GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

$[q,p]=i\hbar$

Master's Thesis

Perspektiven für VBF $H \to \tau_{\mathbf{lep}}\tau_{\mathbf{had}}$ Messungen am High-Luminosity LHC mit dem ATLAS Detektor

Prospects for VBF $H \to \tau_{\mathbf{lep}}\tau_{\mathbf{had}}$ measurements at the High-Luminosity LHC with the ATLAS detector

prepared by

Tobias Fitschen

from 27404, Zeven

at the II. Physikalischen Institut

ii

Abstract

A search for the VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$ process with the ATLAS detector at the HL-LHC with an integrated luminosity of 3000 fb^{-1} is presented. The expected detector performance is simulated and applied to Monte-Carlo generated samples corresponding to signal process and associated background processes. The harsher pileup conditions expected for the HL-LHC compared to the current LHC's layout are simulated by overlaying pileup-jets and inserting additional missing transverse energy as well as by degrading momentum resolutions for identified objects in the detector. A cut-based analysis is performed yielding an expected significance of $\Sigma_{\text{b}}^{\text{exp}} = 8.63 \pm 0.66_{\text{stat}}$. Additionally, the implications of the HL-LHC's harsher pileup conditions on the performance of the ATLAS jet calibration process is investigated and strategies for improving the pileup jet correction are presented.

iv

Contents

Contents

1 Introduction

The Large Hadron Collider (LHC), envisioned as the largest particle accelerator of its time, has already held to its promise by providing the necessary experimental setup for the discovery of the Higgs boson by the ATLAS and the CMS collaboration in 2012 [1][2] while providing numerous important precision measurements [3]. Motivated by these achievements, the LHC will continue to serve as a valuable resource for particle physicists in the future. Repeated upgrades to the accelerator promise to help accomplish this goal [4–9]. To be able to judge how such an upgrade should be ideally implemented, comprehensive studies of the performance of the experiments with different updated detector layouts are needed. The work presented in the following chapters aims to provide studies based on such a simulation to give prospects on the ability of future analyses taking advantage of the ATLAS detector in its newly designed form after the upgrade to the High-Luminosity LHC (HL-LHC).

The particular channel which is considered in this thesis is the decay of Higgs bosons produced by vector boson fusion (VBF) into two τ leptons which is further constrained by the requirement that one (τ_{lep}) decays leptonically while the other one (τ_{had}) decays hadronically. After the Higgs boson was discovered utilising primarily its bosonic decay modes, this channel has the advantage of providing valuable insights into the fermionic coupling to the Higgs sector.

This document is structured as follows: Section 2 gives a short overview on the background underlying modern particle physics in general and the Higgs sector of the Standard Model of particle physics in particular. The following section 3 describes the experimental setup which consists of the LHC, the ATLAS detector, and their proposed High-Luminosity upgrades. Section 4 aims to give an overview on jet phenomenology describing hadronization, jet algorithms, and the calibration of jets in the ATLAS detector. The experimental results of this thesis are presented in Sec. 5 and Sec. 6 which aim to ascertain prospects for the sensitivity of the VBF $H \to \tau\tau$ process and the jet pileup calibration with the ATLAS detector at the HL-LHC, respectively. Finally, section 7 gives a summary and conclusion of the above and an outlook based on the results.

2 Theoretical Background

Since its initiation by the discovery of the electron by J. J. Thomson at the end of the 19th century [10], elementary particle physics has significantly grown in complexity and accuracy. Nowadays, the Standard Model (SM) of particle physics [11] is able to describe all discovered elementary particles in one self-consistent theory.

Sec. 2.1, gives a brief overview on the fundamental concepts and the particle content of the Standard model. Sec. 2.2 describes how electroweak symmetry breaking and the Brout-Englert-Higgs (BEH) mechanism allow for massive gauge bosons and fermions in the Standard Model and how the existence of the Higgs boson emerges from this. Finally, Sec. 2.3 serves as an overview of the production and decay properties of the Higgs boson.

2.1 The Standard Model of Particle Physics

The particle content of the Standard Model can be divided into a bosonic- and a fermionic sector; while the former encompasses the force carrying gauge bosons and the Higgs boson, the latter can be further subdivided into a leptonic- and a quark sector. The leptonic sector, in addition to the electron, includes the electrically neutral electron neutrino, and for each of them respectively, two generations that have the same quantum numbers but differ in mass. Similar to the electron being paired to a corresponding neutrino, the down-quark is in an analogous relation to the up-quark. Like the leptons, these two quarks occur in two additional generations identical in all quantum numbers as well. The generations are ordered in ascending order of the masses of the charged leptons and quarks. A schematic overview of the Standard Model's particle content can be found in Fig. 2.1.

Quantum numbers which determine to which extent the corresponding particle is affected by the three fundamental forces of the SM are called charges. Charges are a consequence of the same phenomenon which is responsible for the existence of the gauge bosons as well: continuous gauge symmetries. The Noether theorem relates continuous symmetries to conserved quantities of a physical system [12, 13]. In its special case for continuous fields, it can be used to derive conserved charges corresponding to local gauge symmetries. The Standard Model of particle physics utilises this phenomenon by assuming that its un-

derlying Lagrangian density respects a local $SU(3)_{\mathrm{C}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}}$ gauge symmetry. As a consequence, new fields corresponding to the generators of these groups must be introduced. The particles associated with these fields are the gauge bosons: g, W, Z, and γ. They mediate the three fundamental forces associated to the respective conserved charges: Colour (strong force), weak isospin (weak force), and weak hypercharge (electromagnetism).
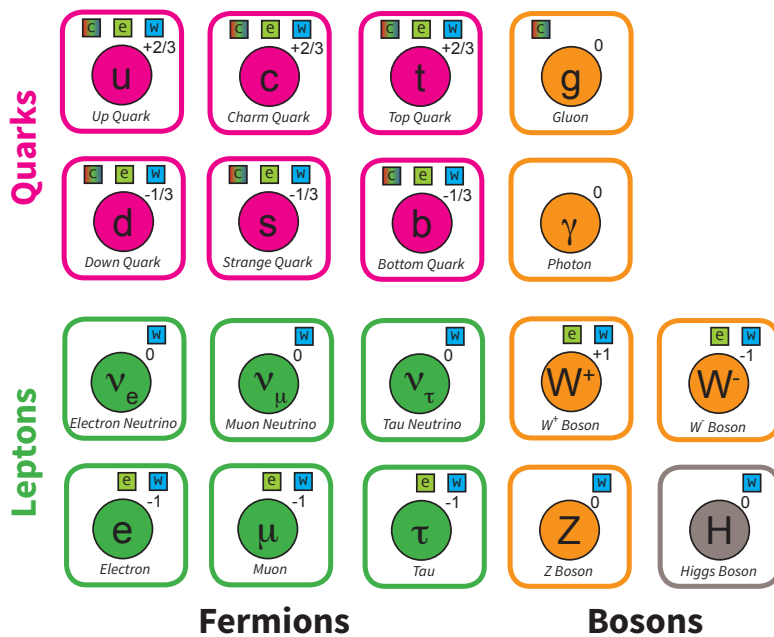
Figure 2.1: Particles of the Standard Model: Leptons (green), quarks (magenta), gauge bosons (orange) and the Higgs boson (grey). The numbers on the upper right of each particle are its electromagnetic charge. The boxes on the upper right indicate the fields to which the particle couples: colour (c), electromagnetic (e) and weak (w) [14].

The fourth fundamental force in modern physics is gravity. Despite many attempts, it is not possible yet to include it into the SM in a similar fashion to the other three forces. Additionally, due to its weak strength compared to the other fundamental forces its effect is negligible in most phenomena under study in particle colliders. At the moment, general relativity serves as the best descriptive theory for this particular force.

In addition to the assumption of the symmetries mentioned above, there are 19 real parameters in the SM that are not dictated by theory but must be measured experimentally. The observation of neutrino oscillation demands the addition of non-vanishing neutrino masses and mixing parameters to this list.

The unification of the electromagnetic and the weak force by the Glashow-Weinberg-Salam (GWS) model [11] utilising the $SU(2) \times U(1)$ gauge group was one of the milestones in the development of the SM. This was done by introducing four bosonic fields $\vec{W}^\mu = (W_1^\mu, W_2^\mu, W_3^\mu)$ and B^μ to the Lagrangian

$$\mathcal{L}_{\mathrm{D}} = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - m\bar{\Psi}\Psi. \tag{2.1}$$

which describes fermionic fields with the corresponding masses m in the form of doublets of spinors Ψ. $\mathcal{L}_{\mathrm{D}}$ is invariant under global but not under local $U(1) \times SU(2)$ transformations. This manifests in additional terms when applying such local transformations on $\mathcal{L}_{\mathrm{D}}$. By replacing the fourgradient ∂_μ in Eq. 2.1 by the covariant derivative

$$D_\mu = \partial_\mu + ig_W\frac{\vec{\sigma}}{2}\vec{W}_\mu + ig'\frac{Y}{2}B_\mu \tag{2.2}$$

the bosonic fields B^μ and $\vec{W}^\mu$ are introduced. Y is called the weak hypercharge of the GWS model, g_W and g' are the coupling constants of the newly introduced fields and $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ contains the Pauli matrices. The transformation properties of B^μ and $\vec{W}^\mu$ under local $U(1) \times SU(2)$ transformations introduce additional terms to $\mathcal{L}_{\mathrm{D}}$ which exactly cancel the terms introduced by the local transformations of the fermionic field Ψ. Consequently, $\mathcal{L}_{\mathrm{D}}$ is invariant under such transformations. However, since the addition of the corresponding mass terms of $\vec{W}^\mu$ and B^μ to the Lagrangian density would spoil the local gauge symmetry, these must be massless states. A similar argument can be made for the fermionic states. The discovery of the massive W and Z bosons [15][16] (and the fermion masses) illustrated that this description of the electroweak interaction could not be complete. This problem was solved by the Brout-Englert-Higgs (BEH) mechanism [17] which, with the introduction of the Higgs field to the SM, allows for spontaneous symmetry breaking of the electroweak gauge symmetry. A more detailed description of this mechanism is given in the following Sec. 2.2.

2.2 Electroweak Symmetry Breaking and the BEH-Mechanism

The Brout-Englert-Higgs (BEH) mechanism was independently proposed among others by Brout and Englert [18], by Kibble, Hagen, and Guralnik [19], and by Higgs [20]. Together with electroweak symmetry breaking, it provides an explanation of the masses of the W and Z boson while still being consistent with the unification of the weak and

electromagnetic forces through a local $SU(2) \times U(1)$ gauge symmetry (i.e. the Glashow-Weinberg-Salam (GWS) model [11]). This is achieved by introducing a bosonic field described by a weak isospin doublet

$$\Phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{bmatrix}. \tag{2.3}$$

with complex scalar components to the Lagrange density of the Standard Model. The corresponding additional term in the Lagrange density can be described by

$$\mathcal{L}_H = (\partial_\mu \Phi)^\dagger (\partial^\mu \Phi) - V(\Phi) \text{ with } V(\Phi) = \mu^2 (\Phi^\dagger \Phi) + \lambda (\Phi^\dagger \Phi)^2 \tag{2.4}$$

where $\lambda \in \mathbb{R}$ and μ are free parameters of the potential $V(\Phi)$. For the potential $V(\Phi)$ to be bounded from below, the value of λ must be positive. In this case, μ parametrizes the shape of $V(\Phi)$ in the following way: for $\mu^2 > 0$, $V(\Phi)$ has a single global minimum at $\phi^0 = \phi^+ = 0$. In the case that $\mu^2 < 0$, $V(\Phi)$ has a set of infinite global minima at $(\phi^0)^2 + (\phi^+)^2 = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}$ and a local maximum at $\phi^0 = \phi^+ = 0$. This case is illustrated in Fig. 2.2.
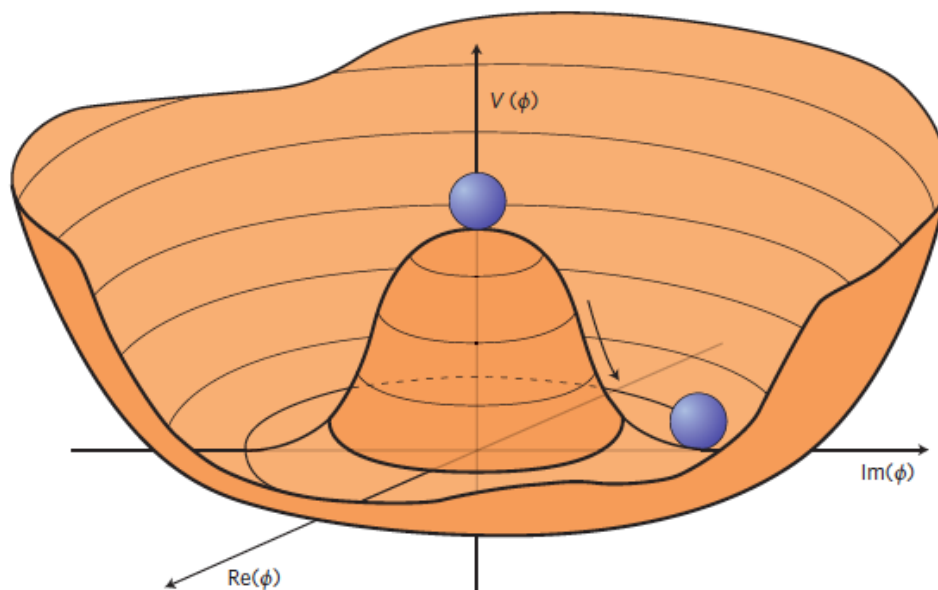
Figure 2.2: An illustration of the Higgs potential $V(\Phi)$ for a complex scalar field $\Phi^0 = \phi_1 + i\phi_2$ in the case that $\mu^2 < 0$ and $\lambda > 0$ [21].

Here, v is called the vacuum expectation value of the field Φ. In the local maximum, $V(0,0)$ is invariant under a local $SU(2) \times U(1)$ symmetry transformation. Transitioning from this state into one of the global minima constitutes the breaking of this symmetry.

2 Theoretical Background

The set of global minima corresponds to an infinite set of possible vacuum states of the field. Choosing the vacuum state to be in the real direction of ϕ^0, i.e.

$$\Phi_v = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v \end{bmatrix} \tag{2.5}$$

and expanding the field around this vacuum state yields

$$\Phi = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1(x) + i\phi_2(x) \\ v + \phi_3(x) + i\phi_4(x) \end{bmatrix} \tag{2.6}$$

with a scalar field h with non-vanishing vacuum expectation value v. This is in accordance with the Nambu-Goldstone theorem which says that the breaking of a continuous symmetry introduces a massless scalar field called Goldstone boson (here ϕ_i) for each generator of the symmetry. The theorem was first introduced by Nambu [22] in the context of superconductivity in solid state theory and further developed and applied to quantum field theories by Goldstone, Weinberg, and Salam [23][24].

By applying the unitary gauge (i.e. $\phi_1 = \phi_2 = \phi_4 = 0$) [25], replacing the four-gradient ∂_μ in Eq. 2.4 by the covariant derivative from Eq. 2.2, and by introducing the mass eigenstates of the experimentally discovered $W^\pm$, Z and γ bosons as

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \tag{2.7}$$

$$\begin{bmatrix} \gamma \\ Z \end{bmatrix} = \begin{bmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{bmatrix} \begin{bmatrix} B \\ W_3 \end{bmatrix}, \tag{2.8}$$

Eq. 2.4 can be written as

$$\begin{aligned} \mathcal{L}_H = &\frac{1}{2}\partial_\mu h \partial^\mu h - \frac{1}{4}\lambda h^4 - \lambda v h^3 \underbrace{-\lambda v^2 h^2 + \frac{g_W^2 v^2}{4}W_\mu^- W^{+\mu} + \frac{v^2}{8(g_W^2 + g'^2)}Z_\mu Z^\mu}_{h,\ W,\ \text{and}\ Z\ \text{mass terms}} \\ &\underbrace{+\frac{g_W^2 v}{2}W_\mu^- W^{+\mu}h + \frac{v}{4(g_W^2 + g'^2)}Z_\mu Z^\mu h + \frac{g_W^2}{4}W_\mu^- W^{+\mu}h^2 + \frac{1}{8(g_W^2 + g'^2)}Z_\mu Z^\mu h^2}_{\text{couplings of}\ h\ \text{to}\ W\ \text{and}\ Z} . \end{aligned} \tag{2.9}$$

Here, θ_W is the Weinberg angle defined by $\tan\theta_W = \frac{g'}{g_W}$. Note that the W and Z bosons

acquire mass terms with the corresponding masses

$$m_W = \frac{g_W v}{2} \text{ and } m_Z = \frac{v}{2\sqrt{g_W^2 + g'^2}} \tag{2.10}$$

while the photon γ remains massless and does not interact with the higgs field h. Due to the application of the unitary gauge, the Goldstone bosons do not explicitly appear in Eq. 2.9. The corresponding degrees of freedom were absorbed to allow for massive gauge bosons. The newly introduced higgs field h also acquired a mass term resulting in a mass of the corresponding Higgs boson H of $m_H = \sqrt{2\lambda v^2}$. Fermion masses are allowed in this model due to Yukawa coupling terms

$$\mathcal{L}_Y = -\frac{\lambda_f}{\sqrt{2}}(v + h)(\bar{f}_L f_R + \bar{f}_R f_R) \tag{2.11}$$

between fermions f and the newly introduced scalar field h. f_L and f_R represent the left- and right-handed components of the fermionic fields. The corresponding fermion mass is therefore given by $m_f = \frac{\lambda_f v}{2}$ where λ_f is the Yukawa coupling constant of the fermion f to the Higgs field h. The scalar field h can be related to a massive scalar particle, the Higgs boson.

2.3 Higgs Boson Production and Decay

In 2012, the ATLAS and the CMS collaboration announced the discovery of a previously unknown boson with a mass of 126.0 ± 0.6 GeV (ATLAS [1]) and 125.3 ± 0.6 GeV (CMS [2]) respectively. Together with later measurements of the bosons properties, in particular the measurements of its spin to be 0 [26, 27], it was confirmed to be consistent with the properties of the Higgs boson of the Standard Model.

The Higgs boson can couple to every massive particle in the Standard Model. In the following, its four dominant production processes at the conditions of the LHC are described. Since the High-Luminosity LHC (see Sec. 3.3) is expected to provide a centre of mass energy of $\sqrt{s} = 14$ TeV which is only slightly higher than the current LHC's value of $\sqrt{s} = 13$ TeV, their relative rate will remain almost unchanged. A small difference in the rates however is to be expected since the increase in $\sqrt{s}$ lowers the suppression of modes including heavy final state particles (i.e. t quarks and W and Z bosons). Feynman diagrams of the four dominant modes can be seen in Fig. 2.3.

The Higgs boson's coupling strength is directly proportional to the particle's mass. Accordingly, processes in which the Higgs boson couples to top quarks (the heaviest particles

in the SM) are highly favoured. As a consequence of this, gluon fusion (see Fig. 2.3 [top left]) which most commonly includes this coupling has the highest cross-section. While the involvement of a loop of top quarks in this process corresponds to the highest probability, it is also possible to have lighter quarks in the loop instead.

A separate, although significantly less probable, process is $t\bar{t}H$ production (see Fig. 2.3 [top right]). This process is suppressed because of the presence of the heavy particles in its final state.
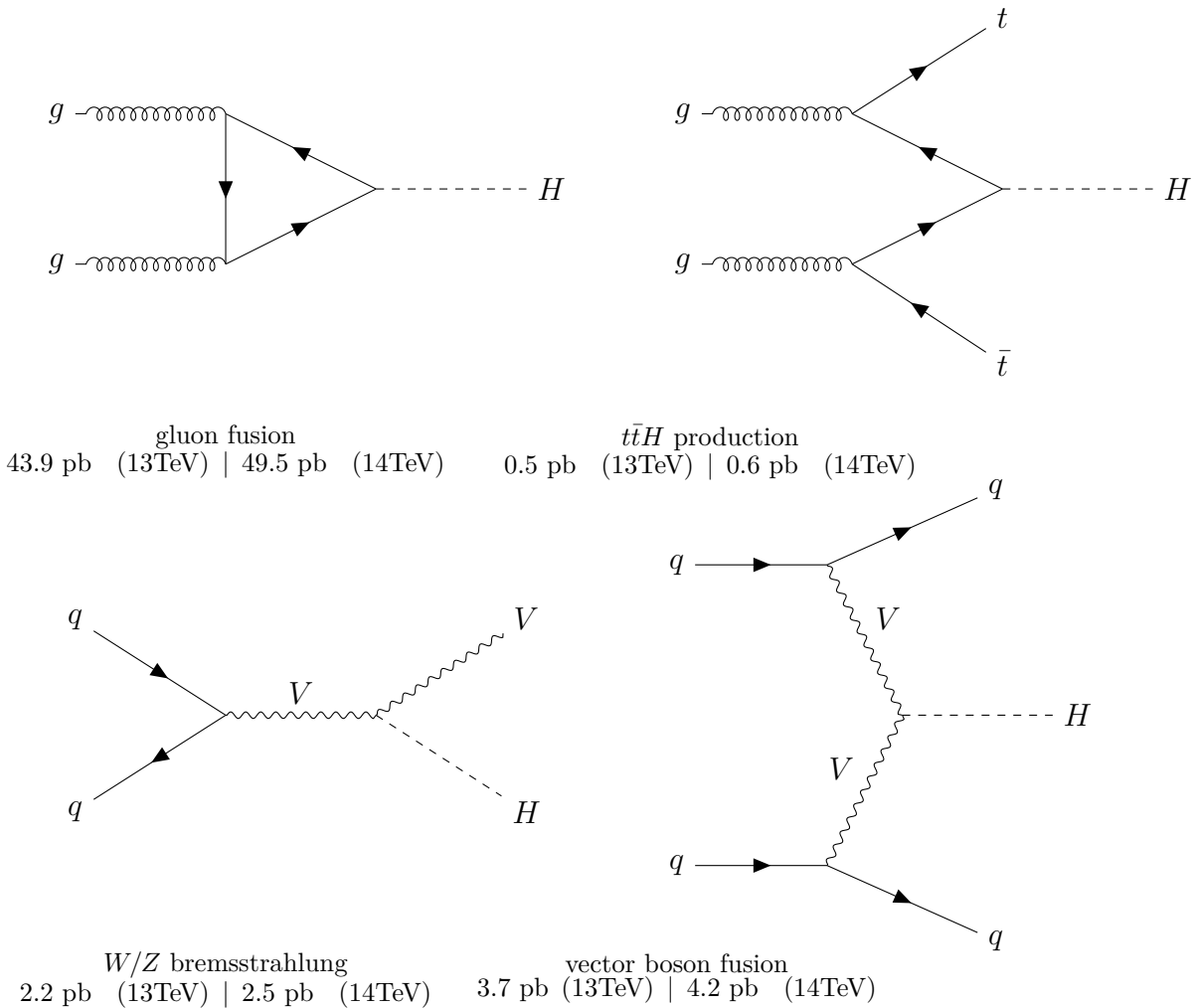
Figure 2.3: Dominant Higgs production modes at the LHC with the corresponding cross-sections with centre of mass energy of 13 TeV and 14 TeV. Values from [28].

The Z and W bosons being the third and fourth heaviest particles in the SM (only after t and H) are also involved in Higgs boson production modes with significant cross-

sections. One notable candidate in this respect is the W/Z bremsstrahlung process (see Fig. 2.3 [bottom left]). The presence of the heavy vector boson in this production mode's final state disfavours it.

Higgs boson production through vector boson fusion (VBF) (see Fig. 2.3 [bottom right]) is the main focus of this thesis. This mode possesses the feature of the existence of two quarks alongside the Higgs boson in the final state. In events where the signature of these quarks can successfully be reconstructed in the form of jets (see Sec. 4), their characteristic kinematic features can be used as a tool to separate processes involving this production mode from other processes. These VBF jets tend to exhibit large values of absolute pseudorapidity $|\eta|$ (i.e. are forward) and tend to lie in opposite hemispheres of the detector.

Due to the Higgs boson's property to be able to couple to every massive particle in the Standard Model, it can decay in many different ways. Since its coupling strength is proportional to the particle's mass, only modes in which a Higgs boson couples to heavy particles must be considered. The branching ratios for other channels are negligible. The Higgs boson's dominant decay modes by branching ratio are listed in Tab. 2.1. Despite the fact that photons and gluons are massless, decays into $\gamma\gamma$ and gg final states are possible. In these cases the Higgs boson does not couple directly to the massless particles but indirectly via a loop containing massive particles.

Decay channel	Branching ratio [%]
$b\bar{b}$	58
WW	21
gg	8.2
$\tau\tau$	6.3
$c\bar{c}$	2.9
ZZ	2.6
$\gamma\gamma$	0.2

Table 2.1: Branching ratios for the decay of Higgs bosons into various final states. Values are taken from Ref. [29].

The Higgs boson's discovery in 2012 was achieved by primarily utilising bosonic decay modes, namely decays into photons or vector bosons. This study on the other hand aims to provide prospects for investigating decays involving fermionic couplings. This would also be the case for $H \rightarrow \gamma\gamma/gg$ however, since here the Higgs boson couples to the photons/gluons only indirectly through a fermionic loop (interfering with a vector boson loop), it is not possible to obtain a direct handle on fermionic couplings utilising these modes.

Judging from the branching ratio alone, the most promising fermionic mode for this appears to be the decay into b quarks. Unfortunately this mode suffers from the difficulty of distinguishing b jets from light jets. The large abundance at the LHC of $t\bar{t}$ or single t quark production with b quarks in the final state also constitutes a vast background for this decay mode. In contrast, τ leptons (especially leptonically decaying ones) exhibit a signature that allows for better background rejection and triggering abilities. Therefore this study focuses on the Higgs bosons decay into τ leptons (more specifically: VBF $H \to \tau\tau$).

The τ lepton differs from electrons and muons by its large mass and the consequentially short lifetime. Thus, most of its decays take place already within the beam pipe and the τ lepton must be detected in an indirect fashion from its decay products. There are two different types of τ decays: leptonic (τ_{lep}) and hadronic (τ_{had}). Leptonic τ decays are decays into ν_τ, either e and μ, and the corresponding neutrino. Hadronic τ decays involve a ν_τ and a shower of hadronic particles such as pions. The corresponding branching ratios for the dacay of a pair of τ leptons are listed in Tab. 2.2.

Decay channel	Branching ratio [%]
$\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	46
$\tau_{\mathrm{had}}\tau_{\mathrm{had}}$	42
$\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$	12

Table 2.2: Branching ratios for the decay of two τ leptons into various final states differentiated by whether the τ leptons decay leptonically (lep) or hadronically (had). Values are taken from Ref. [3].

Out of the possible decays of a pair of τ leptons, the mode in which one decays leptonically while the other decays hadronically is the one with the highest branching ratio (see Tab 2.2). It has the advantage of having only 3 instead of 4 (as for $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$) neutrinos, which can only be reconstructed indirectly from the missing transverse energy in a detector, but still having one charged lepton in its final state (as opposed to $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$). A (leading order) Feynman diagram of the full VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ process can be seen in Fig. 2.4.
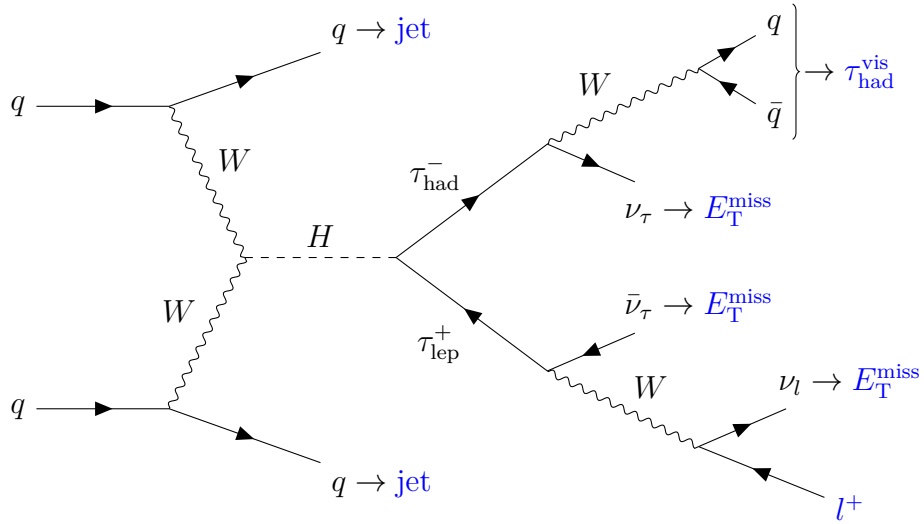
Figure 2.4: Example Feynman diagram at leading order for the VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$ process which is the signal process studied in this thesis. The objects by which the final state particles of this process might be reconstructed in the ATLAS detector are indicated in blue.

3 The ATLAS Experiment

The ATLAS detector is a multipurpose detector situated at the Large Hadron Collider (LHC). The LHC is described in sec. 3.1 while an overview of the ATLAS detector itself is presented in Sec. 3.2. Due the planned upgrade of the LHC to the High-Luminostity LHC (HL-LHC), the ATLAS detector will have to be upgraded as well. This is described in Sec. 3.3.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a proton-proton synchrotron which is able to provide the highest centre of mass energy $\sqrt{s} = 13$ TeV of any particle accelerator ever built. One of its main purposes was to enable the first observation of the Higgs boson. A schematic view of the LHC's several accelerator facilities and experiments can be found in Fig. 3.1. The four largest detector experiments: ATLAS [30], CMS [31], ALICE [32], and LHCb [33] are situated at the four interaction points in the LHC's main ring. ATLAS and CMS are general-purpose detectors probing a wide range of physics while ALICE and LHCb study more specialised processes such as quark-gluon plasma in heavy-ion processes and b-physics. The TOTEM, LHCf, and MoEDAL experiments are smaller detectors which are installed at the interaction points of CMS, ATLAS, and LHCb as well. The LHC utilises smaller accelerator rings as pre-accelerators for the particles before they reach the main accelerator ring. The Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) were predecessors of the current LHC and were repurposed to be part of this pre-accelerator chain.

The substantial energy loss by synchrotron radiation encountered when accelerating electrons, was one reason for the decision to design the LHC as a proton accelerator to be able to provide a high enough centre of mass energy. Because of the inverse proportionality [34]

$$E_{\mathrm{syn}} = \frac{Q^2 E_{\mathrm{beam}}^4}{3\epsilon_0 R (mc^2)^4} \tag{3.1}$$

of the emitted energy by synchrotron radiation E_{syn} per full orbit to the accelerator's radius R and the particles mass m, the choice of the proton instead of the electron, allowed for the construction of the LHC inside of the already existing tunnel of its predecessor, the Large Electron-Positron Collider (LEP). Here, c represents the vacuum speed of light, ϵ_0 is the vacuum permittivity, Q is the particle's electric charge, and E_{beam} is the beam energy. As a result of this, the LHC has a circumference of 27 km.
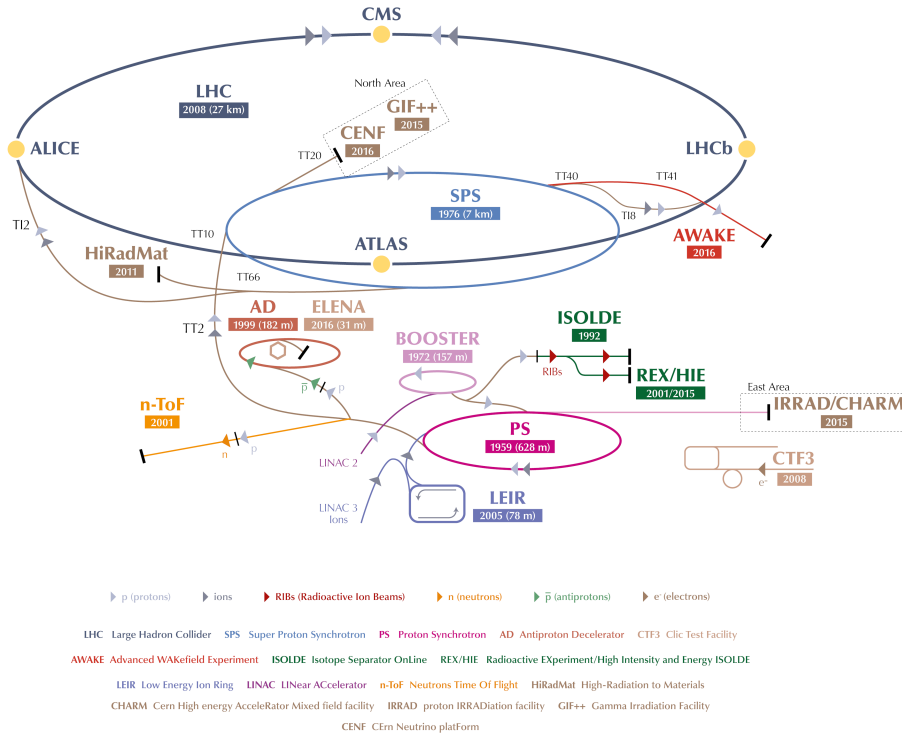
Figure 3.1: Schematic image of the the LHC accelerator complex with the locations of several experiments (source: [35]).

The use of hadrons (i.e. protons) instead of leptons (i.e. electrons) comes at a cost since hadrons are not elementary particles: While it is possible to provide a reasonably consistent centre of mass energy for the fundamental collision by adjusting the beam energies of leptonic accelerators, this is not possible for hadronic accelerators since only a stochastic proportion of the hadron's energies is made available as centre of mass energy. This is a consequence of the substructure of hadrons. On the other hand, this allows for the study of phenomena at various energy ranges at constant beam energy. An additional complication for hadronic accelerators comes from the fact that a large proportion of the events in such experiments are produced by the strong force which is notoriously difficult to model in simulations. In addition to the ability for proton collisions, the LHC can also be used for heavy ion collisions.

3.2 The ATLAS Detector

The ATLAS detector is a multi-purpose particle detector situated at one of the four interaction points of the LHC (see Fig. 3.1). A sketch of the detector layout can be seen in Fig. 3.2. A detailed description can be found in Ref. [30]. It is build in the form of several concentric layers serving specialised purposes for the detection of different kinds of objects originating from proton or heavy ion collisions.
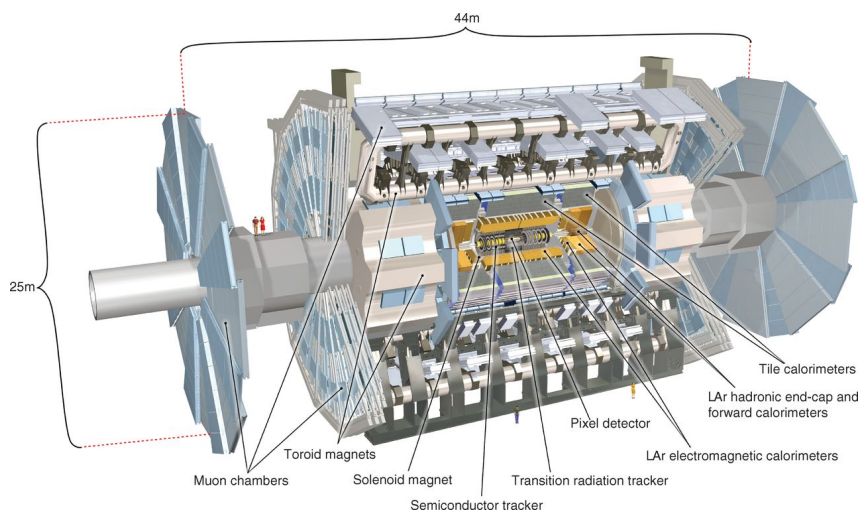
Figure 3.2: Schematic illustration of the ATLAS detector [36].

3.2.1 The ATLAS Coordinate System

Throughout this thesis, a right-handed orthogonal coordinate system with its origin situated at the reconstructed interaction point in the ATLAS detector is used. The positive x-axis points in the direction of the LHC's centre and the positive y-axis points upwards. The z-axis points along the direction of the beam. The transverse momentum of a particle in this coordinate system is defined as

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}. \tag{3.2}$$

Due to the cylindrical shape of the ATLAS detector, cylindrical coordinates are used as well. For this purpose the azimuthal angle ϕ and the polar angle θ are defined with respect to the z-axis.

The rapidity of a particle with energy E and three-momentum $\vec{p}$ is defined as

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right). \tag{3.3}$$

Differences of rapidity are invariant under Lorenz boosts along the z-axis (i.e. along the beam axis) of the coordinate system. In the limit where a particle's mass m is negligible with respect to its momentum $|\vec{p}|$, the rapidity can be approximated by the pseudorapidity

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{3.4}$$

The angular distance between two particles p_1 and p_2 is defined as

$$\Delta R = \sqrt{(\eta_2 - \eta_1)^2 + (\phi_2 - \phi_1)^2} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.5}$$

3.2.2 Inner Detector (ID)

Starting from the inside, the first segment of the ATLAS detector is the inner detector (ID), a tracking system immersed in a 2 T magnetic field whose purpose it is to bend tracks from charged particles in such a way that it is possible to infer their momentum from the curvature. It covers a pseudorapidity range of $|\eta| < 2.5$. The ID can be subdivided into a barrel region ($|\eta| < 2$) and an endcap region ($|\eta| > 1$) which are named according to their geometry. The former consists of three separate layers of pixel tracking detectors, four layers of silicon micro-strip detectors and a transition radiation tracker. The transition radiation tracker is able to provide tracking information over the whole pseudorapidity range of the ID by utilising drift tubes. Additionally, it can be used to identify electrons due to X-rays produced by interleaved radiators. The two end-cap regions are composed of three pixel discs, nine silicon micro-strip discs and 80 straw tube planes each.

A solenoid magnet creates a 2 T magnetic field in the region of the ID. This allows for the measurement of the transverse momentum p_{T} of charged particles by the resulting curvature of their tracks. The relative momentum resolution of the ID as a function of transverse momentum p_{T} in the barrel ($0 < |\eta| < 1.05$), transition ($1.05 < |\eta| < 1.7$), and

end-cap $(1.7 < |\eta| < 2.0)$ region for muons is given by [37]

$$\left(\frac{\sigma_p}{p}\right)_{\text{barrel}} = 0.016 \oplus 0.42 p_{\text{T}}[\text{TeV}] \tag{3.6}$$

$$\left(\frac{\sigma_p}{p}\right)_{\text{transit.}} = 0.026 \oplus 0.80 p_{\text{T}}[\text{TeV}] \tag{3.7}$$

$$\left(\frac{\sigma_p}{p}\right)_{\text{end-cap}} = 0.033 \oplus 0.99 p_{\text{T}}[\text{TeV}] \tag{3.8}$$

3.2.3 Calorimeters

Encompassing the ID and the solenoid field, the liquid-argon calorimeter (LAr) is designed to register energy depositions from particles. It consists of sampling calorimeters, possesses full azimuthal symmetry, and can be subdivided into one barrel region and two endcap regions. The innermost part of the LAr is an electromagnetic calorimeter which contains a barrel part $(|\eta| < 1.475)$ and an endcap part $(1.375 < |\eta| < 3.2)$ with three separate layers each, distinguished in granularity. In between the transition region of barrel and endcap, crack scintillators are installed $(1.2 < |\eta| < 1.6)$ with the purpose of countering the effect of the missing tagging instrumentation in this regions. Additionally, in a region of $|\eta| < 1.8$, a presampler consisting of active liquid-argon calorimetry is situated which can be used to estimate the energy lost before the other calorimeters. In the forward region, the LAr includes a hadronic calorimeter consisting of a copper liquid-argon hadronic endcap calorimeter $(1.5 < |\eta| < 3.2)$ and a copper/tungsten liquid argon forward calorimeter $(3.1 < |\eta| < 4.9)$ in the region closest to the beam pipe. The former has four, while the latter has three layers.

The LAr is surrounded by the hadronic tile calorimeter $(|\eta| < 1.7)$ built out of plastic scintillator tiles and steal absorbers. In radial direction it is segmented into three layers. The pseudorapidity region of $1.0 < |\eta| < 1.2$ of the hadronic tile calorimeter exhibits a gap of about 60 cm which is necessary for cabling and service of ID and LAr. Gap scintillators are installed in this region to compensate the corresponding loss in calorimeter resolution. Together, the hadronic part of the LAr and the hadronic tile calorimeter are referred to as HCAL.

