Jet-Smearing-based Multijet Background Determination for SUSY Searches with the ATLAS Detector

Florian Beisiegel

Masterarbeit in Physik angefertigt im Physikalischen Institut

vorgelegt der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

April 2017

I hereby declare that the work presented here was formulated by myself and that no sources or tools other than those cited were used.

Bonn,Date

Signature

Gutachter: Prof. Dr. Klaus Desch
 Gutachter: Prof. Dr. Jochen Dingfelder

Acknowledgements

I would like to thank Klaus Desch for the opportunity to write my thesis in his group and also for his support and all the helpful comments and ideas. I would also like to thank Philip Bechtle for his support, the great ideas and the inspiring discussions.

A big thanks goes to Steffen Schaepe and Oliver Ricken who both helped me a lot in my daily work with their ideas and by answering all my questions. Moreover they proof-readed this thesis and gave me the opportunity to give a rehearsal talk for my colloquium. Also a great thanks for this. In this context, I want to continue to thank Michael Hübner and Lea Krüger who both helped me to further improve my talk and my thesis, respectively.

In addition, I want to thank the whole working group for the nice and warm-hearted atmosphere.

A special thanks goes also to my friends with whom one can have a lot of fun and who helped me to turn off my head and come to other thoughts.

Finally, the biggest thanks goes to my whole family for their enduring support and advice in all possible situations, even when things are not going so well.

Contents

Acknowledgements						
1	Intro	oduction	1			
2	Theoretical Foundations					
	2.1	The Standard Model of Particle Physics	3			
		2.1.1 Electromagnetic interaction	4			
		2.1.2 Weak interaction	5			
		2.1.3 Electroweak unification	6			
		2.1.4 Strong interaction	7			
		2.1.5 Formation of jets	9			
		2.1.6 Tau leptons	9			
		2.1.7 Shortcomings of the Standard Model	11			
	2.2	Supersymmetry	14			
		2.2.1 Gauge mediated SUSY breaking	15			
		2.2.2 Simplified models	15			
3	Experimental Setup 17					
	3.1	The Large Hadron Collider	17			
	3.2	The ATLAS detector	19			
		3.2.1 Coordinate system	20			
		3.2.2 Inner detector	21			
		3.2.3 Calorimeter	22			
		3.2.4 Muon system	23			
		3.2.5 Trigger system	24			
4	Obje	ect Reconstruction	27			
	4.1	Jets	27			
	4.2	B-jets	28			
	4.3	Tau leptons	28			
	4.4	Missing transverse momentum	29			
	4.5	Muons	31			
	4.6	Electrons	31			
	4.7	Overlap removal	32			
5	Event Simulation 33					
	5.1	Simulation of electroweak backgrounds	33			
	5.2	Normalization to recorded luminosity	34			

	5.3 5.4	Re-weighting	34 35	
6	.let	Smearing	37	
Č	Motivation for let Smearing	37		
	6.2	Seed event selection	39	
	6.3	Quantification of energy fluctuations in the detector	40	
	6.7	Modification of the jet response function	40 42	
	0.4	6.4.1 Dijet analysis	42 42	
		6.4.2 "Moreodos" analysis	42	
		6.4.2 Intercedes analysis	43	
	65	0.4.5 Sinearing of the azimutal angle	44	
	0.3		45	
7	Eve	nt Selection	47	
	7.1	Definitions of important variables	47	
	7.2	Baseline selection	48	
	7.3	Signal, control and validation regions	48	
	,			
8	Dev	elopment of a Multijets Control Region	53	
	8.1	Basic cuts of the multijets control region	53	
	8.2	Enlargement of the seed event statistics	56	
	8.3	Enrichment of the multijets background fraction	61	
•	N.A 14	vijete Reekaveund Estimate for the Analysis with the full 26 fb=1 Detect	67	
9		Treasure of the statistical uncertainties	67	
	9.1		6/	
	9.2	Results	69	
10	Pos	sible Improvements and Outlook	77	
	10.1	Rebalance-and-smear method	77	
	10.2	Tau lepton smearing	77	
	10.3	Determination of systematic uncertainties	83	
11	Sun	nmary and Conclusion	85	
Bibliography				
Α	Add	itional Figures	95	
	A.1	Distributions after different combinations of cuts on new variables	95	
	A 2	Pre-fit distributions in the control and signal regions of the analysis with the full 36 fb^{-1}	20	
	1 1.4	dataset	102	
		uuusee	102	
Lis	List of Figures			
Lie	List of Tables			
-				

CHAPTER 1

Introduction

For a long time, physicists deal with the question what the universe constists of and how it works. In order to answer these questions they performed many experiments over time. Today, particle physicists build large accelerators and detectors to find and investigate the elementary building blocks of matter and the interactions between them.

The currently most accepted model to explain the observed phenomena and the results of the performed measurements is the Standard Model of Particle Physics (SM). It describes the known elementary particles and the interactions between them. It is very successful and is tested to a high precision over a large energy range. One example is the SM Higgs boson, the exitation of the Higgs field. The latter was predicted in 1964 by P. Higgs, F. Englert and R. Brout [1, 2] and discovered in 2012 by the collaborations ATLAS [3] and CMS [4]. Nevertheless it has also some shortcomings, which cannot be solved within the SM. Two of the most prominent ones are the so-called "hierarchy problem" and the existence of dark matter. The first one describes the discrepancy between the experimentally measured and the theoretically expected mass of the SM Higgs boson, which deviate orders of magnitudes from each other. The existence of dark matter follows from different astrophysical observations like the rotation velocity of stars around the center of their galaxy [5] which is not compatible with the SM prediction.

One famous solution for these shortcomings is Supersymmetry (SUSY) [6], an extension of the SM which predicts a symmetry between fermions and bosons. In SUSY models this is realized by doubling the SM particle content. If SUSY exists, these new particles could be found by analyzing data from particle collisions. This is done by searching the data for signatures predicted by the model under investigation. However, such signatures can also be generated by already known SM processes. Therefore it is important to have an accurate modelling of all relevant backgrounds, which can be compared to the data in order to draw any conclusions. The data analyzed in this thesis come from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV. They were provided by the Large Hadron Collider (LHC) at CERN in Geneva and were recorded by the ATLAS detector. The most abundant background in proton-proton collisions are events containing a large number of jets which are the result of hadronization processes due to the strong interaction. Based on the high jet multiplicity in those events, this background is called *multijets background*.

The goal of this thesis is to provide a proper multijets background estimate for an analysis searching for SUSY. Usually background events are generated in Monte Carlo (MC) simulations. Multijets events for SUSY searches, however, cannot be simulated in MC in sufficient quality. This is caused by the extreme requirements on different event quantities in such analyses. One approach for the generation of multijets events for SUSY searches is the *Jet Smearing* technique, which is used in this thesis.

The theoretical foundations for this work are explained in chapter 2. It starts with the Standard Model and its shortcomings and ends with an introduction into SUSY. In this context also the relevant signal model for this thesis is presented. Chapter 3 describes the experimental setup, i.e. the LHC and the different components of the ATLAS detector. Chapter 4 focusses on the reconstruction of all for this thesis relevant objects and explains how these can be distinguished from each other. In chapter 5, the basic principles of event simulations with Monte Carlo are presented. Moreover an overview of the tools is given which are used for the simulation of the electroweak backgrounds in this thesis. A motivation and detailed description of the Jet Smearing technique follows in chapter 6. The most important variables for this analysis as well as the different signal, control and validation regions are defined in chapter 7. Chapter 8 describes the development of a dedicated multijets control region which is then used for the multijets background. Chapter 9 presents and discusses the final results. This includes the calculation of the multijets normalization factor and the obtained background yields in the different control and signal regions, as well as the most relevant uncertainties. Finally, chapter 10 focusses on possible improvements for the Jet Smearing technique and discusses the determination of different systematic uncertainties.

CHAPTER 2

Theoretical Foundations

For every experiment and every analysis it is essential to understand the theoretical foundations of the related topics. Without this knowledge it is very challenging to design an appropriate experiment or to achieve any reliable results. The relevant theories in the context of this thesis are the Standard Model of Particle Physics (SM) and Supersymmetry (SUSY).

