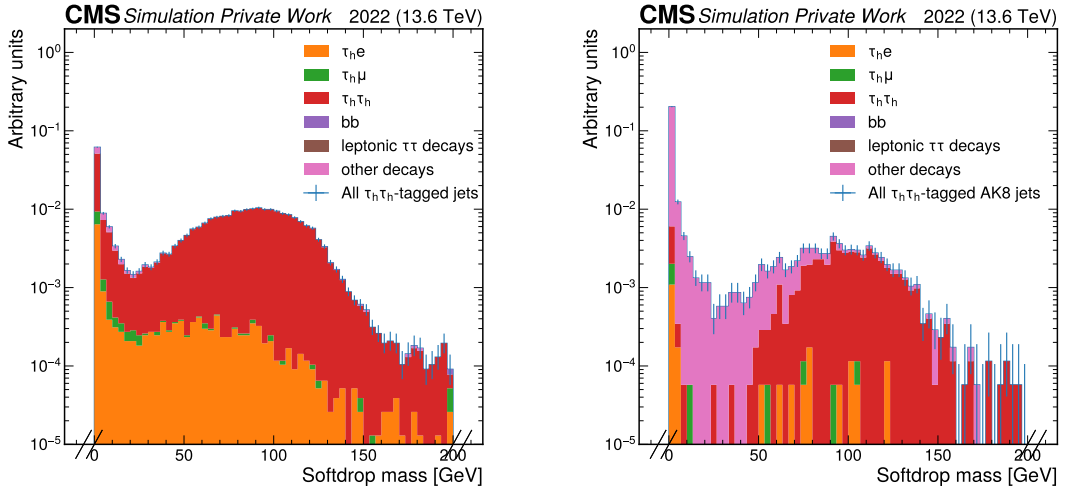


(d) Tagged as bb decay



(e) Tagged as QCD event

Figure 6.7: Distribution of AK8 jet softdrop mass for jets tagged as $\tau_h e$, $\tau_h \mu$, $\tau_h \tau_h$, bb decay or QCD event. The different colors indicate the generator-level decays.



(a) $M_X = 4000$ GeV (b) $M_X = 550$ GeV

Figure 6.8: Comparison of distribution of softdrop mass for AK8 jets tagged as $\tau_h\tau_h$ decay with logarithmic y scale for samples with $M_X = 3500$ GeV and $M_X = 550$ GeV. The different colors indicate the generator-level decays.

6.2.3 $\tau\tau$ pair tagging efficiency distributions

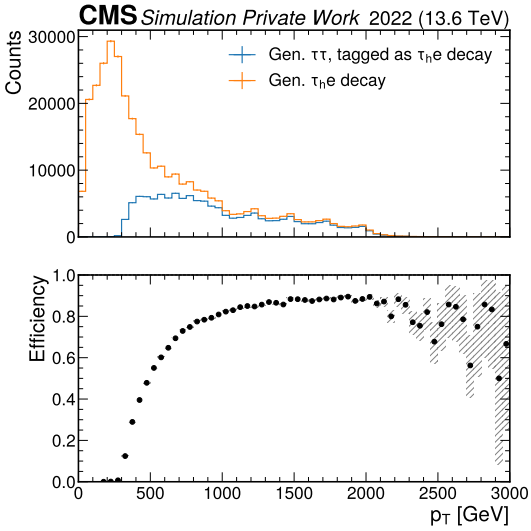
Another interesting piece of information is how efficient the $\tau\tau$ pair jet tagging is for different properties. In this section, the efficiencies are presented as functions of three properties, namely the transverse momentum of both the generator-level $\tau\tau$ pairs and the reconstructed AK8 jets, as well as the pseudorapidity of the reconstructed AK8 jets. All signal samples for the different M_X hypotheses analyzed in this thesis are used and added up in the histograms. The histograms show all generator-level $\tau\tau$ pairs or AK8 jets that are matched to them for the different $\tau\tau$ decays in their respective figure. Additionally, only the $\tau\tau$ pairs or AK8 jets that are identified correctly as the respective decay are shown. By dividing the number of entries in each bin of the two histograms, the tagging efficiency for each bin is calculated. The efficiency is plotted below the distribution, Clopper-Pearson intervals [47] are used to determine their uncertainty.