The energy-deposition resolution for measurements of the ATLAS calorimeter differs for

electromagnetic (ECAL) and hadronic (HCAL) calorimeter:

$$\left(\frac{\sigma_{E_\mathrm{T}}}{E_\mathrm{T}}\right)_\mathrm{ECAL} = \frac{0.1}{\sqrt{E_\mathrm{T}[\mathrm{GeV}]}} \oplus 0.007 \tag{3.9}$$

$$\left(\frac{\sigma_{E_\mathrm{T}}}{E_\mathrm{T}}\right)_\mathrm{HCAL} = \frac{0.5}{\sqrt{E_\mathrm{T}[\mathrm{GeV}]}} \oplus 0.03 \tag{3.10}$$

3.2.4 Muon Spectrometer

The outermost part of the ATLAS detector is the muon spectrometer which is needed since most muons penetrate the ID and calorimeters without significant loss in p_T. It consists of three large air-core toroidal magnets, a barrel, and an endcap. The muon scintillator is comprised of three layers of tracking chambers and a dedicated trigger system. Like in the ID, the muon chambers utilise the curvature of charged particles due to the applied magnetic field to measure the corresponding transverse momentum. It is able to measure transverse momenta in a region up to $|\eta| < 2.7$ and is designed to trigger on particles up to $|\eta| < 2.4$. The relative momentum resolution of the muon spectrometer as a function of transverse momentum p_T in the barrel $(0 < |\eta| < 1.05)$, transition $(1.05 < |\eta| < 1.7)$, and end-cap $(1.7 < |\eta| < 2.0)$ region is given by [37]

$$\left(\frac{\sigma_p}{p}\right)_\mathrm{barrel} = 0.033 \oplus 0.17 p_\mathrm{T}[\mathrm{TeV}] \oplus \frac{0.25}{p_\mathrm{T}[\mathrm{TeV}]} \tag{3.11}$$

$$\left(\frac{\sigma_p}{p}\right)_\mathrm{transit.} = 0.065 \oplus 0.34 p_\mathrm{T}[\mathrm{TeV}] \tag{3.12}$$

$$\left(\frac{\sigma_p}{p}\right)_\mathrm{end\text{-}cap} = 0.038 \oplus 0.20 p_\mathrm{T}[\mathrm{TeV}] \tag{3.13}$$

3.2.5 Trigger System

The ATLAS trigger system is divided into a hardware-based level one (L1) trigger and a software-based higher level trigger (HLT). It is needed to reduce the amount of recorded events in such a way that it is possible to store them despite their rapid succession.
For this purpose, the L1 trigger reduces the rate form the LHC's bunch crossing rate of approximately 30 MHz to only a few hundred events per second that are not rejected by the HLT [38]. Those events are then stored on disk an can be used for analyses.

3.2.6 Monte Carlo Event Generation at ATLAS

Studies in particle physics cannot only rely on real data stemming from detectors situated at particle accelerators but are also in need for simulated samples which can be used to be compared to the data. For this purpose, Monte Carlo (MC) generators like HERWIG [39], PYTHIA [40], and SHERPA [41] where envisioned. These programs have the purpose of simulating the final state of particle interactions as close as possible to what one would observe in real experiments and to assign probabilities to such simulated events that correspond to the approximate probability of that event to occur in a real experiment with the given conditions. Non-trivial final states involving a large number of particles, thresholds and resolutions of real detectors, and the difficulty of modeling QCD interactions complicate this task significantly. Therefore this process is usually divided into several steps an illustration of which can be seen in Fig 3.3:
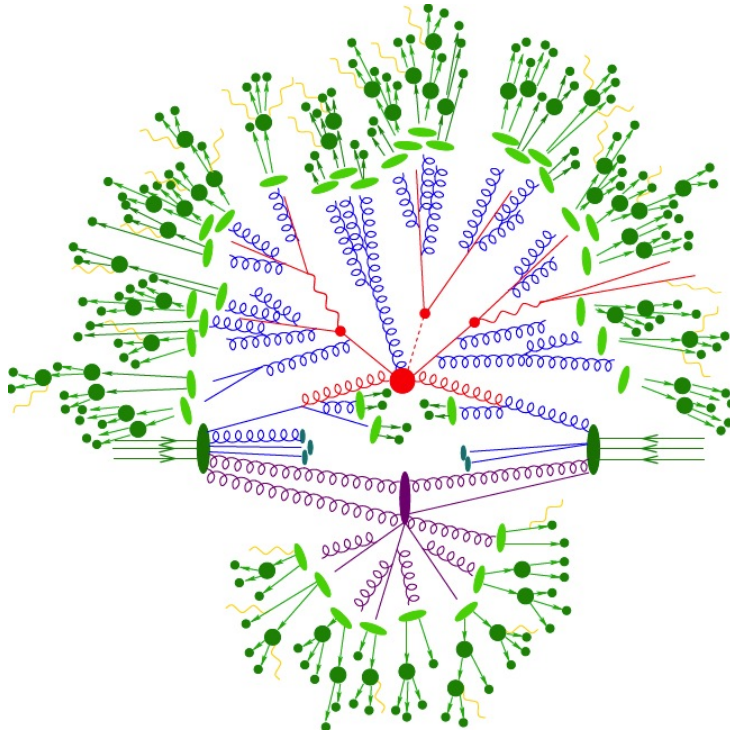
Figure 3.3: Schematic illustration of the different steps performed by Monte Carlo Generators: Hard scattering process via matrix element method (red), parton showers (blue), secondary interactions (purple), hadronization (light green), unstable hadron decay (dark green), and QED bremsstrahlung (yellow) [42].

First, the matrix element for the fundamental hard scatter interaction is calculated. This is exact up to the applied order in perturbation theory. The resulting particles may undergo radiation of softer particles which tend to be collinear with their emitters. This simulation step is called parton showering. Thereafter, the particles undergo a secondary

step of decays which are again first calculated using the matrix element method and then refined by parton showers. The hadronization binds the coloured partons into colourless hadrons. The decays of unstable hadrons is considered thereafter and as a last step the resulting particles may undergo bremsstrahlung in the form of QED processes. As circumstances require, simulations for the ATLAS detector (i.e. GEANT4 [43]), jet algorithms (see Sec. 4.3), and other subsequent steps can be applied to the resulting output to gain a sample which then can be compared to data from a real detector.

3.3 The High-Luminosity LHC

Many searches for rare processes done with the help of the LHC's data suffer from the lack of a sufficient number of events produced which leads to low significances that in some cases are mainly of statistical and to a lesser extent of systematic nature. In these cases a higher integrated luminosity promises to help to provide the necessary event yields consequently lessening the statistical uncertainties. The High-Luminosity LHC (HL-LHC) is a planned upgrade for the LHC that is scheduled to begin operation in 2026 and will increase the instantaneous luminosity of the accelerator by a factor of nearly eight compared to the current nominal value of $\mathscr{L} \approx 1 \times 10^{35}$ cm^{-2} s^{-1} [44]. Its goal is to provide an integrated luminosity of 3000 fb^{-1} by the end of its run in 2035 (see Fig. 3.4). In Run 3, which will occur before the HL-LHC's run, the centre of mass energy will be raised from 13 TeV to 14 TeV and will remain at this value for the HL-LHC.
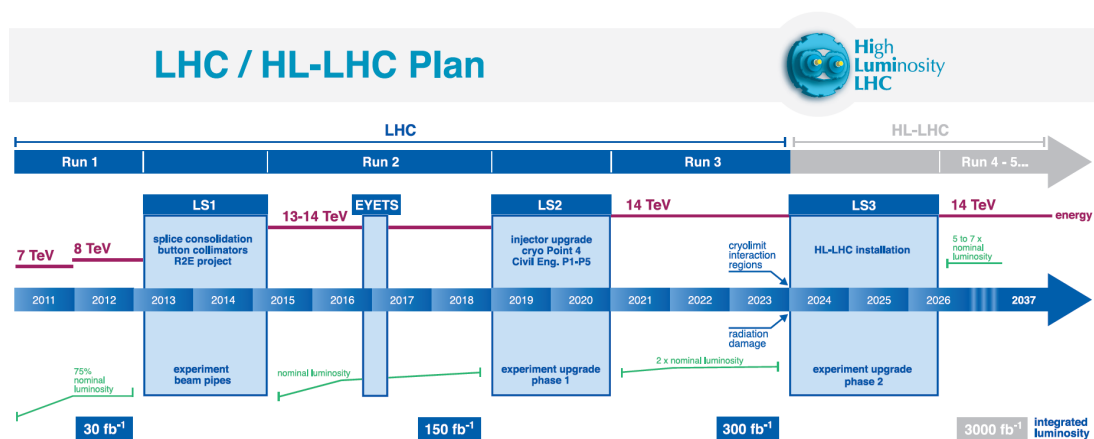
Figure 3.4: Timeline for the LHC's runs up to the planned upgrade to the High-Luminosity LHC [44].

Unfortunately, with an increase in instantaneous luminosity, there is also an increase in pileup effects. Pileup is a phenomenon that concerns the contemporaneous presence of

detector signatures originating from several collision events taking place in close temporal proximity. There are two kinds of pileup: In-time- and out-of-time pileup. While the former is initiated by various collisions taking place during the same bunch crossing as the event of interest, the latter stems from slow readout times of certain detector components that lead to the presence of signatures from previous bunch crossings overlapping with signatures of the current one. The average pileup value $\langle\mu\rangle$ which is defined as the average number of particle collisions per bunch crossing is expected to rise from 40 in the current LHC layout to 200 in the HL-LHC. Coping with the effects of the increased pileup will be one of the main challenges for studies concerning the upgraded LHC.

3.4 ATLAS at the HL-LHC

The harsher pileup conditions and subsequently increased radiation intensities of the High Luminosity LHC demand that the ATLAS detector also is upgraded to cope with the resulting need for better tracking and radiation hardness. For this purpose the individual detector elements which are described in Sec. 3.2 will be upgraded individually. For many parts, the design details of the final layout has not been fully determined. Consequently, this must be taken into account when making predictions for the expected performance of the detector.

3.4.1 Inner Tracker (ITk)

The inner detector (ID) will be completely replaced by the Inner Tracker (ITk) which will be an all-silicon tracking system. The reason for this is that the instrumentation with silicon detector elements allows for a higher tracking resolution which is needed for a better rejection of pileup tracks. It also allows for a faster readout to deal with the harsher pileup conditions at the HL-LHC. Additionally, this type of detector can endure more radiation than the straw tubes which are used in the current ATLAS detector.

The ITk is divided into a strip detector and a pixel detector. The planed layout is described in detail in the corresponding technical design reports in Ref. [4] (strip) and Ref. [5] (pixel). All values presented in this section stem from these documents as well. It is expected that the ITk strip detector will be exposed to a radiation level of approximately $1.2 \times 10^{15} \frac{n_{eq}}{\mathrm{cm}^2}$ and 50 MRad which is ten times the level the strip detector in the ID is currently exposed to. For the innermost part of the ITk pixel detector, the radiation level is expected to reach an even higher level of $1.2 \times 10^{16} \frac{n_{eq}}{\mathrm{cm}^2}$ and 1.7 GRad. The ITk is designed to provide tracking information in a region up to $|\eta| = 4$ while the current ID's acceptance stops at $|\eta| = 2.7$. This allows for the utilisation of tracking for pileup

rejection even in the very forward η region of the detector and will therefore play an important role for event signatures like VBF which include jets or other objects in that region.

The ITk has been designed with the goal of maintaining or outperforming the track parameter resolutions of the current ID despite the harsher pileup conditions of the HL-LHC. Fig. 3.5 shows simulation results of momentum and angular resolution for single muons which suggest that this goal will be reached:
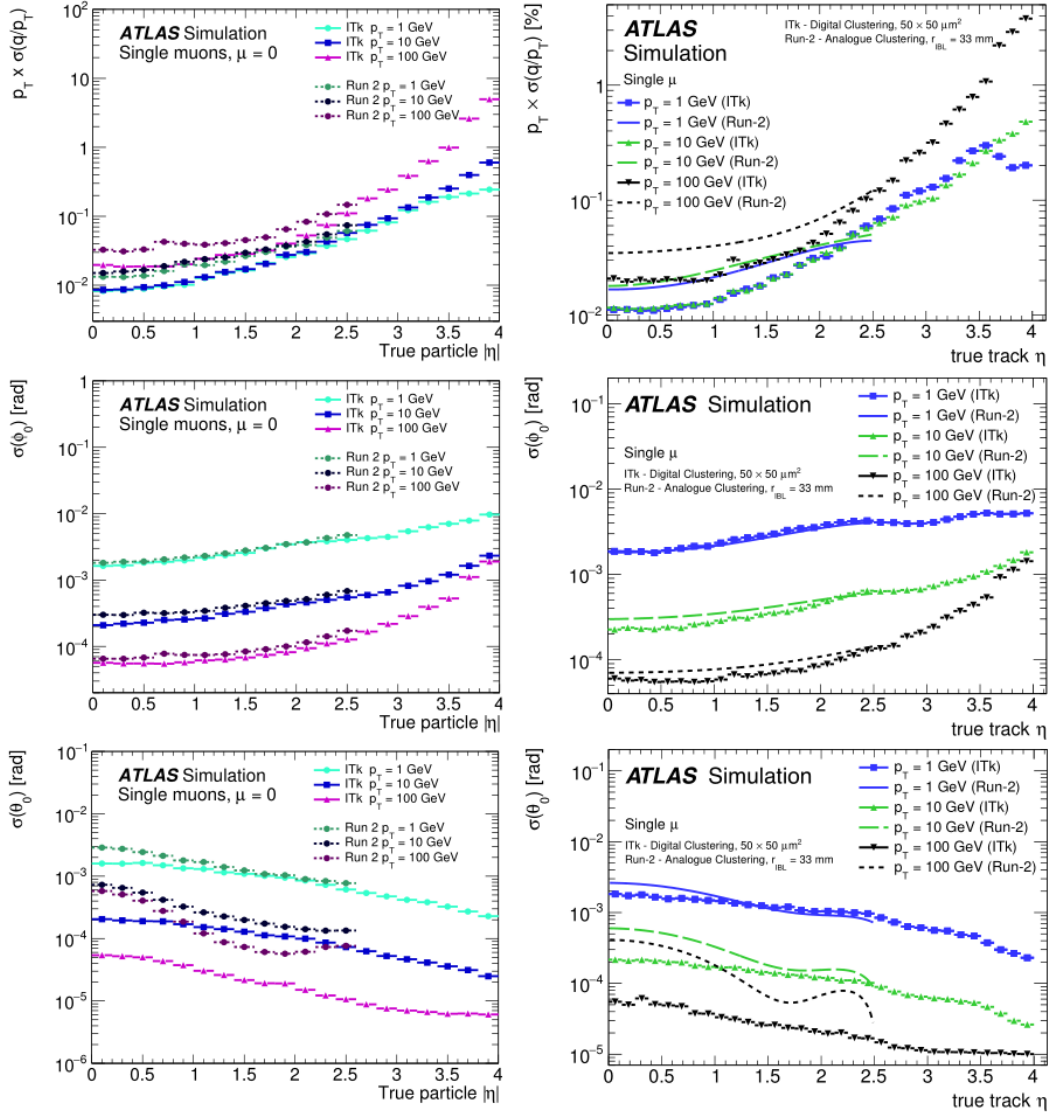
Figure 3.5: Transverse momentum p_T and angular θ_0 and ϕ_0 resolutions for the ITk strip [left] and pixel [right] detector compared to the Run 2 ID detector [4, 5].

The ITk exhibits smaller transverse momentum p_T and angular θ and ϕ resolutions

than the Run 2 ID in almost all η and p_T regions. Only in the region of $2.0 \lesssim |\eta| \lesssim 2.5$, does the ID appear to outperform the performance of the ITk pixel detector in terms of p_T, θ, and ϕ resolutions for low p_T muons. The ITk strip detector, however exhibits consistently better performance than the ID in this region. In particular the p_T, and ϕ resolutions for high p_T muons are significantly lower for both the ITk pixel and the ITk strip detectors compared to the ID. The most prominent feature of the ITk which can be seen in this plots however, is the drastically increased acceptance up to $|\eta| = 4$ compared to the ID's acceptance of $|\eta| = 2.7$. The current solenoid magnet providing a magnetic field of 2 T is planned to remain in the detector for the HL-LHC's run. Since the magnetic field available in the ITk will not be higher than the one used for the ID, it is expected that in the forward region ($\eta > 2.7$), momentum measurements by track curvature by the ITk will only be possible with large uncertainties [4, 5]. In Fig. 3.5, this can be seen by the p_T resolutions that yield values of order ≈ 1 and even higher. Hence, in this region, the ITk will be of use primarily due to its tracking and vertexing abilities which can be used to identify pileup objects.

3.4.2 Calorimeters

In accordance to the current layout of the ATLAS calorimeter system, the HL-LHC's ATLAS calorimeters consist of a scintillator tile calorimeter and a LAr calorimeter. The proposed upgrades for the LH-LHC are described in detail in the corresponding technical design reports [7] (tile) and [6] (LAr). The LAr calorimeter cells are expected to maintain operational capacity in their current design form. Nevertheless, upgrades to their read out electronics and powering system must be made to cope with the increased radiation levels. The planned upgrade of the ATLAS triggering and data acquisition (TDAQ) system will not be compatible with the current LAr readout electronics furthering the need for its upgrade. The current trigger scheme uses analogue signals corresponding to calorimeter towers build from calorimeter cells. In the upgraded triggering system, information for each individual calorimeter cell will be available in digital form. This requires a much higher bandwidth than the current layout but is made feasible due to technological advancements of the new ATLAS trigger system [9].

3.4.3 Muon Spectrometer

As for the calorimeter system, the muon spectrometer's readout and trigger electronics will be replaced for the HL-LHC update to make it compatible with the upgraded ATLAS trigger layout. The current powering system is not compatible with the expected radiation

levels of the HL-LHC and will be replaced as well. Additionally, a larger number of resistive plate chambers (RPC) will be installed to increase the acceptance and robustness of the muon trigger. The current thin gap chambers (TGC) in the barrel end-caps which are installed in the form of doublets will be replaced by triplets to counter the increased probability of random coincidences due to the harsher pileup conditions. This would otherwise lead to extremely high trigger rates. The proposed upgrades of the muon spectrometer for the LH-LHC are described in detail in the corresponding technical design report [8].

3.4.4 High Granularity Timing Detector (HGTD)

The High Granularity Timing Detector (HGTD) is a proposed additional detector segment for the ATLAS detector at the HL-LHC. Its layout, properties, and expected performance will be summarized in the corresponding technical design report which is not yet publicly available. The corresponding technical proposal can be found in Ref. [45]. If realized, the HGTD will be situated in the gap region between the two LAr cryostats and will be composed of Silicium Low Gain Avalanche Diodes (LGAD) with a time resolution of $\approx 30\frac{\text{ps}}{\text{mip}}$. Its proposed acceptance region of $2.4 < |\eta| < 4.0$ would make it suitable for improving the rejection of pileup jets in the forward region due to timing information. Fig. 3.6 shows the pileup-jet rejection with respect to hard-scatter jet efficiency in the HGTD's acceptance region compared for a layout of the ATLAS detector with the HGTD and a layout without the HGTD. The layout with HGTD shows consistently higher rejection efficiencies.
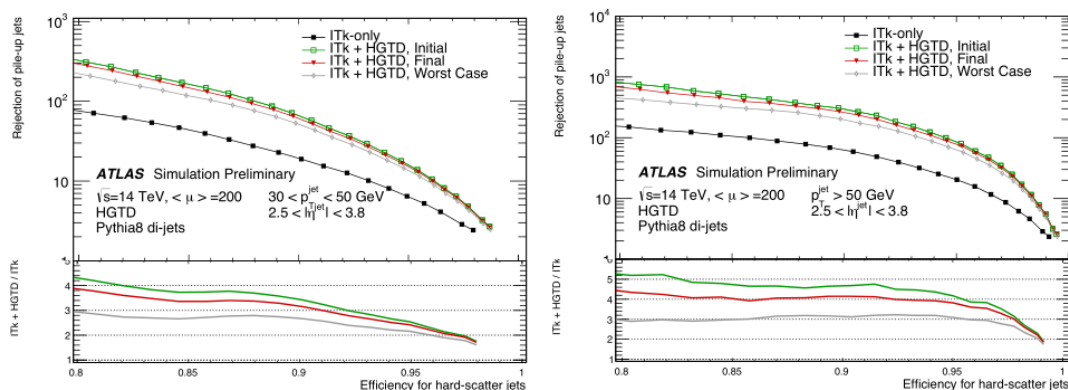
Figure 3.6: Pileup-jet rejection as a function of hard-scatter jet efficiency in the $2.4 < |\eta| < 4.0$ region in the ATLAS detector at the HL-LHC for layouts with and without the HGTD. The plots have been produced for jets with [left] $30\text{ GeV} < p_{\text{T}} < 50\text{ GeV}$ and [right] $p_{\text{T}} > 50\text{ GeV}$ [45].

4 Phenomenology of Jets

Quarks and gluons are particles that leave notoriously complicated detector signatures. They cannot simply be reconstructed from a single track in a particle detector. Due to confinement, they produce a shower of hadronic decay products that in turn produce a variety of different detector signals. This process is called hadronization. This is further described in Sec. 4.1. Since the particle of interest for most analyses is the one at the beginning of such decay chains, the need of reconstructing its properties from the variety of signatures present in the detector arises. For this purpose, algorithms for clustering and jet-finding were developed. Although, in strictly theoretical sense, the objects created by these algorithms do not stand in a one-to-one correspondence with the underlying physical particle, they represent a realistically obtainable and useful parametrization thereof. A short overview on different examples of such algorithms is given in Sec. 4.2 and 4.3. The outcome of such algorithms is called a jet.

4.1 Hadronization

The presence of quarks and gluons is an important property of many decay processes under study at the LHC. Nevertheless, it is not possible to directly detect these particles due to colour confinement. This phenomenon concerns all colour charged particles and has the effect that these particles group together to form hadrons (i.e. hadronize) before they decay or interact with the detector material. The only exception to this in the SM is the top quark which, due to its large total decay width of $\Gamma_t = 1.3$ GeV [46] has an average lifetime of $\tau_t = \frac{\hbar}{\Gamma_t} = 5 \times 10^{-25}$ s which is shorter than the average hadronization timescale $\tau_{\mathrm{hadronization}} = \frac{\hbar}{\Lambda_{\mathrm{QCD}}} = 3 \times 10^{-24}$ s [47].

Colour confinement is a consequence of the non-abelian nature of $SU(3)$, which is the underlying gauge symmetry of the strong force. Consequently, gluons are able to couple to themselves. Phenomenologically, this can be understood by invoking the concept of QCD flux tubes. When two coloured particles are separated, these string-like excitation states of the gluon field form between them and acquire more tension the larger the separation grows. As a result, QCD interactions become asymptotically weaker as the energy scale

increases or the length scale decreases. This phenomenon is called asymptotic freedom. Besides $SU(3)$, one other non-abelian symmetry group is part of the Standard Model: $SU(2)$. Consequently, one might expect asymptotic freedom and subsequently confinement also to occur in processes involving the weak force. This is prevented by the masses of the corresponding gauge bosons W and Z, which do not allow the formation of QCD-like flux tubes out of these bosons. Since no analytically exact theory of hadronization is available, it is necessary to use phenomenological models. There are two of such models which are utilised in most Monte Carlo event generators: The Lund String model [48] describes hadronization by the formation of a string of uniform energy per unit length. As the coloured particles at the end of the string separate, the string gains more energy until it eventually breaks down forming colourless bound states of quarks (i.e. hadrons). The string must fragment in such a way that the hadrons thus formed have the correct masses. This kind of model is used in the PYTHIA Monte Carlo generator [40].

The Cluster fragmentation model [49] starts by splitting gluons into $q\bar{q}$ pairs. These quarks are collected into several clusters which are overall colourless. In each individual cluster, hadrons are formed out of the quarks. The Cluster fragmentation model is used in the HERWIG Monte Carlo generator [39]. A schematic comparison of both models can be seen in Fig. 4.1.
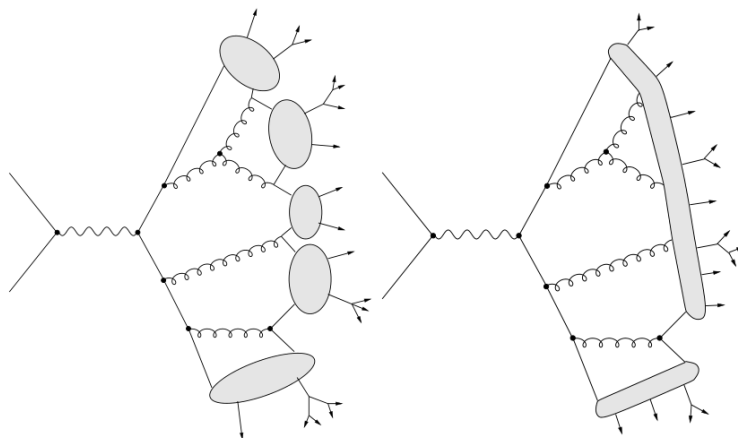
Figure 4.1: Schematic illustration of Cluster fragmentation [left] and the Lund String model of fragmentation [right] [50].

4.2 Calorimeter Clusters and Particle Jets

Two different kinds of jets are of particular interest for this analysis: EM-scale jets and particle-level jets. The former are also called calo jets because they are constructed from topological clusters [51] of energy depositions in calorimeters. EM-scale refers to the cal-

ibration of the calorimeter cells in which the non-compensation [52] of the calorimeter response on hadrons relatively to electrons is not accounted for.

In the first step of the clustering process (see Fig. 4.2) as done by the ATLAS collaboration, all cells registering an energy E_{cell} corresponding to at least four times the average noise threshold σ of the calorimeter cells are searched for. These are called seed cells. As the second step, all neighbouring cells to seed cells are added to the cluster if an energy exceeding 2σ from the noise threshold was deposited in them. As the last step, all cells with positive energy that are directly next to the neighbours are added. The resulting collection of cells is called a topological cluster. A detailed description of topological clusters can be found in Ref. [51]. By utilizing a jet algorithm (see Sec. 4.3), out of such clusters EM-scale calo jets are formed.

Particle-level- or truth jets are reconstructed from the four-momenta of stable truth particles in Monte Carlo generated samples. As a consequence, it is expected that the energy and momentum of these jets constitutes an approximation of the energy and momentum of the underlying initiating particle.
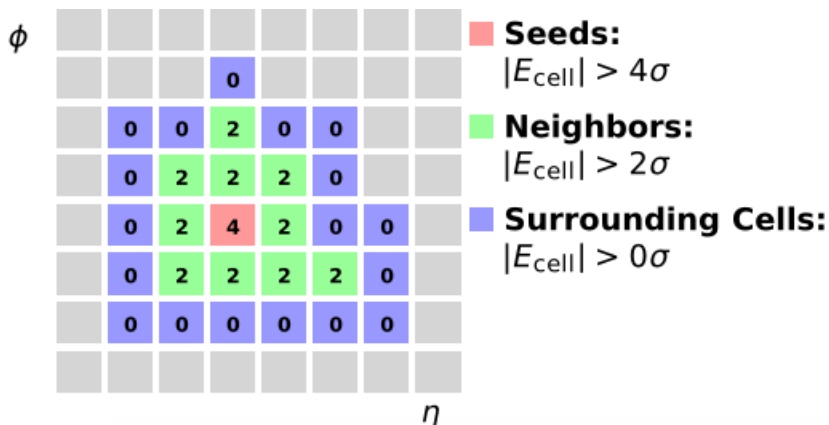
Figure 4.2: Two-dimensional Illustration of the necessary clustering steps to construct topo-cluster calo jets in the ATLAS experiment. The actual topo-clusters are three-dimensional objects.

4.3 Jet-Algorithms

Jet-algorithms are designed to cluster Monte Carlo generated truth particles together in such a way that the outcome corresponds in direction and energy as close as possible to the parton that initiated these signatures. Apart from that, it is possible to use objects constructed from tracks or clusters of energy depositions in calorimeters (e.g. topo-clusters as described in Sec. 4.2) from Monte Carlo generated samples or from data as input as

well. A more detailed description and comparison of the jet algorithms presented in the following can be found in Ref. [53].

To constitute a full jet definition, not only an algorithm has to be specified, but also a recombination scheme should be specified. It defines how the momentum of the jet should be calculated by the momenta of the objects of which it consists. In the case of iterative algorithms, the recombination scheme is applied at each individual iteration. The simplest recombination scheme is the four-vector sum. It is used in all studies presented throughout this thesis as recommended by Ref. [54]. In this scheme, the sum of the four-momenta of the combined objects is assigned to the resulting jet.

Most jet algorithms can either be classified as a cone algorithm or a sequential recombination algorithm. One of the most simple examples of the first kind is the iterative cone algorithm with progressive removal (IC-PR): First, a seed particle i is set in some initial direction. Often, the particle with highest transverse momentum is defined to serve as the seed. Then, all particles $\{j\}$ with an angular distance ΔR_{ij} smaller than a dimensionless parameter R are added to the new jet. The four-momentum is calculated with the respective recombination scheme. Following, the direction of the new jet is used as the new seed particle i. This procedure is repeated until the direction of i no longer changes. The result is then called a jet and all contained particles are removed from the list of particles. From the remaining particles, the one with highest transverse momentum is used as the new seed and the procedure is repeated. The algorithm stops as soon as no particles are left.

The k_T algorithm [55] is a sequential recombination algorithm. Together with a variation of it, the anti-k_T algorithm [56], it is used in most of the studies presented in this thesis. The individual steps are defined as follows:

1) For all particles $\{i\}$ and two-particle combinations $\{ij\}$ calculate the quantities

$$d_{ij} = \min(p_{\mathrm{T}\,i}^{2\lambda},\ p_{\mathrm{T}\,j}^{2\lambda})\frac{\Delta R_{ij}}{R^2} \text{ and } d_i = p_{\mathrm{T}\,i}. \tag{4.1}$$

2) Find the minimum of all d_{ij} and d_i.

3) If it is a d_{ij}, recombine i and j into a single object and return to step 1.

4) Otherwise (i.e. it is a d_i), declare i as a final-state jet, remove it from the list of particles, and return to step 1.

5) Stop when no particles remain.

Like in the IC-PR algorithm, R is a dimensionless quantity and can be varied as a free parameter of the algorithm. Throughout this thesis $R = 0.4$ is used. For the k_T algorithm, the parameter λ is defined to be 1. With the same set of steps, it is possible to define two other widely used algorithms: The anti-k_T algorithm [56], which is defined by $\lambda = -1$, and the Cambridge-Aachen algorithm [57], which is defined by $\lambda = 0$.

Eq. 4.1 consists of two factors:

$$F_1 = \min(p_{\mathrm{T}}^{2\lambda}{}_i, \ p_{\mathrm{T}}^{2\lambda}{}_j) \text{ and} \tag{4.2}$$

$$F_2 = \frac{\Delta R_{ij}}{R^2}. \tag{4.3}$$

F_1 concerns the transverse momenta of the particles while F_2 concerns their geometrical positions in the detector. Since λ is only present in F_1, it can be interpreted as a parameter giving a relative weight on the significance of momentum compared to the significance of angular direction in the respective algorithm. Hence, the anti-k_T algorithm ($\lambda = -1$) is more sensitive to angular direction than the k_T algorithm $\lambda = 1$. The former is therefore on average not as susceptible to additional transverse momentum contributions from pileup effects. The Cambridge-Aachen algorithm ($\lambda = 0$) does not take the transverse momentum into account at all.

Infrared and collinear (IRC) safety is an important feature for a jet-algorithm to be useful. It is a property that ensures, that the set of hard final-state jets found in an event does not change due to the addition of collinear splitting or soft radiation. Collinear unsafe algorithms are susceptible to variations due to the splitting of a hard particle. The same problem arises for infrared unsafe algorithms when a soft gluon is added to the system (e.g. by radiation). In fixed-order perturbative QCD calculations, soft emissions and collinear splittings correspond to divergent tree-level matrix elements. In calculations utilizing IRC-save algorithms, these divergencies cancel out due to diverging loop matrix elements that contribute with opposite sign. In IRC-unsafe algorithms however, the tree-level splittings and the loop-diagrams may lead to two different sets of jets [53]. This voids the cancellation and leads to unphysical, infinite results. It can be shown that the IC-PR algorithm does not fulfill the condition of IRC-safety, while all of the three sequential recombination algorithms presented above do [53].

5 The VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ Process at the HL-LHC

Since the τ lepton is the heaviest lepton of the Standard Model, its coupling strength to the Higgs boson is the highest of all leptonic Higgs couplings. Hence, measurements of $H \to \tau\tau$ decays play an important role for studying the leptonic part of the Higgs sector and can give indications on whether the boson observed by the ATLAS and CMS collaborations in 2012 [1, 2] can be identified as the Higgs boson of the Standard Model. The $H \to \tau\tau$ process has been independently observed by the ATLAS and CMS collaboration with an observed (expected) significance of 6.4 (5.4) and 5.9 (5.9) standard deviations [58, 59]. No significant deviation of the corresponding cross-section to the predicted Standard Model cross-section was observed.

The expected integrated luminosity of 3000 fb^{-1} at the end of the HL-LHC's run promises to increase the sensitivity to processes as $H \to \tau\tau$ and therefore to enable more precise measurements concerning the exclusive properties of this decay mode. But the changed layout of many parts of the ATLAS detector at the HL-LHC and the harsher pileup conditions due to the increased instantaneous luminosity have to be taken into account.

The following sections present an analysis on VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ prospects at the HL-LHC. It has the goal of providing an estimate for the sensitivity to this process that could be reached with the ATLAS detector using a cut based optimisation. The simulations concerning detector performance were calculated using the same tools and values as used in the technical design reports (Ref. [5] and [4]) of the ITk.

Sec. 5.1 describes the expected detector signatures of both the signal and the individual background processes. The following Sec. 5.2 contains a list of all Monte Carlo samples that have been used in the analysis to model these channels and describes the process of simulating the expected conditions at the ATLAS detector at the HL-LHC. Sec. 5.3 presents a cut based analysis in which the expected sensitivity of the ATLAS detector to the VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ process is derived.

5.1 Signal and Background Event Signatures

The signal process investigated in this analysis is VBF $H \rightarrow \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$. A leading order Feynman diagram of this process can be seen in Fig. 2.4. The final state objects of this process are one light lepton (e or μ) stemming from the leptonically decaying τ lepton (τ_{lep}), one hadronically decaying τ lepton (τ_{had}), two jets from the VBF production process, and missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ from three neutrinos (one from τ_{had} and two from τ_{lep}). In the following, the background processes which were considered in this thesis are described. They were chosen because they exhibit a similar signature of final state objects as the signal process. Two categories were considered: Backgrounds which contain a real τ lepton and electron or muon in their final states and fake backgrounds in which at least one of these objects was faked by a jet. The process of jets faking other objects is described in Sec 5.2.3. In the following, such objects are called fake τ lepton/e/μ candidates.

The Higgs decay into two τ leptons exhibits a very similar signature to the $Z \rightarrow \tau\tau$ processes. The significantly lower mass of $(91.1876 \pm 0,0021)$ GeV [3] of the Z boson compared to the Higgs boson's mass of (125.09 ± 0.24) GeV [3] promises to be a good discriminant to separate this kind of background from the signal processes. This can be achieved by reconstructing the initial mass of the particle decaying into the pair of τ leptons from its decay products.

There are two additional possible background processes involving Z boson decays: $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$. While both of them do not contain a τ lepton in their final state, it is possible for such events to exhibit a similar signature to signal events when one of the leptons is not reconstructed while an initial state jet is misidentified as a τ. Such erroneous τ leptons will be called fake τ candidates in the following. Like τ leptons, electrons are reconstructed from energy depositions in the calorimeter. Hence, for the $Z \rightarrow ee$ process in particular, it is also possible for one of the electrons to be misidentified as a τ lepton. This was not taken into account in the analysis presented in this thesis since the corresponding fake rates were not available (see Sec. 5.2.3).

Fake τ leptons also are possible in other background processes. One such background is the W+jets process where the W boson decays leptonically, leaving a similar signature to a leptonically decaying τ, while the jet is misidentified as a fake τ candidate.

Background from diboson events can also contribute since similarly to τ leptons, vector bosons can decay either leptonically or hadronically. The same case can be made for $t\bar{t}$ and single top quark processes since top quarks decay dominantly into a b quark and a W boson of which the latter can further decay either leptonically or hadronically. Due to the top quark's high branching ratio of $(95.7 \pm 3.4)\%$ [3] for decays into a b quark, the

identification of jets initiated by b quarks (i.e. b tagging) plays an important role for this kind of background.

One kinematic property that helps with the separation of the signal process VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ from all of the background processes are the unique jet kinematics of the VBF production process for the Higgs boson as described in Sec. 2.3: Additionally to the Higgs boson's decay products, the VBF H production process involves two quarks. These quarks hadronize and can therefore be reconstructed in the form of two jets (in the following called VBF jets). Since the LHC is a symmetric collider (both beam energies are approximately the same) this results in a topology in which both VBF jets propagate in approximately the same direction as the beams. Accordingly, the VBF jets j_i^{VBF} tend to have large values of absolute pseudorapidity $|\eta(j_i^{\mathrm{VBF}})|$ and tend to point in opposite directions in η (i.e. large separation $\Delta\eta(j_1^{\mathrm{VBF}}, j_2^{\mathrm{VBF}})$). As a consequence of this, selection criteria using the kinematic parameters of the leading jets in an event can be used to filter out background events.

5.2 Simulated Samples and Detector Simulation

All samples used in this study are truth level simulations without detector inefficiencies and resolutions applied to them. Given that the exact layout and properties of certain parts of the detector are still not known with complete certainty, these uncertainties will later have to be taken into account in this study comparing different possible layouts. The detector resolutions and efficiencies used in the following sections have bean derived by modeling particle interactions in different layouts of the upgraded ATLAS detector with GEANT4 [43]. This process is presented in the technical design reports of the individual segments of the ATLAS detector at the HL-LHC [4–9].

The Monte-Carlo simulated samples used for modeling the signal VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ process and the corresponding background processes in this study are presented in Sec. 5.2.1. The following Sec. 5.2.2 presents the application of the simulated detector resolutions to the truth samples and 5.2.3 introduces the concepts of reconstruction efficiencies and fake rates. The individual steps are summarized in the form of a flow chart in Fig. 5.1.
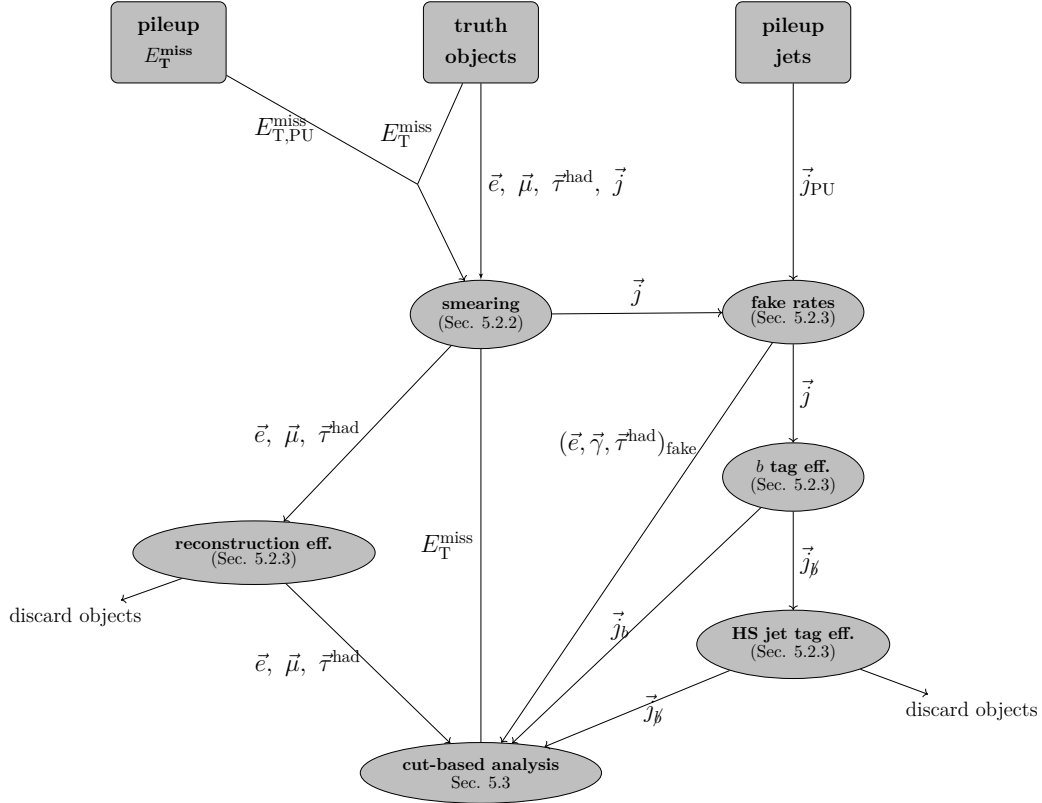
Figure 5.1: Flow chart representing the individual steps of the process of event and detector simulation for the ATLAS detector at HL-LHC as described in Sec. 5.2. The objects being passed from step to step are indicated by labels on the arrows. A vector arrow on top of an object's name indicates that it is a set of objects.

5.2.1 Monte Carlo Samples

Since the HL-LHC will begin its operations in 2026 at the earliest, no real HL-LHC data are available yet. Consequently, only samples simulated via Monte Carlo (MC) generators are used in this analysis. A full list of the samples used for both signal- and background estimation including information on the Monte Carlo generator used for their production and the corresponding cross-section σ at $\sqrt{s} = 13$ TeV and 14 TeV can be seen in Tab. 5.1. Since no complete set of samples corresponding to a centre of mass energy of $\sqrt{s} = 14$ TeV was available at the time, 13 TeV samples were used. The difference in $\sqrt{s}$ was accounted for by scaling their respective cross-sections by the factor obtained from dividing the expected cross-section at 14 TeV by the one for 13 TeV. For the background processes, the cross-sections for both centre of mass energies correspond to the MC generator cross-sections of the used 13 TeV samples and corresponding samples produced at 14 TeV. The cross-sections for the signal process were taken from Ref. [60].