This chapter explains the SM and its interactions whereas the focus is on the strong interaction and its consequences since this is the important part for this thesis. Moreover also the shortcomings of the SM are described and how these are solved in SUSY. In this context also the for this thesis relevant SUSY model is presented.

2.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics (SM) [7–11] is the presently valid model to describe the elementary particles and the interactions between them. Although it is a very successful theory tested with high precision in several experiments, there exist some shortcomings both on the experimental and theoretical side which cannot be described by the SM.

The particle content of the SM is illustrated in Fig. 2.1. It can be divided into six quarks (purple), six leptons (green), four gauge bosons (red) and one Higgs boson (yellow). Quarks and leptons are matter particles. They are fermions carrying a half-integer spin. The gauge bosons are the mediators of the interactions between particles. They have an integer spin (s = 1), just like the Higgs boson (s = 0) which is the excitation of the Higgs field. By interacting with this field the particles acquire their bare masses. As depicted by the three different columns in Fig. 2.1, the fermions can be split into three generations. The first generation is built by the u- and the d-quark, the electron and the electron neutrino. The second and third generation are just heavier copies of the first one, which means that all quantum numbers are equal for the respective copy but the masses are different. While the particles of the first generation are the building blocks of our known matter¹, the particles of the other two generations can only be generated by collisions of high-energetic particles like cosmic particles or at particle accelerators. They are unstable and decay into members of the first generation.

Within the SM, three of the four fundamental forces are described: the electromagnetic interaction, the weak interaction and the strong interaction. Gravitation is not included in the theory, but, however, it

¹ For example, the proton consists of two u- and one d-quark and the neutron of one u- and two d-quarks. Together with electrons protons and neutrons can form atoms.



Figure 2.1: Overview of the elementary particles included in the standard model. The quarks are coloured in purple, the leptons in green, the gauge bosons in red and the Higgs boson in yellow. For every particle, its mass, the electric charge and its spin is specified. The coloured shades illustrate which particles are affected by which forces. The red shade depicts the strong interaction, the purple one the electromagnetic interaction and the green shade the weak interaction. Picture taken from [12].

is also not relevant on this scale. The electromagnetic interaction is mediated by the photon and acts between electrically charged particles. Hence all quarks, down-type leptons and the W^{\pm} bosons are affected by this force. The gauge bosons of the weak interaction are the Z boson and the two W bosons whereas the latter only interact with particles carrying *weak isospin*. These are left-handed² particles and right-handed antiparticles. The Z boson, however, couples to all particles carrying weak isospin or a quantity called *weak hypercharge* (see Eq. (2.1) for definition), i.e. also to right-handed particles and left-handed antiparticles. The exchange particles of the strong force are the gluons. This interaction affects all particles with a colour charge, i.e. the quarks and the gluons themselves. All particles and interactions can be described by locally gauge invariant Lagrange densities.

2.1.1 Electromagnetic interaction

The electromagnetic force can be considered as the exchange of a photon between two electrically charged particles. The respective gauge theory is called quantum electrodynamics (QED). The contributing particles can be described by fields Ψ in the Lagrange density \mathcal{L} which is the field-theoretical equivalent to the Lagrangian L in classical mechanics. Analogously to classical mechanics the equations of motion of a particle can be derived with the Euler-Lagrange equation by replacing the variables in L with the respective fields in \mathcal{L} :

$$\frac{\partial \mathcal{L}}{\partial \Psi} - \partial_{\mu} \frac{\mathcal{L}}{\partial \left(\partial_{\mu} \Psi\right)} = 0$$

² Left- and right-handed denote the helicity of a particle. Helicity is defined as the projection of the spin of a particle onto the direction of its momentum [13]. Right-handed means that both are parallel, left-handed means they are anti-parallel.

Here ∂_{μ} denotes the partial derivative. For QED the free Lagrangian is given by

$$\mathcal{L} = i\overline{\Psi}\gamma^{\mu}\partial_{\mu}\Psi - m\overline{\Psi}\Psi$$

with *m* being the particle mass and γ^{μ} the Dirac matrices. The Lagrangian has to be invariant under local U(1) phase transformations

$$\Psi(x) \to \Psi'(x) = e^{i\alpha(x)}\Psi(x) = \Psi(x)$$

where $\alpha(x)$ is a space-time dependent phase. This can be achieved by replacing ∂_{μ} with the covariant derivative

$$D_{\mu} = \partial_{\mu} - ieA_{\mu}$$
.

At the same time the vector field A_{μ} has to transform as

$$A_{\mu} \rightarrow A_{\mu} + \frac{1}{e} \partial_{\mu} \alpha$$
.

So far the resulting Lagrangian does not contain any propagation of a particle. To include this feature and to preserve the invariance of the Lagrangian under a U(1) transformation, an additional term is introduced using the field strength tensor

$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} \; .$$

With this the whole QED Lagrangian is given by

$$\mathcal{L}_{\text{QED}} = \overline{\Psi}(i\gamma^{\mu}\partial_{\mu} - m)\Psi + e\overline{\Psi}\gamma^{\mu}A_{\mu}\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

2.1.2 Weak interaction

The theory of weak interaction was postulated to explain the beta decay. It is mediated by currents which can be split up into neutral currents (NC) and charged currents (CC). The former are exchanged via Z bosons, the latter via W^+ and W^- bosons which couple to the third component T_3 of the weak isospin of a particle. This is a quantum number assigned to all left-handed particles and right-handed antiparticles. For instance, up-type fermions have a value of $T_3 = +1/2$, down-type ones of $T_3 = -1/2$. Although T_3 is conserved within all currently known fundamental interactions, it is changed by ± 1 for particles involved in CC interactions. This results in a change in flavour of the respective particles which is a unique feature of the weak interaction. The transition probabilities from an up-type weak eigenstate of a quark to a down-type one and vice versa are summarized in the so-called *CKM matrix*³. The weak interaction is special compared to the other ones due to the following reasons:

- it violates parity and CP symmetry
- it can change the flavour of particles
- the gauge bosons are massive

³ "CKM" stands for "Cabibbo-Kobayashi-Maskawa", the developers of the theory.

The fact that the gauge bosons have a non-zero mass leads to different effects. First of all, the lifetime of these bosons is very short, being 3.16×10^{-25} s for the W^{\pm} bosons and 2.64×10^{-25} s for the Z boson⁴, respectively. Thus the range of the weak interaction is small. Additionally, the mass influences the coupling strength of the bosons to other particles which is weaker than for the other two interactions. This is the reason why it is called "weak". It also implies that the weak interaction proceeds slower and is therefore suppressed compared to the electromagnetic and the strong interaction.

2.1.3 Electroweak unification

The electromagnetic and the weak interaction are both low-energy approximations of the more general electroweak interaction which is valid at higher energies. Below a certain threshold, the underlying $SU(2) \times U(1)$ symmetry undergoes a spontaneous symmetry breaking, resulting in the two different low-energy theories. The electroweak model was developed by Glashow, Salam and Weinberg in the 1960s [7–9, 11].

The new theory has to account for the properties of both the electromagnetic and the weak interaction. As mentioned already, the latter couples only to left-handed particles and right-handed antiparticles and the electromagnetic interaction includes all particles with an electric charge. Therefore the particles are grouped in left-handed doublets and right-handed singlets, which are in the lepton sector

$$\left(\begin{array}{c} v_L \\ l_L \end{array} \right)$$
 and $\left(l_R \right)$.

The singlet comes from the fact that no right-handed neutrinos exist in the SM. The weak interaction can be represented by a SU(2) symmetry of the weak isospin, while the electromagnetic part underlies a U(1) symmetry acting on the weak hypercharge which is defined as

$$Y = 2(Q - T_3) \tag{2.1}$$

with *Q* being the electric charge and T_3 the third component of the weak isospin. Then the overall symmetry is $SU(2)_L \times U(1)_Y$.