The efficiency distribution over the generator-level $\tau\tau$ pair p_T is shown in Figure 6.9. Since there are a lot more samples with a low M_X there are more $\tau\tau$ pairs with a small transverse momentum. Below a $\tau\tau$ pair p_T of 300 GeV no AK8 jets are tagged. This may be due to the AK8 jet $p_T \geq 200$ GeV threshold. Reconstructed AK8 jets do not contain the neutrinos from the τ decays, so the generator-level $\tau\tau$ pair p_T is higher than the reconstructed AK8 jet p_T. Low-p_T $\tau\tau$ pairs can also be assigned to resolved topologies which are not reconstructed with AK8 jets.
At a $\tau\tau$ pair p_T above 300 GeV the efficiency is very low but climbs up rapidly until approximately 700 GeV for the τ_he decay channel, where it reaches an efficiency of 70%. From there on the efficiency keeps on increasing slowly, at approximately $p_T = 1500$ GeV it reaches 90%. This efficiency is in agreement with the efficiency matrix in Figure 6.5a where a tagging efficiency of roughly 90% is stated. Since the number of $\tau\tau$ pairs with very high p_T is very low, efficiencies in this region have a high uncertainty.
The distribution of the tagging efficiency for the $\tau_h\mu$ decay has the same form as the distribution of the tagging efficiency for the τ_he decay but reaches higher efficiencies. For a high $\tau\tau$ pair p_T it has an efficiency close to one. This aligns with the efficiency matrix, according to which the $\tau_h\mu$ decay is correctly tagged roughly 95% of times.

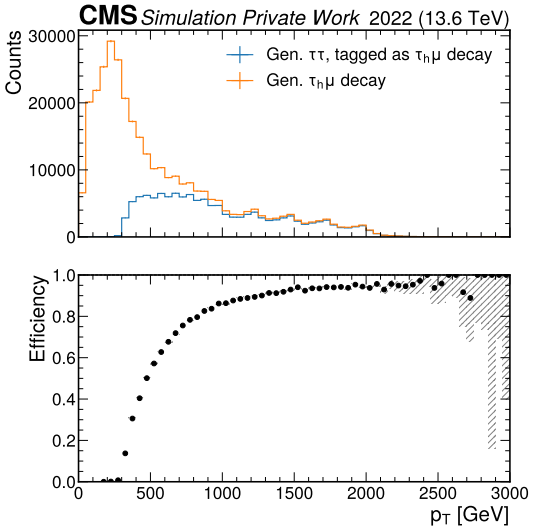
The efficiency for the $\tau_h\tau_h$ decay channel also increases rapidly between a p_T of 300 GeV and 700 GeV, reaching an constant efficiency of 70% for higher p_T values. This again is expected when comparing to the results of the efficiency matrix.
While the τ_he and $\tau_h\mu$ decays contain an easily reconstructible particle, the electron and muon, the $\tau_h\tau_h$ decay is made up of two hadronic decays that are more challenging to reconstruct. This leads to a big difference in their tagging efficiency. Overall the jet tagging efficiency for a $\tau\tau$ pair $p_T \geq 700$ GeV is high. For a $\tau\tau$ pair $p_T < 700$ GeV a higher efficiency is desired.

Figure 6.10 presents the tagging efficiency of matched AK8 jets as a function of the p_T of the AK8 jets.
The efficiency for the tagging of the τ_he decay is constantly above 90%, with an overall small uncertainty, besides for very high momenta, where statistical uncertainties appear.
The efficiency distribution for the $\tau_h\mu$ decay is constant as well at an efficiency of around 95%.
The efficiency of the $\tau_h\tau_h$ tagging on the other hand decreases approximately linearly for a higher AK8 jet p_T, starting at an efficiency of 80% down to 65% for high AK8 jet p_T. A possible reason for this decrease is that high p_T $\tau\tau$ pairs are boosted more strongly. A stronger boost makes the identification of the decay significantly more challenging. A high $\tau\tau$ pair p_T roughly translates to a high AK8 jet p_T. Interestingly, this drop cannot be observed in the previous Figure 6.9c. Potentially, there is a different reason for this decrease, a future study might analyze this behavior.