Sample	Events	σ(13 TeV) [pb]	σ(14 TeV) [pb]	$\int \mathcal{L}^{\text{eff}}$(14 TeV)dt [fb^{-1}]
VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$	800000	0.01913336	0.02163983	36968.867
ggF $H \to \tau_{\text{lep}}\tau_{\text{had}}$	3.00×10^6	0.2302695	0.2590532	11580.633
di-boson	1.35×10^8	90.20987	97.42666	1385.658
W+jets	5.60×10^8	58767.3	63468.68	8.82325
$Z \to \tau\tau$	1.53×10^8	19011	20575	7.436209
$Z \to \mu\mu$	2.32×10^8	19011	20575	11.27582
$Z \to ee$	3.12×10^8	19011	20575	15.16403
$t\bar{t}$	4.68×10^8	729.77	824.5102	567.6097
single t	3.30×10^7	140.8279	159.1105	207.403

Table 5.1: *Monte Carlo (MC) generated samples used to model the signal (VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$) and background channels. Shown are the number of events in the samples, the cross-sections σ at $\sqrt{s} = 13$ TeV and 14 TeV, and the corresponding effective integrated luminosity $\int \mathcal{L}^{\text{eff}}$(14 TeV)dt which was calculated by dividing the number of events in the sample by σ(14 TeV). The corresponding dataset IDs are listed in Tab. A.1.*

Sample	Generator	UEPS Model	Tune	PDF set
VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$	Powheg v2	Pythia v8.212	AZNLO	CTEQ6L1
ggF $H \to \tau_{\text{lep}}\tau_{\text{had}}$	Powheg v2	Pythia v8.212	AZNLO	CTEQ6L1
di-boson	Powheg v2	Pythia v8.210	AZNLO	CTEQ6L1
W+jets	Powheg v1	Pythia v8.186	AZNLO	CTEQ6L1
$Z \to \tau\tau$	Powheg v1	Pythia v8.186	AZNLO	CTEQ6L1
$Z \to \mu\mu$	Powheg v1	Pythia v8.186	AZNLO	CTEQ6L1
$Z \to ee$	Powheg v1	Pythia v8.186	AZNLO	CTEQ6L1
$t\bar{t}$	Powheg v2	Pythia v8.230	A14	NNPDF23LO
single t	Powheg v1	Pythia v6.428	Perugia2012	CTEQ6L1

Table 5.2: *MC generators, UEPS models, sets of tuned parameters, and PDF sets of all samples used in this analysis.*

The vector boson fusion (VBF) signal process and the gluon gluon fusion (ggF) background process were simulated at next-to-leading order (NLO) accuracy in QCD with POWHEG v2 [61–64] using the MiNLO approach [65]. They are the same samples as the ones used in the ATLAS Run 2 analysis of the $H \to \tau\tau$ process [58]. The CTEQ6L1 [66] PDF set and the AZNLO [67] set of tuned parameters was used. Photos++ v3.52

[68] was used to model QED emissions from electroweak vertices and charged leptons. Parton shower, hadronization and underlying event (UEPS) were modeled with PYTHIA v8.212 [40]. The MC generators, UEPS models, sets of tuned parameters, and PDF sets of all samples used in this analysis are summarized in Tab. 5.2. Additionally to the tools referenced above, this list contains the A14 [69] and Perugia [70] set of tuned parameters and the NNPDF23LO PDF set [71].

5.2.2 Detector Resolutions

The approach to modeling the detector resolution that is taken in this study is a drastically simplified one. This process is called smearing: It is assumed that all resolutions are of purely Gaussian nature and that at least for electrons, muons, τ leptons, and jets, the angular resolution is negligible compared to the energy/momentum resolution. Furthermore, it is assumed that the Gaussian widths of these resolutions depend only on the energy E and pseudorapidity η of the object in question. In the case of hadronic τ leptons, it also depends on the number of prongs, i.e. the number of charged tracks of its decay products. Whereas the momenta of electrons and muons itself were smeared, for τ leptons a different approach was chosen. This is a consequence of the τ lepton's short lifetime because of which such leptons can only be reconstructed by their decay products. Consequently, for τ leptons, only the visible momenta were smeared, rather than he full momenta. The visible momentum of leptonically decaying τ leptons is represented by the momentum of the charged lepton in the τ decay while for hadronically decaying τ leptons, the visible momentum is defined as $p_T(\tau^{vis}) = p_T(\tau) - p_T(\nu)$ where $p_T(\tau)$ is the τ lepton's momentum and $p_T(\nu)$ the momentum of the τ-neutrino into which it decays.

Since it is not possible to directly detect neutrinos in the ATLAS detector, there is no need to apply the smearing process to them directly either. Instead, the missing transverse energy E_T^{miss} within the detector is smeared. This was done independently from the smearing process of the other objects. One important feature in this regard is that prior to the smearing of E_T^{miss}, additional missing transverse energy is added due to pileup. For the jets, a similar approach is followed, i.e. a pileup-overlay of numerous (already smeared) jets is added to the jets that are present in the hard-scatter interaction[1]. Fig. 5.2 shows the resulting resolutions for the smearing of electrons, muons, τ leptons, jets, and E_T^{miss} produced with either all signal samples or all background samples combined.

---

[1]Objects and quantities before smearing will be called "truth-" in the following, while the resulting objects/quantities after the smearing process will be called "smeared-".
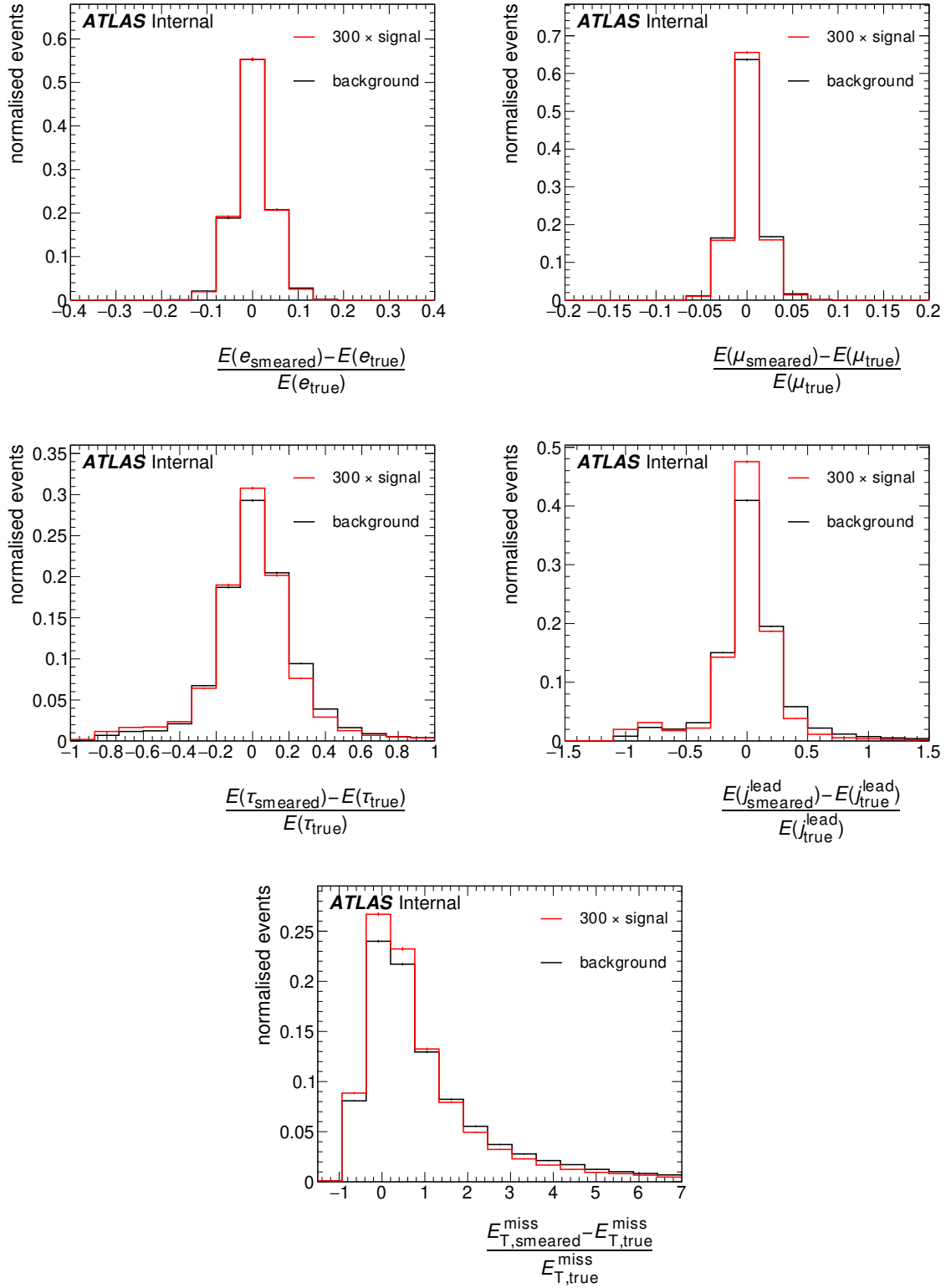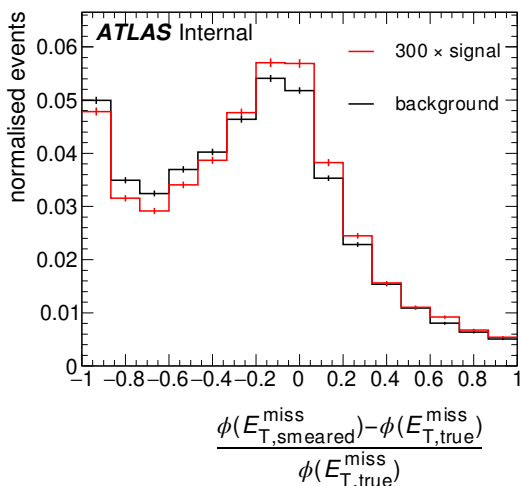
Figure 5.2: Energy resolutions for different objects in the VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ signal and the background after the application of the smearing process as described in Sec. 5.2.2. The represented objects are (from top left to bottom right): Electrons, muons, hadronically decaying τ leptons, the jet j^{lead} with highest transverse momentum, and missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$.

The width of the resolution for muons is by far the smallest. The reason for this is that for the reconstruction of muons, information from a dedicated detector element, the muon chambers (see Sec. 3.2.4) is used additionally to the information of the Inner Detector. This is due to the muon's ability to traverse most of the detectors material without great loss of energy and allows for a particularly clean signature. Electrons, which are reconstructed from their energy deposition in the calorimeters in combination with tracking information from the Inner Detector exhibit a slightly worse resolution. Hadronically decaying τ leptons and jets are reconstructed from energy depositions in calorimeters. The calorimeter's momentum resolution is only smaller than the ID's momentum resolution for particles with large transverse momenta. Hence, the energy resolution of hadronically decaying τ leptons and jets is on average larger than for electrons and muons.

By far the worst resolution can be seen for the missing transverse momentum E_T^{miss}. Not only does it exhibit the largest width; The distribution is also drastically asymmetric. This is due to the fact that in the smearing process of E_T^{miss}, additional E_T^{miss} from pileup is added. Its value is therefore on average larger after smearing than it was before. Due to the harsh pileup conditions at the HL-LHC, this can lead to situations in which the smeared E_T^{miss} is dominated by the contribution from the additional pileup E_T^{miss}. In this case, the angular direction $\phi(E_T^{miss})$ of the missing transverse momentum's vector in the transverse plane may be drastically different to its value before smearing. The corresponding resolution of $\phi(E_T^{miss})$ can be seen in Fig. 5.3.

Figure 5.3: Resolution of the angular direction $\phi(E_T^{miss})$ of the missing transverse momentum's vector in the transverse plane in the VBF $H \to \tau_{lep}\tau_{had}$ signal and the background after the application of the smearing process as described in Sec. 5.2.2.

As expected, it shows a wide spread. This is an important issue when calculating the reconstructed di-τ mass with the Missing Mass Calculator (MMC) algorithm. A detailed description of this is presented in Sec. 5.3.2.

5.2.3 Reconstruction Efficiencies and Fake Rates

Efficiencies are probabilities of objects to be identified as candidates in the analysis. Similarly, fake rates describe the probabilities of jets being misidentified as a different object. In this study, four different kinds of such efficiencies and rates were considered:
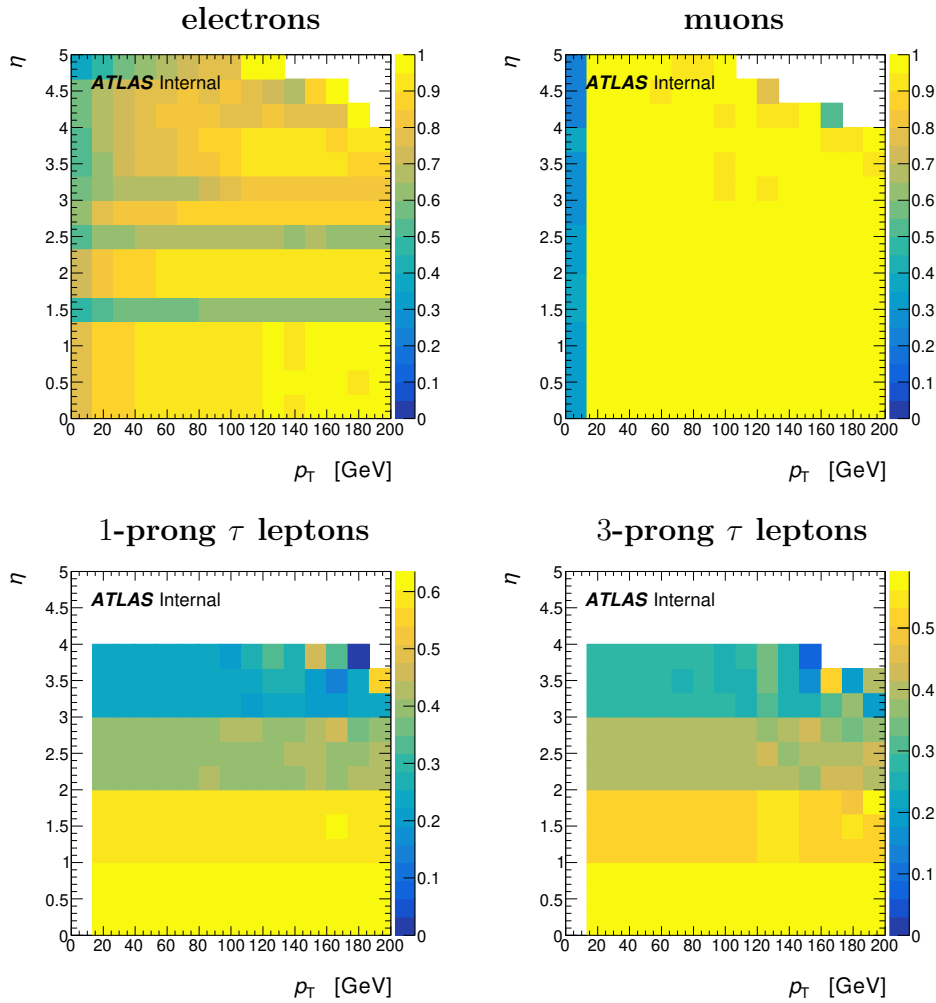
- The efficiencies of electrons, muons, or τ leptons to be reconstructed and identified

- b tagging efficiencies, i.e. the probability of a jet being identified as a b-jet

- HS jet tagging efficiencies, i.e. the probability of a jet being identified as a hard scatter (HS) jet (i.e. not a pileup (PU) jet)

- Fake rates, i.e. the probability of a jet being identified as either e or τ lepton

These efficiencies and fake rates depend on η and p_T, in the case of jets on whether it is a true pileup jet or not, and in the case of τ leptons on their number of prongs. Each efficiency is compared to a randomly generated number from a uniform distribution between 0 and 1 and depending on it being higher or lower the corresponding action is applied: The particle is discarded, or in the case of fake rates, an e or τ with the four momentum calculated by the initial jets four momentum is added. The four different kinds of efficiencies are discussed in the following:

Reconstruction efficiencies: Fig. 5.4 shows the reconstruction efficiencies for electrons, muons, 1 prong τ leptons, and 3 prong τ leptons. It is apparent that muons possess on average the highest reconstruction efficiency in all η and p_T regions. Electrons exhibit a slightly worse reconstruction efficiency. One prominent feature is the occurrence of drastically lower efficiencies especially in the pseudorapidity region around $|\eta| \approx 1.5$. This corresponds to the transition region between barrel and endcap LAr calorimeters of $1.37 \lesssim |\eta| \lesssim 1.52$ which contains significantly less instrumentation. In the object definition of electrons described in Sec. 5.3.1, electrons in this region are therefore excluded. Similarly, hadronically decaying τ leptons (also reconstructed from the calorimeter signatures) in this region are excluded as well. The reconstruction efficiencies for τ leptons are drastically lower than the ones for electrons and muons. This is the case because most τ leptons already decay inside of the beam pipe and can therefore not be detected

directly but can only be reconstructed from their decay products. The discrete change of the τ lepton reconstruction efficiency at certain values of pseudorapidity is an artificial effect due to the implementation in the tool used for the simulation which only considers integer values of pseudorapidity η. This was fixed in a later version of the tool which then allowed for arbitrary values of η but due to time constraints this was not implemented in the results presented in this thesis. The reconstruction efficiencies of jets are implemented in the form of HS jet tagging- and b tagging efficiencies and are described further below.

Figure 5.4: Reconstruction efficiencies for several different objects. Represented are (from top left to bottom right): The efficiencies of an electron, a muon, a 1 prong τ lepton, and a 3 prong τ lepton to be reconstructed calculated by the number of $e/\mu/\tau^{1\text{ prong}}/\tau^{3\text{ prong}}$ not being discarded in the corresponding analysis step by the number of $e/\mu/\tau^{1\text{ prong}}/\tau^{3\text{ prong}}$ entering this step. All signal and background samples combined have been used in the creation.
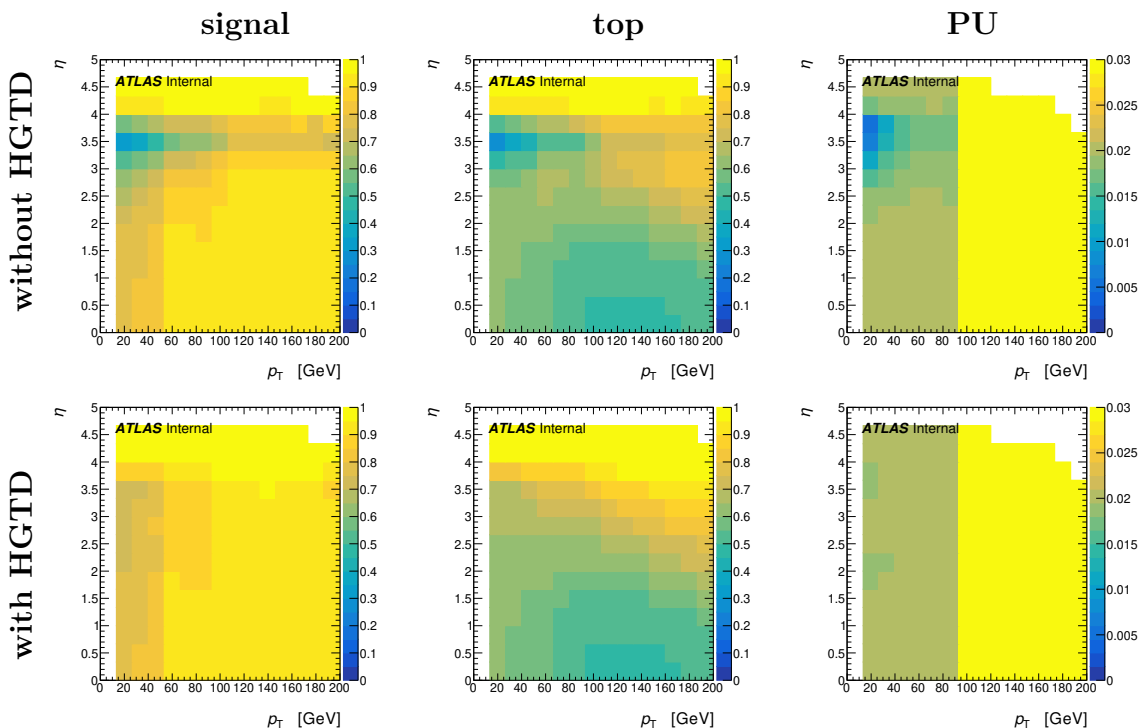
b **tagging efficiencies:** Fig. 5.5 shows the *b* tagging fractions for hard scatter- (HS) jets in the VBF $H \to \tau\tau$ signal sample and the top background samples (i.e. $t\bar{t}$ and single t) and for pileup (PU) jets from all signal and background samples combined.
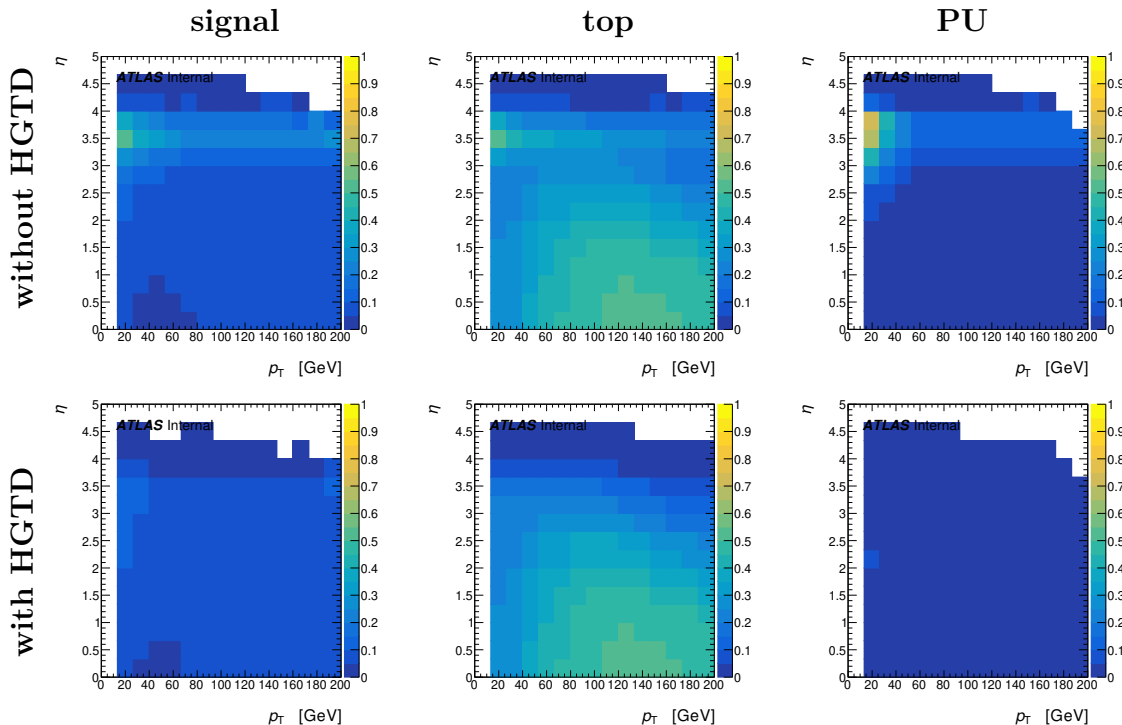
Figure 5.5: *b* tagging fractions for hard scatter- (HS) jets in the VBF $H \to \tau\tau$ signal sample [left], the combined top background samples (i.e. $t\bar{t}$ and single t) [middle] and for pileup (PU) jets from all signal and background samples combined [right]. They were calculated by the number of jets being identified as *b* jet divided by the number of jets entering the corresponding analysis step. For the [top] row of figures a detector layout without the HGTD was used while for the [bottom] row the same layout but with HGTD was used.

While the *b* tagging was simulated in the form of *b* tagging efficiencies (i.e. the probability for a *b* jet to be labeled as a *b* jet), the *b* tagging fractions yield information about the probability for any jet to be labeled as a *b* jet. The former differs only weakly for the different signal and background samples, while the latter can give information on the differences between the channels. The simulated *b* tagging was done choosing an expected performance of the MV1 *b* tagging algorithm [72] with a working point of 85% *b*-tagging efficiency. The *b* tagging fractions for $|\eta| > 2.5$ are very similar for both HS and PU jets and both signal and top background. In the region $p_T > 20$ GeV and $|\eta| < 2.5$ however, the *b* tagging fractions of HS jets in the top background exhibit much higher values than the other distributions shown in Fig. 5.5.

The same kind of plots are shown but for a detector layout which includes the High Granularity Timing Detector (HGTD) (see Sec. 3.4.4). By comparing these results to the ones without the HGTD, it becomes apparent, that especially in the problematic region of $\eta > 2.5$ a clear improvement can be seen due to the HGTD. With the HGTD, a clear difference in the b tagging fraction for the top backgrounds compared to the signal can also be seen in this region. Hence, with the HGTD installed, the η region for a b-veto cut can be increased to include jets with even higher η eventually yielding better separation of the signal channel to the top background channels.

Figure 5.6: HS jet tagging efficiencies for hard scatter- (HS) jets in the VBF $H \rightarrow \tau\tau$ signal sample [left], the combined top background samples (i.e. $t\bar{t}$ and single t) [middle] and for pileup (PU) jets from all signal and background samples combined [right]. They were calculated by the number of jets being identified as HS jet (i.e. not PU jet) divided by the number of jets entering the corresponding analysis step. For the [top] row of figures a detector layout without the HGTD was used while for the [bottom] row the same layout but with HGTD was used. The efficiencies for PU jets with $p_{\mathrm{T}} > 100$ GeV are always one in the simulated samples. The colour palette in the corresponding plots was chosen to end at a value of 0.03 to make the efficiencies in the range $p_{\mathrm{T}} < 100$ GeV visible despite of that.

HS jet tagging efficiencies: Fig. 5.6 shows the HS jet tagging efficiencies (i.e. efficiencies for a non-b-tagged jet to be labeled as a hard scatter jet) for hard scatter- (HS)

jets in the VBF $H \to \tau\tau$ signal sample and the top background samples (i.e. $t\bar{t}$ and single t) and for pileup (PU) jets from all signal and background samples combined. Since the implementation of these kind of efficiencies in the tool used for the simulation has been derived for jets that are not b jet candidates, the values shown in Fig. 5.6 represent only such jets that have not been labeled as b jets in the process described above.

The unique signature of the jet kinematics in the VBF $H \to \tau\tau$ signal channel (see Sec. 5.1) yields a good rejection of pileup jets especially in the forward region (i.e. high $|\eta|$) desirable. A HS jet tagging working point of 98% has been chosen as suggested in the technical design reports of the ITk [4, 5].

Similarly to the b tagging efficiency presented above, the HS jet tagging efficiency breaks down in the forward η region if a detector layout without the HGTD is chosen. Comparing the corresponding plots with and without the HGTD in Fig. 5.6 shows that including the HGTD into the detector layout can compensate this effect drastically.

Fake rates: Fig. 5.6 shows the fake rates (i.e. efficiencies for a jet to be misidentified as a different object) for hard scatter- (HS) and pileup (PU) jets.
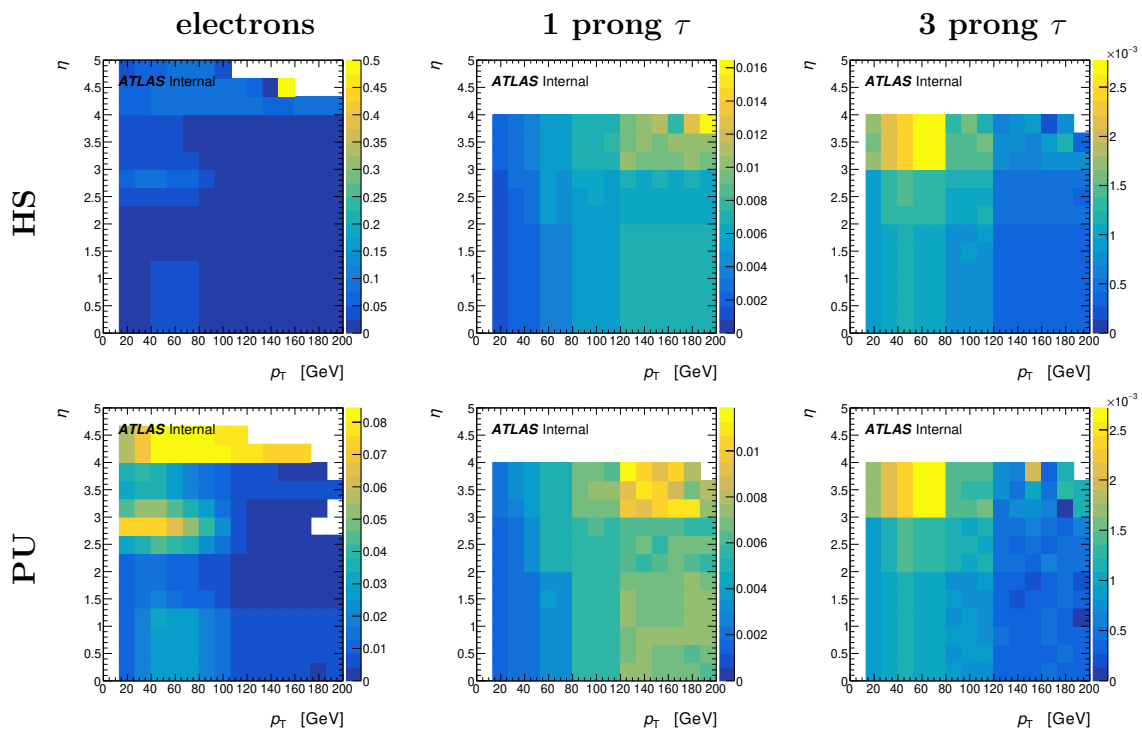
Figure 5.7: Fake rate efficiencies for hard scatter [top] and pileup [bottom] jets for faking electrons [left], 1 prong [middle], or 3 prong [right] τ leptons calculated by the number of jets that did fake the corresponding object divided by the total number of jets entering this analysis step.

It is apparent, that the fake rates do not differ greatly between HS and PU jets. Implemented in this simulation are only the fake rates for jets faking electrons or (1 prong or 3 prong) τ leptons. The possibility of electrons faking τ leptons is not taken into account because no such fake rates were available at the time. This implies that the impact of the $Z \to ee$ background channel will be underestimated in this analysis.

5.3 Cut Based Analysis

The following sections present a cut based analysis in which the expected sensitivity of the ATLAS detector at the HL-LHC to the VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$ process is derived. It was performed on the signal and background samples listed in Tab. 5.1 on which the smearing process and the different efficiencies as described in Sec. 5.2 for a layout of the HL-LHC without HGTD was performed. The criteria for object selection and overlap removal for a selection of VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$ events are listed in Sec 5.3.1. Subsequently, the calculation of the mass of the reconstructed di-τ system utilising the Missing Mass Calculator (MMC) algorithm is described in Sec. 5.3.2. The optimization process of the cuts for event selection is described in Sec. 5.3.3. Finally, the resulting expected sensitivity of the ATLAS detector to the VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$ process is described in Sec. 5.3.4.

5.3.1 Object Selection and Overlap Removal

Each object considered in this study has to pass certain criteria which are motivated by the selection criteria used in the ATLAS Run 2 analysis of the $H \to \tau\tau$ process [58]. If at least one is not fulfilled, the object gets discarded. The object selection criteria are:

Electrons:

- loose isolation working point
- $p_{\text{T}} > 15$ GeV
- $|\eta| < 2.47$ and $\notin [1.37, \ 1.52]$

Muons:

- loose isolation working point
- $p_{\text{T}} > 10$ GeV
- $|\eta| < 2.5$

τ leptons:

- Hadronic decay with 1 or 3 prongs

- p_T > 20 GeV

- $|\eta|$ < 2.5 and $\notin$ [1.37, 1.52]

- absolute electric charge $|Q| = 1$

- No electron candidate with p_T > 5 GeV within $\Delta R = 0.4$ cone

Jets:

- p_T > 20 GeV

- $|\eta|$ < 4.5

The isolation criteria stated above require that no additional high p_T tracks are found in a cone around the lepton's track and that no significant amount of energy is deposited in a cone around the calorimeter cluster corresponding to the lepton. The sizes of these cones depend on the transverse momentum of the lepton. The individual isolation working points are defined in Ref. [73] for electrons and Ref. [74] for muons.

The pseudorapidity criterion of $\eta \notin$ [1.37, 1.52] for electrons and τ leptons is motivated by the transition region between the barrel and endcap LAr calorimeters of the ATLAS detector. Due to its lack of instrumentation, this region does not allow for a good reconstruction and identification efficiency of particles from calorimeter clusters (see 5.2.3).

Due to the possibility of jets faking electrons or τ leptons introduced in Sec. 5.2.3, it is necessary to discard jets that are likely to be responsible for such a fake e or τ. To accomplish this, an overlap removal was performed in which jets j with a separation of $\Delta R(j, \tau) < 0.2$ or $\Delta R(j, e) < 0.4$ to a τ or e candidate that passes all object selection criteria are discarded

5.3.2 Ditau-Mass Calculation

The calculation of the mass of the di-τ system, is done by utilising the Missing Mass Calculator (MMC) algorithm [75]. The operational principle of this algorithm is to require that the mutual orientations of the decay products are consistent with the mass and kinematics of a decay of a tau lepton. A likelihood function is minimized in the kinematically allowed phase space region. The resulting mass m_MMC can be used as an indicator for the separation of the signal channel from background since the average outcome for signal events should be close to the Higgs boson's mass and for $Z \to \tau\tau$ background close to the Z boson's mass. This can be seen in the m_MMC distributions shown in Fig. 5.8.

As expected, the distributions peak at the respective mother particle's rest masses. By comparing the plots produced with and without prior application of the smearing process as described in Sec. 5.2.2, one can clearly see that the added resolutions drastically lessen the separation power that can be achieved by this discriminator.
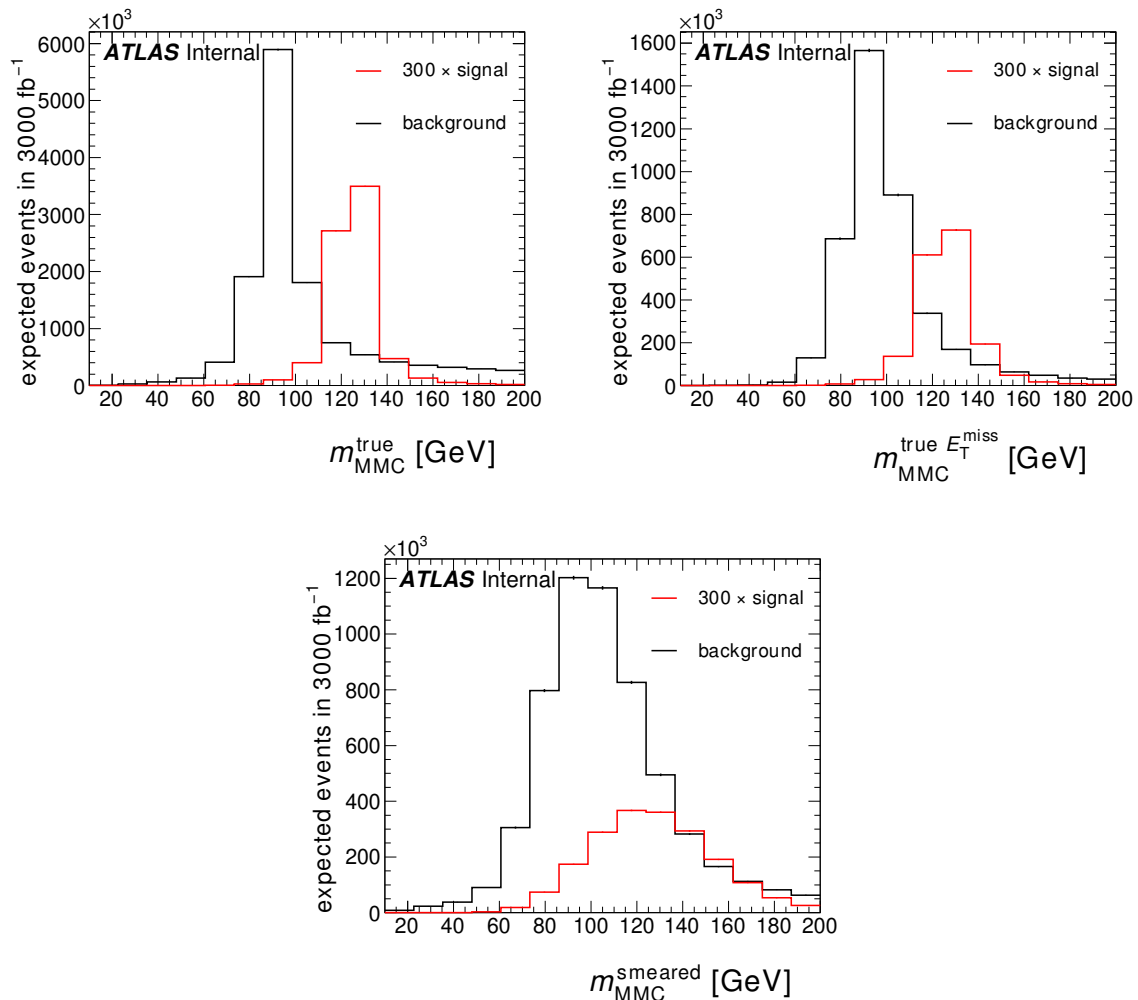
Figure 5.8: Di-τ mass m_{MMC} after preselection calculated by the Missing Mass Calculator (MMC) algorithm for the VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ signal channel and the combined background channels as described in Sec. 5.3.2. The plots were produced with different options regarding the smearing process as described in Sec. 5.2.2 (from top left to bottom): No smearing, smearing of all objects except $E_{\mathrm{T}}^{\mathrm{miss}}$ (i.e. no addition of pileup $E_{\mathrm{T}}^{\mathrm{miss}}$ as well), and smearing of all objects. Preselection cuts as described in Sec. 5.3.3 but no additional signal region cuts have been applied.

Further comparing the fully smeared version of the m_{MMC} distribution with the one where all objects but the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ are smeared, it is apparent that

the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution plays the main role here. This is likely an effect of the addition of pileup $E_{\mathrm{T}}^{\mathrm{miss}}$ in the smearing process. As a result, not only the amplitude of $E_{\mathrm{T}}^{\mathrm{miss}}$ but also its direction (i.e. its azimuthal angle ϕ) can change drastically leading to distorted values of m_{MMC} (see Sec. 5.2.2). Consequently the more severe pileup conditions in the HL-LHC constitute one of the most prominent challenges for future analyses. The reason for this is that the MMC algorithm assumes the orientations of the momenta of the objects used as input are consistent with the decay kinematics of a τ lepton. It then minimizes a likelihood function defined in the kinetically allowed phase space region. Since $E_{\mathrm{T}}^{\mathrm{miss}}$ is used as a proxy for the neutrino four-momenta, a badly reconstructed direction may lead the MMC algorithm to produce inaccurate results or to fail to calculate a possible mass at all.

5.3.3 Event Selection and Cut Optimization

The goal of this analysis is to make a prediction on the expected significance of signal with respect to background at a integrated luminosity of $3000\mathrm{fb}^{-1}$. This is done by utilizing a cut based analysis. There are two sets of cuts applied on event level implemented: Preselection cuts and VBF signal region cuts. They are described below.

Preselection: The first set of cuts, the preselection, aims to select events loosely based on the expected final state objects for the $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ decay: one hadronically decaying τ lepton, one lepton from the leptonically decaying τ lepton, and three neutrinos. A leading order Feynman diagram of the signal process can be seen in Fig. 2.4. The preselection cuts are defined by:

- Exactly one e or μ (from now on called light lepton l) candidate
- Exactly one τ_{had} candidate
- $|\eta(\tau_{\mathrm{had}})| < 2.4$
- Opposite electric charge of l and τ_{had}
- Event fired single lepton trigger (SLT)
- $p_{\mathrm{T}}(\mu) > 21$ GeV
- $p_{\mathrm{T}}(e) > 25$ GeV
- b-jet veto: No b-jet candidate with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$
- At least two jets that are not b-jet candidates

The single lepton trigger (SLT) is a combination of simulations of the single electron trigger (SET) and the single muon trigger (SMT) which are described in detail in the technical design report [9] of the proposed ATLAS trigger system for the HL-LHC. The SLT triggers if either the SET or the SMT triggers but not both.

The restriction of the b-jet veto on a region of $p_T > 20$ GeV and $|\eta| < 2.5$ is motivated by the worse b tagging efficiency outside of this region as discussed in Sec. 5.2.3.

VBF signal region cuts: The second set of event level cuts, VBF signal region cuts, is intended to serve as a mechanism to separate signal from background by exploiting the special signature of the VBF production process and its unique jet kinematics (see Sec. 5.1). The values of the VBF signal region cuts are the result of a cut optimisation aiming to maximize the significance. It was performed on the signal and background samples listed in Tab. 5.1 on which the smearing process as described in Sec. 5.2 for a layout of the HL-LHC ATLAS detector without HGTD was performed. For this purpose, six different cut values c_i were varied in three individual phases of cut optimization. The cut values are defined as

- $p_T(j^{\text{lead}}) > c_1$

- $p_T(j^{\text{sublead}}) > c_2$

- Reconstructed di-τ mass $c_3 < m_{\text{MMC}} < c_4$

- Combined jet mass $m(j^{\text{lead}}, j^{\text{lead}}) > c_5$

- $\Delta\eta(j^{\text{lead}}, j^{\text{sublead}}) > c_6$

Here, j^{lead} (j^{sublead}) refers to the jet with the (second) largest value of transverse momentum in the event. The combined jet mass $m(j^{\text{lead}}, j^{\text{sublead}})$ is defined by the mass of the four-vector sum of these two jets. The reconstructed di-τ mass was calculated with the Missing Mass Calculator algorithm which is described in Sec. 5.3.2.