The SU(2) symmetry has three degrees of freedom and the U(1) symmetry one. These come along with three and one additional gauge fields, respectively. The fields are named W^1_{μ} , W^2_{μ} , W^3_{μ} and B_{μ} . The W_{μ} couple to the weak isospin and B_{μ} to the weak hypercharge. However, these fields do not correspond to the known SM gauge bosons γ (represented by A_{μ}), W^+_{μ} , W^-_{μ} and Z_{μ} . This issue is solved by spontaneous breaking of the SU(2) × U(1) symmetry which makes the gauge bosons to be superpositions of the four electroweak gauge fields:

$$\begin{pmatrix} A_{\mu} \\ Z_{\mu} \end{pmatrix} = \begin{pmatrix} \cos \theta_{W} & \sin \theta_{W} \\ -\sin \theta_{W} & \cos \theta_{W} \end{pmatrix} \cdot \begin{pmatrix} B_{\mu} \\ W_{\mu}^{3} \end{pmatrix}$$
$$W_{\mu}^{\pm} = \frac{1}{\sqrt{2}} \left(W_{\mu}^{1} \mp i W_{\mu}^{2} \right)$$

Here, θ_W denotes the weak mixing angle with $\sin^2 \theta_W \approx 0.23$.

The compositions of the W and Z bosons imply that the W bosons only couple to particles carrying weak isospin, while the Z boson additionally couples to all particles with weak hypercharge. These are also

⁴ Lifetimes were calculated from the decay widths given in [14].

right-handed particles and left-handed antiparticles.

A similar representation can be established for quarks. Here, the doublets are built out of the up-type quarks and the weak eigenstates of the down-type ones where the weak eigenstates are connected to the mass eigenstates via the CKM matrix. The coupling constant of the $SU(2)_L$ symmetry, g, and the coupling constant of $U(1)_Y$, g', are related to each other via

$$e = g \sin \theta_W = g' \cos \theta_W$$

where e denotes the electromagnetic coupling constant.

2.1.4 Strong interaction

The strong interaction is responsible for the structure of baryons and mesons and ultimately for the formation of atomic nuclei. It is mediated by gluons which couple to the so-called *colour charge* of a particle. This quantum number has three possible values, "red", "green" and "blue". It was postulated as the consequence of different observations, for example to explain the existence of the Δ^{++} resonance, a bound state consisting of three u-quarks. Since the gluons also carry colour charge themselves, they couple to each other, too, leading to triple and quartic gauge boson vertices. This self-coupling of gauge bosons is a unique property of the strong interaction.

As well as the other ones, also the strong interaction can be described by a gauge theory called *quantum chromodynamics* (QCD) which is based on a SU(3) symmetry acting on the colour charge [10]. From the SU(3) group structure it follows that there exist eight gluons $(3 \times 3 - 1)$ since the ninth would be colour-neutral and does therefore not exist in nature due to confinement⁵.

Running of the coupling constant The strong interaction is unique compared to the other ones in terms of evolution of its coupling constant α_S with energy. While for the electromagnetic and the weak interaction the corresponding coupling constant grows with increasing energy, α_S behaves in the opposite way as depicted in Fig. 2.2. This means that at low energy scales the coupling constant diverges



Figure 2.2: Evolution of the strong interaction coupling constant α_{s} with energy. Picture taken from [14].

⁵ *Confinement* refers to the phenomenon that only colour-neutral objects can be observed in nature. As a consequence it is not possible to observe single quarks or gluons.

which makes the calculation of higher order correction terms impossible. Hence perturbation theory is not applicable and matrix element calculations for Monte Carlo (MC) simulations have to be done phenomenologically.

Another consequence of the shape of α_S is the form of the potential of the strong force which is given by [15]

$$V(r) = -\frac{4}{3} \frac{\alpha_{\rm S}(r) \hbar c}{r} + k \cdot r$$

with $k \approx 1 \text{ GeV/fm}$. This potential is negative up to a certain distance *r* and then becomes positive with a constant slope. This leads to an increasing energy between two colour-charged objects moving away from each other. When the energy is high enough⁶ a quark-antiquark pair can emerge from vacuum. It forms bound states with the original quarks in order to build colour-neutral objects. This process is called *hadronization*. It is schematically illustrated in Fig. 2.3.

For protons this means that besides the valence quarks (*uud*) it also consists of "sea quarks", produced out of the vacuum due to confinement, and gluons which are exchanged between the quarks and other gluons. In general only one of these partons is involved in a collision, the other ones are not affected and continue their movement along the beam pipe of an accelerator. This phenomenon is called the "underlying event" which is depicted in Fig. 2.3. It causes a severe problem for hadron colliders like the LHC because the center-of-mass energy of a proton is distributed over all particles in it. The momentum fraction of a colliding parton can only be calculated from *parton distribution functions* (PDFs) which are determined from measurements.



Figure 2.3: Schematic illustration of a proton-proton collision. The hard interaction is coloured in red, the parton showering in blue, the hadronization process in green and the underlying event in purple. Figure taken from [16].

⁶ The energy has to be as high as the sum of the masses of both quarks.

2.1.5 Formation of jets

When two particles collide, for example at hadron colliders, new particles are produced. In case of quarks or gluons those will carry a net colour charge. As a result of confinement they undergo hadronization as explained in section 2.1.4. This procedure usually requires several steps in which a number of new particles is produced until only colour-neutral particles remain. Since all particles generated in this process have roughly the same direction of motion, bunches of particles arise which are called *jets*. Jets are the dominant objects in hadron colliders because there, the colliding particles are not elementary but consist of quarks and gluons (c.f. section 2.1.4) which mainly interact via the strong force. While jets produced in the hard interaction process are well described in simulation, the modelling of low momentum jets arising from the underlying event is very challenging. Jets can also emerge from gluons which are emitted from initial or final state particles resulting in initial and final state radiation (ISR and FSR). The simulation of ISR and FSR is also difficult which often leads to an imprecise modelling of the

2.1.6 Tau leptons

number of jets for high jet multiplicities.

With a mass of 1777.86 MeV [14], the tau lepton is the heaviest lepton in the standard model. Due to its high mass it has a very short lifetime of approximately 290 fs [14] which means that it decays right after production in an environment such as the LHC. The high mass is also the reason why it can decay both leptonically and hadronically. In detail, the tau lepton decays via the weak interaction into a tau neutrino and a *W* boson. The latter then decays again either into a lepton (electron or muon) and the respective anti-neutrino or into a quark and an antiquark. The corresponding Feynman diagram is depicted in Fig. 2.4. In case of hadronic decays the quarks form a bound state decaying further, mostly into an odd number of charged pions and a number of neutral ones. The branching fractions of the main decay channels can be reviewed in Fig. 2.5. With a branching ratio of roughly 35 %, the tau lepton



Figure 2.4: Feynman diagram of the decay of a tau lepton. The tau lepton can either decay leptonically into an electron or muon plus the respective anti-neutrino or into a quark and an antiquark. Picture taken from [17].

decays leptonically and in approximately 65 % of all cases hadronically [14]. Since the tau decays after a distance of $c\tau = 87 \,\mu\text{m}$ [14], i.e. already in the beam pipe, the leptons produced in the decay cannot be distinguished from primary ones. Hence leptonically decaying tau leptons are impossible to reconstruct which is why only hadronic decays are considered. These, however, look like jets from quarks or gluons since both originate from hadronized objects. They can be distinguished by their spread in the detector which is broader for jets than for tau leptons (c.f. Fig. 2.6) because the latter are colour-neutral in the initial state while quarks or gluons resulting in jets have a net colour. Hence these have to undergo more steps of hadronization in order to become colour-neutral.



Figure 2.5: Branching fractions for the main decay channels of tau leptons. The values for the leptonic decays are nearly equal. For the hadronic channels the branching ratio for one charged pion and one neutral pion is the highest. Values taken from [14].



Figure 2.6: Difference between the spread of jets and of tau leptons. Picture taken from [18].