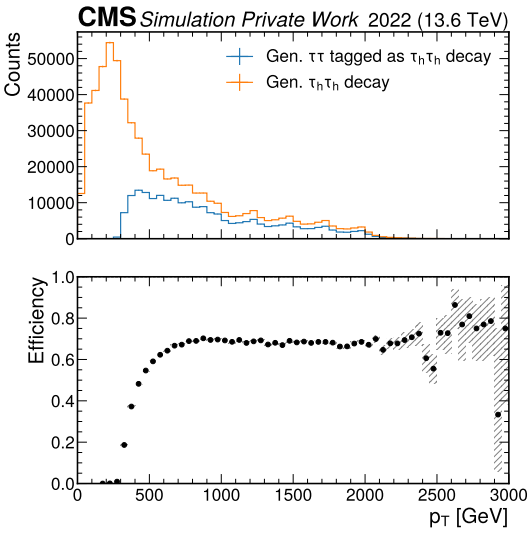
Lastly, the $\tau\tau$ pair jet tagging efficiency as a function of the AK8 jet η is featured in Figure 6.11. This efficiency showcases potential differences in the tagging efficiency with increasing polar angle. All three efficiencies are constant for $|\eta| \leq 1$ and drop slightly at higher η. This is the expected behavior because it is a cylindrical detector which is best at detecting objects if they pass through the layers orthogonally. In the endcaps of the CMS detector, different conditions apply and different sensors are used compared to the barrel region [3]. This leads to a small decrease in the tagging efficiency.
Again, the tagging efficiencies for the τ_he and $\tau_h\mu$ decay channel are above 90%. The tagging efficiency for the $\tau_h\tau_h$ decay channel is lower at around 75%, which reproduces the values from the efficiency matrix in Figure 6.5a.