The reconstructed Higgs boson $p_T(H)$, defined as the transverse momentum of the four-vector sum of the hadronically decaying τ lepton, the light lepton (i.e. electron or muon) and the missing transverse momentum E_T^{miss} in the event, has also been studied as a potential cut parameter. This was motivated by the corresponding Run 2 $H \to \tau\tau$ analysis [58] in which this variable was used. Here, different signal regions are defined which are either enriched in events in which a Higgs boson was produced via vector boson fusion or via gluon fusion. Since gluon fusion Higgs production is more likely to result in boosted Higgs bosons, the Higgs boson's transverse momentum $p_T(H)$ is an important indicator when such approach is taken.
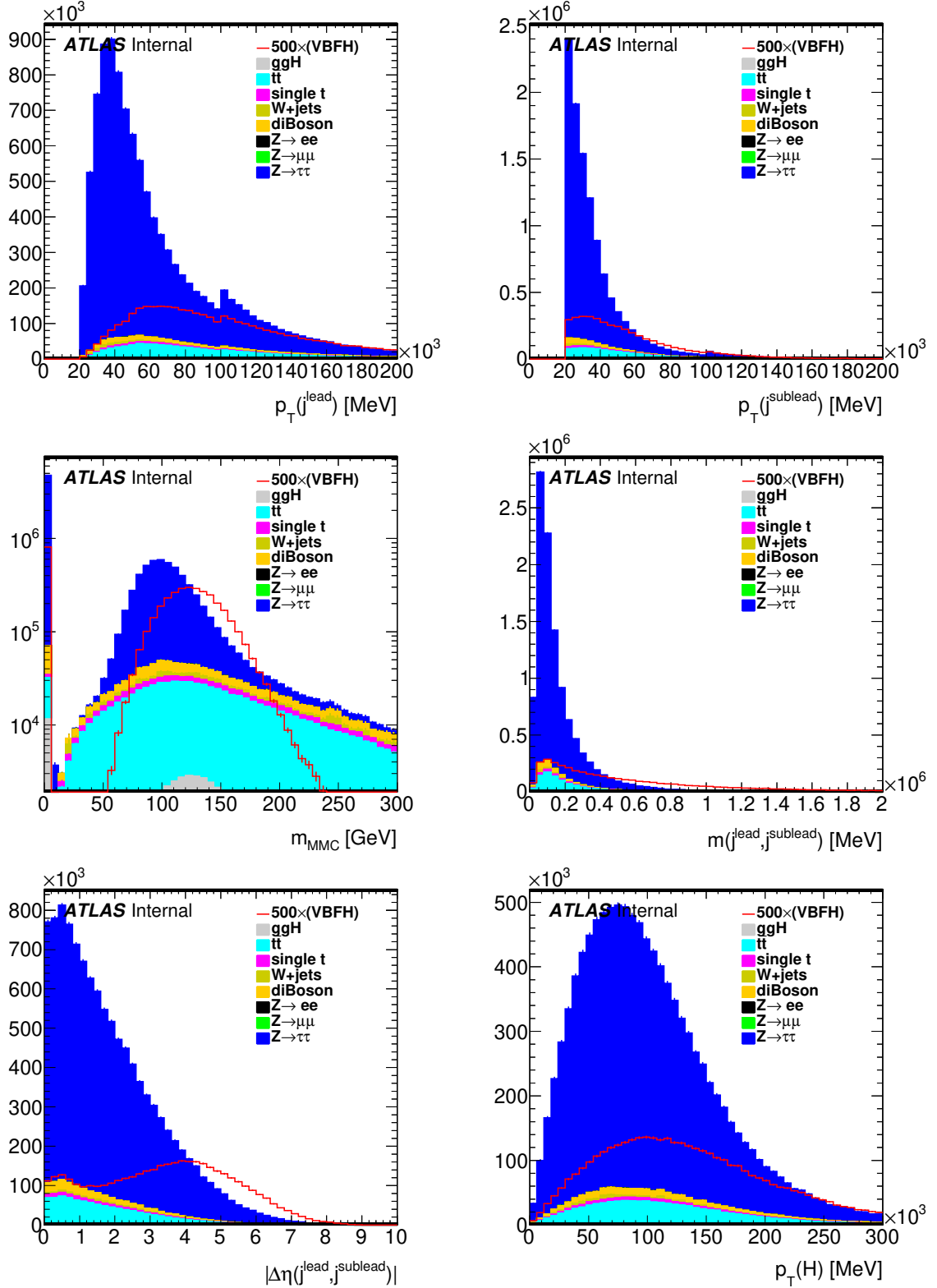
Figure 5.9: Distributions after preselection of the parameters x_i used for signal region cuts. The shoulders at 100 GeV in the distributions of $p_T(j^{\text{lead}})$ and $p_T(j^{\text{sublead}})$ are an artifact of the simulated HS jet tagging efficiencies which is further described in Sec. 5.2.3. The large number of events with $m_{\text{MMC}} = 0$ correspond to events in which the MMC algorithm did not yield a valid result (i.e. no kniematically possible configuration of the τ lepton's decay products could be found). This is described further in Sec. 5.3.2. Here, a logarithmic presentation was chosen to increase the visibility.

In the studies presented in this thesis however, a simplified approach was taken in which only Higgs boson production via vector boson fusion is considered as signal. No gain in significance was observed by applying cuts on $p_T(H)$. Hence, it is not considered in the following. Fig. 5.9 shows the distributions of the physical parameters x_i corresponding to the cut values c_i at preselection level.

Significance: In the following, two different types of significance with respect to signal-to-background rejection are used: The unbinned significance Σ_u and the binned significance Σ_b. They are calculated using the distribution of the reconstructed di-τ mass m_{MMC} for signal and background events after the respective set of cuts has been applied. An example of such m_{MMC} distributions can be seen in Fig. 5.9 (in this case only the preselection cuts have been applied). The underlying definition for both the binned and unbinned significances, is the Asimov-significance [76]

$$\Sigma = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right) - s\right)}. \tag{5.1}$$

The corresponding uncertainty given by Gaussian error propagation is

$$\sigma_\Sigma = \sqrt{\frac{1}{\Sigma}\left[\left(\sigma_s \ln\left(1+\frac{s}{b}\right)\right)^2 + \left(\sigma_b\left(\ln\left(1+\frac{s}{b}\right) - \frac{s}{b}\right)\right)^2\right]}. \tag{5.2}$$

Here, s is the number of signal events and b is the number of background events . To calculate the unbinned significance Σ_u, all bins in the m_{MMC} distributions for signal and background are summed and the results are used as s and b in Eq. 5.1. To calculate the binned significance Σ_b, first the Asimov significance is calculated for each bin of the m_{MMC} distribution individually. A binning of $\Delta m_{\mathrm{MMC}} = 6$ GeV was chosen. Here, s and b are the bin contents in the signal and background distributions respectively. Then, the significances for each bin are summed in quadrature to yield Σ_b. The cut optimization presented below aims to maximize Σ_b and was performed in three phases:

Phase 1: All preselection cuts were applied. Then, the signal region cuts were tested individually: For each, the corresponding cut value c_i was varied and the significance for the different values of c_i was calculated. The results of this phase of cut optimization can be seen in Fig. 5.10. Also shown are the efficiencies of signal and background events in the respective m_{MMC} distributions, i.e. the number of events passing the cut corresponding to the cut value c_i divided by the total number of events before that cut. The cut parameters c_i which yielded the highest significance are listed in Tab. 5.3.
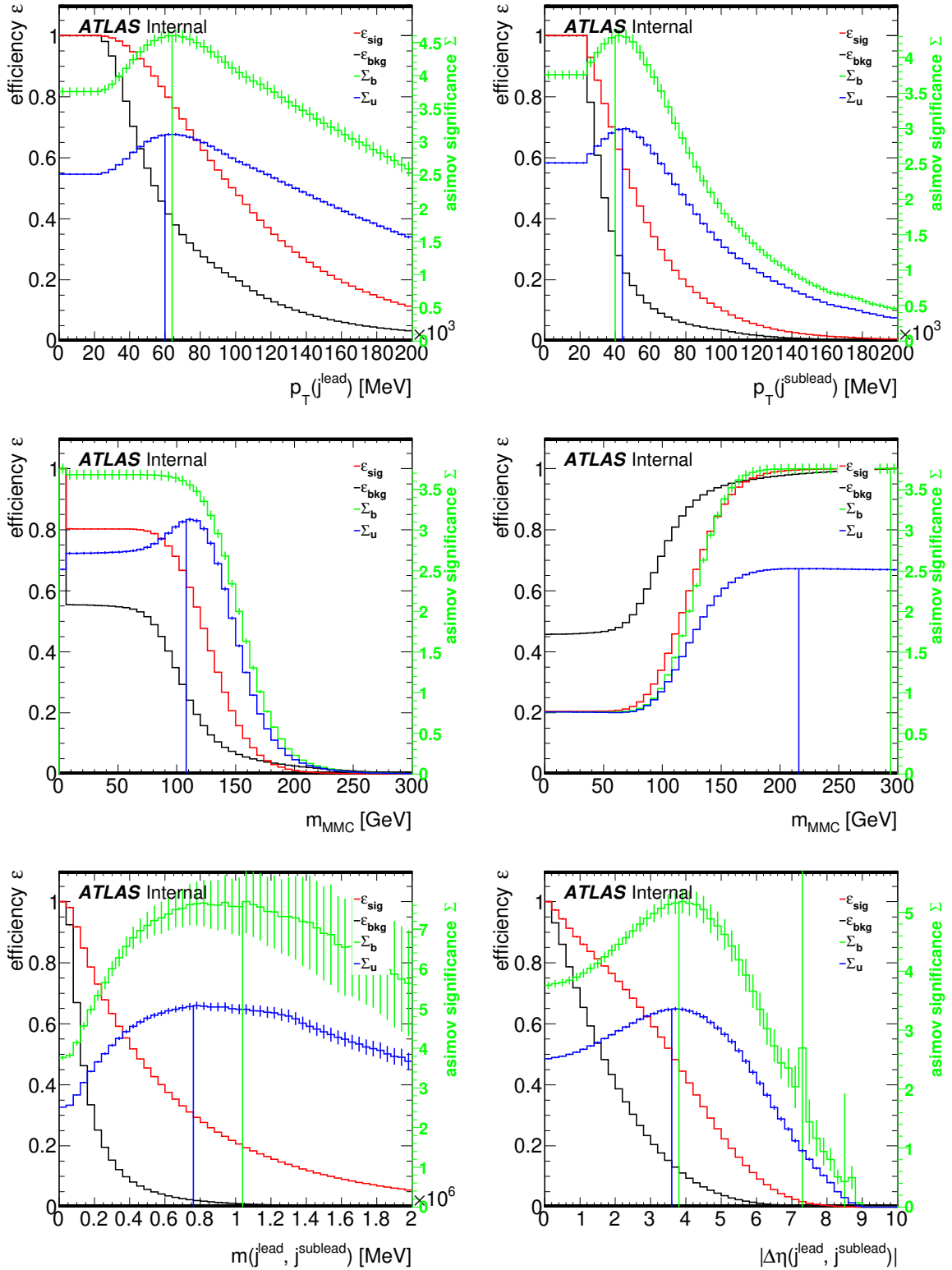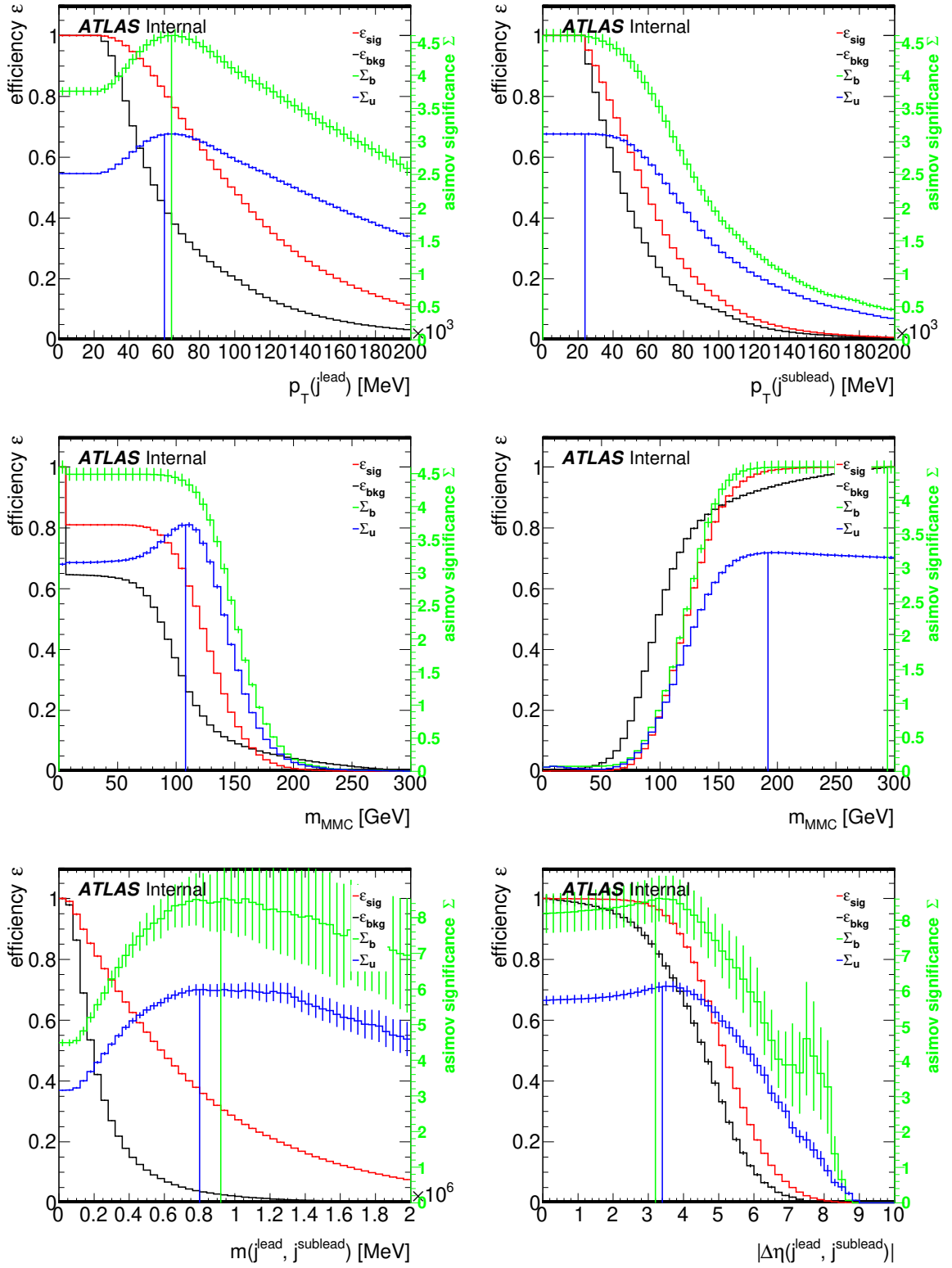
Figure 5.10: Results of phase 1 of cut optimization. Shown are the binned and unbinned Asimov significances Σ_{b} and Σ_{u} (see Eq. 5.1). The global maxima of these curves are indicated by a vertical line. Also shown are the corresponding efficiencies $\varepsilon_{\mathrm{sig}}$ and $\varepsilon_{\mathrm{bkg}}$ of signal and background versus the cut value c_i. The corresponding signal distributions after preselection with respect to one of the signal region cut variables x_i versus the reconstructed di-τ mass m_{MMC} are shown in Fig. B.1 and Fig. B.2.

Phase 2:   The results of this phase can be seen in Fig. 5.11. It was performed similarly to what has been done in phase 1 with only one exception: While in phase 1 all cuts have been investigated independently, in this phase the cuts were applied subsequently: During the optimization of cut parameter c_i not only were the preselection cuts applied but also all signal region cuts c_j with $j < i$. Here, the value of c_j were chosen according to the result of the phase 2 cut optimization of c_j that was done before the optimization of c_i. Most of the time the cut value which yields the maximum binned significance Σ_b was chosen. One exception to this rule are the values for c_3 and c_4 which concern cuts on m_{MMC}. Since the significance is calculated on m_{MMC} distributions, cutting on m_{MMC} cannot lead to higher binned significances than not cutting on m_{MMC}. Therefore, the choice was made to only exclude events were the MMC algorithm did not yield a valid solution (i.e. $m_{\mathrm{MMC}} = 0$) and to not apply any further cuts on m_{MMC}. The chosen cut values are referred to as c_i^{nom}.

For the cut on the combined jet mass $m(j^{\mathrm{lead}}, \ j^{\mathrm{sublead}})$ the cut value $c_5^{\mathrm{nom}} = 600$ GeV was chosen which differs from the value $c_5(\Sigma_b^{\mathrm{max}}) = 920$ GeV yielding maximum binned significance Σ_b. This was done because otherwise to many events would be lost due to this cut which would lead to a large statistical uncertainty on Σ_b. The value of $c_5^{\mathrm{nom}} = 600$ GeV was chosen because, judging by the curve in Fig. 5.11, this corresponds to only a small decrease in Σ_b compared to $c_5(\Sigma_b^{\mathrm{max}}) = 920$ GeV while still reducing the statistical error drastically. All choices for the cut parameters together with the values which yield maximum Σ_u and Σ_b are listed in Tab. 5.3.

It should be noted that the cut optimization for the parameter $x_2 = p_{\mathrm{T}}(j^{\mathrm{sublead}})$ yields $c_2 = 0$ in this phase, i.e. no separation power can be gained from $x_2 = p_{\mathrm{T}}(j^{\mathrm{sublead}})$ after the cut $x_1 = p_{\mathrm{T}}(j^{\mathrm{lead}})$ was applied with the chosen value c_1 as stated in Tab. 5.3. This is in contrast to the result from phase 1 which showed that this cut does have some separation power if applied independently. One reason for this could be that the distribution of $x_2 = p_{\mathrm{T}}(j^{\mathrm{sublead}})$ peaks at low values, even below the cut-value of 20 GeV dictated by object selection (see Sec. 5.3.1). The more important reason however, might be the high correlation between $x_1 = p_{\mathrm{T}}(j^{\mathrm{lead}})$ and $x_2 = p_{\mathrm{T}}(j^{\mathrm{sublead}})$ which can be seen in Fig. 5.12. Consequently, varying the cut parameters c_1 and c_2 at the same time should give insight on this. This is done in phase 3.

Figure 5.11: Results of phase 2 of cut optimization. Shown are the binned and unbinned Asimov significances Σ_b and Σ_u (see Eq. 5.1). The global maxima of these curves are indicated by a vertical line. Also shown are the corresponding efficiencies ε_{sig} and ε_{bkg} of signal and background versus the cut value c_i. The corresponding signal distributions after preselection with respect to one of the signal region cut variables x_i versus the reconstructed di-τ mass m_{MMC} are shown in Fig. B.3 and Fig. B.4.
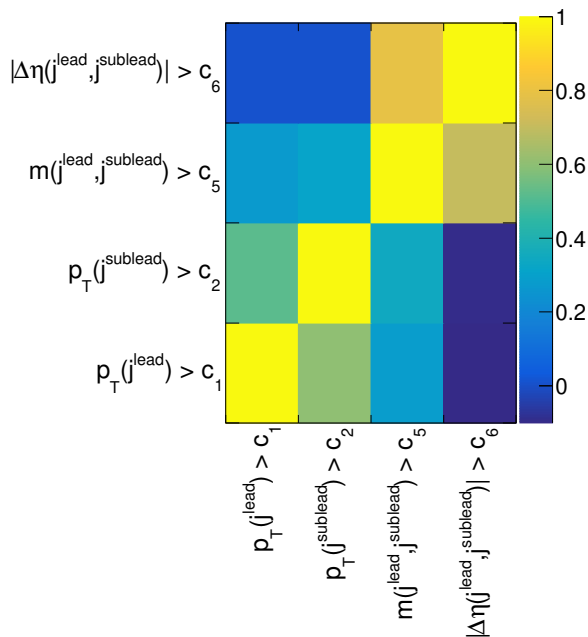
Figure 5.12: Correlations of the variables x_i corresponding to the signal region cut parameters c_i in signal and background to each other. Entries in the top left half were calculated using the signal sample and entries in the bottom right half were calculated using the background samples. The corresponding correlation plots can be seen in Fig. B.5-B.8. The correlations for cut parameters x_i which concern the reconstructed di-τ mass m_{MMC} are not shown. This is because the calculation of the significances Σ_u and Σ_b was done on m_{MMC} histograms.

Phase 3: Firstly, those parameters x_i were identified which show high correlations to each other. Fig. 5.12 shows the correlations of the six cut parameters to each other. The two pairs of parameters x_i which show the highest correlation are

$$(x_1, \ x_2) = (p_{\mathrm{T}}(j^{\mathrm{lead}}), \ p_{\mathrm{T}}(j^{\mathrm{sublead}})) \text{ and} \tag{5.3}$$

$$(x_5, \ x_6) = (m(j^{\mathrm{lead}}, \ j^{\mathrm{sublead}}), \ |\Delta\eta(j^{\mathrm{lead}}, \ j^{\mathrm{sublead}})|). \tag{5.4}$$

For both of these pairs the following process was done: The binned and unbinned significances Σ_b and Σ_u were calculated in the same way as in phase 1 of the cut optimization. In contrast to what has been done in phase 1 however, all preselection cuts and all signal region cuts were applied simultaneously. For this purpose, the signal region cut values c_i resulting from phase 2 were chosen. They can be found in Tab. 5.3 and are referred to as the nominal values c_i^{nom}. This was done for each of the two pairs $(x_i, \ x_j)$ stated in Eq. 5.3 and 5.4. The values of the cut parameters $(c_i, \ c_j)$ corresponding to the pair

(x_i, x_j) were varied simultaneously. The resulting significances Σ_u and Σ_b can be seen in Fig. 5.13. The purpose of this phase of cut optimization was to make sure that no drastically better set of cut values was missed due to correlations between the six cut variables x_i or due to the order of applying the corresponding cuts. Would this have been the case, it would have been noticeable in the form of a higher local maximum separate from the one in approximately the middle in the histograms shown in Fig. 5.13. Since no such additional maximum is apparent, the nominal cut values c_i^{nom} that have been derived in phase 2 were chosen for the final calculation of significance presented in Sec. 5.3.4.
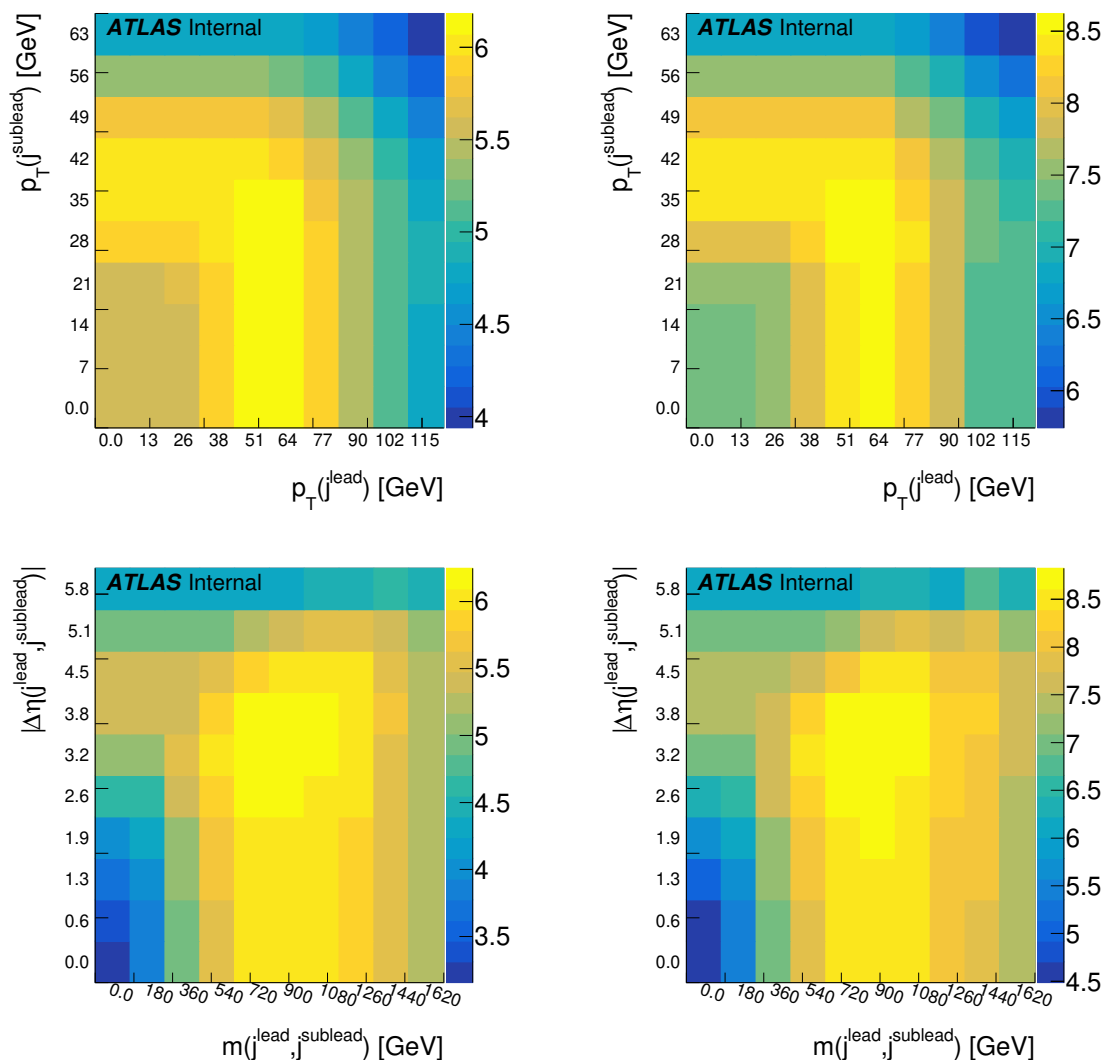
Figure 5.13: Results of phase 3 of cut optimization. Shown are the unbinned significance Σ_u [left] and the binned significance Σ_b [right] with respect to the two pairs of cut parameters given in Eq. 5.3 and 5.4.

Results: Tab. 5.3 summarises the results of the cut optimization process described above. The final set of cut values used for the calculation of the expected significance in Sec. 5.3.4 is shown in the last column of this table.

	phase 1		phase 2		
cut	$c_i(\Sigma_u^{\max})$	$c_i(\Sigma_b^{\max})$	$c_i(\Sigma_u^{\max})$	$c_i(\Sigma_b^{\max})$	**choice: $c_i^{\mathbf{nom}}$**
$p_{\mathrm{T}}(j^{\mathrm{lead}}) > c_1$	60 GeV	64 GeV	60 GeV	64 GeV	**64 GeV**
$p_{\mathrm{T}}(j^{\mathrm{sublead}}) > c_2$	44 GeV	40 GeV	24 GeV	-	**-**
$c_3 < m_{\mathrm{MMC}}$	108 GeV	-	108 GeV	-	**-***
$m_{\mathrm{MMC}} < c_4$	216 GeV	-	192 GeV	-	**-**
$m(j^{\mathrm{lead}}, j^{\mathrm{lead}}) > c_5$	760 GeV	1.04 TeV	800 GeV	920 GeV	**600 GeV**
$\Delta\eta(j^{\mathrm{lead}}, j^{\mathrm{sublead}}) > c_6$	3.6	3.8	3.4	3.2	**3.2**

Table 5.3: Optimized cut values c_i resulting from phase 1 and phase 2 of cut optimization. Shown are those values $c_i(\Sigma_u^{\max})$ and $c_i(\Sigma_b^{\max})$ which yielded the maximum unbinned significance Σ_u and binned significance Σ_b. The corresponding curves can be seen in Fig. 5.10 and Fig. 5.11. The nominal cut values c_i^{nom} constitute the set of cuts that were applied to derive the final significance in Sec. 5.3.4. The * indicates that instead of applying an explicit cut on m_{MMC}, all events for which the MMC algorithm did not yield a valid solution (see Sec. 5.3.2) were excluded.

Tab. 5.4 summarises the event yields of the individual channels before and after the application of both preselection cuts and the final VBF signal region cuts. The corresponding cut flow and cut efficiencies are shown in Fig. 5.14. Note that no events of the $Z \to \mu\mu$ background are present after applying the VBF signal region cuts. The reason for this might be the low number of events in the corresponding Monte-Carlo generated truth sample as shown in Tab. 5.1. The corresponding effective Luminosity for that sample is with 11.3 fb^{-1} much lower than the luminosity of 3000 fb^{-1} to which the sample was scaled to correspond to the HL-LHC conditions. The final significance calculated utilising the optimized set of cuts is presented in Sec. 5.3.4.

sample	initial	Preselection		VBF signal region	
VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	64920	8304	(12.79 %)	2322	(3.58 %)
ggF $H \to \tau\tau$	775246	44497	(5.74 %)	1521	(0.20 %)
di-boson	2.92×10^{8}	315546	(0.11 %)	7948	$(2.72 \times 10^{-3}$ %)
W+jets	1.90×10^{11}	78159	$(4.10 \times 10^{-05}$ %)	834	$(4.38 \times 10^{-7}$ %)
$Z \to \tau\tau$	6.17×10^{9}	9.58×10^{6}	(0.16 %)	101676	$(1.65 \times 10^{-3}$ %)
$Z \to \mu\mu$	6.17×10^{9}	451	$(7.31 \times 10^{-6}$ %)	0	(0 %)
$Z \to ee$	6.17×10^{9}	2429	$(3.94 \times 10^{-5}$ %)	79	$(1.28 \times 10^{-6}$ %)
$t\bar{t}$	1.35×10^{9}	914823	(.07 %)	35835	$(2.66 \times 10^{-3}$ %)
single t	4.77×10^{8}	131846	(0.03 %)	3397	$(7.12 \times 10^{-4}$ %)

Table 5.4: Event yields for signal (VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$) and background channels after preselection and VBF signal region at HL-LHC conditions with an integrated luminosity of 3000 fb^{-1}. For the yields after each individual cut see Fig. 5.14.
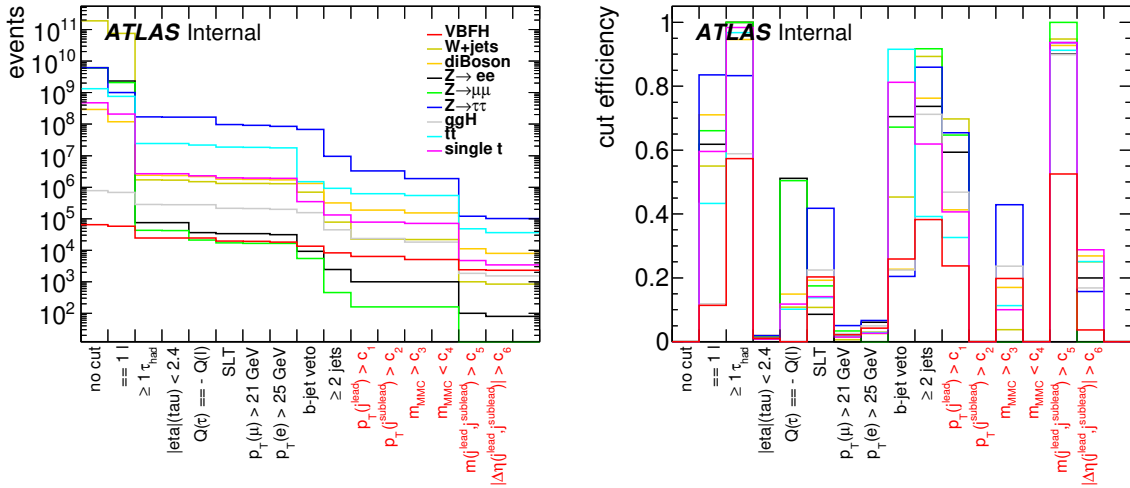
Figure 5.14: Event yields [left] and cut efficiencies [right] for signal (VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$) and background channels after each individual cut of preselection (black) and VBF signal region (red) at HL-LHC conditions with an integrated luminosity of 3000 fb^{-1}. The represented values can be found in the appendix in Tab. B.1. For the yields after the full set of preselection and signal region cuts see also Tab. 5.4.

5.3.4 Expected Significance
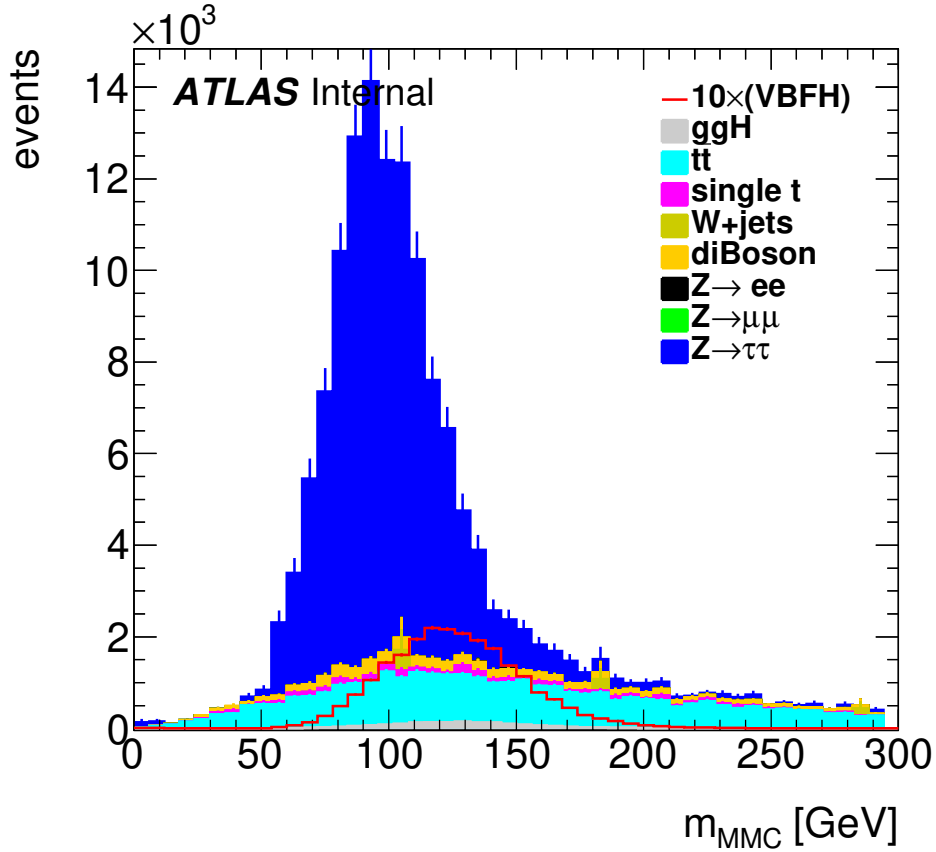
Figure 5.15: Distribution of reconstructed di-τ mass m_{MMC} which was used to calculate the final expected significance. An integrated luminosity of 3000 fb^{-1} was assumed. All preselection and signal region cuts as defined in Sec. 5.3.3 have been applied.

The final expected significance was calculated as the binned Asimov significance as it is defined in Eq. 5.1. For this purpose the full list of preselection and signal region cuts described in Sec. 5.3.3 has been used. The signal region cuts are the result of the cut optimization described in Sec. 5.3.3 above. Fig. 5.15 shows the distribution of reconstructed di-τ mass m_{MMC} which was used to calculate the significance. This resulted in a value of

$$\Sigma_{\mathrm{b}}^{\mathrm{exp}} = 8.63 \pm 0.66_{\mathrm{stat}}. \tag{5.5}$$

With the same set of cuts, the significance was calculated for a detector layout including the HGTD (see Sec. 3.4.4) for comparison. This resulted in

$$\Sigma_{\mathrm{b}}^{\mathrm{exp,HGTD}} = 9.40 \pm 0.71_{\mathrm{stat}}. \tag{5.6}$$

The higher value of $\Sigma_{\mathrm{b}}^{\mathrm{exp,HGTD}}$ compared to $\Sigma_{\mathrm{b}}^{\mathrm{exp}}$ was expected due to the better performance in HS jet tagging efficiency in the forward detector region gained by including the HGTD which was shown in Sec. 5.2.3. Hence, including the HGTD lowers the number of pileup jets passing the HS jet tagging and therefore reduces the yields for background events passing the signal region cuts due to pileup jets faking the characteristic kinematic structure of VBF jets.

6 Jet Calibration in the Forward Detector Region

Since the VBF $H \rightarrow \tau\tau$ signal channel possesses the very characteristic signature of forward jets (see Sec. 5.1), a reliable jet calibration is of paramount importance. For this reason, the following sections aim to provide insights to the challenges of the jet calibration in the ATLAS collaboration as it is done now and to give suggestions on how these challenges could be addressed. The main focus here is the pileup calibration since the harsher pileup conditions will be one of the main challenges for the HL-LHC.

Sec. 6.1 lists the Monte Carlo generated samples and explains the general structure of the analysis. In Sec. 6.2 the jet calibration procedure used by the ATLAS collaboration is presented. The following Sec. 6.3 introduces the jet area A and the pileup density ρ which are used in the jet pileup calibration scheme. In Sec. 6.4 the different jet pileup calibration methods considered in this study are explained and their individual performance is investigated. It starts with the Jet-Area correction and the Residual correction which are both currently used in the ATLAS jet calibration. It then introduces the Gradient Jet-Area correction which is a variation of the two former methods and is not currently used. Sec. 6.5.2 then compares the performance of the different methods with respect to closure of truth p_{T} to reconstructed p_{T} after the subsequent absolute Jet Energy Scale (JES) calibration has been applied additionally to the pileup calibration. Sec. 6.5.3 describes an issue concerning jets that have negative energy after the pileup calibration which is present in all of the pileup calibration methods stated above, albeit with differing levels of severity.

6.1 Monte Carlo Samples and Structure of Analysis

The Monte Carlo samples used in this part of the analysis have been produced with the PYTHIA [40] Monte Carlo generator and correspond to di-jet events at a centre of mass energy of $\sqrt{s} = 14$ TeV. The samples contain information about truth jets j_{true} and calo-cluster jets j_{calo} as defined in Sec. 4.2. The general structure of the analysis can

be seen in Fig. 6.1. The individual pileup calibration steps, which are described in the sections below, are performed only on the j_{calo} jets. In the following, the calibrated jets will be called j_{corr} while the uncalibrated jets will be called j_{calo}. Out of the j_{true} jets, only
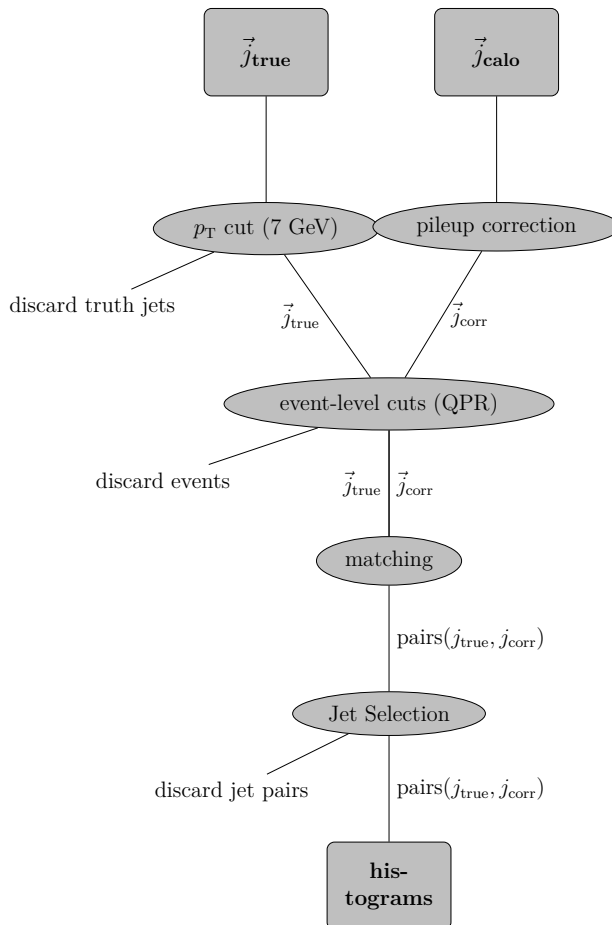
Figure 6.1: Flow chart representing the main steps of the jet calibration prospects study.

those with transverse momentum $p_{\text{T}} > 7$ GeV are selected. Following, an event selection concerning the quality-p_{T}-ratio (QPR)

$$QPR = \frac{p_{\text{T}}(j_{\text{corr}}^{\text{lead}}) + p_{\text{T}}(j_{\text{corr}}^{\text{sublead}})}{2p_{\text{T}}(j_{\text{true}}^{\text{lead}})} \tag{6.1}$$

is applied. Only events with $0 < QPR < 1.4$ are kept. Subsequently, each j_{true} is matched to the j_{corr} which is closest in ΔR. Out of the thus created pairs $(j_{\text{true}},\ j_{\text{corr}})$ only those are kept which fulfill the following conditions:

- $\Delta R(j^{\text{true}},\ j^{\text{corr}}) < 0.3$
- j^{true}-isolation: No other j^{true} with $p_{\text{T}} > 7$ GeV within $\Delta R > 1.0$
- j^{corr}-isolation: No other j^{corr} with $p_{\text{T}} > 7$ GeV within $\Delta R > 0.3$

6.2 Jet Calibration

The energy of jets reconstructed from calorimeter clusters (see Sec. 4.2) does not correspond on average to the energy of the particle initiating the jet. Reasons for this include the non-compensating nature of the hadronic calorimeters [52] and pileup effects as described in Sec. 3.3.