The similar looking signatures of tau leptons and jets lead to a severe problem. On reconstruction level both objects are treated as jets (c.f. section 4). A distinction between both object types is first drawn in the identification procedure where dedicated variables are used for the tau identification (tau ID) to separate them from jets (c.f. section 4.3). Nevertheless, there is a non-negligible probability for jets to be mis-identified as hadronically decaying tau leptons. Those jets are called "fake- τ ". The probability to fake a tau lepton depends on the origin of the jet (quark or gluon), its momentum, the number of tracks of the reconstructed tau candidate and the used identification algorithm [19]. For a center-of-mass energy of 13 TeV, there are no values for the tau lepton mis-identification probability of jets available. For $\sqrt{s} = 7$ TeV, however, it is between 0.1 % and 10 % [19].

2.1.7 Shortcomings of the Standard Model

The Standard Model is a very successful theory which is tested over a high energy range and with an impressive precision. As an example, Fig. 2.7 shows the total and the fiducial production cross-section of several SM processes. They are distributed over nine orders of magnitude ranging down to approximately 10^{-3} pb. Nevertheless, there are also some phenomena that cannot be explained by the SM. This section focusses on three of them: the "hierarchy problem", the existence of dark matter and the unification of forces.



Figure 2.7: Summary of the total and fiducial production cross-sections for several standard model processes measured with ATLAS. Fig. taken from [20].

The hierarchy problem The term "hierarchy problem" describes the issue that the experimentally measured mass of the Higgs bosond and the theoretically predicted one differ by approximately 16 orders of magnitude [21]. Experimentally the mass has been found to be roughly 125 GeV [3, 4] but the theoretical value is near the Planck Scale at 2.4×10^{18} GeV [21]. The problem is caused by the non-zero vacuum expectation value of the Higgs field *H* which is connected to the Higgs mass m_H via [21]

$$\langle H \rangle = \sqrt{-m_H^2/2\lambda}$$

The Higgs mass receives large corrections from higher order loop effects from every particle which couples to it. For instance, for every fermion coupling to the Higgs boson via $-\lambda_f H \bar{f} f$, the mass gets a correction of

$$\Delta m_H^2 = -\frac{\left|\lambda_f\right|^2}{8\pi^2}\Lambda_{\rm UV}^2 + \dots$$

with Λ_{UV} being an ultraviolet momentum cutoff of the order of the Planck scale. This would lead to a Higgs boson mass at the same order. To obtain the measured mass, a so-called "fine-tuning" would be necessary. Fig. 2.8 shows such loop corrections caused by the coupling to a fermion and a scalar. A natural cancellation of the loop corrections could be achieved by introducing a symmetry in which for



Figure 2.8: Loop corrections to the Higgs mass caused by the coupling to (a) a fermion and (b) a scalar.

every SM fermion two complex scalars exist [21] since every coupling $-\lambda_S |H|^2 |S|^2$ of a scalar S to the Higgs would lead to a correction

$$\Delta m_H^2 = \frac{\lambda_S}{16\pi^2} \Lambda_{\rm UV}^2 + \dots \; .$$

If $\lambda_s = |\lambda_f|^2$, the corrections of both particle types would cancel to zero. One theory where this is the case is called *supersymmetry* which is briefly introduced in section 2.2.

Dark Matter Another phenomenon that cannot be explained within the SM is the existence of dark matter which is predicted by several observations and measurements, like the observed rotation curves of galaxies [5], gravitational lensing measurements [22–24] or measurements of the cosmic microwave background composition [25]. As Fig. 2.9 shows, the percentage of ordinary (or baryonic) matter in the universe is very small with approximately 5% [26]. Dark energy is expected to be the main part with roughly 68%, followed by dark matter with around 27% [26]. While the SM was developed to describe ordinary matter, it is not valid for dark matter. The latter is a form of matter which could not be observed so far. This implies that it interacts at most weakly with SM particles. Supersymmetry provides a solution for this issue since the lightest supersymmetric particle is a candidate for dark matter as will be explained in section 2.2.



Figure 2.9: Fractions of ordinary matter, dark matter and dark energy in the universe. Values taken from [26].

Unification of forces The successful unification of electromagnetic and weak interaction into one electroweak theory encouraged physicists to find a theory wherein all three fundamental interactions are unified. Such a *Grand Unified Theory* (GUT) implies that the couplings of all forces have the same value at a certain energy scale. Below this GUT scale the respective symmetry breaks down to the electroweak and the strong force. As depicted in Fig. 2.10, the couplings will never meet at the same point within the standard model. Supersymmetry would change the couplings in a way that they meet at approximately 10^{16} GeV [21].



Figure 2.10: Evolution of the coupling constants of the three SM forces with energy. Within the SM they never have the same strength all at once (dashed lines). This issue can be fixed in supersymmetry models (coloured lines). Figure taken from [21].

2.2 Supersymmetry

Supersymmetry (SUSY) [6] is a symmetry between fermions and bosons that introduces bosonic superpartners for standard model fermions and vice versa. All quantum numbers of the SUSY particles, except for the spin, are identical to those of their SM equivalents.

The minimal extension to the SM with the lowest number of new particles is the so called "Minimal Supersymmetric Standard Model" (MSSM) [21, 27, 28] which postulates one SUSY partner for each SM particle. This means fermions have two bosonic superpartners since the SM fermions are treated as Weyl fermions which have two weak eigenstates with one SUSY partner for each [21]. SM bosons instead have only one fermionic superpartner. A SM particle and its supersymmetric partner form a supermultiplet. For the Higgs boson two chiral supermultiplets with two Higgs bosons each have to be introduced, one with weak hypercharge Y = 1/2 and the other with Y = -1/2, since with only one supermultiplet, gauge anomalies would affect the electroweak gauge symmetry [21]. Another reason for the need of two Higgs multiplets are the Yukawa couplings to the up-type quarks whereas the supermultiplet with Y = -1/2 is responsible for the Yukawa couplings to the down-type quarks [21].

The naming of SUSY particles follows the convention that fermionic ones get an "ino" as suffix and bosonic ones an "s" as prefix. For instance the supersymmetric partner of an electron is a "selectron" and the partner of a gluon a "gluino". In general neutral gauginos are called *neutralinos* ($\tilde{\chi}^0$) and charged ones *charginos* ($\tilde{\chi}^{\pm}$).

Experimentally, SUSY particles can be produced in collisions of SM particles in accelerators like the LHC. After production they decay via a cascade of other SUSY particles into the lightest supersymmetric particle (LSP). In each decay also SM particles are emitted which leads to certain signatures that can be searched for. A direct decay into SM particles is not possible for SUSY particles due to a quantum number called "R-parity" which is defined by

$$P_R = (-1)^{3(B-L)+2s}$$

with *B* and *L* being the baryon number and the lepton number, respectively, and *s* the spin. R-parity has been introduced because supersymmetry would lead to *B* and *L* conservation violating interactions which are forbidden in the standard model and have not been observed yet. Moreover the proton would no longer be stable⁷ in this case. By assigning $P_R = +1$ to SM particles and $P_R = -1$ to SUSY particles, R-parity conservation forbids mixing between SM and SUSY particles and thus solves both problems described above. The fact that SUSY particles cannot decay into SM particles implies that the LSP is stable. Since it has not been observed yet it can only interact weakly with ordinary matter and is thus a candidate for dark matter [21]. Another consequence of R-parity conservation is that SUSY particles can only be produced in an even number in experiments [21].

If SUSY would be an exact symmetry, the masses of the supersymmetric particles would be equal to those of their SM partners and SUSY particles would have to be observed already. Since this is not the case the superpartners have to be heavier and SUSY must be broken. Over time several models for breaking mechanisms have been invented. In the MSSM, the breaking is usually achieved by adding a Lagrangian $\mathcal{L}_{soft}^{MSSM}$ to the SUSY Lagrangian which contains all possible soft breaking terms. A detailed explanation can be found in [21].

⁷ Here "stable" means a lifetime of more than 10^{29} years [14], since up to now no proton decay has been observed.