(a) τ_he decay

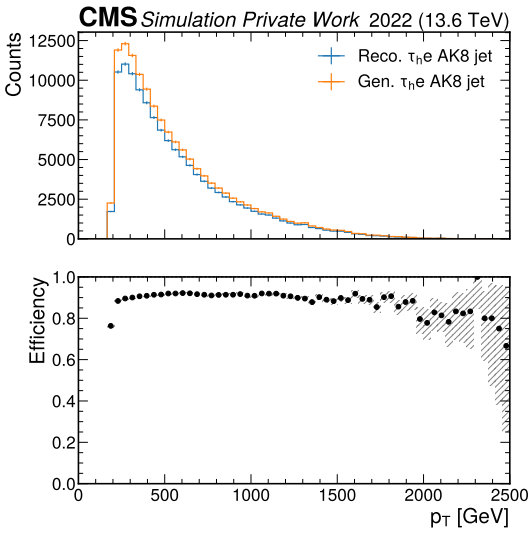


(b) $\tau_\mathrm{h}\mu$ decay

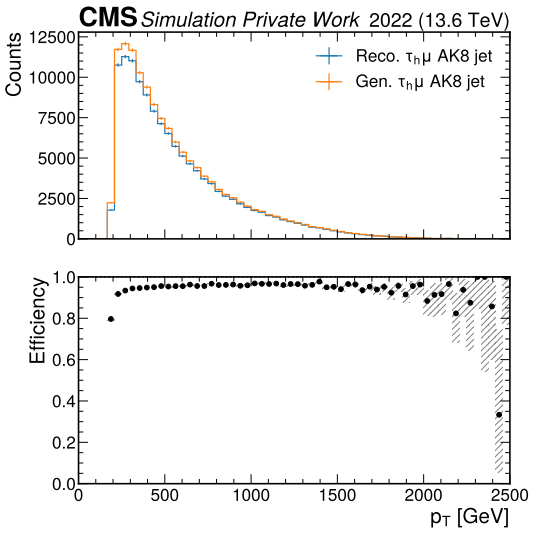


(c) $\tau_\mathrm{h}\tau_\mathrm{h}$ decay

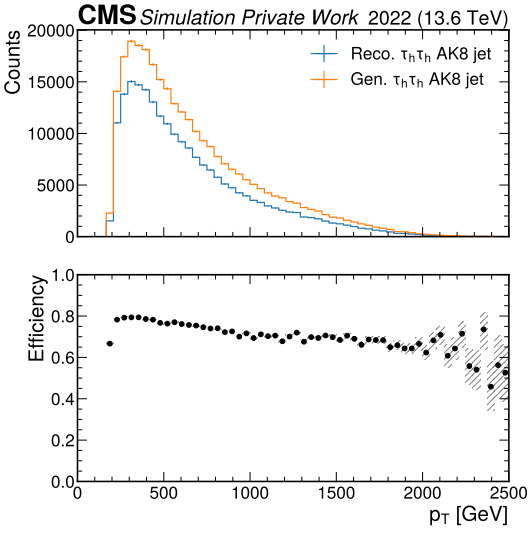
Figure 6.9: $\tau\tau$ pair decay tagging efficiency for $\tau\tau$ pair p_T on generator-level. The upper panel shows histograms of the generator-level $\tau\tau$ pairs with the respective generator-level decays (τ_he,$\tau_\mathrm{h}\mu$, $\tau_\mathrm{h}\tau_\mathrm{h}$) and all generator-level $\tau\tau$ pairs where the resulting AK8 jet was tagged correctly. The lower panel shows the efficiency by dividing the number of entries in the bins of the two histograms.



(a) τ_he decay

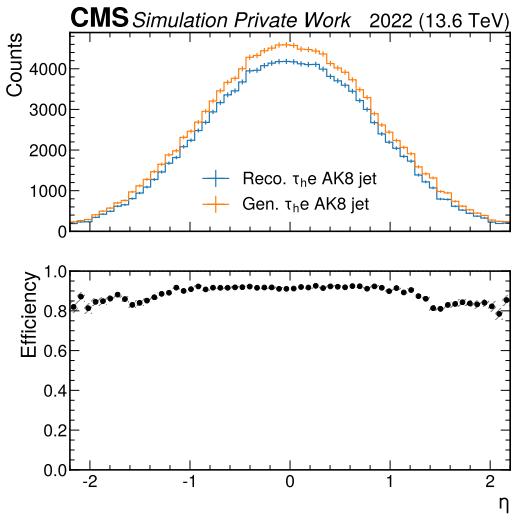


(b) $\tau_h\mu$ decay.

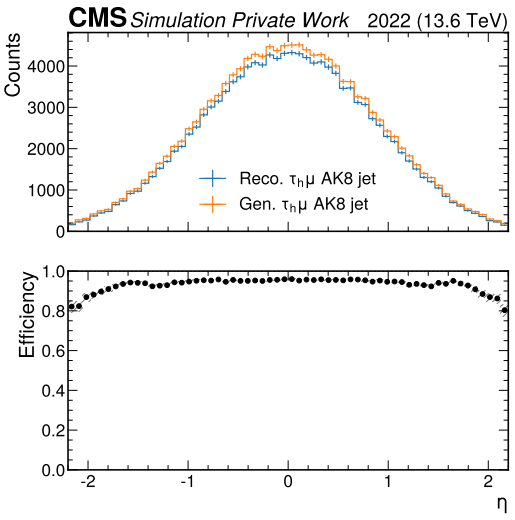


(c) $\tau_h\tau_h$ decay

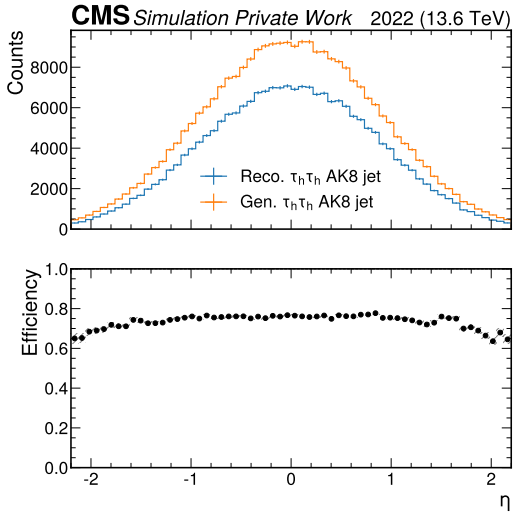
Figure 6.10: $\tau\tau$ pair decay tagging efficiency for AK8 jet p_T. The upper panel shows histograms of the AK8 jets matched to the respective generator-level $\tau\tau$ decays (τ_he,$\tau_h\mu$, $\tau_h\tau_h$) and of all AK8 jets that are tagged correctly. Only AK8 jets matched to a generated $\tau\tau$ pair are considered. The lower panel shows the efficiency by dividing the number of entries in the bins of the two histograms.