The goal of the jet calibration process is to find a functional dependence with which it is possible to translate the momenta of calorimeter jets to the momenta of truth particle jets (see Sec. 4.2). The subsequent jet calibration steps as they are performed in the ATLAS collaboration can be seen in Fig. 6.2 and are described in more detail in Ref. [77].
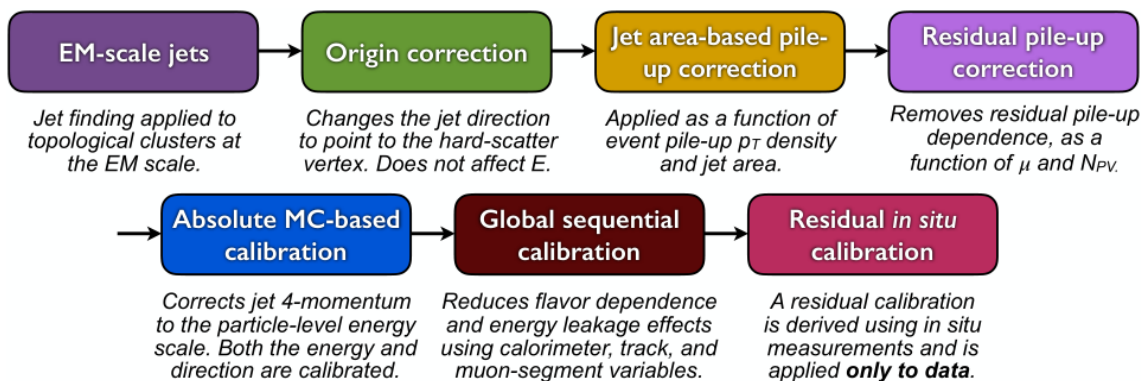
Figure 6.2: Calibration steps for topo-cluster calo jets in the ATLAS experiment [77].

The first step in the calibration process is called origin correction. Here the direction of the jet's four momentum is changed in such a way that it points to the hard-scatter vertex. The energy of the jet is not affected by this process.

The next step is pileup correction which can be subdivided into Jet-Area correction and Residual correction. During the Jet-Area correction, the uncalibrated momentum $p_\mathrm{T}^\mathrm{calo}$ of the jet is corrected by a term depending on the (active) jet area A (Eq. 6.6) and the pileup density ρ (Eq. 6.7):

$$p_\mathrm{T}^\mathrm{corr1} = p_\mathrm{T}^\mathrm{reco} - \rho A \tag{6.2}$$

Both variables will be described in detail in Sec. 6.3. The resulting momentum $p_\mathrm{T}^\mathrm{corr1}$ is then further calibrated during the Residual correction:

$$p_\mathrm{T}^\mathrm{corr2} = p_\mathrm{T}^\mathrm{corr1} - \alpha N_\mathrm{PV} - \beta \langle \mu \rangle \tag{6.3}$$

where α and β are fit parameters derived form Monte-Carlo-simulated samples. The two variables used in this step are the number of primary vertices N_{PV} and the average number of interactions per bunch crossing $\langle \mu \rangle$ which are used as an indicator of how many additional collisions are likely to be present in a given event. Since the pileup calibration was the main focus of this part of the analysis presented in this thesis, its structure and performance is described in more detail in Sec. 6.4.

After pileup correction, the next step is the absolute jet energy scale (JES) calibration. For this purpose calo jets are associated to truth jets by their proximity in ΔR. From these jet pairs the energy response $R(E, \eta)$ defined as the mean of a Gaussian fit to $E^{\text{reco}}/E^{\text{truth}}$ is derived. This is done in several bins of $|\eta|$ and jet energy. The inverse of the response is the Jet-calibration-factor $F(E, \eta)$. The calibration factor that was derived in this manner can be used to calibrate the energy of calo jets to the energy of a corresponding truth particle jet.

The following Global-Sequential calibration uses variables from the calorimeters, the trackers and the dedicated muon chambers of the ATLAS detector to quantify and account for variations in the calorimeter response due to varying particle composition in the jets. Quark-initiated jets are more likely to contain hadrons with a higher fraction of transverse momentum with respect to the whole jet momentum whereas gluon-initiated jets tend to contain more soft particles and therefore to have a wider profile in the transverse direction [77]. Due to the non-compensating nature of the calorimeters [52], this effect has to be accounted for. Jets with a larger hadronic fraction compared to their electromagnetic content are more affected by this phenomenon. Gluon-initiated jets are on average wider. This is used to distinguish them from quark-initiated jets, taking into account the different response of the non-compensating calorimeters.

The last step in the jet calibration process of the ATLAS collaboration is the In-Situ calibration. It accounts for the difference in the jet response between data and Monte Carlo simulations and is therefore only performed on data. The differences are quantified by balancing the transverse momentum of a jet against the transverse momentum of a reference object that can be measured with high accuracy. Examples of such reference objects are Z bosons, photons, or multi-jet systems [77].

Due to the vastly different conditions in the ATLAS detector after the planned upgrade to the HL-LHC, it is expected, that significant changes to this calibration mechanism must be made to ensure a calibration with as least bias as possible. Two factors that must be considered here are the new ATLAS layout, especially of the inner detector, and the harsher pileup conditions.

In the studies presented in this thesis, the calibration process was applied up to the abso-

lute jet energy scale (JES) calibration. The last two steps of the calibration process (i.e. Global-Sequential- and In-Situ calibration) were not investigated.

6.3 Jet Area and Pileup Density

Jets in all-purpose detectors like the ATLAS detector are affected by pileup effects in two fundamentally different ways: On the one hand, additional tracks or calorimeter deposits from pileup events can lead to the formation of jets that would not have been formed from the signature of the primary event alone. These pileup-dominated jets will be called pileup jets in the following. On the other hand, the momenta of jets formed by the signatures stemming from the primary event can be altered because of pileup contamination. The first phenomenon can be countered by pileup-jet tagging, which classifies jets as pileup- or non-pileup jets by exploiting vertexing information from the inner detector. For the latter phenomenon, a specific correction (i.e. pileup correction) is introduced in the jet calibration scheme. Therefore, it is necessary to find physical variables that can be used to quantify the amount of pileup on both event-by-event and jet-by-jet bases.

The first step of the ATLAS jet calibration pileup-removal scheme as described in 6.2 utilises two such variables: the jet area A, which is defined for each individual jet and the pileup density ρ, which is defined on event-by-event basis. ρ is an indicator for the amount of diffuse uniformly distributed pileup noise, while A is used as an indicator for each jet's susceptibility to contamination from such noise. Both variables are calculated during the jet-finding algorithm which is described further in Sec. 4.3.

One might think of a jet's area as the area in angular space coordinates of a detector, in which all the jets substructure is contained. This naïve definition gives a first intuitive understanding on why A can be used as a quantitative indicator for the susceptibility of a jet's momentum to pileup-effects: Under the assumption that the additional calorimeter energy depositions from pileup events are distributed approximately in a uniform manner, it is obvious that jets with larger area would include more of the additional energy.

Although the active area (defined further below) has been used in the studies presented here, it is sensible to start with a definition of the simpler passive area. Both are described in detail in Ref. [78].

Everything described below can also directly be translated into a setting where calorimeter clusters instead of truth particles are used. Assume that during jet-finding, a set of particles $\{p_i\}$ are clustered into a set of jets $\{J_i\}$ by an arbitrary infrared-safe jet algorithm. Due to the infrared-safety in the jet-reconstruction algorithm, adding one particle g with infinitesimally small momentum beforehand does not change the outcome except for two

possibilities: Firstly, one of the jets could contain g but its momentum and direction would be unchanged. Secondly, an additional jet only containing g might be added to the set of jets. Such a particle g is called ghost particle. The passive area of a jet J is then defined as

$$A_{\text{passive}}(J) = \int \mathrm{d}y\mathrm{d}\phi f(g(y,\phi), J) \tag{6.4}$$

where $f(g, J)$ is 1 if J contains g and is 0 otherwise. Integration over the whole rapidity y and azimuthal angle ϕ space is implied. For each numerical integration step in y and ϕ only a single ghost particle g is present. It is situated at the detector position (y, ϕ).

The active jet area is the one used in all of the following studies presented in this thesis. In contrast to the passive area, in the active area's calculation process there are more than one ghost particles present at once. For this purpose, a set $\{g_i\}$ of ghost particles with random positions in y and ϕ is added. Since in this definition of area, the ghosts can cluster not only with one of the jets but also with each other, they take a more active role during jet-finding. For a given set $\{g_i\}$ of ghost particles the active jet area for the jet J is defined as

$$A(J|\{g_i\}) = \frac{N_g(J)}{\nu_g}. \tag{6.5}$$

Here, N_g is the number of ghost particles contained in the jet J while ν_g is the ghost density, i.e. the average number of ghost particles per unit area in y-ϕ space. An illustration of the active areas of several hard scatter jets calculated in this way can be seen in Fig. 6.3. Since the result depends on the individual positions of the ghost particles in the detector, the process has to be repeated several times with different randomly generated sets $\{g_i\}$. From this, a set-independent value can be calculated by the average

$$A(J) = \langle A(J|\{g_i\})\rangle_g \tag{6.6}$$

over all obtained values of $A(J|\{g_i\})$. For this definition to be stable and sensible, sufficiently large ghost densities v_g and number of repetitions of calculating $A(J|\{g_i\})$ for different sets $\{g_i\}$ have to be implemented. Since this is a computationally expensive process, a trade-off between number of repetitions and calculation time has to be made. The pileup density ρ which serves as an indicator for the amount of pileup present in an
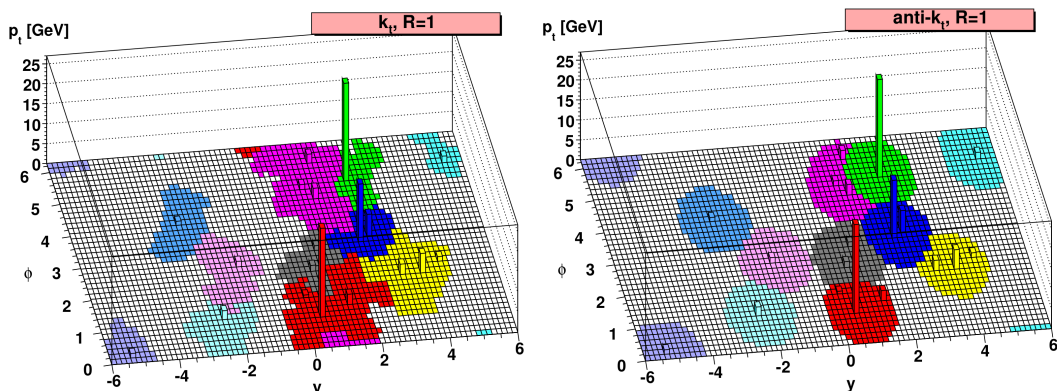
Figure 6.3: Truth particles from a sample parton-level event together with a set of random ghost particles $\{g_i\}$ clustered with the k_T algorithm [left] and the anti-k_T algorithm [right] [53]. The shaded areas correspond to ghost particles g_i that are clustered into a jet. They therefore constitute an illustration of the active jet area $A(J|\{g_i\})$ of the corresponding jet J. Jets that only contain ghost particles are not shaded. The more circular shape of the anti-k_T jet's areas compared to k_T jet's areas is due to the anti-k_T algorithm's higher priority on angular position as opposed to momentum as described in Sec. 4.3.

individual event is defined as [79]

$$\rho = \text{median}\left[\frac{p_{\text{T}}(j)}{A(j)}\right]_j . \tag{6.7}$$

Here $\{j\}$ is a set of soft jets. For this purpose, jet-finding is performed with the k_T algorithm . The k_T algorithm is used because of its higher susceptibility to pileup effects compared to the anti-k_T algorithm as explained in Sec. 4.3. As a consequence of this, a large sample of jets with low momenta is part of the set of jets created. This sample is enriched in jets that do not correspond to any hard scatter interaction. Hence, it is dominated by pileup effects and can therefore serve as an indicator thereof. This constitutes the set of soft jets $\{j\}$ used in Eq. 6.7.

The reasoning to use ρ and A as indicators for pileup in the jet calibration process is the assumption that for pure pileup jets J the condition

$$\rho \approx \frac{p_{\text{T}}(J)}{A(J)} \tag{6.8}$$

holds on average. This has been studied and verified by Ref. [79] for simulated events with 22 pileup interactions. Whether or not this assumption holds also for the HL-LHC remains to be seen.

6.4 Methods for Pileup Calibration

The following sections describe the different methods for jet pileup calibration, which have been studied in the scope of this thesis. The pileup calibration is only one of many steps of the complete jet calibration procedure in the ATLAS collaboration. The full calibration chain was described in Sec. 6.2. The purpose of the pileup calibration is to ensure that the calibrated jet's properties (i.e. momenta) are not affected by the individual pileup condition of the corresponding event. Consequently, a low correlation of the calibrated jet's momenta to variables which quantify the amount of pileup is desirable.
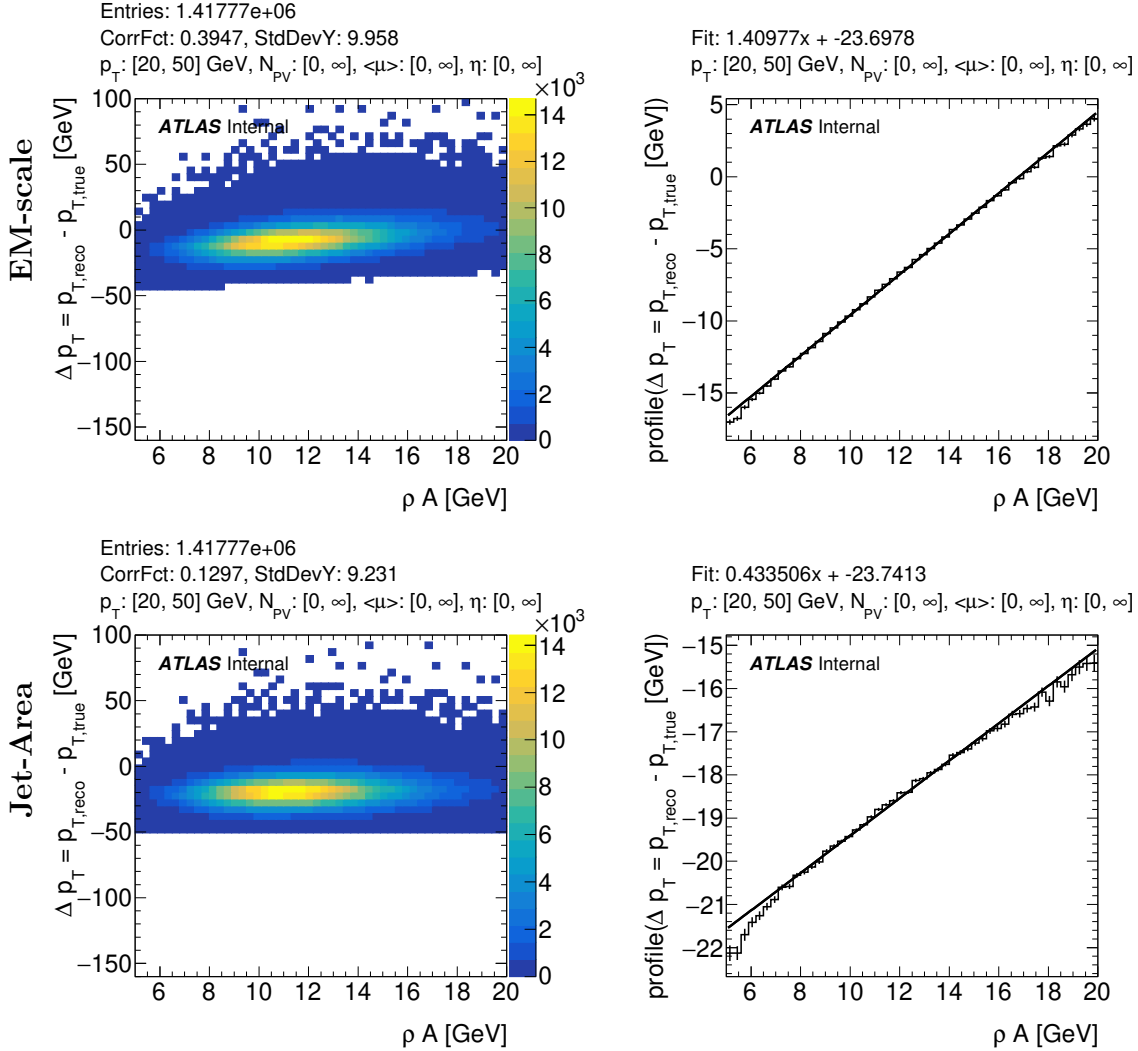
6.4.1 Jet-Area Correction

The Jet-Area correction is the first step of two in the jet pileup calibration as it is currently performed in the ATLAS jet calibration. It is performed on jets j^{reco} that have not been pileup corrected before. In this study, EM-scale jets j^{calo} reconstructed from calorimeter topo-clusters (see Sec. 4.2) were used. EM-scale referes to the reconstruction process of the topo-clusters in which the non-compensating nature of the calorimeter has not been taken into account. The jet momentum after Jet-Area correction is given by

$$p_{\text{T}}^{\text{corr1}} = p_{\text{T}}^{\text{reco}} - \rho A. \tag{6.9}$$

It utilizes two variables serving as indicators for the amount of pileup: The jet area A and the pileup density ρ (see Sec. 6.3). ρ serves as an indicator for the amount of pileup in an individual event, while A serves as an indicator for how susceptible an individual jet is affected by the pileup in the event. Consequently, the product $\rho A(j^{\text{reco}})$ can be used as an indicator of the effect of pileup on the jet j^{reco} .

Fig. 6.4 shows the correlation of the difference Δp_{T} of truth jet and calo jet p_{T} with respect to ρA before and after Jet-Area correction. Only jet pairs (j_{reco}, j_{true}) with $4.0 < |\eta(j_{\text{true}})| < 4.1$ were used. It is apparent that by applying the Jet-Area correction, the linear correlation factor between Δp_{T} and ρA can be lowered drastically from 0.39 to 0.13 and the slope of the corresponding profile can be lowered from 1.41 to 0.43. The same kind of studies were performed using only jet pairs (j_{true}, j_{reco}) in the central, medium, or forward η region. They can be found in Fig. C.1, Fig. C.2, and Fig. C.3. The resulting correlation factors and slopes of the fits to the profiles are summarized in Sec. 6.5.1.

Figure 6.4: [Left]: Correlation of the difference Δp_T of truth jet and reconstructed jet p_T with respect to ρA at EM-scale (i.e. no pileup calibration) and after Jet-Area correction. The linear correlation factors (CorrFct) can be seen on the top of the plots. Only jet pairs (j_{true}, j_{reco}) with $20 <$ GeV $p_T(j_{\text{true}}) < 50$ GeV have been considered. [Right]: Profiles of the corresponding plots on the left. A linear fit has been applied. The resulting slope and offset can be seen on top of the plots.
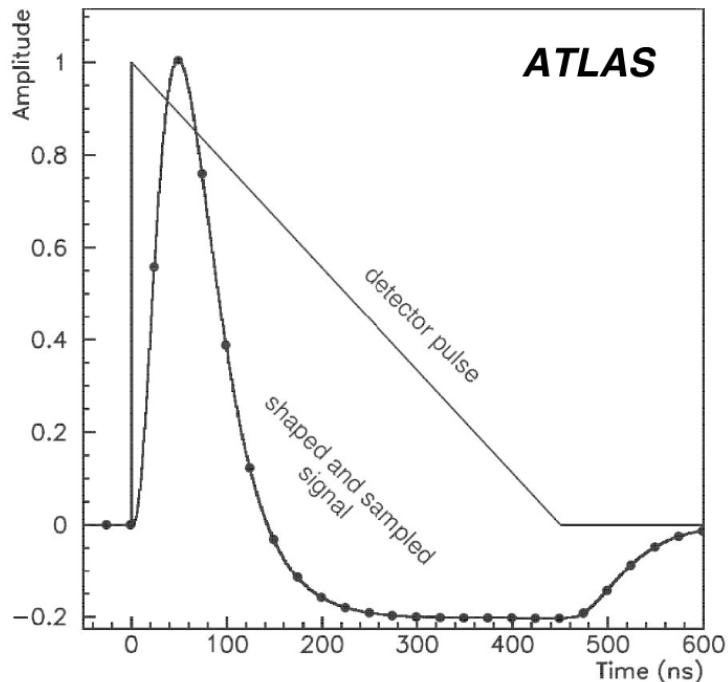
6.4.2 Residual Correction

The Residual correction is the following step in the jet pileup correction, as it is currently performed in the ATLAS jet calibration. It is performed on jets that have already been calibrated by the Jet-Area correction. In the following, they are referred to as j^{corr1}. The jet momentum after Residual correction is given by

$$p_{\text{T}}^{\text{corr2}} = p_{\text{T}}^{\text{corr1}} - \alpha N_{\text{PV}} - \beta \langle \mu \rangle \tag{6.10}$$

where α and β are fit parameters derived form Monte Carlo samples. In this study, values of α and β were determined using HL-LHC conditions. In this step the number of reconstructed primary vertices in the detector N_{PV} and the average pileup value $\langle \mu \rangle$ are used as indicators for the amount of pileup. Both are defined on an event-by-event basis: N_{PV}, similar to the pileup density ρ, gives a handle on the amount of in-time pileup. The number of reconstructed primary vertices in the detector is directly related to the number of additional proton-proton collisions in an event (i.e. in-time pileup). It is defined as the number of all reconstructed vertices that have at least two associated tracks.

The average pileup value $\langle \mu \rangle$ is defined as the number of proton-proton collisions per event averaged over many events (typically corresponding to one luminosity block [80]). Hence, $\langle \mu \rangle$ is an indicator for both in-time pileup and for out-of time pileup. The reason for this is that by taking the average over several events, $\langle \mu \rangle$ is susceptible to the amount of collisions in events earlier than the current one. A higher number of collisions in earlier events corresponds to a larger amount of detector signatures of such events that are still detectable in the current event (i.e. out-of-time pileup). This is the case because the read-out time of the calorimeters is higher than the time between two bunch crossings. While in-time pileup has an increasing effect on the energy measured by the calorimeters, out-of time pileup on average decreases the energy. The reason for this is the shape of the current pulse of the LAr calorimeters (see the schematic graph of the shaped and sampled signal in Fig. 6.5). Its amplitude becomes negative after an elapsed time of ≈ 150 ns. This time is larger than the average bunch-crossing time of 25 ns of the LHC. Therefore, if a pulse from an earlier event overlaps with one from the current event this can lead to a decrease in the amplitude of the pulse. Since the Jet-Area correction is already able to eliminate the effect of a large part of the in-time pileup, the Residual correction mainly corrects for out-of-time pileup. Hence, on average it increases the energy of the reconstructed jets.

Fig. 6.6 and 6.7 show the correlation of the difference Δp_{T} of truth jet and calo jet p_{T} with respect to N_{PV} and $\langle \mu \rangle$ before and after Residual correction. Only jet pairs $(j_{\text{true}}, j_{\text{reco}})$

Figure 6.5: Shapes of the LAr calorimeter current pulse in the detector and of the signal output [81].

with $20 < \mathrm{GeV}\ p_\mathrm{T}(j_\mathrm{true}) < 50\ \mathrm{GeV}$ have been considered. It is apparent that by applying the Residual correction the linear correlation factor between Δp_T and N_PV can be lowered from 0.08 to 0.01 and the slope of the corresponding profile can be lowered from 0.11 to 0.01. The absolute value of the correlation factor between Δp_T and $\langle\mu\rangle$ increases from 0.0006 to 0.002 and the absolute value of the slope of the corresponding profile increases from 0.001 to 0.004. This unwanted behaviour suggests that the Residual correction as it is implemented currently is not suitable for the HL-LHC's conditions. Consequently, the values of α and β have to be redetermined. The value of the correlation factor and the slope before Residual correction is negative because of the negative impact of out-of-time pileup on the jet's energy as described above. The same kind of studies were performed using only jet pairs $(j_\mathrm{true},\ j_\mathrm{reco})$ in the central, medium, or forward η region. They can be found in Fig. C.5, Fig. C.6, and Fig. C.7 for N_PV and in Fig. C.9, Fig. C.10, and Fig. C.11 for $\langle\mu\rangle$. The resulting correlation factors and slopes of the fits to the profiles are summarized in Sec. 6.5.1.
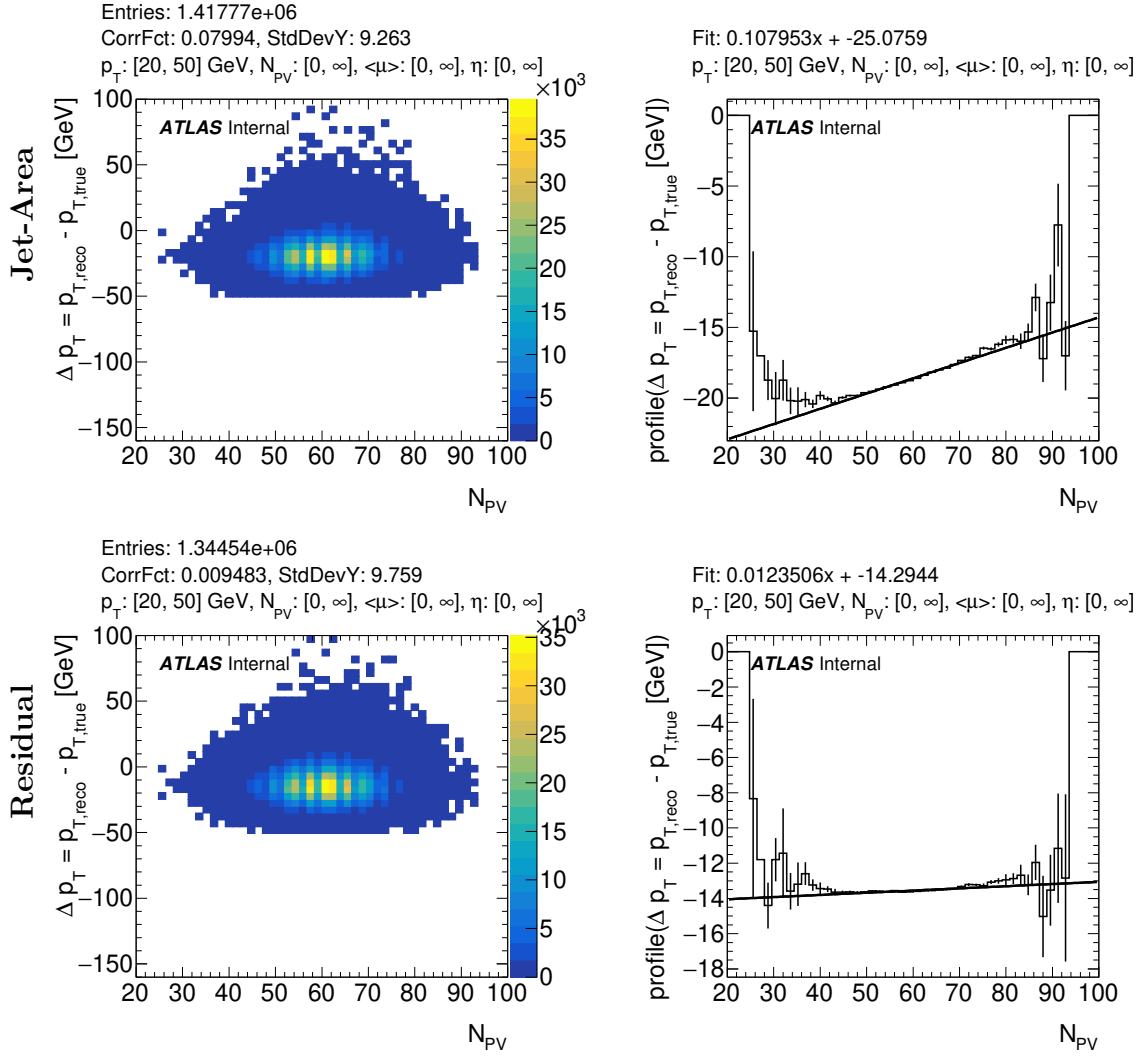
Figure 6.6: [Left]: Correlation of the difference Δp_T of truth jet and reconstructed jet p_T with respect to N_{PV} after Jet-Area correction and after subsequently applying the Residual correction. The linear correlation factors (CorrFct) can be seen on the top of the plots. Only jet pairs (j_{true}, j_{reco}) with $20 <$ GeV $p_T(j_{true}) < 50$ GeV have been considered. [Right]: Profiles of the corresponding plots on the left. A linear fit has been applied. The resulting slope and offset can be seen on top of the plots.
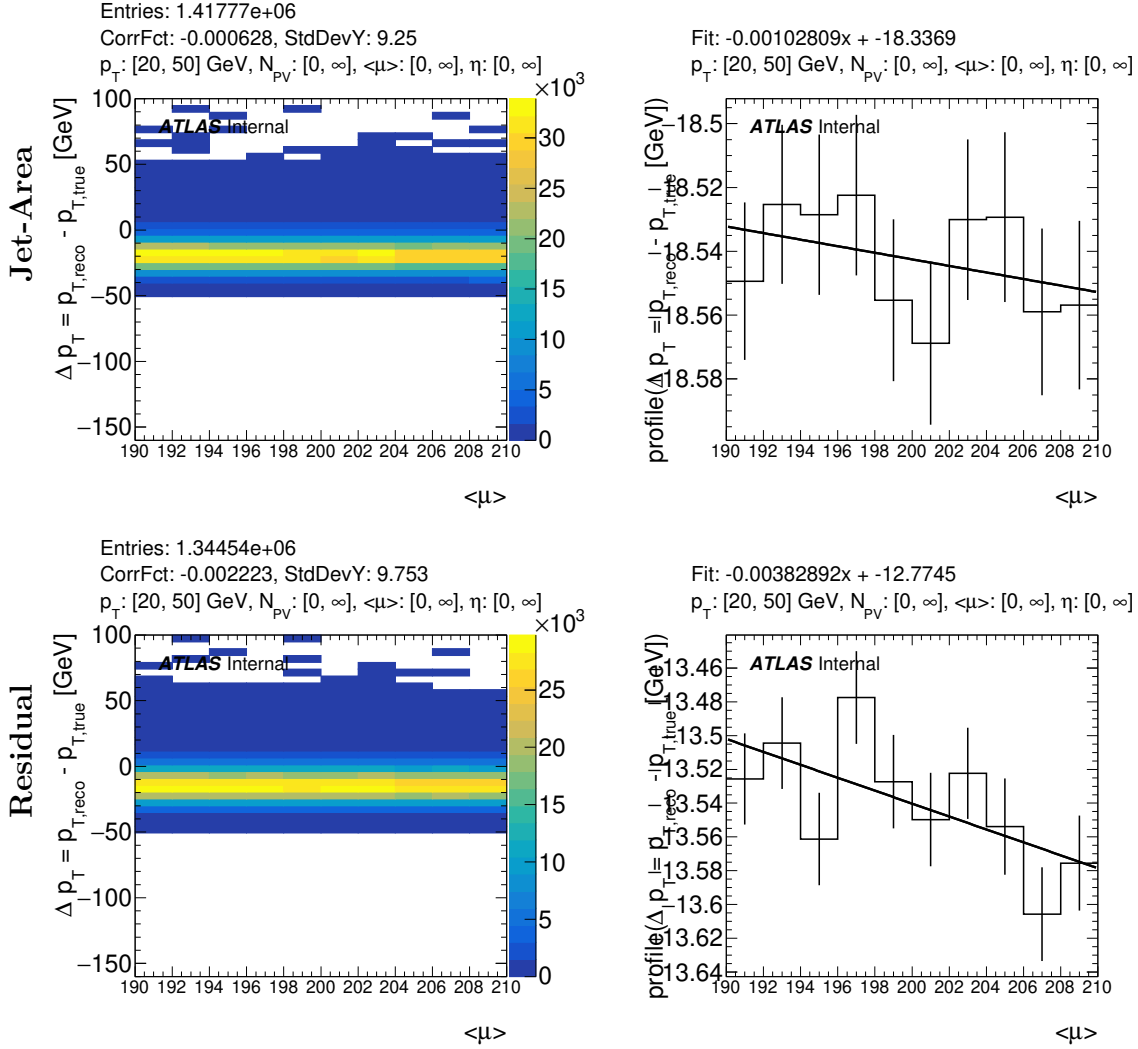
Figure 6.7: [Left]: Correlation of the difference Δp_T of truth jet and reconstructed jet p_T with respect to $\langle\mu\rangle$ after Jet-Area correction and after subsequently applying the Residual correction. The linear correlation factors (CorrFct) can be seen on the top of the plots. Only jet pairs (j_{true}, j_{reco}) with $20 <$ GeV $p_T(j_{\text{true}}) < 50$ GeV have been considered. [Right]: Profiles of the corresponding plots on the left. A linear fit has been applied. The resulting slope and offset can be seen on top of the plots.

6.4.3 Gradient-Jet-Area Correction

The Gradient-Jet-Area correction is an altered version of the Jet-Area Correction. While the latter is defined by

$$p_T^{\text{corr1}} = p_T^{\text{reco}} - \rho A, \tag{6.11}$$

the former includes an additional factor $a(p_T, \eta)$ yielding

$$p_T^{\text{corr}} = p_T^{\text{reco}} - a(p_T, \eta)\rho A. \tag{6.12}$$

The use of this factor is motivated by the observation that the Jet-Area correction allows a reduction of the slope of the profile of Δp_T with respect to ρA in all regions by a value close to one (as can be seen in Fig. 6.4 and Tab. 6.1). Since this slope is greater than one in most regions, the introduction of a factor $a(p_T, \eta)$ which is also on average greater than one should yield better results. For this purpose, $a(p_T, \eta)$ is simply defined as the value of the respective slope before pileup calibration. Since this slope is given by

$$a(p_T, \eta) = \frac{\Delta p_T}{\rho A} = \frac{p_T^{\text{reco}} - p_T^{\text{true}}}{\rho A}, \tag{6.13}$$

inserting it in Eq. 6.12 yields

$$p_T^{\text{corr}} = p_T^{\text{reco}} - \frac{p_T^{\text{reco}} - p_T^{\text{true}}}{\rho A}\rho A = p_T^{\text{true}}. \tag{6.14}$$

Hence, on average, the Gradient-Jet-Area corrected jet p_T should correspond to the true jet p_T. Since only the slope and not the offset of the dependence of Δp_T to ρA was taken into account, the actual calibrated p_T^{corr} can still on average be different from p_T^{true} by a constant offset. However, as described in Sec. 6.2, the purpose of the pileup calibration is not to get a one-to-one correspondence of the reconstructed- and truth jet's p_T but only to remove the correlation of p_T^{reco} to the individual pileup conditions of the respective event. A scaling of the already pileup-calibrated p_T^{reco} to be as close as possible to the true value is done at a later stage in the absolute JES calibration step. Hence, a constant offset is not a concern here. One can see such a constant offset also when looking at the respective profile fits for the other pileup calibration methods (i.e. Jet-Area and Residual). A more detailed discussion on the effect of this offset can be found in Sec. 6.5.3. The values for $a(p_T, \eta)$ were calculated in bins of p_T and η. The result can be seen in Fig. 6.8.
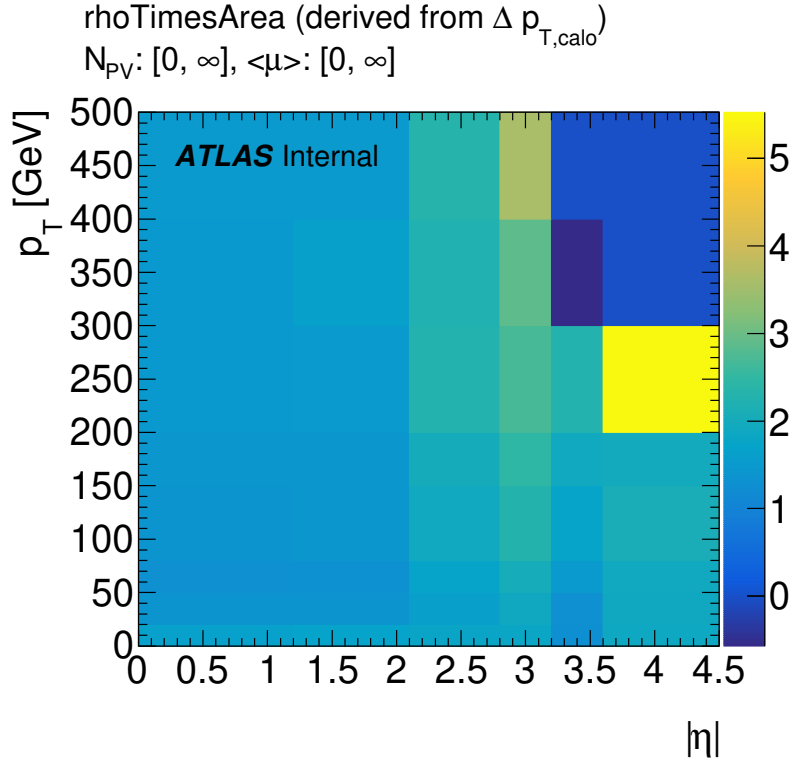
Figure 6.8: Values for the factor $a(p_\text{T},\ \eta)$ used for Gradient-Jet-Area correction as defined in Eq. 6.12. They were calculated binned in p_T and η. The large variation of the values in regions of high p_T and η is due to statistical fluctuations caused by the small number of jets in this region.

Fig. 6.9 shows the correlation of the difference Δp_T of truth jet and calo jet p_T with respect to ρA before and after Gradient-Jet-Area correction. Only jet pairs $(j_\text{true},\ j_\text{reco})$ with $20 < \text{GeV } p_\text{T}(j_\text{true}) < 50$ GeV have been considered in these plots. It is apparent that by applying the Gradient-Jet-Area correction the linear correlation factor between Δp_T and ρA can be lowered drastically from 0.38 to 0.01 and the slope of the corresponding profile can be lowered from 1.45 to 0.05. This constitutes a significant improvement compared to the Jet-Area method (not using $a(\eta, p_\text{T})$) which resulted in a correlation factor of 0.13 and a slope of 0.43 (see Sec. 6.4.1).

The same kind of studies were performed using only jet pairs $(j_\text{true},\ j_\text{reco})$ in the central, medium, or forward η region. They can be found in Fig. C.1, Fig. C.2, and Fig. C.3. The resulting correlation factors and slopes of the fits to the profiles are summarized in Sec. 6.5.1.
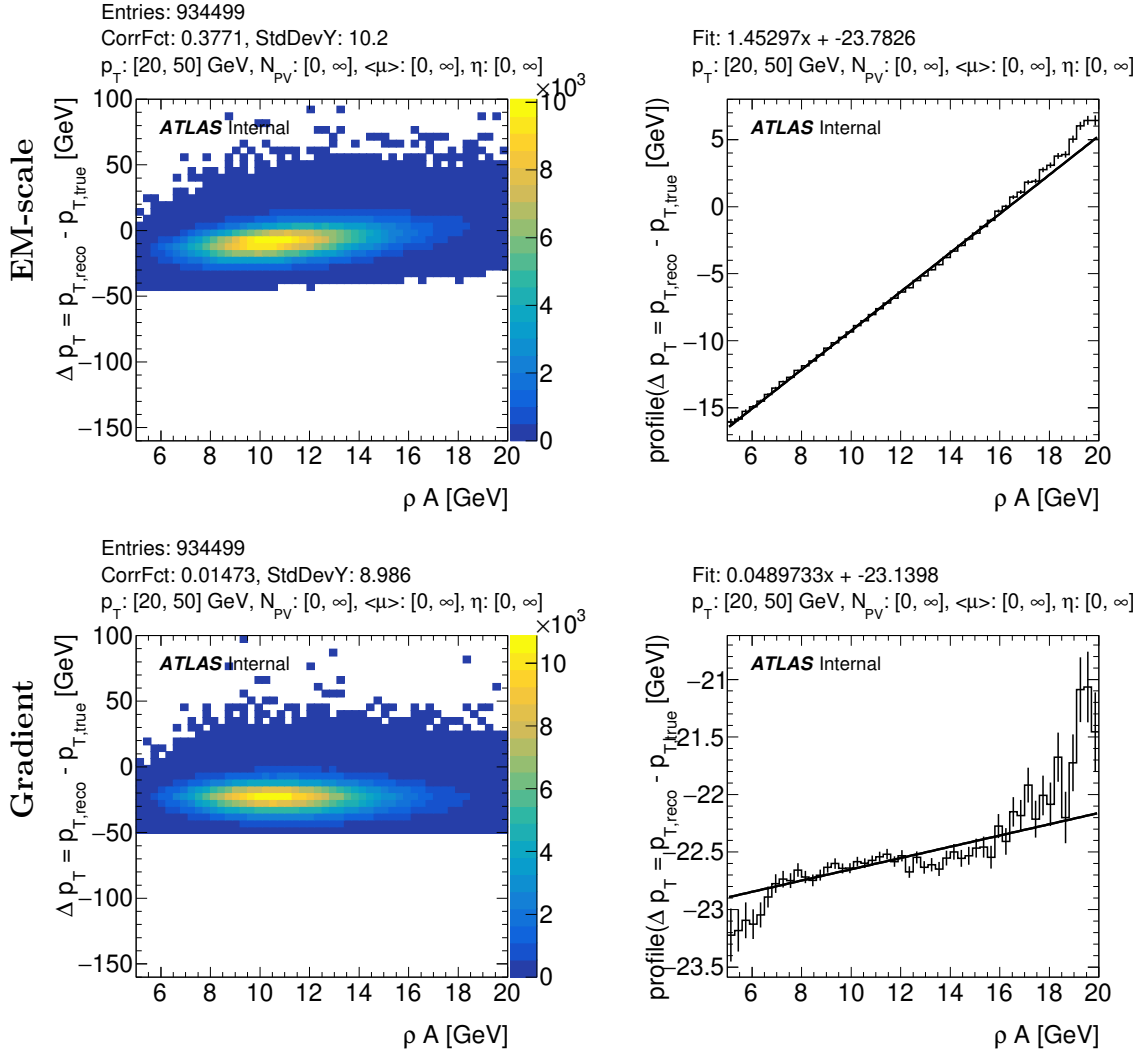
Figure 6.9: [Left]: Correlation of the difference Δp_T of truth jet and reconstructed jet p_T with respect to ρA at EM-scale (i.e. no pileup calibration) and after Gradient-Jet-Area correction. The linear correlation factors (CorrFct) can be seen on the top of the plots. Only jet pairs $(j_{\text{true}}, j_{\text{reco}})$ with $20 <$ GeV $p_T(j_{\text{true}}) < 50$ GeV have been considered. [Right]: Profiles of the corresponding plots on the left. A linear fit has been applied. The resulting slope and offset can be seen on top of the plots.