2.2.1 Gauge mediated SUSY breaking

One example for SUSY breaking models are "gauge-mediated supersymmetry breaking" (GMSB) models, where SM gauge interactions are responsible for the symmetry breaking. In these models, additional chiral supermultipletts l, \bar{l}, q and \bar{q} , consisting of messenger quarks and leptons as well as scalar quarks and leptons, couple indirectly to the particles of the MSSM via gauge boson and gaugino interaction [21]. In most GMSB models, tau leptons play an important role since the mass hierarchy of the sleptons is inverted in such models, meaning that the stau lepton $\tilde{\tau}$ is the lightest slepton. Hence, in the decay chain of SUSY particles, the stau lepton is often the next-to-lightest SUSY particle (NLSP). It then decays further into the LSP which is usually the gravitino \tilde{G} . In this decay, a SM tau lepton is emitted which is a key signature in GMSB models.

2.2.2 Simplified models

Full supersymmetry breaking models like GMSB have the advantage that they include the whole set of relevant parameters which allows precise predictions and consequently exclusions of a large parameter space. The disadvantage of such models is the large number of different models which all have to be tested.

In current analyses, a common approach are *simplified models*. These models do not specify a particular model with a certain breaking mechanism. Instead general topologies are assumed which are relevant in several full models and have only a few free parameters by making general assumptions.

The relevant simplified model for this thesis is shown in Fig. 2.11. It is an R-parity conserving model of gluino pair production. The gluinos decay in three steps into the LSP which is a $\tilde{\chi}_1^0$ in this model. In the first step the gluinos either decay into a $\tilde{\chi}_2^0$ or a $\tilde{\chi}_1^{\pm}$ under emission of two jets in both cases. Both then decay further into either a stau lepton and or a tau sneutrino. In the first case an additional tau lepton is emitted for the $\tilde{\chi}_2^0$ and a tau neutrino for the $\tilde{\chi}_1^{\pm}$. For the second case it is vice versa. The stau finally decays into the LSP and a tau while for the tau sneutrino the LSP and a tau neutrino are produced.

Since there are no experimental or theoretical hints for the mass splitting in each decay step, it is assumed that the mass is equally splitted in the decay, i.e. both daughter particles carry half of the mother particle's mass.

The defining signature of the described simplified model are tau leptons and numerous jets with high transverse momenta. Since neither the neutrinos nor the LSPs can be detected, a large $\not\!\!E_T$ is required as well. Similar signatures can also be featured by different electroweak backgrounds (c.f. section 5.1) and multijets events (c.f. section 5.4).



Figure 2.11: The for this thesis relevent simplified model. Figure taken from [29].

CHAPTER 3

Experimental Setup

The data analyzed in this thesis come from proton-proton collisions at a center-of-mass energy of 13 TeV. The collisions were provided by the Large Hadron Collider (LHC) at CERN and the final events were recorded with the ATLAS detector.

This chapter describes the experimental setup. It stats with an overview of the LHC and its technical specification. Afterwards the ATLAS detector and its different components with their main features are explained. In the end the handling of the huge amount of provided data is discussed.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) at CERN in Geneva is the largest particle accelerator worldwide with a circumference of about 27 km [30]. It is located around 100 m under ground as illustrated in Fig. 3.1. It accelerates lead ions and protons¹ in two seperated beam pipes with contrariwise propagating beams. After a series of pre-accelerators, the LHC brings the particles to the respective collision energy. For protons this was a centre-of-mass energy \sqrt{s} of 7 TeV and 8 TeV, respectively, in Run-I and is 13 TeV in Run-II. In the next years the LHC will be upgraded to its design collision energy of 14 TeV [30].

The protons are accelerated using electric fields inside cavities. The sinusodial form of the waves leads to a differential acceleration depending on the position of the particle with respect to the wave. Thereby the protons are grouped to bunches of roughly 10^{11} particles per bunch. This is necessary since protons are too small to collide single particles. The probability of a collision is much higher with bunches of many particles. A measure to quantify the particle flux is the so-called *luminosity* [30]

$$L = \frac{N_b^2 n_b f_{\rm rev} \gamma_r}{4\pi\epsilon_n \beta^*} F$$

with N_b being the number of particles per bunch, n_b the number of bunches per beam, γ_r the relativistic gamma factor, f_{rev} the revolution frequency of the beams, ϵ_n the normalized transverse emmitance of the beam and β^* the beta function at the collision point. *F* is a geometrical reduction factor due to the crossing angle of the beams.

For the design peak luminosity of 10^{34} cm⁻²s⁻¹, the LHC contains 2808 bunches with a bunch spacing time of 25 ns which corresponds to a bunch crossing rate of 40 MHz [30].

¹ Since the data used for this thesis is from proton-proton collisions, this section describes only those.



Figure 3.1: The LHC accelerator and its experiments. Picture taken from [31].

The number of events per second is given by

$$\dot{N}_{\text{events}} = L \cdot \sigma$$

where σ is the proton-proton interaction cross section, a theoretical probability of an interaction to take place. The total number of events and therefore the amount of recorded data can be calculated by multiplying the cross section with the (time) integrated luminosity which is thus the commonly used measure for the available amount of data. The peak luminosity by fill delivered by the LHC is illustrated in Fig. 3.2(a) for 2015 and in Fig. 3.2(b) for 2016.

Since the LHC is a ring collider the protons need to be deflected. This is achieved with dipole magnets which use the Lorentz force to bend the particles. Therefore a magnetic field of 8.33 T needs to be generated [30]. Due to the fact that protons are electrically charged, the bunches are broadened by the Coulomb force. To countersteer this, quadrupole magnets are integrated into the LHC.

Around the LHC there are four interaction points where the four experiments ALICE, LHC-b, CMS and ATLAS are located (c.f. Fig. 3.1). At these points, the two beams are brought to collision. The data used for this thesis was recorded by ATLAS.

Pile-up The formation of single protons to bunches also has a big disadvantage. The large particle number and the high density result in a high probability of multiple interactions per bunch crossing. Beside the hard interaction process, several soft interactions like scattering can take place as well. This phenomenon is called "pile-up". Fig. 3.2(c) shows the pile-up profile for the collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV in 2015 and 2016 separately as well as the combined profile. It should be noted that the mean value for 2015 is much lower than the corresponding value for 2016 with $\langle \mu \rangle = 13.7$ and $\langle \mu \rangle = 24.2$, respectively. This discrepancy can be explained with the higher instantaneous luminosities in 2016 (c.f. Fig. 3.2(a) and Fig. 3.2(b)) which result in higher interaction probabilities.



Figure 3.2: Peak luminosities by fill delivered by the LHC in (a) 2015 and (b) 2016 and (c) the number of interactions per bunch crossing (pile-up) for 2015 and 2016 measured with the ATLAS detector. Plots taken from [32].

3.2 The ATLAS detector

The ATLAS experiment is one of the four experiments at the LHC. With a height of 25 m and a length of 44 m (c.f. Fig. 3.3) it is the largest of the four detectors. It is a multi-purpose detector which is used to investigate a broadly based spectrum of physical questions. Fig. 3.3 shows an overview of the whole ATLAS detector. It is a cylindrical detector where the different components are arranged in different layers around the interaction point (IP). In the beam pipe in the middle of the detector, the proton bunches are brought to collision. In this processes new particles are produced which, depending on their lifetime, either propagate directly through the detector or decay already in the beam pipe resulting in decay products which traverse the detector. To detect these particles and measure their properties, different sub-detectors are used. From the inside out these are the inner detector, the calorimeters and the muon spectrometer [34].



Figure 3.3: Overview of the ATLAS detector. Picture taken from [33].

3.2.1 Coordinate system

With the different detector systems, several properties can be measured for each particle and event, respectively, for instance the energy or the momentum of a particle. The ATLAS coordinate system is right-handed with the z-axis pointing along the beam pipe, the x-axis towards the centre of the LHC ring and the y-axis upwards. The origin is in the interaction point [34].