(a) τ_he decay



(b) $\tau_h\mu$ decay



(c) $\tau_h\tau_h$ decay

Figure 6.11: $\tau\tau$ pair decay tagging efficiency for AK8 jet η. The upper panel shows histograms of the AK8 jets matched to the respective generator-level $\tau\tau$ decays (τ_he,$\tau_h\mu$, $\tau_h\tau_h$) and of all AK8 jets that are tagged correctly. Only AK8 jets matched to a generated $\tau\tau$ pair are considered. The lower panel shows the efficiency by dividing the number of entries in the bins of the two histograms.

7 Summary and Outlook

In the search for theories beyond the standard model (SM) the next-to-minimal supersymmetric standard model (NMSSM) was proposed. Within its extended Higgs sector the decay of a heavy Higgs boson X into a 125 GeV Higgs boson H and a light Higgs boson Y is postulated. This di-Higgs process is searched for at the Large Hadron Collider (LHC) [5] at CERN by experiments like the Compact Muon Solenoid (CMS) [3]. Simulated events of proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ based on the response of the CMS detector were analyzed in this thesis. In the case of a high mass of the X boson relative to the mass of the other Higgs bosons, boosted decays occur which make it more challenging to identify the decaying particles. To find better identification methods, this thesis studied the behavior and efficiency of the ParticleNet [14] algorithm for the tagging of boosted $\tau\tau$ and bb pairs in the NMSSM $X \to YH$ analysis.

In the beginning, event samples were selected that are enriched in boosted $\tau\tau$ and bb pair decays. Simulated events using relatively high X boson mass hypotheses and a Y boson mass of 125 GeV were chosen. The objects in these data samples needed to fulfill additional selection criteria to further increase the number of boosted decays.

Studies on the generator level of these data samples have been performed. The decays of the $\tau\tau$ pairs were determined and compared to their theoretically predicted portions. With the results aligning to the expectation, a matching process between AK8 jets and generator-level particles was performed. In this way, for every measured AK8 jet information about its incident particles ($\tau\tau$ or bb pairs) was gained.

With this information, the tagging efficiency of the ParticleNet scores for boosted $\tau\tau$ and bb pair decays was calculated. Confusion matrices of the efficiency and purity show an overall precise tagging process. A significant confusion of the ParticleNet tagger between the τ_he and $\tau_\text{h}\tau_\text{h}$ output classes is observed. With the need of the existence of an electron in the AK8 jet for the τ_he decay channel or a muon for the $\tau_\text{h}\mu$ decay channel, an additional criterion was implemented to improve the tagging efficiency.
Reasons responsible for the misidentification of jets were searched for as well. In general, the efficiency drops for low transverse momenta of the $\tau\tau$ pair. On the other hand, the misidentification between the $\tau_\text{h}\tau_\text{h}$ and τ_he decays mainly happen for higher softdrop masses and transverse momenta of the AK8 jets.

Previous analyses on the $X \to YH$ process used older approaches [13] to explore boosted topologies of the $\tau\tau$ and bb pairs. The ParticleNet algorithm is a modern architecture

and a prime contender to use for jet identification in the future of boosted X $\rightarrow$ YH analyses. The results of this thesis demonstrate the behavior of the ParticleNet algorithm and will help to implement it in upcoming studies on the boosted section of the NMSSM di-Higgs process. Other algorithms, like DeepTau [48] for hadronic τ decay identification, are developed alongside ParticleNet. Future studies must be conducted to be able to compare their individual tagging efficiency and ultimately decide which algorithm will be the best for which use case.

Bibliography

[1] S. Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[2] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[3] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[4] G. Aad et al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[5] O. Brüning et al. *LHC design report*. CERN 2004-003, June 27, 2012.

[6] Y. Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003.