6.5 Comparison of Performance

In the following sections the the performance of the three methods of jet pileup correction introduced in Sec. 6.4 are compared. First, their ability to remove the dependency on pileup effects is summarized. Then, they are compared with respect to closure of p_{reco} to p_{true} after the absolute JES calibration. Lastly, an issue with negative energy after pileup

correction is presented affecting all of these methods to different extents. Approaches to mitigate this problem are presented as well.

6.5.1 Removal of Pileup Dependencies

As described in Sec. 6.4, the removal of the correlation between jet momentum and the individual pileup conditions of the event is the main goal of the pileup correction step of the jet calibration. Therefore, comparing the correlation factors and profile slopes before and after pileup correction for the different methods yields insight on their individual performance. Tab. 6.1 summarizes these values with respect to ρA, N_{PV}, and $\langle \mu \rangle$ for the different methods and divided by η region.

	η-region	EM-scale correl.	EM-scale slope	Jet-Area correl.	Jet-Area slope	Residual correl.	Residual slope	Gradient-Jet-Area correl.	Gradient-Jet-Area slope
ρA	central	0.3808	1.3029	0.0992	0.3267	0.0905	0.2980	0.0273	0.1066
	medium	0.4096	1.4556	0.1466	0.4785	0.0844	0.2537	0.0210	0.0522
	forward	0.4484	1.6964	0.2074	0.7187	0.0757	0.2450	0.0537	0.1632
	inclusive	0.3947	1.4098	0.1297	0.4335	0.0622	0.2192	0.0147	0.0490
N_{PV}	central	0.1880	0.2676	0.0478	0.0625	0.0222	0.0281	0.0213	0.0288
	medium	0.2348	0.3348	0.1089	0.1417	0.0169	0.0178	0.0468	0.0586
	forward	0.3000	0.4667	0.1886	0.2654	0.0429	0.0538	0.1085	0.1387
	inclusive	0.2099	0.3096	0.0799	0.1080	0.0095	0.0124	0.0282	0.0368
$\langle \mu \rangle$	central	0.0457	0.0777	0.0067	0.0104	0.0026	0.0046	−0.0014	−0.0020
	medium	0.0313	0.0509	−0.0069	−0.0111	−0.0066	−0.0111	−0.0222	−0.0336
	forward	0.0154	0.0293	−0.0230	−0.0388	−0.0128	−0.0221	−0.0414	−0.0627
	inclusive	0.0371	0.0648	−0.0006	−0.0010	−0.0022	−0.0038	−0.0168	−0.0265

Table 6.1: Linear correlation factors (correl.) and profile slopes of the difference Δp_{T} of truth- and reconstructed jets with respect to ρA, N_{PV}, and $\langle \mu \rangle$. The values were derived at EM-scale (i.e. no pileup calibration), after Jet-Area correction, after Jet-Area+Residual correction, and after Gradient-Jet-Area correction. The corresponding plots used for the calculation of these values can be found in the appendix (Fig. C.1-C.12). The η regions are defined as $0 < |\eta(j_{\mathrm{true}})| < 1.2$ (central), $2.1 < |\eta(j_{\mathrm{true}})| < 2.8$ (medium), and $3.6 < |\eta(j_{\mathrm{true}})| < 4.5$ (forward). The slopes with respect to N_{PV} and $\langle \mu \rangle$ are given in GeV. All other values are unitless.

It is apparent that the Gradient-Jet-Area correction method performs best in terms of removal of ρA dependence in all η regions. This was expected, because it explicitly uses ρA and the corresponding slope of Δp_{T} versus ρA. Since the Residual correction is performed subsequently to the Jet-Area correction, it is not surprising that it performs

second best in this regard, while the Jet-Area correction alone performs worst.

The Residual correction performs best in removing the NPV dependence. This is expected because it explicitly uses N_{PV} as one of the values to parametrize the amount of pileup. The possibility to also include N_{PV} into the definition of the Gradient-Jet-Area correction was investigated as well, but did not yield promising results. The reason for this is an issue with negative energy after pileup removal which is discussed in further detail in Sec. 6.5.3. The Residual correction performs also best in removing the dependence to $\langle \mu \rangle$. This is also one of the parameters explicitly used in this method of correction.

6.5.2 Closure

Closure in the context of this study is defined by comparing the value of the transverse momentum of the reconstructed jet ($p_{\mathrm{T,reco}}$) to the corresponding value of the truth jet ($p_{\mathrm{T,true}}$) which is matched to it. Fig. 6.10 shows the closure in a highly forward region $4.0 < |\eta(j_{\mathrm{true}})| < 4.1$ for the three different pileup calibration methods presented in Sec. 6.4. It is apparent that the Gradient-Jet-Area method performs best for jets with low transverse momentum 20 GeV $p_{\mathrm{T,true}} < 35$ GeV as it is the only method for which all points lie on the line corresponding to $p_{\mathrm{T,reco}} = p_{\mathrm{T,true}}$ within their respective error bars. Fig. D.1 and Fig. D.2 show the corresponding plots in a central- and a medium η region. Only small differences between the three methods can be seen here.
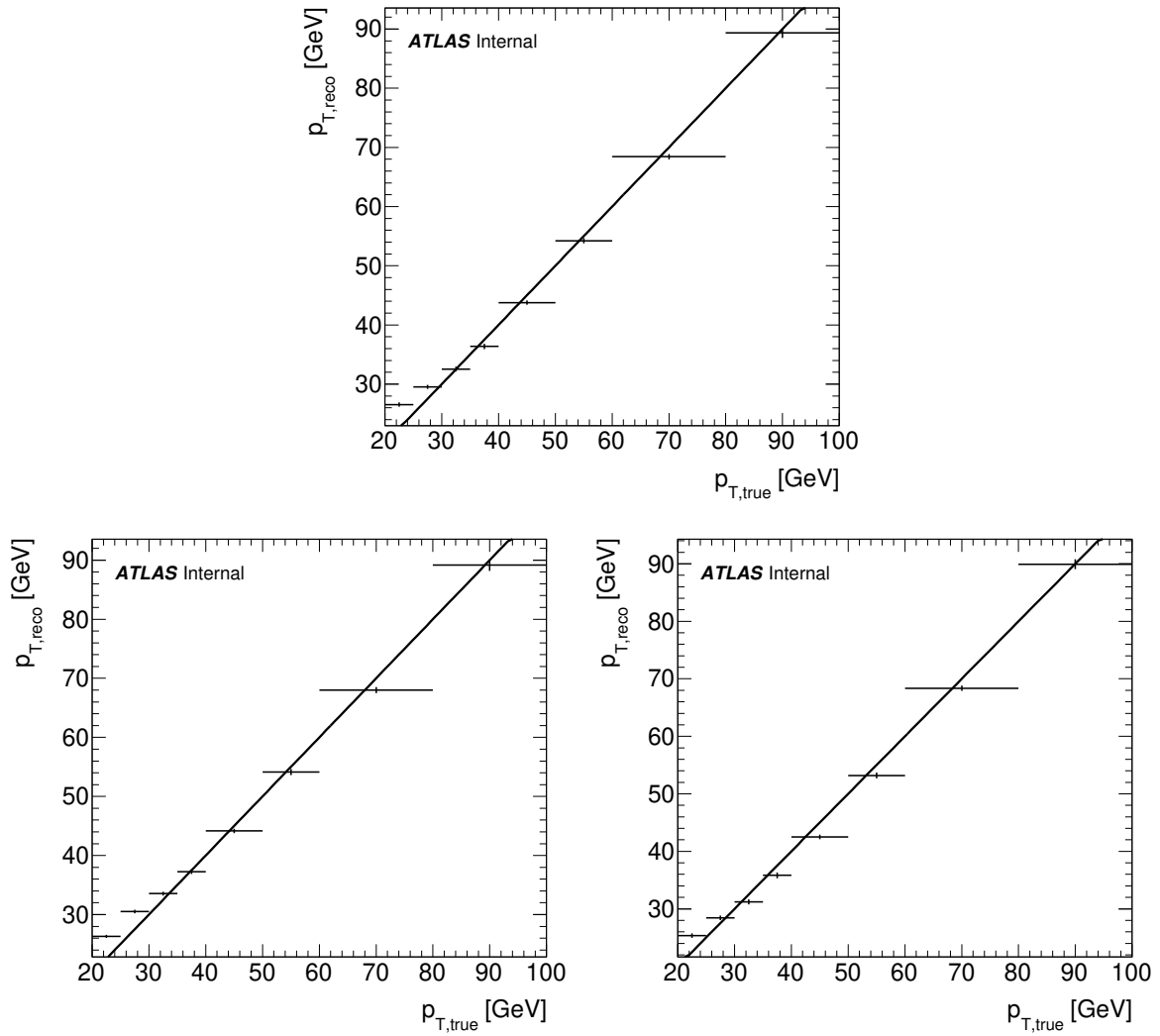
Figure 6.10: Closure plots of the transverse momentum of reconstructed and calibrated jets ($p_{\mathrm{T,reco}}$) with respect to truth level jets ($p_{\mathrm{T,reco}}$). Pileup calibration and absolute JES calibration have been performed. Each plot corresponds to a different method of pileup calibration that was used: Jet-Area [top], Residual [left], and Gradient-Jet-Area [right]. The black line corresponds to $p_{\mathrm{T,reco}} = p_{\mathrm{T,true}}$. Only jet pairs ($j_{\mathrm{reco}}$, j_{true}) with $4.0 < |\eta(j_{\mathrm{true}})| < 4.1$ were used.

6.5.3 Negative Energy

The equations defining the three different pileup calibration methods presented in 6.4 suggest that the calibrated transverse jet momentum can become negative under certain conditions. The defining equations for the three methods are

$$p_\mathrm{T}^\mathrm{corr1} = p_\mathrm{T}^\mathrm{reco} - \rho A \tag{6.15}$$

for the Jet-Area correction,

$$p_\mathrm{T}^\mathrm{corr2} = p_\mathrm{T}^\mathrm{corr1} - \alpha N_\mathrm{PV} - \beta\langle\mu\rangle \tag{6.16}$$

for the Residual correction, and

$$p_\mathrm{T}^\mathrm{corr} = p_\mathrm{T}^\mathrm{reco} - a(p_\mathrm{T},\ \eta)\rho A \tag{6.17}$$

for the Gradient-Jet-Area correction. The conditions for negative p_T (and therefore negative energy as well) after pileup correction are therefore

$$p_\mathrm{T}^\mathrm{reco} > \rho A \tag{6.18}$$

for the Jet-Area correction,

$$p_\mathrm{T}^\mathrm{corr1} > \alpha N_\mathrm{PV} - \beta\langle\mu\rangle \tag{6.19}$$

for the Residual correction, and

$$p_\mathrm{T}^\mathrm{reco} > a(p_\mathrm{T},\ \eta)\rho A \tag{6.20}$$

for the Gradient-Jet-Area correction. Since the Residual correction with respect to $\langle\mu\rangle$ has on average an increasing effect on the energy (see Sec. 6.4.2), the Residual correction is expected not to be affected by this phenomenon as much as the other calibration methods. $a(p_\mathrm{T},\ \eta)$ is on average greater than one (see Fig. 6.8). Therefore, the Gradient-Jet-Area correction is expected to be affected by this phenomenon the most. Fig. 6.11 shows the number of jets with negative energy after pileup calibration with respect to true energy E_true for the three different methods. It confirms the expectations stated above. It was decided to discard jets that have negative energy after pileup calibration for all following steps of the analysis. A different approach would be to assign values of transverse momentum and energy of zero or close to zero to such jets.
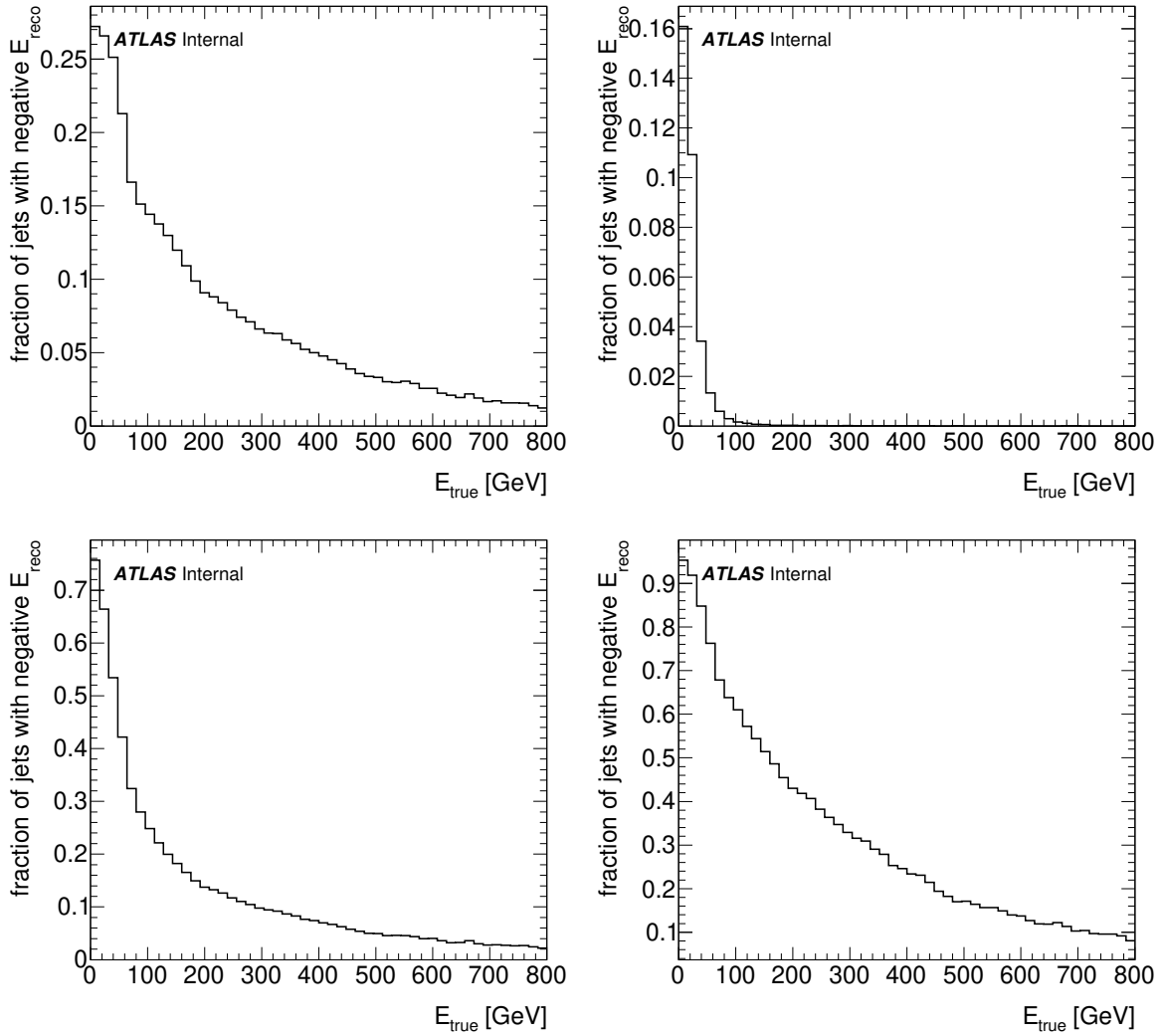
Figure 6.11: Fraction of jets with negative energy after pileup correction with respect to true jet energy E_{true}. Results for the three different pileup correction methods are shown [from top left to bottom right]: Jet-Area, Residual, Gradient-Jet-Area, and Gradient-Jet-Area-NPV.

As mentioned in Sec. 6.4.3, the possibility of including N_{PV} as an additional term similar to the one for ρA into the definition of the Gradient-Jet-Area method (Eq. 6.17) has been investigated. This however, would lead to even more jets having negative energy after the pileup calibration (see Fig. 6.11). In this particular case, the amount of jets that would have to be discarded was so large that no sensible values for a pileup correction could have been derived. It was therefore decided to only include ρA in the Gradient-Jet-Area correction, i.e. to refrain from performing a residual calibration with respect to N_{PV} and $\langle\mu\rangle$.

Possible techniques to compensate the amount of jets with negative energy include taking into account a constant term in the definition of the Gradient-Jet-Area correction. This would not spoil the pileup correction because the actual scaling of the reconstructed momentum to match the true momentum as close as possible is done in the later step of absolute JES calibration. Hence, the additional offset could be compensated by a larger scaling factor in the latter calibration step. A choice for the constant term could be the offset corresponding to the slope which was used as the factor $a(p_{\mathrm{T}},\ \eta)$. Choosing an even higher value might lead to even better results. Including $\langle\mu\rangle$ into the definition of the Gradient-Jet-Area calibration in a similar way as ρA may also provide a partial solution for the problem with negative energy after pileup calibration. The results for the Residual calibration show that the calibration with respect to $\langle\mu\rangle$ increases the energy on average by a high enough amount to effect a large amount of jets in such a way that they gain positive energy after pileup calibration. A different approach for compensating this problem could be made by adapting the implementation of the absolute JES calibration step which in its current form is not able to derive a sensible calibration for jets with negative energy.

7 Conclusion and Outlook

The conclusion and outlook corresponding to the studies presented above is divided into two sections: Sec. 7.1 describes the prospects study of the VBF $H \to \tau\tau$ process at the HL-LHC which has been presented in Sec. 5. Sec. 7.2 describes the studies corresponding to the pileup calibration of jets at the ATLAS detector for the HL-LHC which have been presented in Sec. 6.

7.1 Prospects for VBF $H \to \tau\tau$

The expected performance of the ATLAS detector at the HL-LHC with respect to signatures corresponding to the VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ processes and the associated background processes has been simulated and implemented in Monte-Carlo-generated samples. The results have been used to perform a cut-based analysis to estimate the sensitivity of ATLAS to this process in the scenario of the HL-LHC.

An expected significance for the access of VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ processes over the expected background at the HL-LHC was calculated with the result of $\Sigma_{\mathrm{b}}^{\mathrm{exp}} = 8.63 \pm 0.66_{\mathrm{stat}}$ standard deviations for an integrated luminosity of 3000 fb^{-1}. This corresponds to 2322 signal events compared to 151290 events from all of the considered backgrounds after preselection and VBF signal region cuts (see Tab. 5.4). With a layout of the HL-LHC which includes the High Granularity Timing Detector (HGTD), an even higher significance of $\Sigma_{\mathrm{b}}^{\mathrm{exp,HGTD}} = 9.40 \pm 0.71_{\mathrm{stat}}$ was estimated. This value is drastically higher than the observed (expected) significance of $\Sigma_o = 4.4$ ($\Sigma_e = 4.13$) that has been obtained using Run 2 data corresponding to 36.1 fb^{-1} ($\Sigma_o = 6.4$ ($\Sigma_e = 5.4$) in combination with Run 1 data) and combining all relevant $\tau\tau$ final states (i.e. $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$, $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$, and $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$) [58]. Since the analysis presented in this thesis only concerns the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ final state, the combination of all final states at HL-LHC conditions promises to yield an even higher significance. Further gains could be made by implementing a more involved type of cut optimization such as it has been done in the corresponding Run 2 analysis [58]. This has been achieved by employing a binned likelihood function combining several signal- and control regions. Multivariate analysis methods such as boosted decision trees or artificial neural networks

have already been utilized in many analyses in the ATLAS collaboration and often yielded higher sensitivities than cut based analyses. Hence, incorporating such a method might yield an even higher significance.

The HL-LHC's substantial increase in luminosity is accompanied by the drawback of increased pileup. Consequently, the increase of additional E_T^{miss} and jets from pileup worsen the separation power of variables dependent on these objects: It was shown in this analysis that the resolution of the di-τ mass m_{MMC} suffers drastically due to pileup E_T^{miss} while the additional pileup jets lessen the significance gained by the utilisation of the unique kinematics of VBF jets, i.e. by the VBF signal region cuts.

The incorporation of the High Granularity Timing Detector (HGTD) into the HL-LHC's layout yielded an improved rejection efficiency of pileup-jets especially in the forward detector region. Due to the important role of forward jets in the presented analysis, not only a high rejection efficiency but also a reliable calibration of such jets is desirable. The latter was investigated in Sec. 6 and is further discussed in the following Sec. 7.2.

7.2 Jet Calibration in the Forward Detector Region

Due to the harsher pileup conditions at the HL-LHC it is necessary to reevaluate the individual steps of the jet calibration procedure used in ATLAS. In particular, the pileup calibration consisting of Jet-Area correction and Residual correction will not be suited for the conditions introduced by the upgrade.

A modified approach, the Gradient-Jet-Area correction, is introduced and it shows improvement with respect to the removal of pileup dependencies over the currently used combination of Jet-Area- and Residual correction (see Sec. 6.5). This however comes at the cost of an increased number to jets being assigned a negative (and therefore unphysical) transverse momentum and energy after the correction. This phenomenon is also present in the currently used methods, albeit to a smaller extent. With the current LHC conditions, this phenomenon only occurs for a small number of jets and hence does not constitute a significant threat to the performance of the calibration process. Due to the HL-LHC's harsher pileup conditions however, its impact will increase and suitable solutions have to be found. As suggested in Sec. 6.5.3, the addition of a constant term or a term concerning the pileup density $\langle \mu \rangle$ to the Gradient-Jet-Area correction might help to compensate this effect.

The non-closure of the reconstructed and true jet momentum after the absolute JES calibration is especially prominent for jets with low momentum using the current scheme of the ATLAS jet calibration. In the forward detector region the Gradient-Jet-Area correc-

tion appears to perform better in this regard. Since this method is especially impaired by jets having negative energy after pileup correction, a solution for the latter might further improve this methods performance with regards to closure.

Calorimeter-jets might eventually be replaced by more sophisticated jet definitions. One possibility for this are particle-flow jets [82]. Nevertheless, a consistent and optimized calibration procedure for EM-scale jets will remain an important goal for the ATLAS collaboration. The ability to compare the results from particle-flow jets with results from the already well-understood calorimeter-jets is desirable.

Appendix

A Dataset IDs

Sample	Dataset ID
VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$	345074+345075
ggH $H \to \tau\tau$	345121+345122
di-boson	361600-361611
W+jets	361100-361105
$Z \to \tau\tau$	361108
$Z \to \mu\mu$	361107
$Z \to ee$	361106
$t\bar{t}$	410470
single t	410011-410014
	410025+410026

Table A.1: Dataset IDs of the Monte Carlo (MC) generated samples used to model the signal (VBF $H \to \tau_{\text{lep}}\tau_{\text{had}}$) and background channels for the VBF $H \to \tau\tau$ prospects study presented in Sec. 5. For more information on the individual samples see Tab. 5.1.

Sample	Dataset ID
di-jet	147910-147917

Table A.2: Dataset IDs of the Monte Carlo (MC) generated samples used for the jet calibration studies presented in Sec. 6.

B Cut Optimization

| sample	no cut	== 1 l	≥ 1 τ_{had}	$|\eta(\tau)| < 2.4$	$Q(\tau) == -Q(l)$	SLT	$p_T(\mu) > 21$ GeV	$p_T(e) > 25$ GeV	$b-$ jet veto	≥ 2 jets
W+jets	190406377472 (0.00%)	76405350400 (55.03%)	1706740 (100.00%)	1681424 (1.43%)	1494513 (10.82%)	1318928 (10.78%)	1305011 (0.65%)	1280013 (2.58%)	699234 (45.31%)	78159 (89.29%)
di-boson	292280000 (0.00%)	119959120 (70.99%)	2403069 (94.60%)	2359078 (1.90%)	2189647 (14.93%)	1792699 (19.26%)	1767268 (1.54%)	1716958 (3.03%)	1319478 (22.69%)	315546 (76.28%)
$Z \to ee$	6172499456 (0.00%)	2355125248 (61.84%)	75491 (100.00%)	74523 (1.28%)	36422 (51.13%)	17179 (8.62%)	33282 (0.00%)	31247 (6.11%)	9224 (70.48%)	2429 (73.66%)
$Z \to \mu\mu$	6172498944 (0.00%)	2093962752 (66.08%)	42801 (100.00%)	42004 (1.86%)	20816 (50.44%)	17179 (17.47%)	16595 (3.40%)	16595 (0.00%)	5443 (67.20%)	451 (91.71%)
$Z \to \tau\tau$	6172498944 (0.00%)	1014645760 (83.56%)	169322816 (83.31%)	165995424 (1.97%)	165992592 (0.00%)	96592792 (41.81%)	91698208 (5.07%)	85571384 (6.68%)	68042832 (20.48%)	9578229 (85.92%)
ggH	775246 (0.00%)	683036 (11.89%)	281081 (58.86%)	277943 (1.12%)	277941 (2.59%)	215561 (22.44%)	209967 (2.59%)	199443 (5.01%)	154597 (22.49%)	44497 (71.22%)
$t\bar{t}$	1345204864 (0.00%)	762862912 (43.29%)	24480556 (96.80%)	24180176 (0.82%)	21717770 (10.18%)	18717136 (13.82%)	18382002 (1.79%)	17849966 (2.89%)	1505174 (91.57%)	914823 (39.22%)
single t	477331392 (0.00%)	207967600 (59.58%)	2682775 (98.38%)	2660168 (0.84%)	2309134 (11.78%)	1978655 (14.15%)	1949706 (1.47%)	1896613 (2.68%)	346408 (81.24%)	131846 (61.91%)
VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	64919 (0.00%)	57516 (11.40%)	24549 (57.32%)	24288 (1.06%)	24287 (0.00%)	19362 (20.28%)	18941 (2.17%)	18135 (4.26%)	13439 (25.89%)	8304 (38.21%)

VBF signal region cuts

sample	$p_T(j^{\mathrm{lead}}) > c_1$	$p_T(j^{\mathrm{sublead}}) > c_2$	$m_{\mathrm{MMC}} > c_3$	$m_{\mathrm{MMC}} < c_4$	$m(j^{\mathrm{lead}}, j^{\mathrm{sublead}}) > c_5$	$|\Delta\eta(j^{\mathrm{lead}}, j^{\mathrm{sublead}})| > c_6$
W+jets	22704 (69.73%)	22704 (0.00%)	21590 (3.80%)	21590 (0.00%)	987 (94.74%)	834 (25.00%)
di-boson	187871 (41.29%)	187871 (0.00%)	152461 (16.99%)	152461 (0.00%)	11155 (92.75%)	7948 (26.83%)
$Z \to ee$	988 (59.35%)	988 (0.00%)	988 (0.00%)	988 (0.00%)	99 (90.00%)	79 (20.00%)
$Z \to \mu\mu$	159 (64.71%)	159 (0.00%)	159 (0.00%)	159 (0.00%)	0 (100.00%)	0 (0.00%)
$Z \to \tau\tau$	3308892 (65.45%)	3308892 (0.00%)	1888457 (42.93%)	1888457 (0.00%)	120584 (93.61%)	101676 (15.68%)
ggH	23658 (46.83%)	23658 (0.00%)	18051 (23.69%)	18051 (0.00%)	1829 (89.87%)	1521 (16.83%)
$t\bar{t}$	615941 (32.67%)	615941 (0.00%)	546348 (11.30%)	546348 (0.00%)	47820 (91.25%)	35835 (25.06%)
single t	78262 (40.65%)	78262 (0.00%)	70583 (10.02%)	70583 (0.00%)	4731 (93.68%)	3397 (28.80%)
VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	6334 (23.72%)	6334 (0.00%)	5080 (19.79%)	5080 (0.00%)	2411 (52.55%)	2322 (3.70%)

Table B.1: Event yields and the corresponding cut efficiencies in percent for signal (VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$) and background channels after each individual cut of preselection and VBF signal region at HL-LHC conditions with an integrated luminosity of 3000 fb^{-1}. For a graphical representaion see Fig. 5.14.
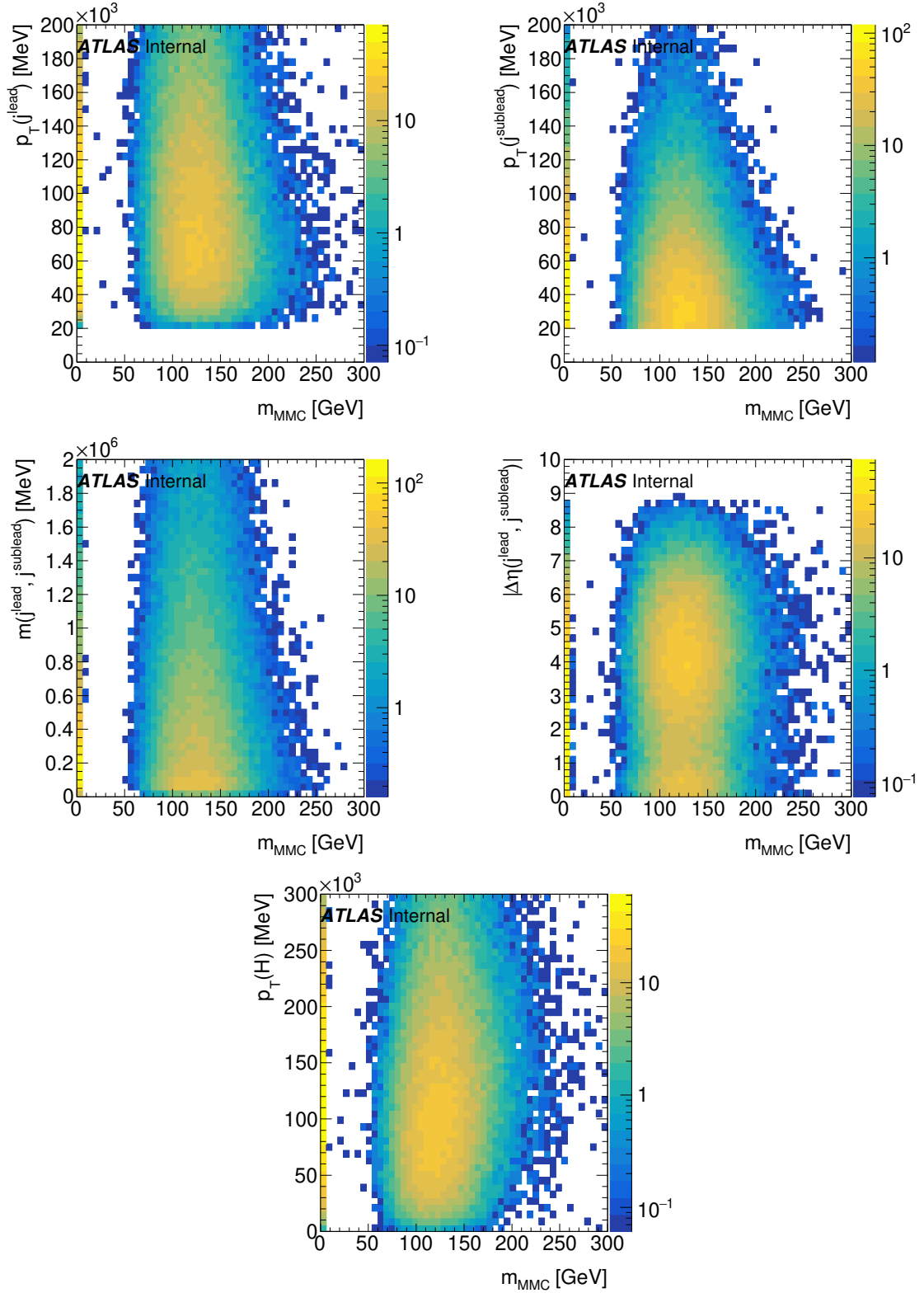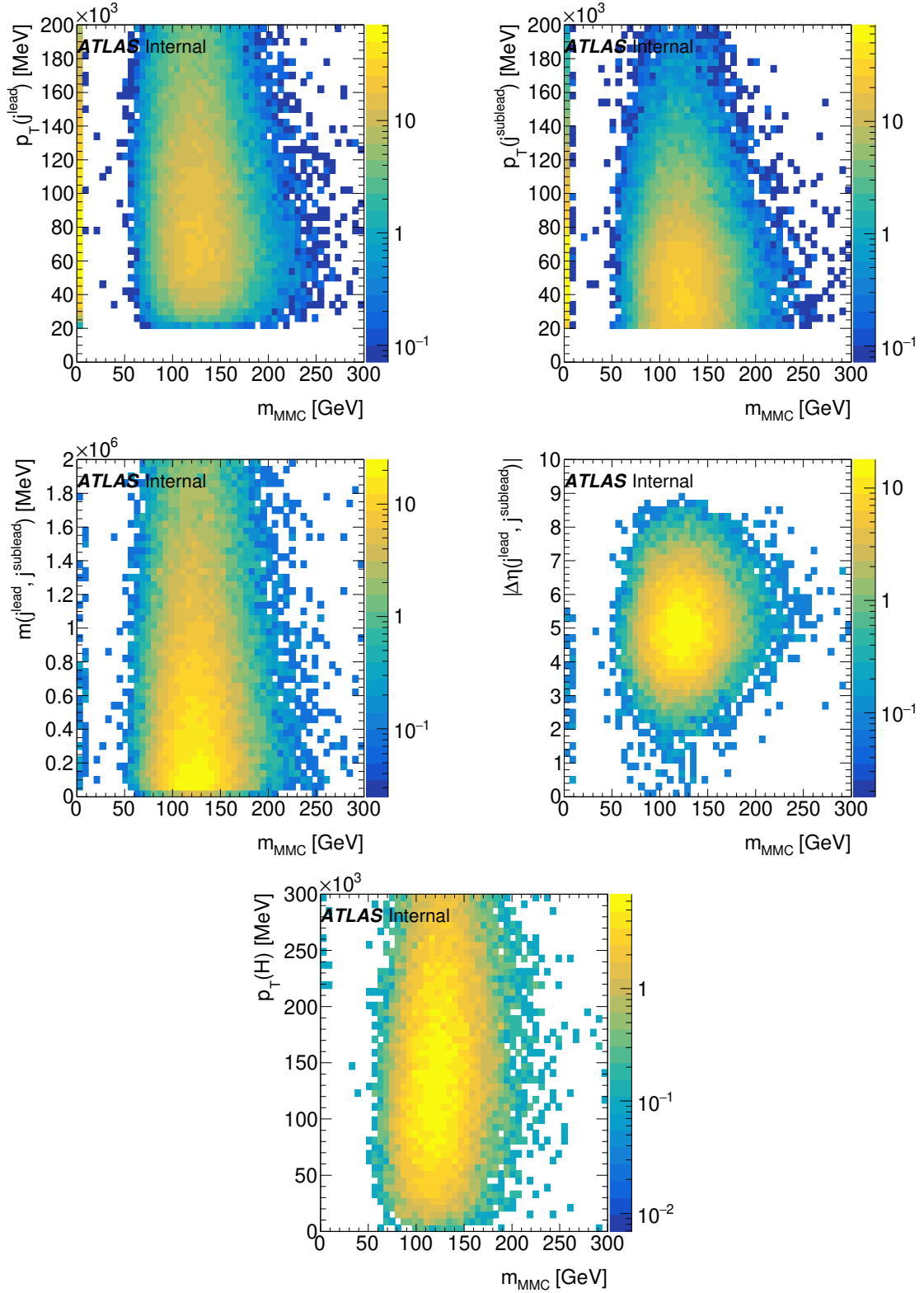
Figure B.1: Signal distributions after preselection with respect to one of the signal region cut variables x_i versus the reconstructed di-τ mass m_{MMC}. The resulting efficiencies of signal and background versus the cut value c_i are shown in Fig. 5.10.
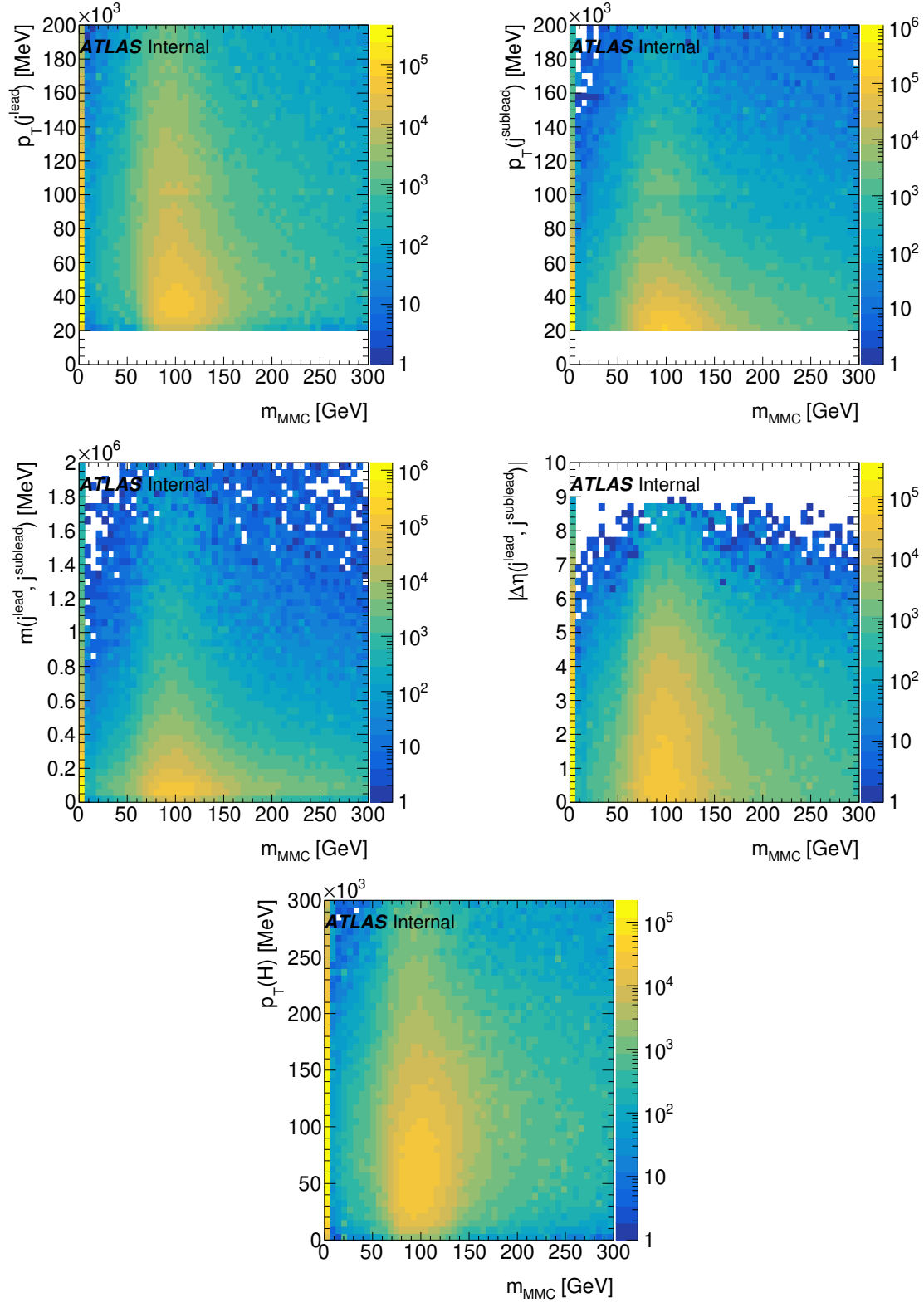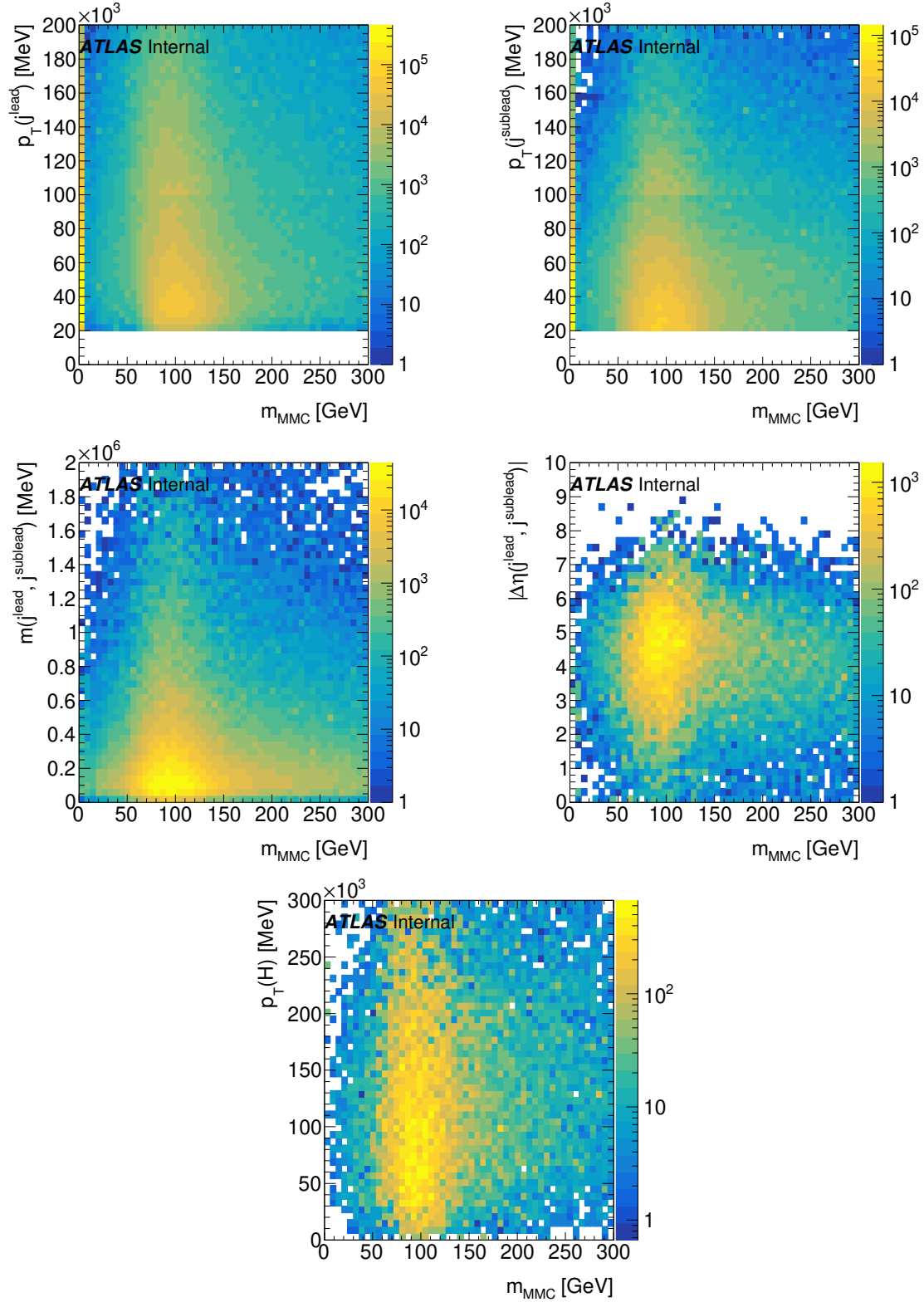
Figure B.2: Background distributions after preselection with respect to one of the signal region cut variables x_i versus the reconstructed di-τ mass m_{MMC}. The resulting efficiencies of signal and background versus the cut value c_i are shown in Fig. 5.10.