Due to the cylindrical symmetry of ATLAS, it is more convenient to use polar coordinates (r, θ, ϕ) with r being the distance to the interaction point, θ the polar angle measured from the beam pipe, and ϕ the azimutal angle which is given in radians. All angles ϕ in this thesis are given in radians as well. Instead of the polar angle, the *pseudorapidity* [34]

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

is used. The big advantage of this quantity is its lorentz-invariance which leads to constant particle fluxes for equidistant pseudorapidity intervals.

In ATLAS, only the transverse components of quantities are used, for example the transverse momentum of a particle or the missing transverse energy in the event. This is a consequence of the fact that the colliding protons are composite particles. Hence the longitudinal proton momentum fractions of the interacting partons are not exactly known. Hence both in the collision involved particles can carry different momenta. This leads to event topologies which are boosted in one direction. However, the initial transverse momentum is zero and consequently the sum of the transverse momenta of all final state particles should be zero as well due to momentum conservation.

3.2.2 Inner detector

In the ATLAS detector, protons collide every 25 ns leading to a huge number of produced particles. In order to disentangle all the tracks and reconstruct their vertices, detector systems with extremely high resolutions are needed. The inner detector (ID) is designed to fulfill these conditions. It consists of the pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT) [34]. All three components are divided into a barrel part and an end-cap part. The setup of the inner detector is sketched in Fig. 3.4. The whole inner detector is surrounded by a solenoid magnet. It creates a magnetic field of 2 T to guarantee a good momentum reconstruction by exploiting the bending of the tracks due to the Lorentz force.

The pixel detector is the innermost detector and encloses the beam pipe directly. It constists of small



Figure 3.4: Setup of the ATLAS inner detector. Picture taken from [35].

cells (pixels) of semiconductor detectors. Originally it was build up of three disks per side in the end-cap regions and three cylindrical layers in the barrel with pixels of size $50 \times 400 \,\mu\text{m}^2$ in $R - \phi \times z$ to ensure a precise track and momentum reconstruction with a resolution of $10 \,\mu\text{m} \times 115 \,\mu\text{m}$ in $R - \phi \times z$ [34]. During the Run-II upgrade an additional "Insertable B-Layer" (IBL) was integrated [36] to further increase the resolution. The pixel detector covers a range of $|\eta| < 2.5$ and has roughly 92 million readout channels in total (80 million in the three original layers [34] and 12 million in the IBL [37]).

The pixel detector is surrounded by the semiconductor tracker which is also based on semiconductor technology but is built in strips rather than in pixels. Hence it has a coarser granularity leading to a resolution of $17 \,\mu\text{m} \times 580 \,\mu\text{m}$ in $R - \phi \times z$ [34]. As for the pixel detector, the pseudorapidity coverage is $|\eta| < 2.5$. The SCT has four layers in the barrel and nine disks in each end-cap region. It has approximately 6.3 million readout channels in total [34].

The third and outermost component of the ID is the transition radiation tracker. In contrast to the other two sub-detectors, the TRT is made of straw tubes and exploits the transition radiation a particle produces when traversing the TRT. There are 73 straw planes in the barrel and 160 in the end-cap regions whereas each straw has a diameter of 4 mm, leading to a resolution of 130 µm in $R - \phi$. The TRT has roughly 350 000 readout channels and covers the region of $|\eta| < 2.0$ [34].

3.2.3 Calorimeter

The inner detector is enclosed by the calorimeter which measures the energy of particles by stopping them and measuring the deposited energy. It is split into two sub-systems, an electromagnetic calorimeter (ECAL) for electromagnetically interacting particles and a hadronic calorimeter (HCAL) for hadronically interacting particles. The latter is placed around the ECAL as depicted in Fig. 3.5. Both parts are made of alternating active and passive layers. Passive layers are in general made of dense materials in order to slow down the particles by forcing them to produce showers. The relevant processes in this context are bremsstrahlung and electron-positron pair creation for the ECAL and elastic and inelastic scattering as well as fission for the HCAL. The function of the active layers is the detection of single shower particles and the measurement of their energy by collecting charges generated in the passive layers through the different processes described above. In ATLAS, these layers are made of liquid argon (LAr) or scintillators. In order to measure the total energy of a particle it has to be stopped within the calorimeter. Therefore several steps of the relevant processes like bremsstrahlung and pair creation have to take place. Thus the calorimeter needs a certain thickness which can be quantified in terms of *radiation lengths*² X_0 for the ECAL and *nuclear interaction lengths*³ λ_0 for the HCAL.

All calorimeter parts together have an η -coverage of $|\eta| < 4.9$ [34]. The different sub-detectors have



Figure 3.5: The ATLAS calorimeter. It is divided into an electromagnetic and a hadronic calorimeter whereas the latter one encloses the electromagnetic calorimeter. Figure taken from [38].

different resolutions depending on their position and function. The electromagnetic calorimeter uses LAr as active material and lead as passive one. It is divided into a barrel ($|\eta| < 1.475$) and an end-cap part (1.375 < $|\eta| < 3.2$) [34] like the inner detector. The thickness of the ECAL is larger than 22 X₀ in the barrel and 24 X₀ in the end-caps [34].

The hadronic calorimeter is built up of three different components. The first one is the *Tile calorimeter* which has a barrel part ($|\eta| < 1.0$) and two extended barrels ($0.8 < |\eta| < 1.7$). It uses scintillators as active material and steel as passive one. The second HCAL component is an LAr end-cap calorimeter

² The radiation length is the average distance after which the electron energy is reduced by 1/e by bremsstrahlung or nearly 7/9 of the mean free path of the pair production process of a high-energetic photon [13].

³ The nuclear interaction length is the mean path between two hadronic interactions for relativistic hadrons [13].

with copper as an absorber. It has a coverage of $1.5 < |\eta| < 3.2$. The last one is the LAr forward calorimeter which covers the region $3.1 < |\eta| < 4.9$. Is consists of three modules whereas the first one is for electromagnetic measurements and the other two for hadronic measurements. The total thickness of the HCAL is 9.7 λ_0 in the barrel and $10 \lambda_0$ in the end-cap regions [34].

Energy mis-measurements Since real detectors are not ideal, every quantity measured with a detector has a non-negligible probability to be mis-measured. For the ATLAS calorimeter, the energy is not exactly measured due to several reasons:

- Dead material or gaps in the calorimeter (e.g. for readout or power supply)
- Defect detector modules
- The material budget in front of the calorimeter (e.g. beam pipe or inner detector) reduces the energy of a particle.
- High-energetic jets can pass the detector without depositing their full energy in the calorimeter. This phenomenon is called "punch-through".

All these sources of mis-measurements lead to a relative energy resolution of the calorimeter of

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E(GeV)}} \oplus b$$

with *a* being the *stochastic term* and *b* the *constant term*. These terms have individual values for the different calorimeter types. For instance the combined LAr and tile calorimeter performance measurements result in $a = (52.0 \pm 1.0) \% \sqrt{\text{GeV}}$ and $b = (3.0 \pm 0.1) \% [34]$.

Neglecting effects like neutrinos which leave the detector without being detected, the transverse energy in an event would be perfectly balanced out assuming an ideal detector. Due to mis-measurements in a real detector most events have a net transverse energy in one direction, leading to "fake- $\not\!\!\!E_T$ " in the opposite direction.

Since neutrinos only interact weakly, they leave the detector unnoticed and thus build another source of missing transverse energy which is called "true- $\not\!\!E_T$ ".

3.2.4 Muon system

The aim of the calorimeter is to stop all traversing particles in order to measure their energy. Since muons are minimally ionizing particles, they leave the calorimeter nearly without being deaccelerated. To measure their momentum, a *Muon System* (MS) is installed around the calorimeter. In addition, a toroid magnet system, composed of one barrel and two end-cap toroids, is embedded in the muon system. It produces magnetic fields of roughly 0.5 T and 1.0 T [34], respectively, to bend the trajectories. The barrel magnet covers a range of $|\eta| < 1.4$ and the end-cap toroids the region $1.6 < |\eta| < 2.7$. In between a superposition of both is used [34].