[7] C. Balazs et al. *A Primer on Dark Matter*. 2024. arXiv: 2411.05062 [astro-ph.CO]. URL: https://arxiv.org/abs/2411.05062.

[8] S. P. MARTIN. "A supersymmetry primer". In: *Perspectives on Supersymmetry*. WORLD SCIENTIFIC, July 1998, pp. 1–98. DOI: 10.1142/9789812839657_0001. URL: http://dx.doi.org/10.1142/9789812839657_0001.

[9] U. Ellwanger, C. Hugonie, and A. M. Teixeira. "The Next-to-Minimal Supersymmetric Standard Model". In: *Physics Reports* 496.1–2 (Nov. 2010), pp. 1–77. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2010.07.001. URL: http://dx.doi.org/10.1016/j.physrep.2010.07.001.

[10] M. MANIATIS. "The next-to-minimal supersymmetric extension of the standard model reviewed". In: *International Journal of Modern Physics A* 25.18n19 (July 2010), pp. 3505–3602. ISSN: 1793-656X. DOI: 10.1142/s0217751x10049827. URL: http://dx.doi.org/10.1142/S0217751X10049827.

[11] A. Tumasyan et al. "Search for a heavy Higgs boson decaying into two lighter Higgs bosons in the $\tau\tau$bb final state at 13 TeV". In: *JHEP* 11 (2021), p. 057. DOI: 10.1007/JHEP11(2021)057. arXiv: 2106.10361 [hep-ex].

[12] N. Shadskiy. "Search for resonant di-Higgs production in bb + final states in pp collisions at s = 13 TeV". PhD thesis. Karlsruhe Institute of Technology (KIT), 2025.

[13] A. Sirunyan et al. "Performance of reconstruction and identification of tau leptons decaying to hadrons and tau neutrinos in pp collisions at s=13 TeV". In: *Journal of Instrumentation* 13.10 (Oct. 2018), P10005–P10005. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/10/p10005. URL: http://dx.doi.org/10.1088/1748-0221/13/10/P10005.

[14] H. Qu and L. Gouskos. "Jet tagging via particle clouds". In: *Physical Review D* 101.5 (Mar. 2020). ISSN: 2470-0029. DOI: 10.1103/physrevd.101.056019. URL: http://dx.doi.org/10.1103/PhysRevD.101.056019.

[15] B. Povh et al. *Particles and Nuclei. An Introduction to the Physical Concepts*. Graduate Texts in Physics. Springer Berlin, Heidelberg, 2015. ISBN: 978-3-662-46320-8, 978-3-662-49583-4, 978-3-662-46321-5. DOI: 10.1007/978-3-662-46321-5.

[16] W. Demtröder. *Experimentalphysik 4 (Kern-, Teilchen- und Astrophysik)*. Springer Spektrum Berlin, Heidelberg, Feb. 2017. DOI: 10.1007/978-3-662-52884-6.

[17] P. Jordan and E. P. Wigner. "About the Pauli exclusion principle". In: *Z. Phys.* 47 (1928), pp. 631–651. DOI: 10.1007/BF01331938.

[18] K. G. Wilson. "Confinement of Quarks". In: *Phys. Rev. D* 10 (1974). Ed. by J. C. Taylor, pp. 2445–2459. DOI: 10.1103/PhysRevD.10.2445.

[19] Q. R. Ahmad et al. "Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory". In: *Phys. Rev. Lett.* 89 (2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301. arXiv: nucl-ex/0204008.

[20] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98.3 (2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[21] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[22] P. W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[23] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[24] Cush. *Standard Model of Elementary Particles*. (Accessed: 6. February 2025). 2019. URL: https://upload.wikimedia.org/wikipedia/commons/0/00/Standard_Model_of_Elementary_Particles.svg.

[25] M. Li et al. "Dark Energy". In: *Communications in Theoretical Physics* 56.3 (Sept. 2011), pp. 525–604. ISSN: 0253-6102. DOI: 10.1088/0253-6102/56/3/24. URL: http://dx.doi.org/10.1088/0253-6102/56/3/24.