Figure B.3: Signal distributions after preselection with respect to one of the signal region cut variables x_i versus the reconstructed di-τ mass m_{MMC}. The resulting efficiencies of signal and background versus the cut value c_i are shown in Fig. 5.10.

Figure B.4: Background distributions after preselection with respect to one of the signal region cut variables x_i versus the reconstructed di-τ mass m_{MMC}. The resulting efficiencies of signal and background versus the cut value c_i are shown in Fig. 5.10.
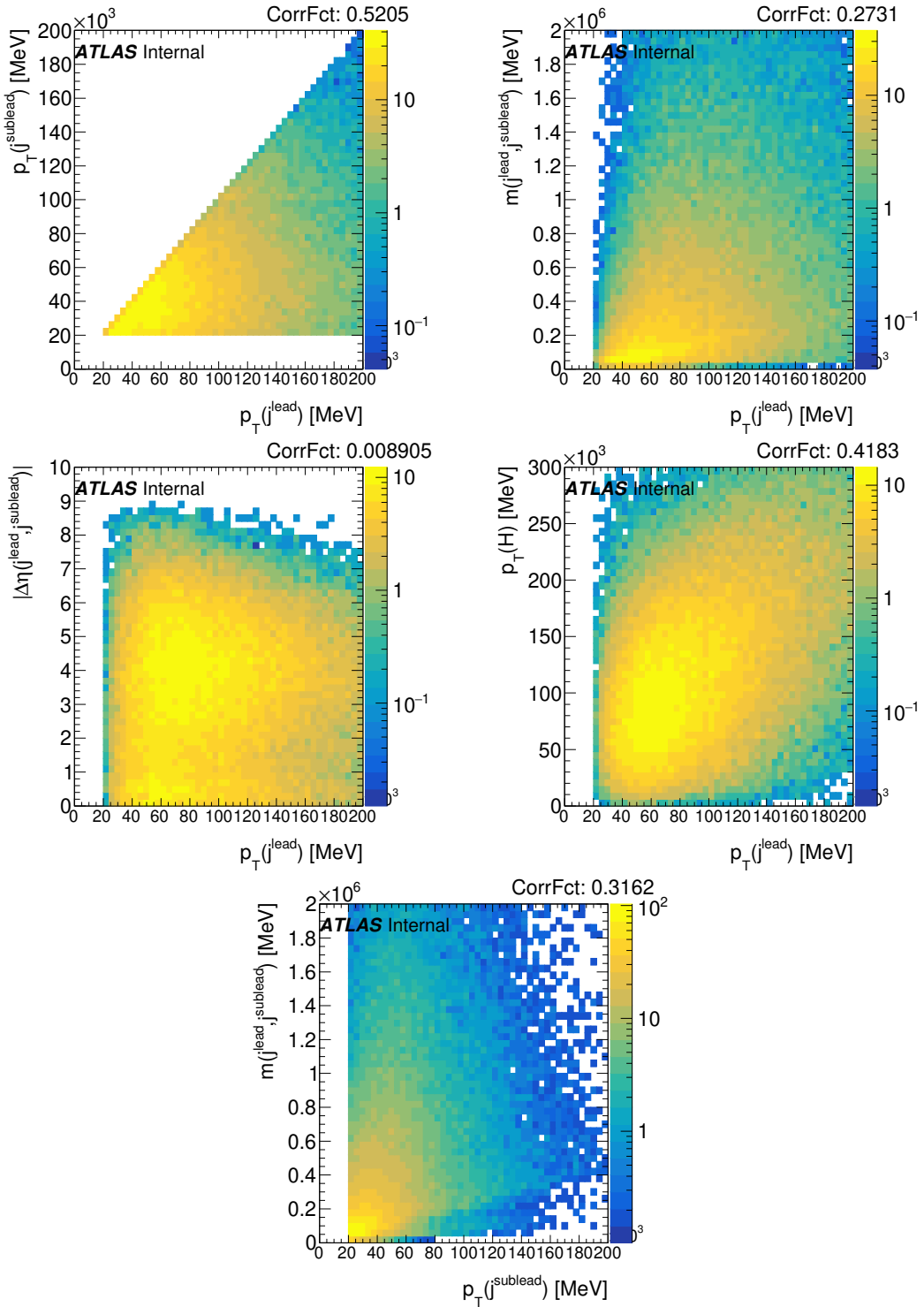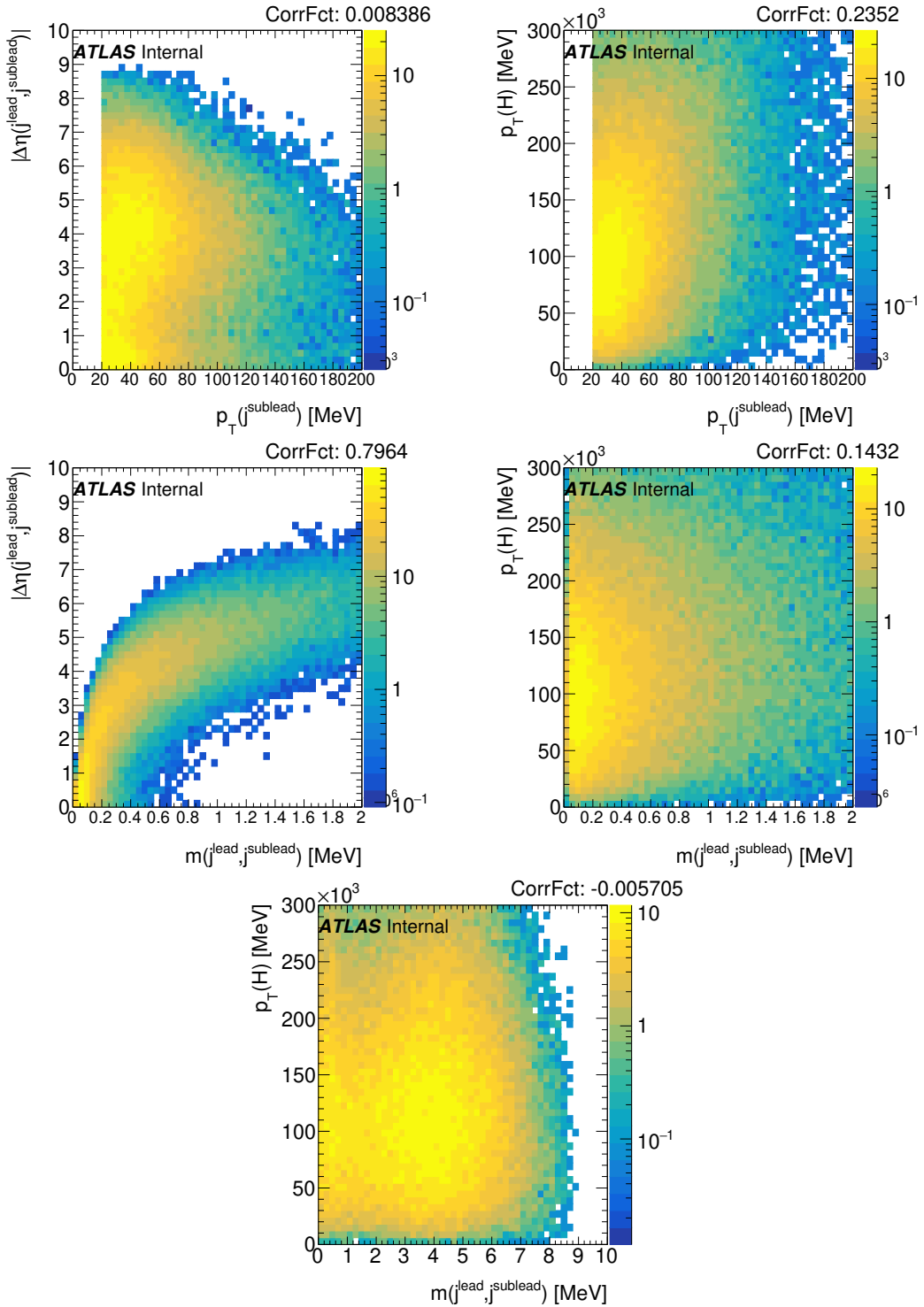
Figure B.5: Correlations of some of the signal region cut parameters x_i to each other using only the signal sample. The discontinuities at 50 and 100 GeV in the distributions of $p_T(j^{\text{lead}})$ and $p_T(j^{\text{sublead}})$ are an artifact of the simulated track confiramtion efficiencies which is further explained in Sec. 5.2.3.
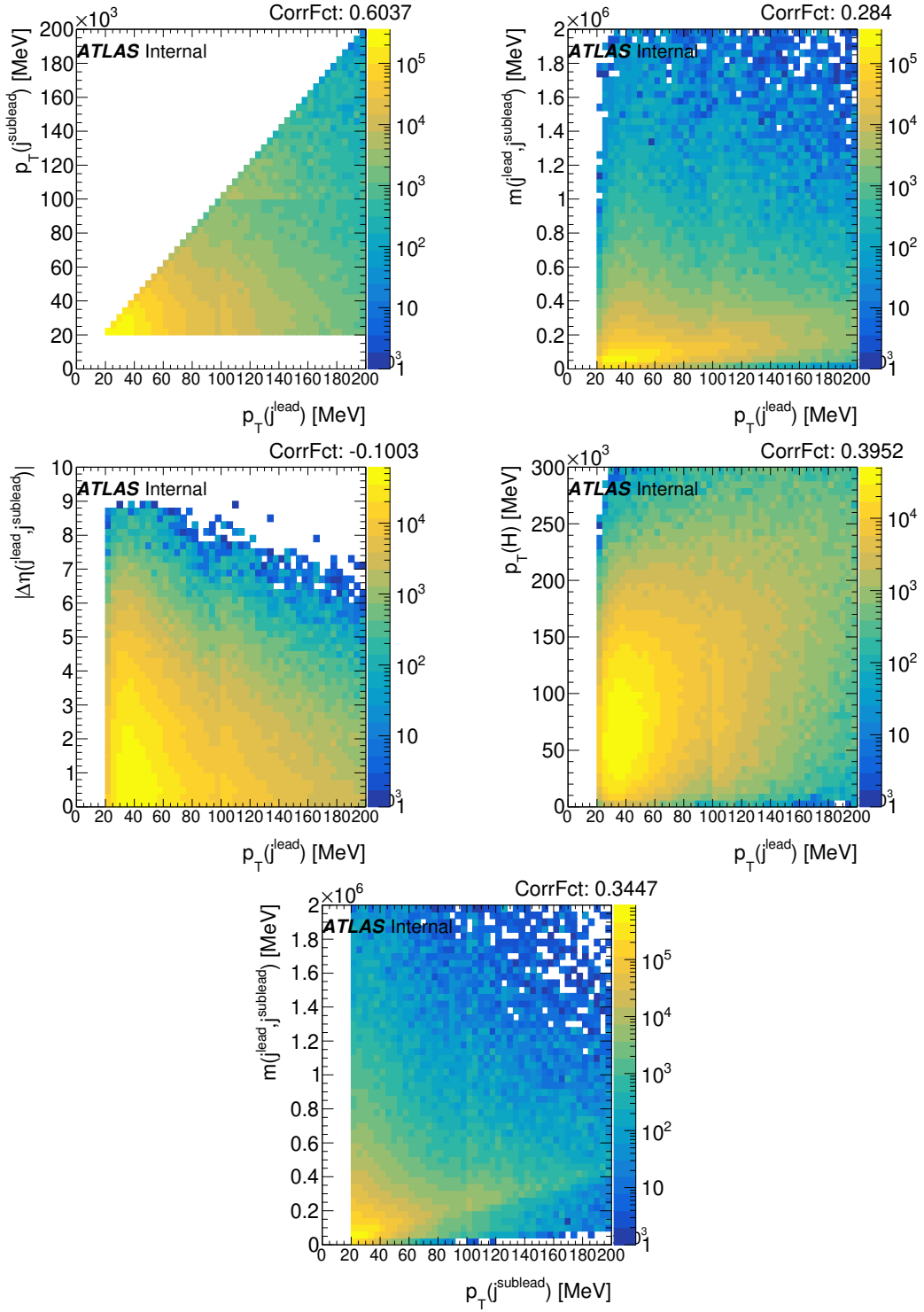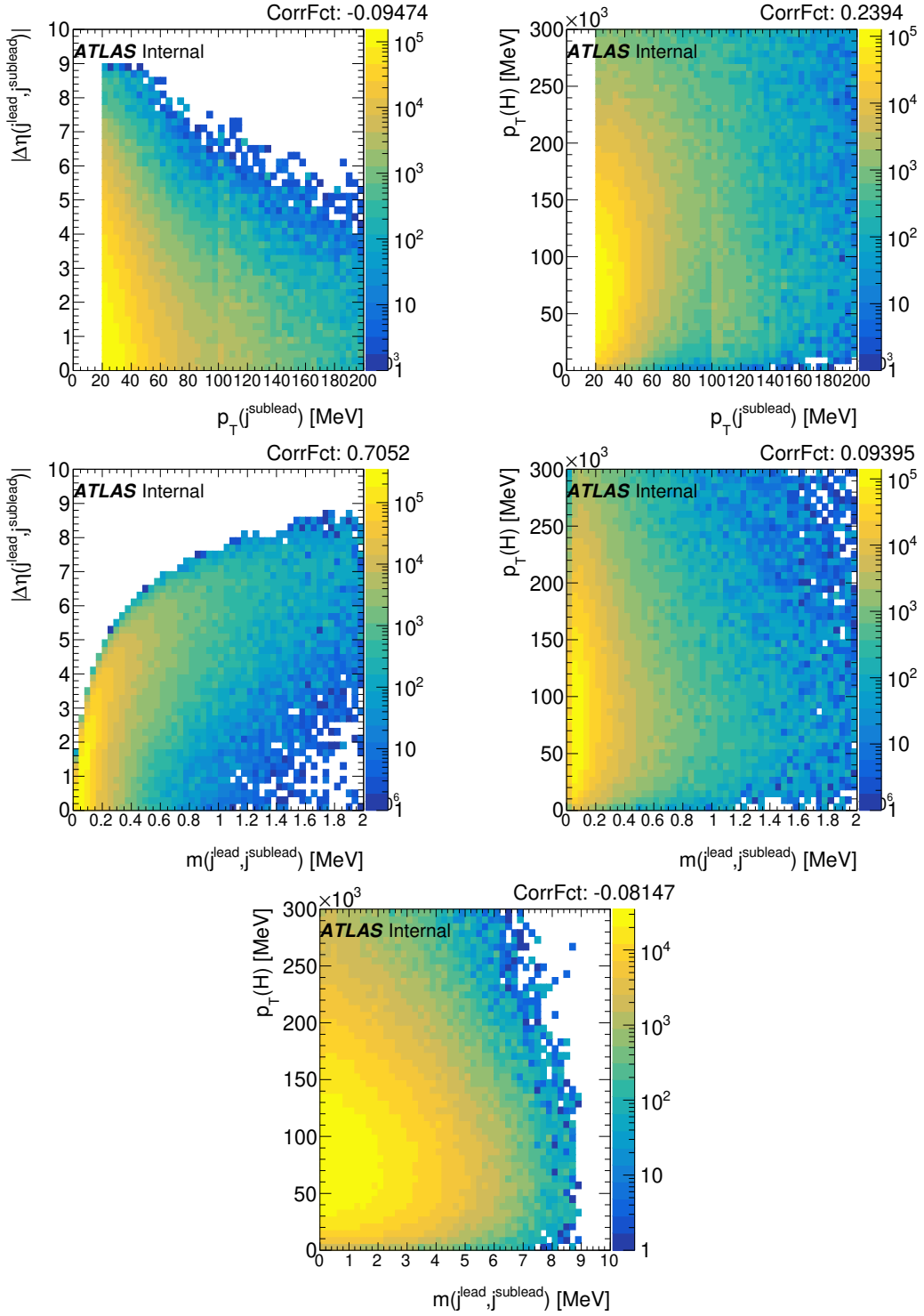
Figure B.6: Correlations of some of the signal region cut parameters x_i to each other using combination of all background samples. The discontinuities at 50 and 100 GeV in the distributions of $p_T(j^{\text{lead}})$ and $p_T(j^{\text{sublead}})$ are an artifact of the simulated track confiramtion efficiencies which is further explained in Sec. 5.2.3.

Figure B.7: Correlations of some of the signal region cut parameters x_i to each other using only the signal sample. The discontinuities at 50 and 100 GeV in the distributions of $p_T(j^{\text{lead}})$ and $p_T(j^{\text{sublead}})$ are an artifact of the simulated track confiramtion efficiencies which is further explained in Sec. 5.2.3.

Figure B.8: Correlations of some of the signal region cut parameters x_i to each other using combination of all background samples. The discontinuities at 50 and 100 GeV in the distributions of $p_{\mathrm{T}}(j^{\mathrm{lead}})$ and $p_{\mathrm{T}}(j^{\mathrm{sublead}})$ are an artifact of the simulated track confiramtion efficiencies which is further explained in Sec. 5.2.3.

C Correlation of $\triangle p_{\mathbf{T}}$ to ρA, $N_{\mathbf{PV}}$, and $\langle\mu\rangle$
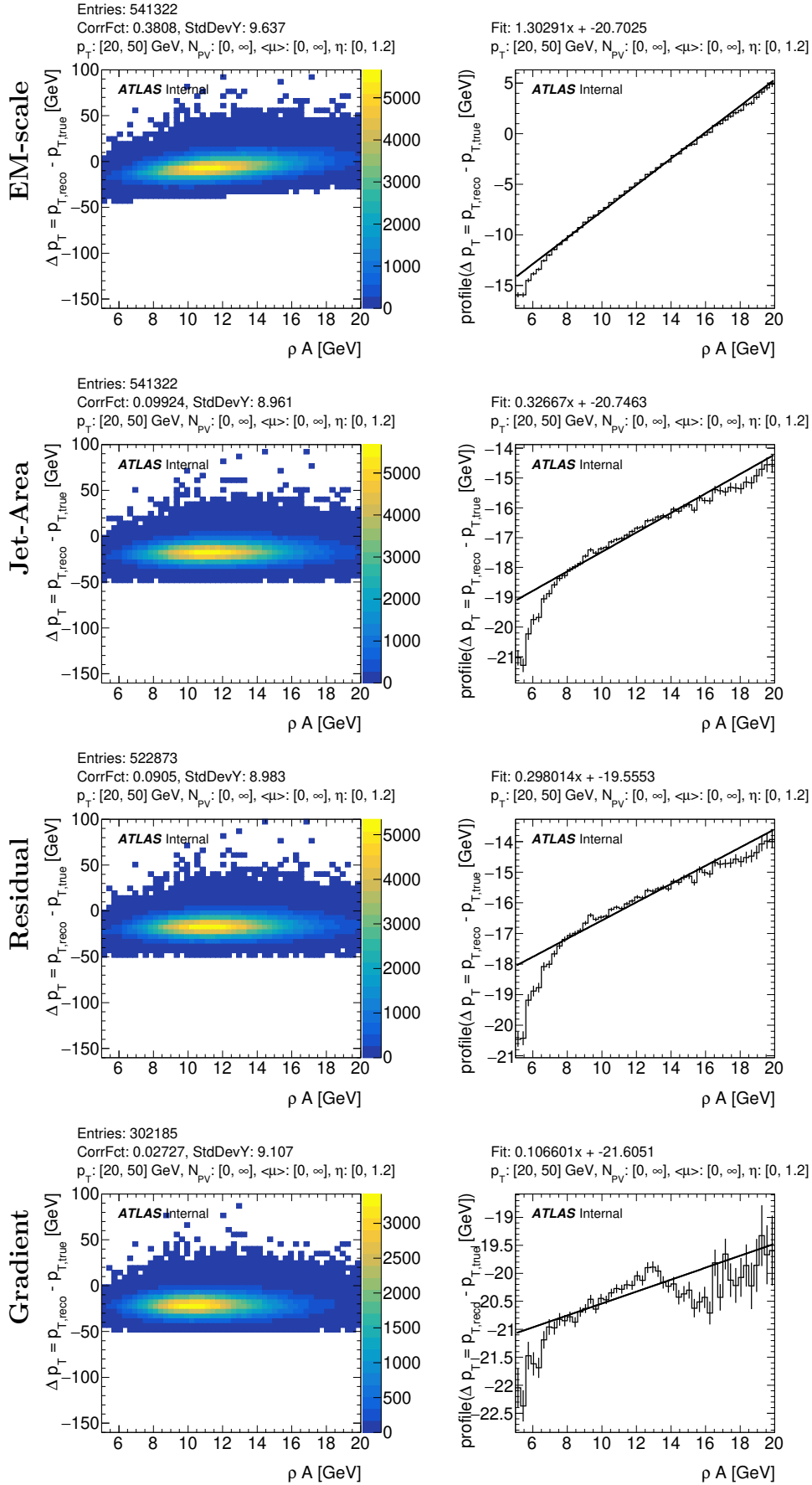
Figure C.1: Δp_{T} with respect to ρA in the central η region.

Figure C.2: Δp_T with respect to ρA in the medium η region.

Entries: 118133
CorrFct: 0.4484, StdDevY: 10.46
p_T: [20, 50] GeV, N_{PV}: [0, ∞], $<\mu>$: [0, ∞], η: [3.6, 4.5]

Fit: 1.69643x + -28.6259
p_T: [20, 50] GeV, N_{PV}: [0, ∞], $<\mu>$: [0, ∞], η: [3.6, 4.5]

Entries: 118133
CorrFct: 0.2074, StdDevY: 9.559
p_T: [20, 50] GeV, N_{PV}: [0, ∞], $<\mu>$: [0, ∞], η: [3.6, 4.5]

Fit: 0.718715x + -28.6528
p_T: [20, 50] GeV, N_{PV}: [0, ∞], $<\mu>$: [0, ∞], η: [3.6, 4.5]

Entries: 108775
CorrFct: 0.07566, StdDevY: 9.246
p_T: [20, 50] GeV, N_{PV}: [0, ∞], $<\mu>$: [0, ∞], η: [3.6, 4.5]

Fit: 0.244979x + -7.2797
p_T: [20, 50] GeV, N_{PV}: [0, ∞], $<\mu>$: [0, ∞], η: [3.6, 4.5]

Entries: 108625
CorrFct: 0.05371, StdDevY: 8.67
p_T: [20, 50] GeV, N_{PV}: [0, ∞], $<\mu>$: [0, ∞], η: [3.6, 4.5]

Fit: 0.163222x + -26.2953
p_T: [20, 50] GeV, N_{PV}: [0, ∞], $<\mu>$: [0, ∞], η: [3.6, 4.5]

Figure C.3: Δp_T with respect to ρA in the forward η region.
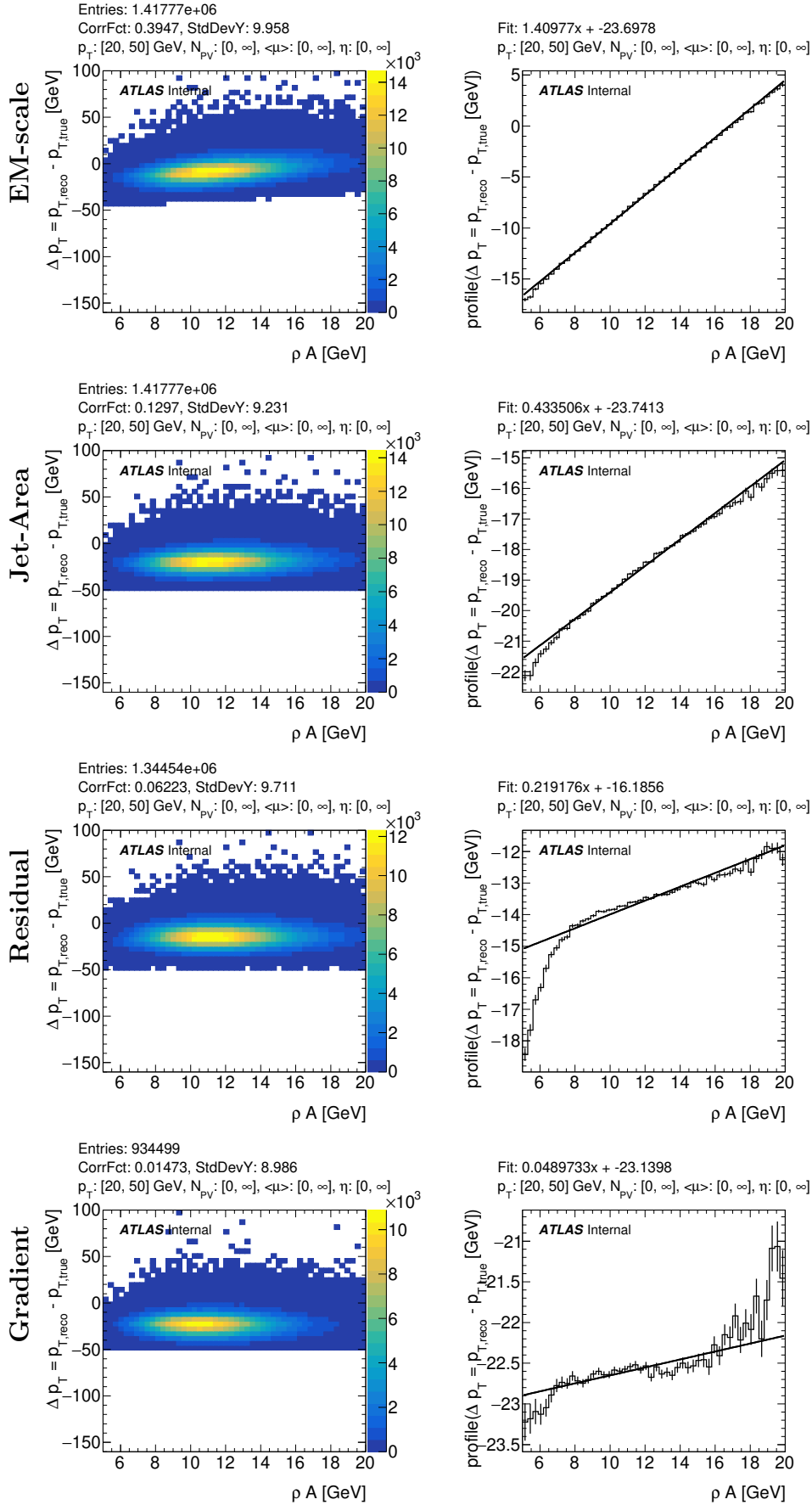
97
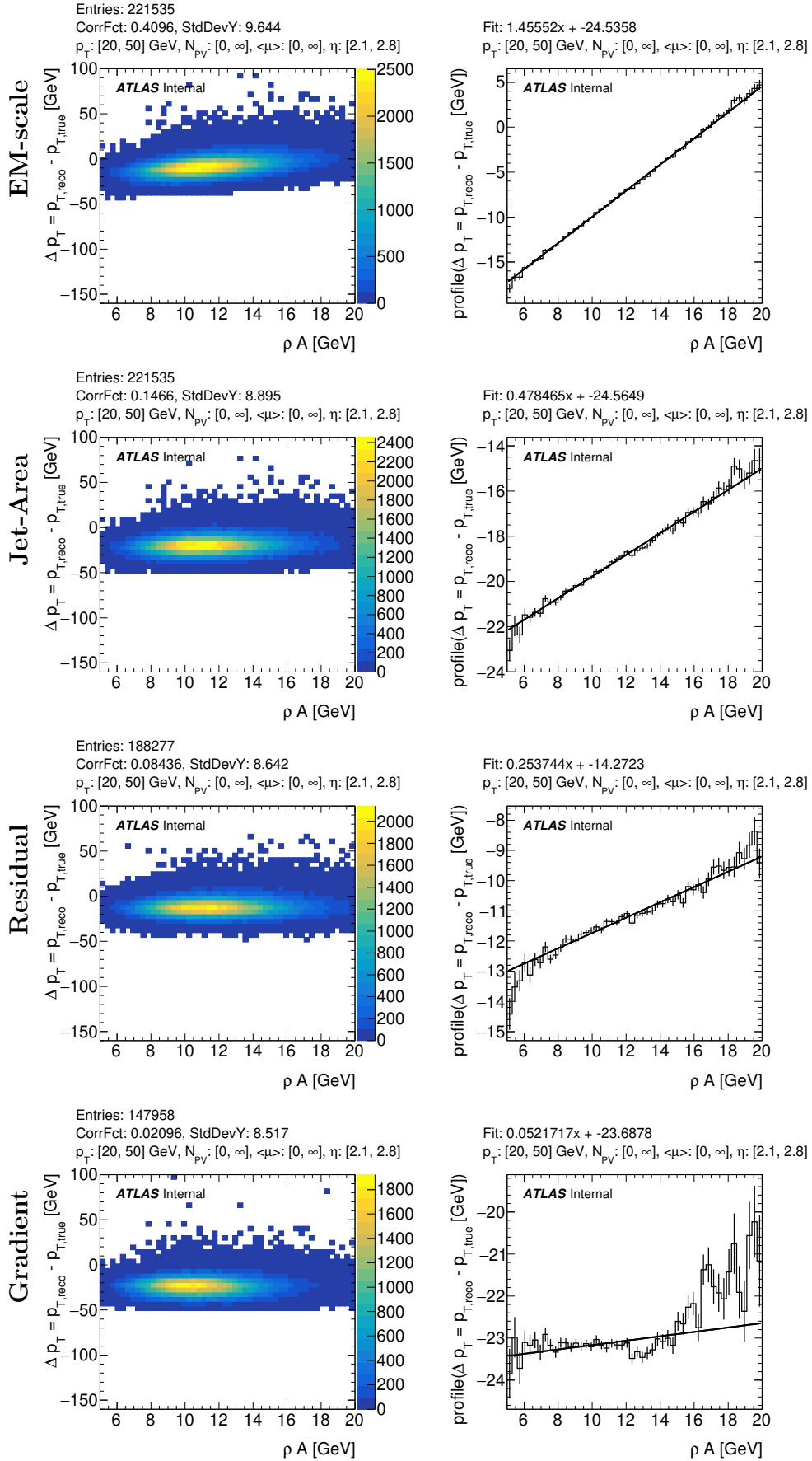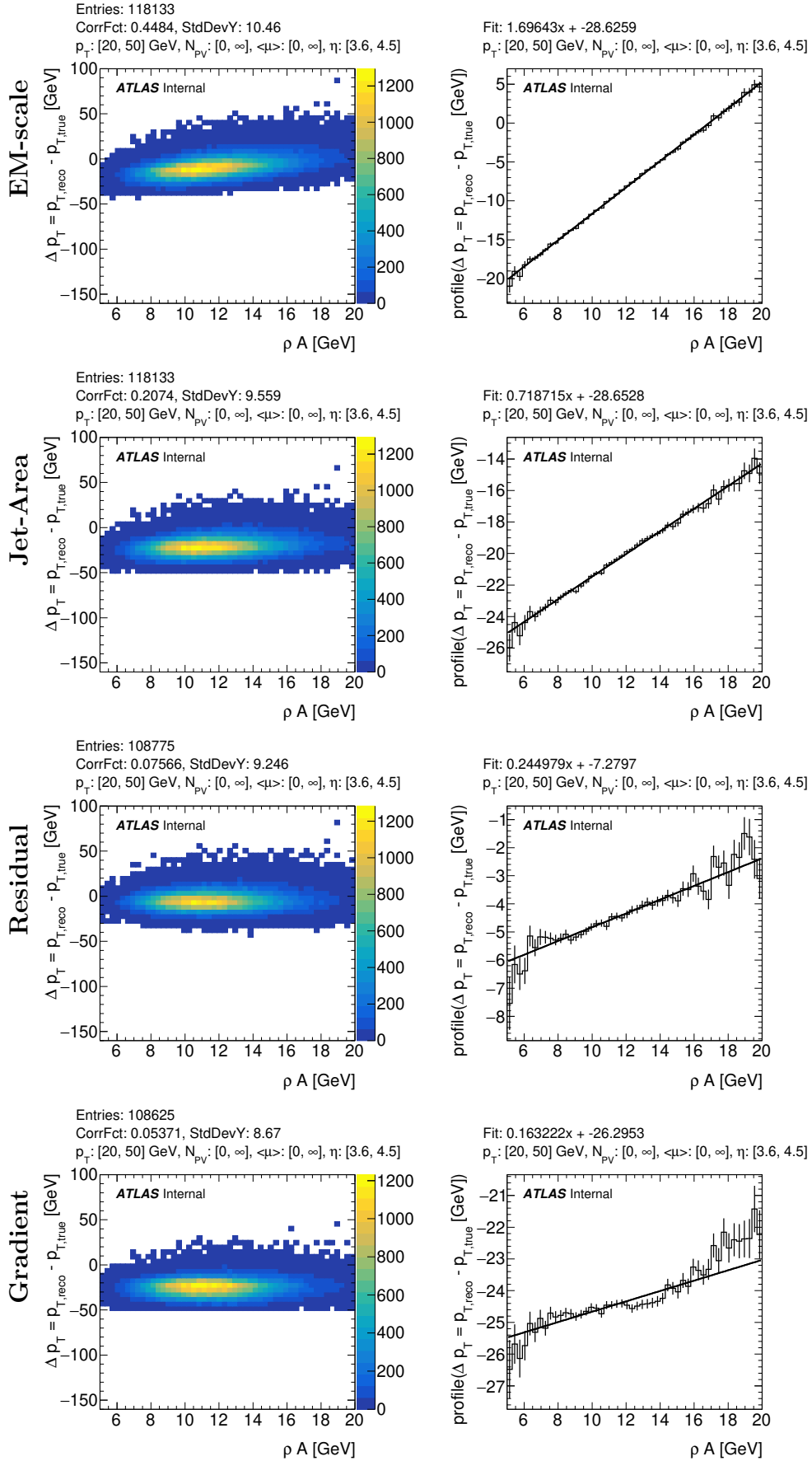
Figure C.4: Δp_{T} with respect to ρA inclusive in η.
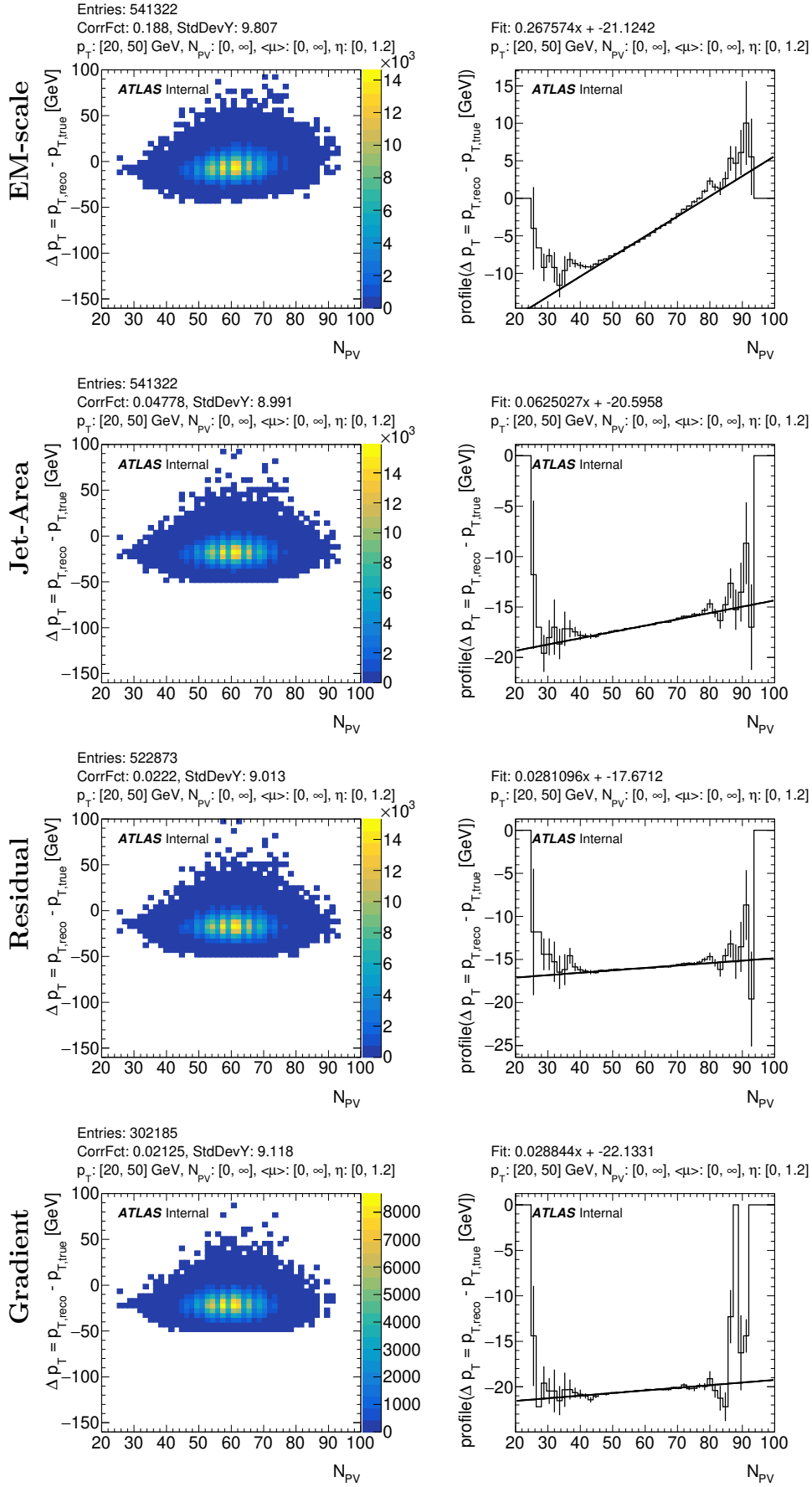
Figure C.5: Δp_T with respect to N_PV in the central η region.
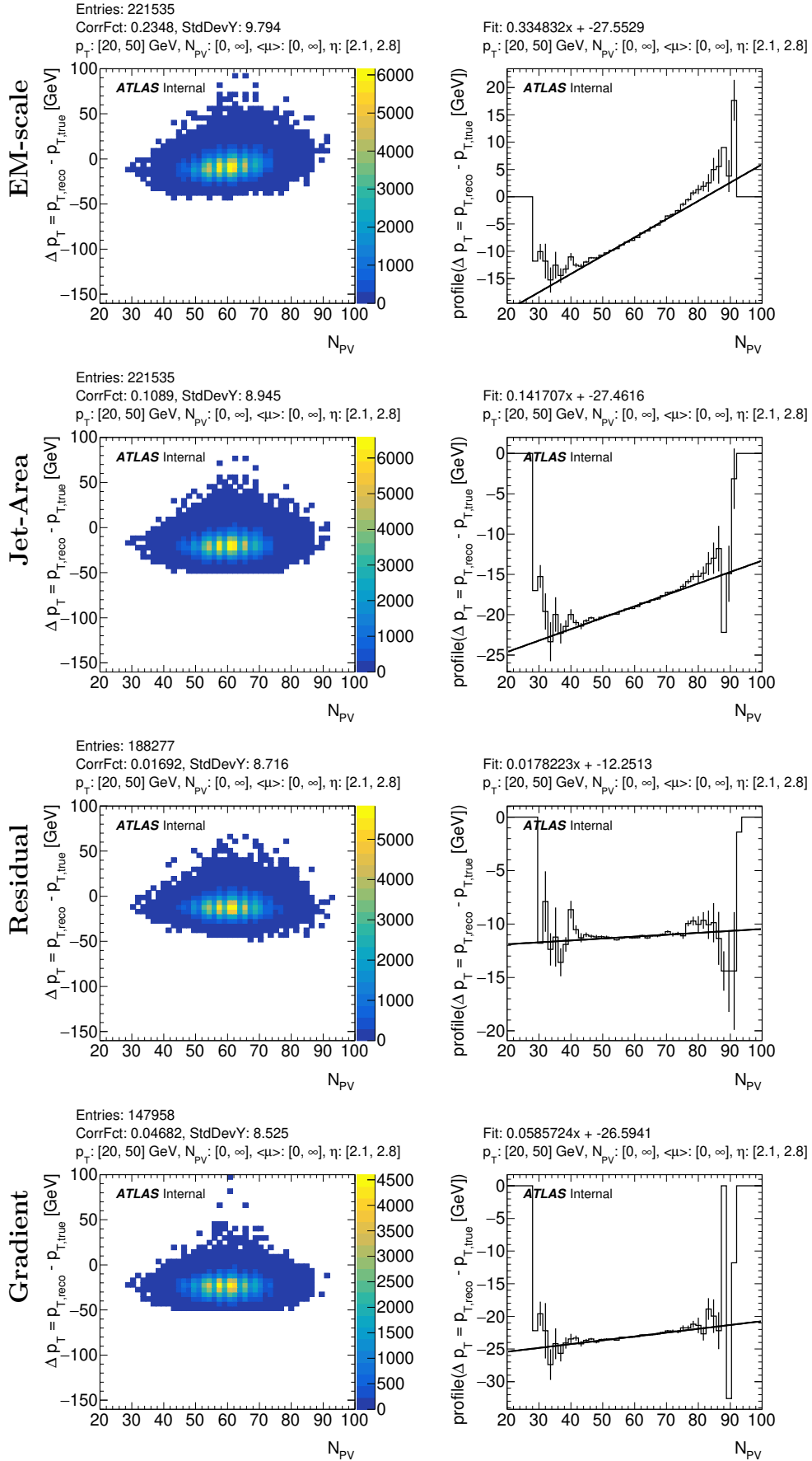
Figure C.6: Δp_T with respect to N_{PV} in the medium η region.