The muon system itself uses four different detector types, two of which are used for precision tracking and the other two for triggering and determination of the second coordinate. In the barrel as well as in the end-caps ($|\eta| < 2.7$), *Monitored Drift Tubes* (MDTs) are installed for the tracking of muons [34]. *Cathode Strip Chambers* (CSCs) are used in the forward directions ($2.0 < |\eta| < 2.7$) since they are more robust against radiation [34]. For triggering purposes and coordinate measurements, *Resistive Plate Chambers*

(RPCs) and *Thin Gap Chambers* (TGCs) are installed in the barrel ($|\eta| < 1.05$) and in the end-cap regions (1.05 < $|\eta| < 2.7$), respectively, whereas the TGCs are only employed for triggering within $|\eta| < 2.4$ [34]. The muon system and the toroid magnets are sketched in Fig. 3.6.



Figure 3.6: Overview of the ATLAS muon system and the toroid magnet system. Picture taken from [39].

3.2.5 Trigger system

The amount of data recorded by ATLAS is huge. At the design luminosity of 10^{34} cm⁻²s⁻¹, collisions take place every 25 ns. This corresponds to an event rate of 40 MHz. Since the available computing infrastructure and the storage capacity is by far not sufficient to store every event, triggers are used to record only physically interesting ones. The ATLAS trigger system consists of a Level-1 trigger (L1) and the High Level Trigger (HLT) which reduce the recording rate from 40 MHz to roughly 1 kHz [40]. Due to the large event rate, the L1 is hardware-based in order to be as fast as possible. It makes decisions based on information from the calorimeter, the muon system and several other sub-systems like the Minimum Bias Trigger Scintillators (MBTS) [40]. In this step not the full detector information are used, for example calorimeter cells are added up just roughly to clusters. These preselected events passing the L1 are then further analyzed by the HLT. This trigger is software-based and uses information from the calorimeter and the muon system. Additionally, the full reconstruction algorithms are applied in the HLT. All events passing the HLT are finally recorded to disk.

Depending on what kind of physics processes is investigated, different so-called "trigger items" can be chosen. For instance one trigger item used in this thesis is the HLT_xE70, which selects only events with a missing transverse energy of more than 70 GeV. Since ATLAS was designed to investigate many different physics scenarios and to perform various measurements, the requirements on the trigger are tremendous. Hence, during data taking periods the trigger menus are changed in order to collect data for all kinds of analyses and measurements.

A characteristic feature of triggers are turn-on effects which are caused by measurement uncertainties and reconstruction issues. This means that the efficiency of a trigger is very low at their offline threshold and

increases for higher values until it reaches approximately 100% at a certain value. Above this value the efficiency stays approximately constant. To ensure that the trigger selects only events above the chosen threshold, it is essential that its efficiency was as high as possible. This can be achieved by imposing "trigger plateau cuts" on the triggered quantity at a value at the beginning of the efficiency plateau.

CHAPTER 4

Object Reconstruction

The direct output information of the ATLAS detector are only electronical signals for every event which are then converted into more practical information like hits, deposited energy or timing. For an analysis, however, it is important to have knowledge about the physical objects (e.g. particles) which produced the signatures and also about their properties. Therefore these objects have to be reconstructed from the provided information mentioned before. From the reconstructed objects also other event properties like the missing transverse momentum can be derived. This chapter focusses on the reconstruction of all for this thesis relevant objects and explains how they can be distinguished from each other.

4.1 Jets

As explained in section 2.1.5, jets are the most frequent objects at hadron colliders like the LHC. In terms of reconstruction, jets are in general just clusters of objects. These can be tracks, truth particles from simulation or topo-clusters [41]. The latter are calorimeter cells with a signal-over-noise ratio above a certain threshold which belong topologically together. Before applying any reconstruction algorithm, every object in the calorimeter is a jet. This principle is especially important for quarks and gluons since they form bundles of particles due to hadronization. These traverse the detector and deposit their energy in many different calorimeter cells which then have to be merged to clusters.

For the reconstruction, it is essential that all tracks belonging to a jet need to be included in the final object. In ATLAS a jet clustering algorithm called "anti- k_t algorithm" [42] is used [41]. Two distance measures are needed for this algorithm. The first one is the distance between two entities *i* and *j* from the list of all detected objects and the second one is the distance between *i* and the beam *B*. The algorithm uses a radius parameter *R* defining the size of the cone in which calorimeter clusters are added to the final jet. This parameter can be chosen individually. For ATLAS, R = 0.4 was determined to be the best value. As starting point of the anti- k_t algorithm, the distances d_{ij} and d_{iB} for all combinations of entities are calculated. Afterwards the distances are sorted by their value and the smallest one is considered. If it is d_{ij} , the entities *i* and *j* are merged together and the list is updated. If d_{iB} is the smallest, the object *i* is considered as jet and will be deleted from the list. The distances are then recalculated and the procedure is repeated iteratively until the entity list is empty.

As explained above, jets are built out of topo-clusters by using calorimeter cells with a signal-to-noise ratio over a certain threshold as seed cells and adding neighbouring cells exceeding another (lower) signal-to-noise ratio [43]. These topo-clusters are reconstructed at the EM scale and have to be calibrated in several steps to the Jet Energy scale (JES) [41, 44].

4.2 B-jets

Jets containing a *b*-quark are called "*b*-jets". In terms of reconstruction they are special compared to lighter jets since *b*-quarks have a longer lifetime compared to lighter quarks due to the CKM suppression. For instance, a *b*-quark with a transverse momentum of 50 GeV travels approximately 3 mm [45] before it decays. This leads to a secondary vertex with a non-negligible displacement with respect to the primary one, which can be used to distinguish *b*-jets from light jets. The pixel detector was upgraded by installing the *Insertable B-layer* [36] in order to achieve a better resolution and separation power for *b*-jets. One approach for *b*-jet identification, also called "*b*-tagging", is to determine the impact parameter in the transverse plane and in the longitudinal direction, d_0 and z_0 , which is the closest distance between the reconstructed jet track and the primary vertex. The impact parameter information are used for the *IP3D* tagging method [45]. Another approach is to reconstruct the secondary vertices directly what is exploited by the *SV algorithm* and the *JetFitter algorithm* [45]. The first one tries to reconstruct an inclusive secondary vertex, whereas the latter one reconstructs the whole *b*-hadron decay chain. All these algorithms are combined in the *MV1 tagger* [45], a neural network differentiating between *b*-jets and light jets.

4.3 Tau leptons

In the context of ATLAS, "tau leptons" denote only hadronically decaying tau leptons. Due to the short lifetime of tau leptons, leptonically decaying ones cannot be destinguished from prompt light leptons coming for example from *W* boson decays. The only reconstructable part of the tau lepton is the "visible" hadronic part of the decay since the neutrino is not detectable in ATLAS. The information about the tau lepton reconstruction and identification used in this section are taken from [46].

The $\tau_{had-vis}$ reconstruction algorithm in Run-II is basically the same as in Run-I described in [47] but with some modifications. In general tau leptons are nothing but "ugly" jets. Jet candidates with $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$, generated with the anti- k_t algorithm with R = 0.4, are used as seeds for the $\tau_{had-vis}$ reconstruction algorithm. For calibration purposes the momentum of the tau candidates have to be scaled to the total energy of the topo-clusters within $\Delta R < 0.2$. Afterwards, a vertex association is done in order to find the real tau vertex (TV). This reduces the impact of pile-up effects resulting in a higher reconstruction efficiency.

Tau leptons decay mainly into one or three charged pions and a number of neutral pions. The charged ones leave tracks in the inner detector. These have to be associated to the $\tau_{had-vis}$ candidates in order to differentiate the decay modes of the tau lepton and reconstruct it correctly. The associated tracks need to have an angular distance within the core region of $\Delta R < 0.2$ around the $\tau_{had-vis}$ direction. In addition, they have to have at least two hits in the pixel detector and at least seven hits together in the pixel detector and the SCT. Furthermore their transverse momentum has to be larger than 1 GeV. To reach a better reconstruction efficiency of 1-prong¹ and 3-prong decays, the tracks have to fulfill requirements on the closest distance to the TV. In the longitudinal direction the condition is $|z_0 \sin \theta| < 1.5$ mm and for the transverse plane it is $|d_0| < 1.0$ mm.