[26] C. CSÁKI. "The minimal supersymmetric standard model". In: *Modern Physics Letters A* 11.08 (Mar. 1996), pp. 599–613. ISSN: 1793-6632. DOI: 10.1142/s021773239600062x. URL: http://dx.doi.org/10.1142/S021773239600062X.

[27] S. Navas et al. "Review of particle physics". In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[28] "LHC Machine". In: *JINST* 3 (2008). Ed. by L. Evans and P. Bryant, S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[29] *The Large Hadron Collider*. Accessed: 18. February 2025. URL: https://home.cern/science/accelerators/large-hadron-collider.

[30] *CERN's accelerator complex*. Accessed: 18. February 2025. URL: https://home.cern/science/accelerators/accelerator-complex.

[31] A. Hayrapetyan et al. "Development of the CMS detector for the CERN LHC Run 3". In: *JINST* 19.05 (2024), P05064. DOI: 10.1088/1748-0221/19/05/P05064. arXiv: 2309.05466 [physics.ins-det].

[32] A. A. Alves Jr. et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[33] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[34] E. Lopienska. *The CERN accelerator complex, layout in 2022*. Accessed: 14. February 2025. Feb. 2022. URL: https://cds.cern.ch/images/CERN-GRAPHICS-2022-001-1.

[35] T. Sakuma. "Cutaway diagrams of CMS detector". In: (2019). URL: https://cds.cern.ch/record/2665537.

[36] V. Karimäki et al. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/368412.

[37] *The Phase-2 Upgrade of the CMS Tracker*. Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.QZ28.FLHW. URL: https://cds.cern.ch/record/2272264.

[38] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/349375.

[39] *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. The following files are from http://uscms.fnal.gov/pub/hcal_tdr and may not be the version as printed, please check the printed version to be sure. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/357153.

[40] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[41] *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Tech. rep. Geneva: CERN, 2009. URL: https://cds.cern.ch/record/1194487.

[42] M. Schröder and on behalf of the CMS collaboration. "Performance of jets at CMS". In: *Journal of Physics: Conference Series* 587.1 (Feb. 2015), p. 012004. DOI: 10.1088/1742-6596/587/1/012004. URL: https://dx.doi.org/10.1088/1742-6596/587/1/012004.

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

[44] C. Li and on behalf of the CMS collaboration. "Boosted jet tagging in CMS". In: *ML4jets*. (Accessed: 23. January 2025). 2021. URL: https://indi.to/mqD7t.

[45] T. Schörner-Sadenius, ed. *The Large Hadron Collider - Harvest of Run 1*. Springer International Publishing AG Switzerland, 2015.

[46] A. J. Larkoski et al. "Soft drop". In: *Journal of High Energy Physics* 2014.5 (May 2014). ISSN: 1029-8479. DOI: 10.1007/jhep05(2014)146. URL: http://dx.doi.org/10.1007/JHEP05(2014)146.

[47] C. J. Clopper and E. S. Pearson. "The use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial". In: *Biometrika* 26.4 (1934), pp. 404–413. DOI: 10.1093/biomet/26.4.404.

[48] A. Tumasyan et al. "Identification of hadronic tau lepton decays using a deep neural network". In: *JINST* 17 (2022), P07023. DOI: 10.1088/1748-0221/17/07/P07023. arXiv: 2201.08458 [hep-ex].

Appendix

A Additional tagging efficiency distributions

A.1 Tagging efficiency over azimuthal angle ϕ

In addition to the presented tagging efficiency distributions in Chapter 6.2.3 the tagging efficiency for different ϕ can be taken into consideration. The CMS detector is build cylindrically around the collision point without a gap. It is therefore expected that the tagging efficiency is constant for every ϕ. This expected behavior indeed takes place as can be seen in Figures A.1, A.2 and A.3.