EM-scale

Entries: 118133
CorrFct: 0.3, StdDevY: 10.66
p_T: [20, 50] GeV, N_{PV}: [0, ∞], <μ>: [0, ∞], η: [3.6, 4.5]

Fit: 0.466745x + -36.4883
p_T: [20, 50] GeV, N_{PV}: [0, ∞], <μ>: [0, ∞], η: [3.6, 4.5]

Jet-Area

Entries: 118133
CorrFct: 0.1886, StdDevY: 9.617
p_T: [20, 50] GeV, N_{PV}: [0, ∞], <μ>: [0, ∞], η: [3.6, 4.5]

Fit: 0.265418x + -36.1015
p_T: [20, 50] GeV, N_{PV}: [0, ∞], <μ>: [0, ∞], η: [3.6, 4.5]

Residual

Entries: 108775
CorrFct: 0.04289, StdDevY: 9.302
p_T: [20, 50] GeV, N_{PV}: [0, ∞], <μ>: [0, ∞], η: [3.6, 4.5]

Fit: 0.0537585x + -7.56516
p_T: [20, 50] GeV, N_{PV}: [0, ∞], <μ>: [0, ∞], η: [3.6, 4.5]

Gradient

Entries: 108625
CorrFct: 0.1085, StdDevY: 8.681
p_T: [20, 50] GeV, N_{PV}: [0, ∞], <μ>: [0, ∞], η: [3.6, 4.5]

Fit: 0.138728x + -32.7744
p_T: [20, 50] GeV, N_{PV}: [0, ∞], <μ>: [0, ∞], η: [3.6, 4.5]

Figure C.7: Δp_T with respect to N_{PV} in the forward η region.

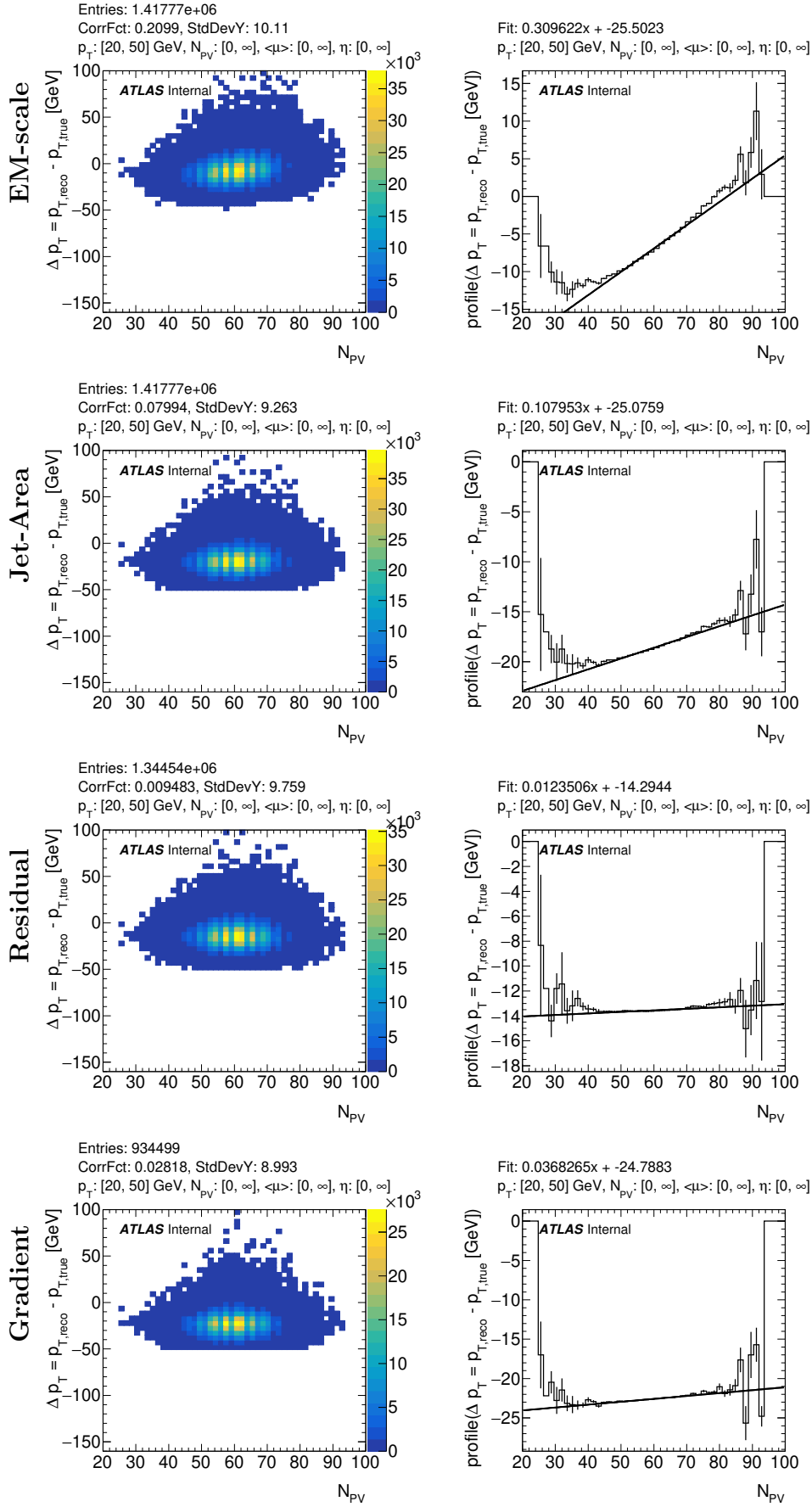***Figure C.8:*** Δp_{T} with respect to N_{PV} inclusive in η.
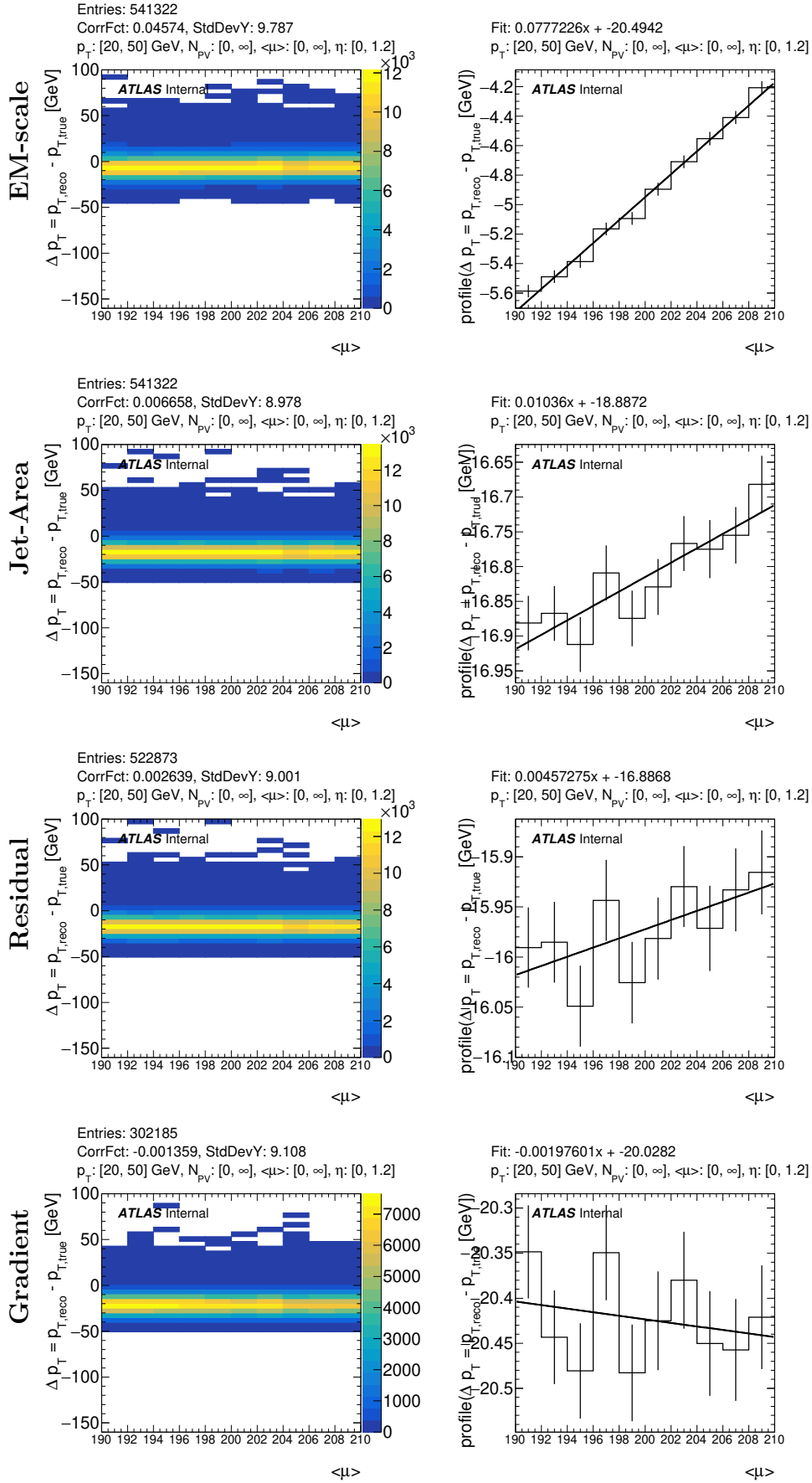
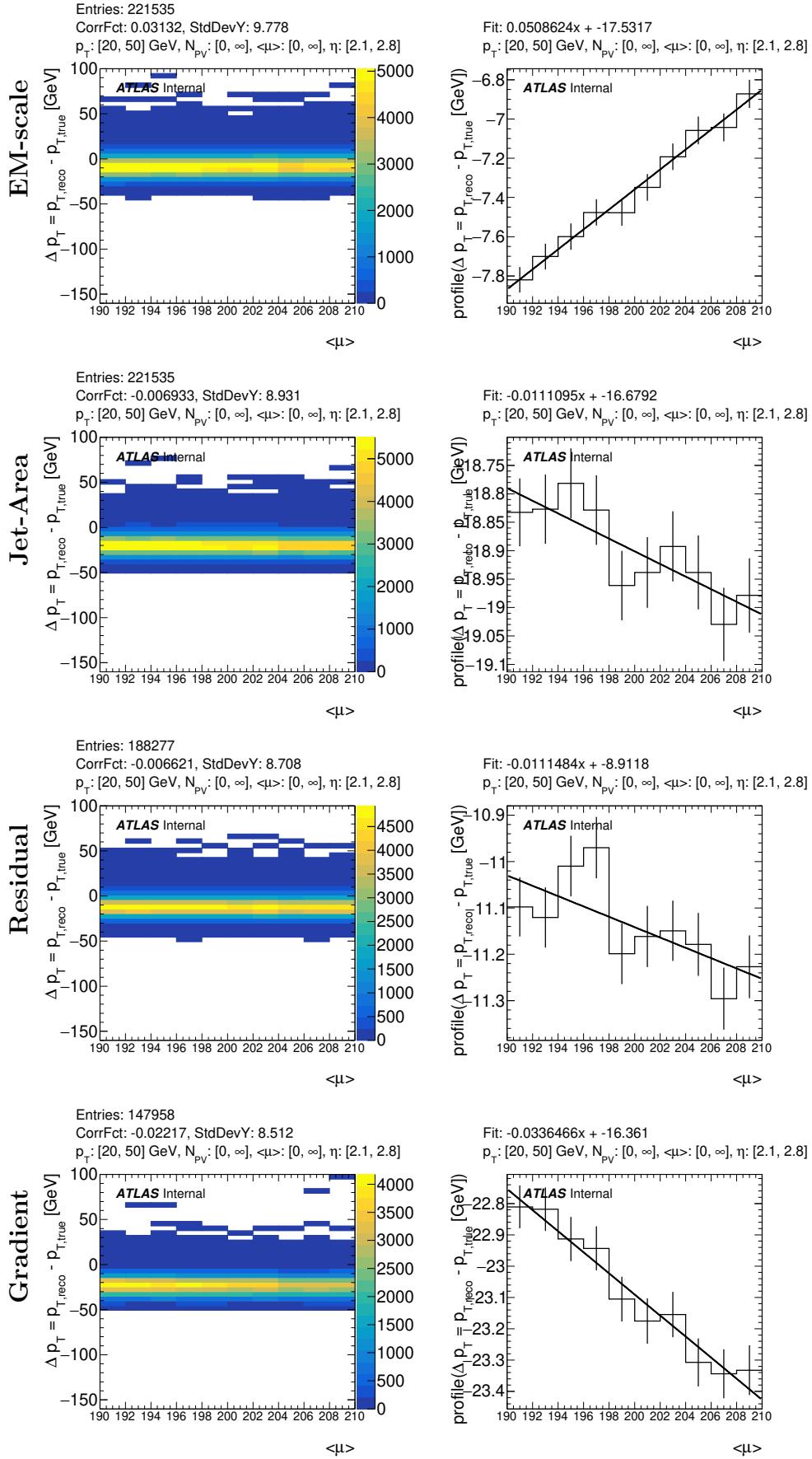**Figure C.9:** Δp_T with respect to $\langle \mu \rangle$ in the central η region.

Figure C.10: Δp_T with respect to $\langle \mu \rangle$ in the medium η region.
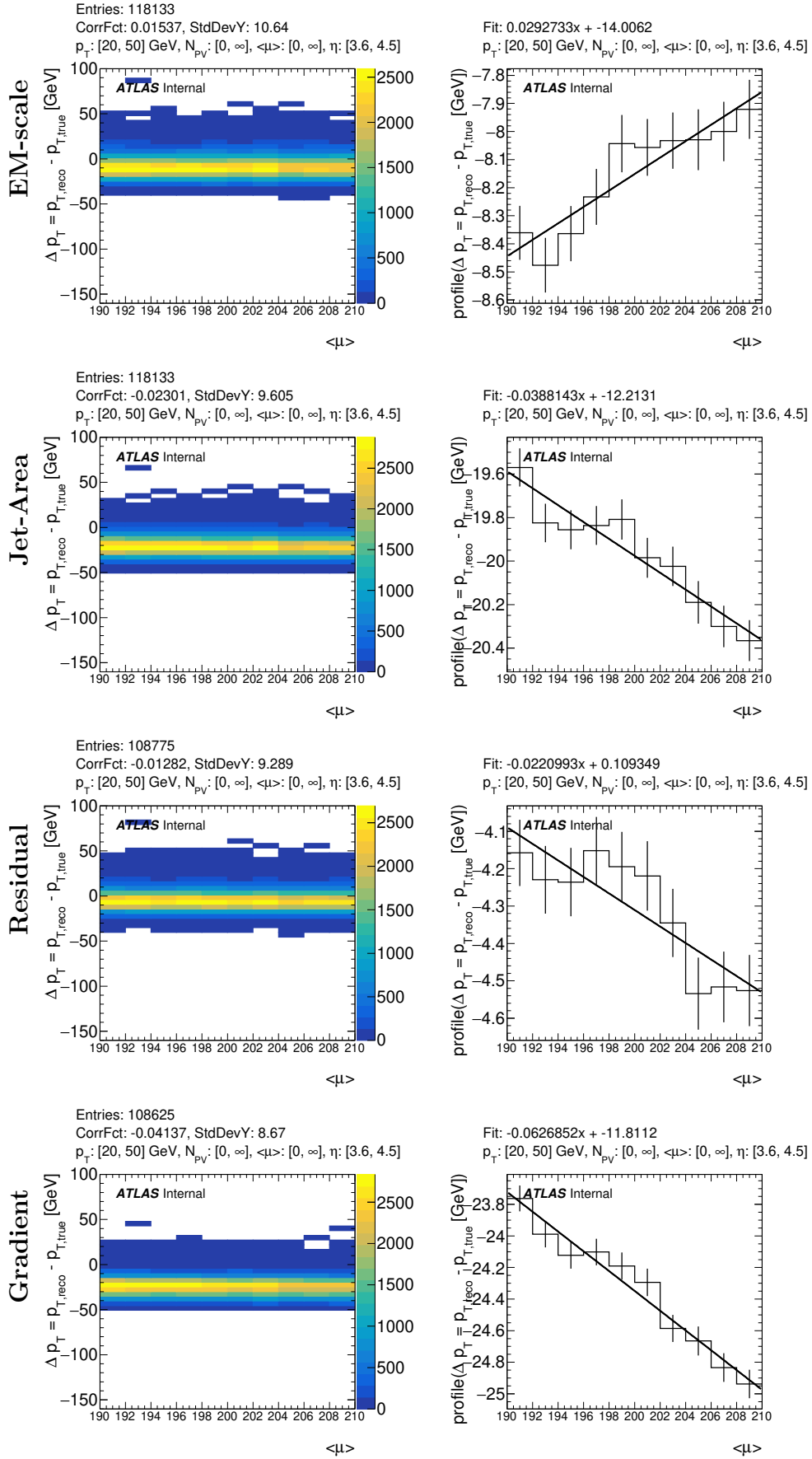
Figure C.11: Δp_{T} with respect to $\langle \mu \rangle$ in the forward η region.
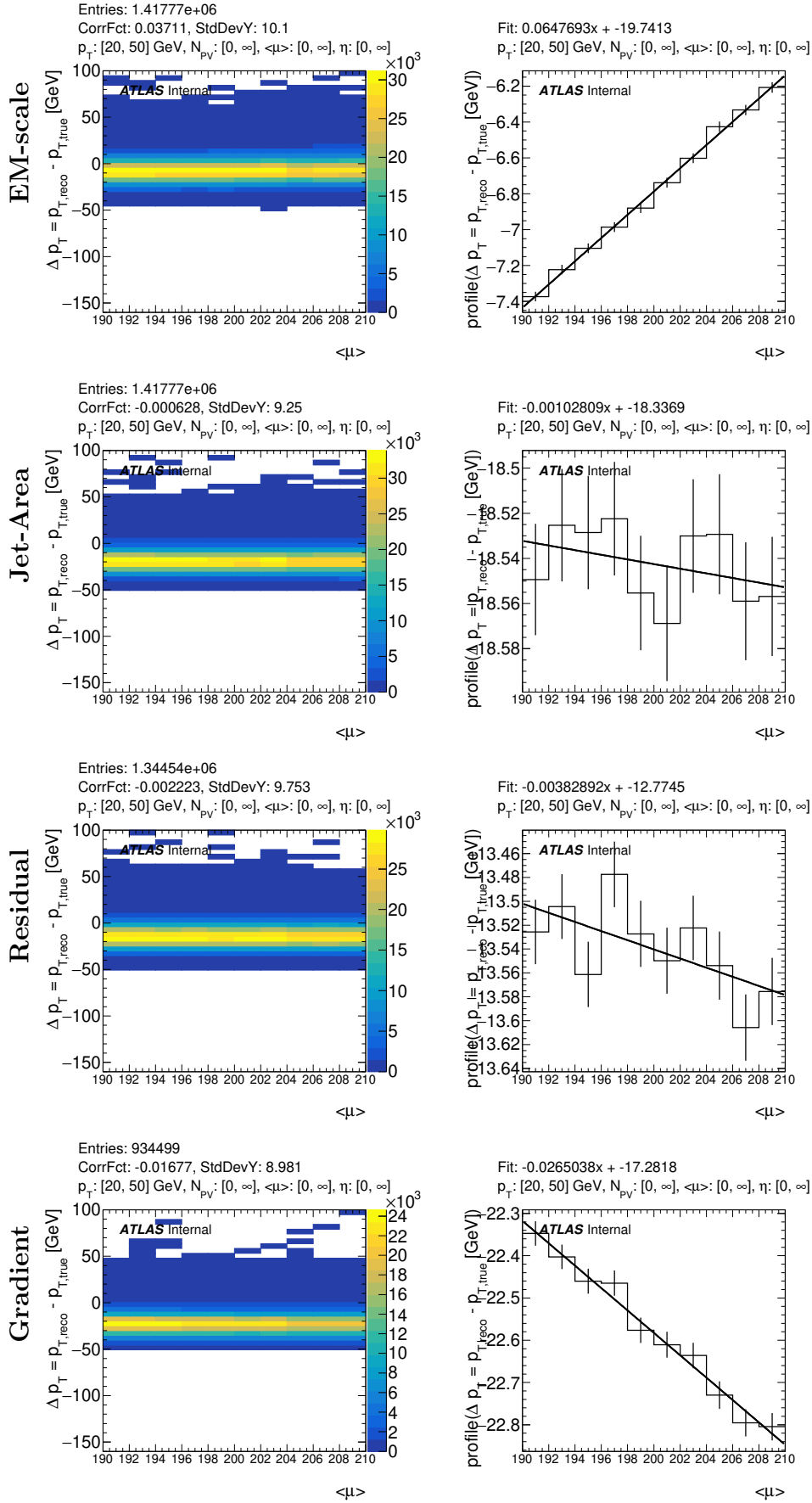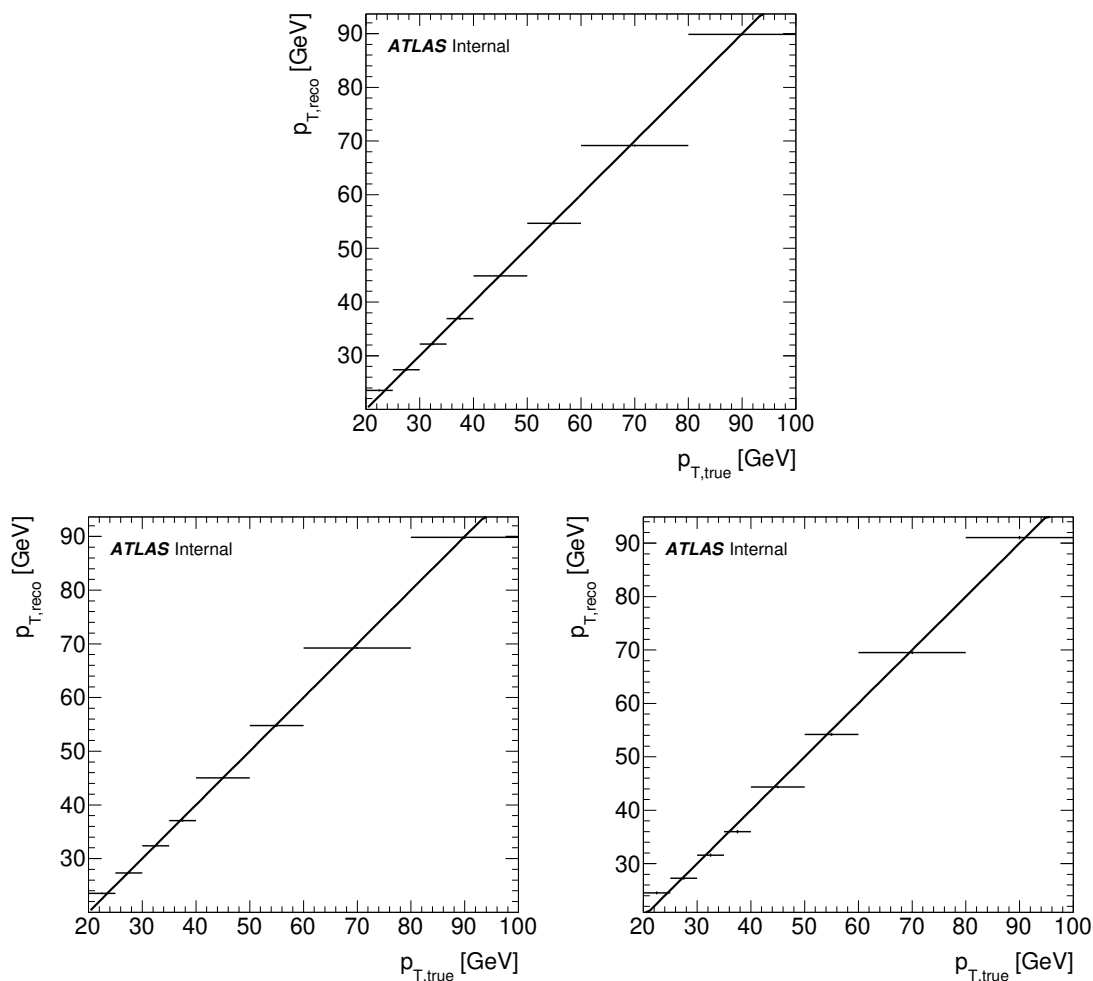
Figure C.12: Δp_T with respect to $\langle \mu \rangle$ inclusive in η.

D Closure Plots

Figure D.1: Closure plots of the transverse momentum of reconstructed and calibrated jets ($p_{\mathrm{T,reco}}$) with respect to truth level jets ($p_{\mathrm{T,reco}}$). Pileup calibration and absolute JES calibration have been performed. Each plot corresponds to a different method of pileup calibration that was used: Jet-Area [top], Residual [left], and Gradient-Jet-Area [right]. The black line corresponds to $p_{\mathrm{T,reco}} = p_{\mathrm{T,true}}$. Only jet pairs ($j_{\mathrm{reco}}$, j_{true}) with $0.0 < |\eta(j_{\mathrm{true}})| < 0.1$ were used. This corresponds to a central η region. Fig. 6.10 and Fig. D.2 show the corresponding plots for a forward- and a medium η region.
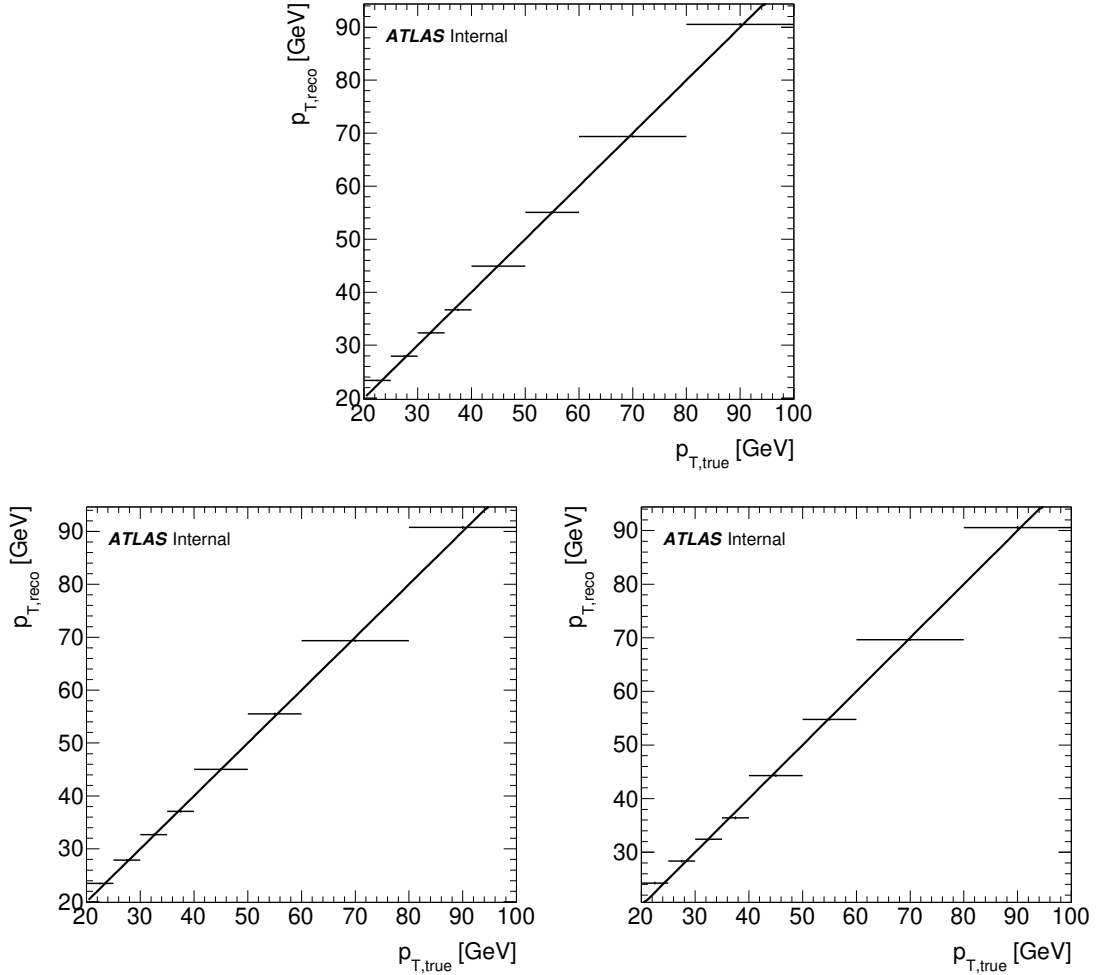
Figure D.2: Closure plots of the transverse momentum of reconstructed and calibrated jets ($p_{\mathrm{T,reco}}$) with respect to truth level jets ($p_{\mathrm{T,reco}}$). Pileup calibration and absolute JES calibration have been performed. Each plot corresponds to a different method of pileup calibration that was used: Jet-Area [top], Residual [left], and Gradient-Jet-Area [right]. The black line corresponds to $p_{\mathrm{T,reco}} = p_{\mathrm{T,true}}$. Only jet pairs ($j_{\mathrm{reco}}$, j_{true}) with $2.0 < |\eta(j_{\mathrm{true}})| < 2.1$ were used. This corresponds to a medium η region. Fig. D.1 and Fig. 6.10 show the corresponding plots for a central- and a forward η region.

Bibliography

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716**, 1 (2012), `1207.7214`

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716**, 30 (2012)

[3] Particle Data Group, *Review of Particle Physics*, Chin. Phys. **C40(10)**, 100001 (2016)

[4] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Strip Detector*, Technical Report CERN-LHCC-2017-005. ATLAS-TDR-025, CERN, Geneva (2017), URL `http://cds.cern.ch/record/2257755`

[5] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, Technical Report CERN-LHCC-2017-021. ATLAS-TDR-030, CERN, Geneva (2017), URL `https://cds.cern.ch/record/2285585`

[6] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS LAr Calorimeter*, Technical Report CERN-LHCC-2017-018. ATLAS-TDR-027, CERN, Geneva (2017), URL `http://cds.cern.ch/record/2285582`

[7] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter*, Technical Report CERN-LHCC-2017-019. ATLAS-TDR-028, CERN, Geneva (2017), URL `http://cds.cern.ch/record/2285583`

[8] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*, Technical Report CERN-LHCC-2017-017. ATLAS-TDR-026, CERN, Geneva (2017), URL `http://cds.cern.ch/record/2285580`

[9] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*, Technical Report CERN-LHCC-2017-020. ATLAS-TDR-029, CERN, Geneva (2017), URL `https://cds.cern.ch/record/2285584`

Bibliography

[10] J. J. Thomson, *Cathode Rays*, Phil. Mag. Ser. 5 (1897)

[11] S. M. Bilenky, J. Hosek, *Glashow-Weinberg-Salam Theory of Electroweak Interactions and the Neutral Currents*, Phys. Rept. **90**, 73 (1982)

[12] E. Noether, *Invarianten beliebiger Differentialausdrücke*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918**, 37 (1918), URL `http://eudml.org/doc/59011`

[13] E. Noether, *Invariante Variationsprobleme*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918**, 235 (1918), URL `http://eudml.org/doc/59024`

[14] B. Lemmer, *Measurement of Spin Correlations in $t\bar{t}$ Events from pp Collisions at $\sqrt{s} = 7$ TeV in the Lepton + Jets Final State with the ATLAS Detector* (2014), II.PHYSIK-UNIGÖ-DISS-2014-02, `1410.1791`

[15] UA1 Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV*, Phys. Lett. **B122**, 103 (1983)

[16] UA1 Collaboration, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider*, Phys. Lett. **B126**, 398 (1983)

[17] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, Phys. Rept. **457**, 1 (2008)

[18] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, 321 (1964)

[19] G. S. Guralnik, C. R. Hagen, T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**, 585 (1964)

[20] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 508 (1964)

[21] J. Ellis, *Higgs Physics* (2013), URL `https://cds.cern.ch/record/1638469`

[22] Y. Nambu, *Quasiparticles and Gauge Invariance in the Theory of Superconductivity*, Phys. Rev. **117**, 648 (1960)

[23] J. Goldstone, *Field Theories with Superconductor Solutions*, Nuovo Cim. **19**, 154 (1961)

[24] J. Goldstone, A. Salam, S. Weinberg, *Broken Symmetries*, Phys. Rev. **127**, 965 (1962)

[25] S. Weinberg, *Physical Processes in a Convergent Theory of the Weak and Electromagnetic Interactions*, Phys. Rev. Lett. **27**, 1688 (1971)

[26] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. **B726**, 120 (2013), `1307.1432`

[27] CMS Collaboration, *Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs*, Phys. Rev. Lett. **110(8)**, 081803 (2013), `1212.6639`

[28] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties* (2013), `1307.1347`

[29] B. Mellado Garcia, P. Musella, M. Grazzini, R. Harlander, *CERN Report 4: Part I Standard Model Predictions* (2016), URL `https://cds.cern.ch/record/2150771`

[30] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, S08003 (2008)

[31] S. Chatrchyan, et al. (CMS), *The CMS Experiment at the CERN LHC*, JINST **3**, S08004 (2008)

[32] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3**, S08002 (2008)

[33] V. Gibson (LHCb), *The LHCb experiment*, in *Proceedings, 22nd Lake Louise Winter Institute: Fundamental Interactions (LLWI 2007): Lake Louise, Alberta, Canada, February 19-24, 2007*, pages 60–79 (2007)

[34] F. Hinterberger, *Physik der Teilchenbeschleuniger und Ionenoptik* (1997)

[35] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN* (2016), general Photo, URL `http://cds.cern.ch/record/2197559`

[36] J. Pequenao, *Computer generated image of the whole* ATLAS *detector* (2008), URL `http://www.atlas.ch/photos/full-detector-cgi.html`

[37] A. Salvucci, *Measurement of muon momentum resolution of the ATLAS detector*, EPJ Web Conf. **28**, 12039 (2012), `1201.4704`

[38] A. R. Martínez (ATLAS), *The Run-2 ATLAS Trigger System*, J. Phys. Conf. Ser. **762(1)**, 012003 (2016)

Bibliography

[39] M. Bahr, et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58**, 639 (2008), `0803.0883`

[40] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, P. Z. Skands, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191**, 159 (2015), `1410.3012`

[41] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, J. Winter, *Event generation with SHERPA 1.1*, JHEP **02**, 007 (2009), `0811.4622`

[42] A. Harmon, *Sherpa and Open Science Grid: Predicting the emergence of jets* (2014), URL `https://sciencenode.org/feature/sherpa-and-open-science-grid-predicting-emergence-jets.php`

[43] GEANT4 Collaboration, *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506**, 250 (2003)

[44] ATLAS Collaboration, *ATLAS Phase-II Upgrade Scoping Document*, Technical Report CERN-LHCC-2015-020. LHCC-G-166, CERN, Geneva (2015), URL `https://cds.cern.ch/record/2055248`

[45] ATLAS Collaboration, *Technical Proposal: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*, Technical Report CERN-LHCC-2018-023. LHCC-P-012, CERN, Geneva (2018), URL `http://cds.cern.ch/record/2623663`

[46] J. Gao, C. S. Li, H. X. Zhu, *Top Quark Decay at Next-to-Next-to Leading Order in QCD*, Phys.Rev.Lett. **110(4)**, 042001 (2013)

[47] I. I. Bigi, Y. L. Dokshitzer, V. A. Khoze, J. H. Kuhn, P. M. Zerwas, *Production and Decay Properties of Ultraheavy Quarks*, Phys.Lett. **B181**, 157 (1986)

[48] B. Andersson, G. Gustafson, G. Ingelman, T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97**, 31 (1983)

[49] G. Marchesini, B. R. Webber, *Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation*, Nucl. Phys. **B310**, 461 (1988)

[50] B. R. Webber, *Fragmentation and hadronization*, Int. J. Mod. Phys. **A15S1**, 577 (2000)

[51] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. **C77**, 490 (2017)

[52] Y. A. Kulchitskii, V. B. Vinogradov, *Non-compensation of the ATLAS barrel tile hadron module-0 calorimeter* (1999)

[53] G. P. Salam, *Towards Jetography*, Eur. Phys. J. **C67**, 637 (2010)

[54] G. C. Blazey, et al., *Run II jet physics*, in *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*, pages 47–77 (2000), `hep-ex/0005012`, URL `http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-00-092`

[55] S. D. Ellis, D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. **D48**, 3160 (1993), `hep-ph/9305266`

[56] M. Cacciari, G. P. Salam, G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04**, 063 (2008), `0802.1189`

[57] Y. L. Dokshitzer, G. D. Leder, S. Moretti, B. R. Webber, *Better jet clustering algorithms*, JHEP **08**, 001 (1997), `hep-ph/9707323`

[58] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying to a pair of tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Technical Report ATLAS-CONF-2018-021, CERN, Geneva (2018), URL `https://cds.cern.ch/record/2621794`

[59] A. M. Sirunyan, et al. (CMS), *Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector*, Phys. Lett. **B779**, 283 (2018), `1708.00373`

[60] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector* (2016), `1610.07922`

[61] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms* (2004), `hep-ph/0409146`

[62] S. Frixione, P. Nason, C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11**, 070 (2007), `0709.2092`

[63] S. Alioli, P. Nason, C. Oleari, E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06**, 043 (2010), `1002.2581`

[64] E. Bagnaschi, G. Degrassi, P. Slavich, A. Vicini, *Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM*, JHEP **02**, 088 (2012), `1111.2854`

Bibliography

[65] K. Hamilton, P. Nason, G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05**, 140 (2015), `1501.04637`

[66] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, W. K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07**, 012 (2002), `hep-ph/0201195`

[67] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09**, 145 (2014), `1406.3660`

[68] N. Davidson, T. Przedzinski, Z. Was, *PHOTOS interface in C++: Technical and Physics Documentation*, Comput. Phys. Commun. **199**, 86 (2016), `1011.0937`

[69] *ATLAS Run 1 Pythia8 tunes*, Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva (2014), URL `https://cds.cern.ch/record/1966419`

[70] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82**, 074018 (2010), `1005.3457`

[71] R. D. Ball, et al., *Parton distributions with LHC data*, Nucl. Phys. **B867**, 244 (2013), `1207.1303`

[72] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11(04)**, P04008 (2016), `1512.01094`

[73] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, Technical Report ATLAS-CONF-2016-024, CERN, Geneva (2016), URL `https://cds.cern.ch/record/2157687`

[74] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in protonproton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. **C76(5)**, 292 (2016), `1603.05598`

[75] A. Elagin, P. Murat, A. Pranko, A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. **A654**, 481 (2011), `1012.4686`

[76] G. Cowan, K. Cranmer, E. Gross, O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71**, 1554 (2011), [Erratum: Eur. Phys. J.C73,2501(2013)], `1007.1727`

[77] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. **D96(7)**, 072002 (2017), `1703.09665`

[78] M. Cacciari, G. P. Salam, G. Soyez, *The Catchment Area of Jets*, JHEP **04**, 005 (2008), `0802.1188`

[79] M. Cacciari, G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. **B659**, 119 (2008), `0707.1378`

[80] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C76(12)**, 653 (2016), `1608.03953`

[81] N. Nikiforou (ATLAS), *Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade*, in *Proceedings, 3rd International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA 2013): Marseille, France, June 23-27, 2013* (2013), `1306.6756`

[82] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. **C77(7)**, 466 (2017), `1703.10485`

Erklärung nach §13(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen: Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

<div align="center">

Göttingen, den 12. November 2018

(Tobias Fitschen)

</div>