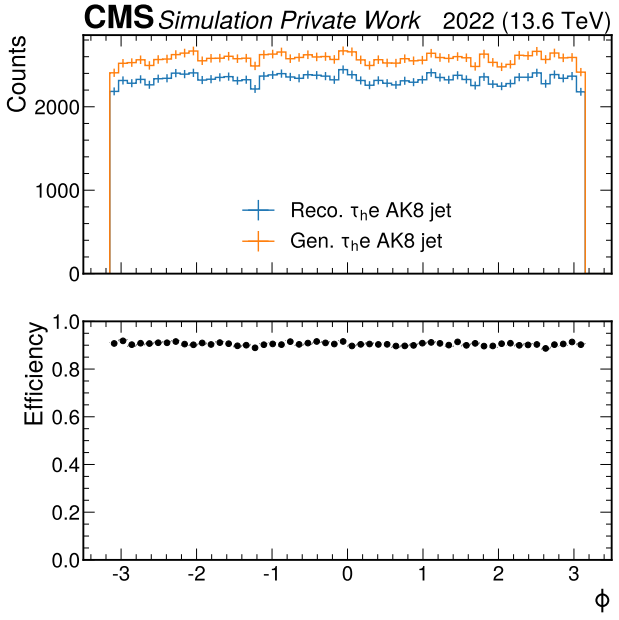


Figure A.1: Jet tagging efficiency of τ_{h}e decay channel over AK8 jet ϕ.

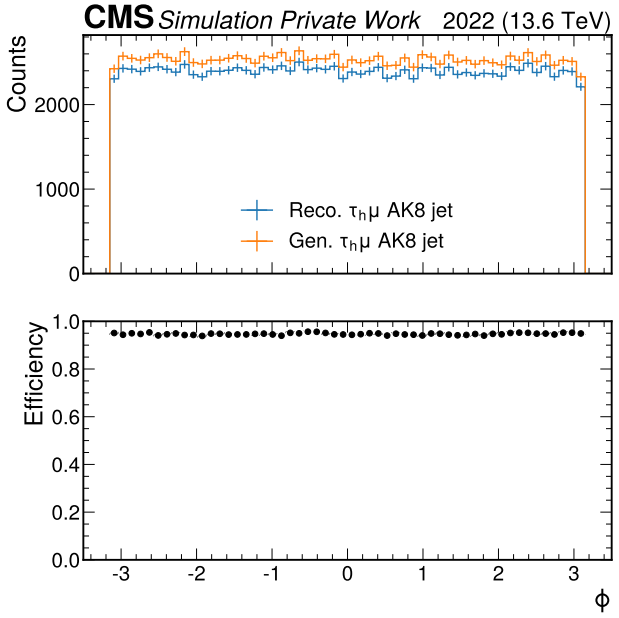


Figure A.2: Jet tagging efficiency of $\tau_h\mu$ decay channel over AK8 jet ϕ.

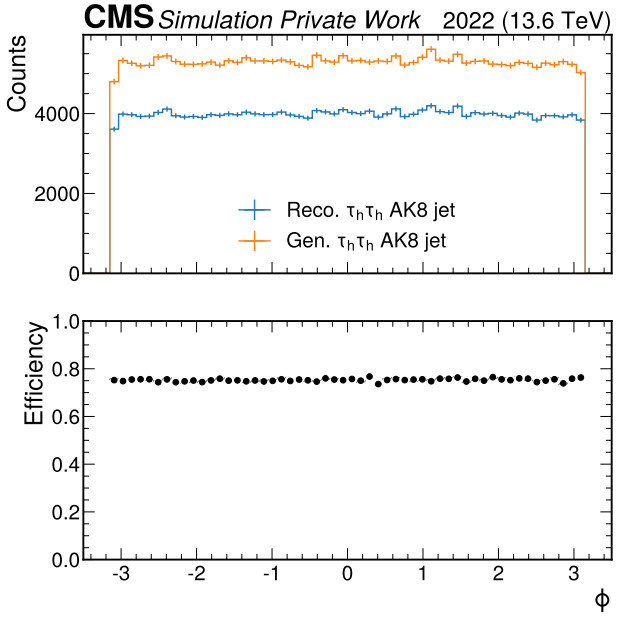


Figure A.3: $\tau\tau$ pair decay tagging efficiency for AK8 jet ϕ. The upper panel shows histograms of the AK8 jets matched to the respective generator-level $\tau\tau$ decays (τ_he,$\tau_h\mu$, $\tau_h\tau_h$) and of all AK8 jets that are tagged correctly. Only AK8 jets matched to a generated $\tau\tau$ pair are considered. The lower panel shows the efficiency by dividing the number of entries in the bins of the two histograms.