The discrimination of $\tau_{had-vis}$ candidates from jets is not sufficient at reconstruction level. So far, tau leptons are jets with one or three tracks. In order to improve the rejection, all $\tau_{had-vis}$ candidates are further analyzed in an identification step based on Boosted Decision Trees (BDTs) [48] trained on $Z/\gamma^* \rightarrow \tau\tau$ events as signal and dijet events as background. This is done seperately for 1-prong and 3-prong candidates. In the following some identification variables are described which are used as

¹ "Prong" denotes the number of visible tracks belonging to the tau candidate.

discriminating variables in the BDTs. They are based on information of tracks and topo-clusters in the core or isolation region ($0.2 < \Delta R < 0.4$) [47]. The whole list can be found in [46].

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

Track radius $(R_{\text{track}}^{0.2})$: p_{T} -weighted ΔR distance of tracks in the core region associated to the direction of the $\tau_{\text{had-vis}}$ candidate.

Maximum ΔR (ΔR_{Max}): Maximum ΔR between tracks in the core region associated with $\tau_{\text{had-vis}}$ candidates and the $\tau_{\text{had-vis}}$ direction.

Track mass (m_{track}) **:** Invariant mass of all tracks in the core and isolation regions. A pion mass is assumed for each track.

Fig. 4.1 shows the distributions of some of these BDT input variables for the signal and the background sample used for the BDT training. All these variables exhibit a clear separation between tau leptons and jets. The central energy fraction (Fig. 4.1(a)) and the track mass (Fig. 4.1(c)) are approximately flat for jets and the events are distributed over the whole range while for tau leptons a peak can be observed. For the maximum ΔR used in the 3-track $\tau_{had-vis}$ candidate (Fig. 4.1(b)), the signal events tend to have lower values and the background events to have higher ones. In reality, however, the distinction of jets and tau leptons is very challenging since due to the much higher production cross-section of jets compared to all other particles (c.f. Fig. 2.7). Hence the jet distribution exceeds the one of the signal events making the separation difficult. This leads to probabilities of up to 10 % [19] for jets to be identified as a tau lepton.

The BDTs have a certain signal efficiency, which is defined as the fraction of 1-track (3-track) true hadronic tau leptons being reconstructed as 1-track (3-track) hadronic tau leptons and which additionally pass the tau identification criteria. Depending on this efficiency, the three working points *loose, medium* and *tight* are defined which are approximately $p_{\rm T}$ -independent [46]. In this thesis, "loose", "medium" and "tight" tau leptons refer to these working points. Besides the jet BDT, also an electron BDT is needed to distinguish tau leptons from electrons. This BDT has also the three working points mentioned above. A tau lepton is referred to as "baseline" if it has a transverse momentum of more than 20 GeV, has one or three tracks and fulfills a kinematic selection with $|\eta| < 2.5$. A "signal" tau lepton has to satisfy also the loose ID criteria.

4.4 Missing transverse momentum

Momentum conservation is one of the fundamental physics laws. At the LHC, protons are collided which are composite particles. Thus the proton momentum fraction of the partons involved in the hard interaction process is unknown. As a consequence it is impossible to know how large the longitudinal momentum is in the final event. The transverse momentum, however, is zero before the collision and thus the sum of the transverse momenta of all particles in the final state has to be zero as well. However, in reality the sum of all reconstructed momenta is usually non-zero. This causes missing transverse energy $\not \!$ which is the magnitude of the missing transverse momentum $\not \!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!$ on the event. This can have



(c) *m*_{track}

Figure 4.1: Distributions of some BDT input variables for the signal (red) and the background (black) sample used for the BDT training for the tau lepton identification. Shown are (a) the central energy fraction (f_{cent}), (b) the maximum ΔR (ΔR_{max}) used in the 3-track $\tau_{had-vis}$ candidate and (c) the track mass (m_{track}). Plots taken from [47].

different reasons: the first one is that the detector cannot measure all particles exactly (c.f. section 3.2.3). Those mis-measurements lead to fake- $\not\!\!E_T$. Another source of momentum imbalances are neutral weakly interacting particles which cannot be detected with ATLAS. Neutrinos are such particles in the SM, but also SUSY predicts particles that cannot be detected directly. Both are sources of true- $\not\!\!E_T$.

The missing transverse momentum reconstruction uses information from tracks in the inner detector and the deposited energy in the calorimeter. Since $\not\!\!E_T$ cannot be reconstructed directly, it has to be determined out of the other particle objects. Missing transverse energy is defined as the negative vectorial sum of the transverse momenta of all reconstructed physics objects in the event and can be calculated via [49]

$$\not\!\!E_{\rm T} = \sqrt{\not\!\!E_x^2 + \not\!\!E_y^2}$$

$$E_{x(y)} = E_{x(y)}^{e} + E_{x(y)}^{\gamma} + E_{x(y)}^{\tau} + E_{x(y)}^{\text{jets}} + E_{x(y)}^{\mu} + E_{x(y)}^{\text{soft}}$$

from the missing energy components of the different objects in the respective direction. Here " τ " denotes the visible part of hadronically decaying tau leptons and $\mathcal{I}_{x(u)}^{\text{soft}}$ is the "soft term" of the missing energy
including all signals not belonging to any physics object [49]. The tracks and deposited energy are assigned to one reconstructed particle object to avoid double-counting.

Another relevant quantity in this context is the sum of transverse energies of all objects in an event which can be calculated analogously to the missing transverse energy by [49]

$$\sum E_{\rm T} = \sum p_{\rm T}^{e} + \sum p_{\rm T}^{\gamma} + \sum p_{\rm T}^{\tau} + \sum p_{\rm T}^{\rm tets} + \sum p_{\rm T}^{\mu} + \sum p_{\rm T}^{\rm soft} \,.$$

4.5 Muons

The muon reconstruction in ATLAS is based on a combination of separate reconstructions in the inner detector and the muon spectrometer [50]. The information used in this section are also taken from this reference.

In the ID, muons are reconstructed like all other charged particles [51, 52]. The reconstruction in the MS starts with a straight-line fit of hits in the MDT and trigger chambers which have to be aligned on a trajectory in the bending plane. RPCs and TGCs are used to measure the perpendicular coordinate and the CSCs to reconstruct segments. After finding the segments, track candidates are generated by matching them. First, only those in the middle layers of the detector are used as seeds, afterwards also those in the inner and outer layers are considered. The segments are then matched to each other and tracks are built out of at least two matched segments. In the transition region between barrel and end-cap, a track can be built using only one well-measured segment. An overlap removal has to be applied since one segment can be used for more than one track. In the end, a global χ^2 fit is performed to check the quality of the association of hits to tracks. Only tracks fulfilling certain criteria on the χ^2 are accepted. Hits with large contributions to χ^2 are removed and the fit is repeated. On the other hand, an algorithm searches for additional hits belonging to the track. If such a hit is found the track candidate is fitted again. The reconstructed muon tracks in the MS are finally combined with those found in the ID. The muons can be grouped into four categories depending on the sub-detectors used for the reconstruction: Combined muons (CB), segment-tagged muons (ST), calorimeter-tagged muons (CT) and extrapolated muons (ME). The definitions of the different categories can be found in [50]. A muon can fulfill the criteria for more than one type and thus one track in the ID can be assigned to more than one muon. To avoid this, the overlap has to be removed. Priority is given to CB, then to ST and in the end to CT muons. For ME muons, the track with the best fit and largest hit multiplicity is chosen.

The four working points *loose*, *medium*, *tight* and *high-p_T* are defined for the muon identification which can be found in [50]. A muon is called "baseline" if it has a transverse momentum of more than 10 GeV and fulfills the "loose" ID conditions. If it has a $p_T > 20$ GeV and satisfies certain isolation criteria, the muon is called "signal". For the isolation two different discriminating variables are defined, one track-based and one calorimeter-based. The first one measures the scalar sum of the p_T of all tracks with $p_T > 1$ GeV in a cone with $\Delta R = \min(10 \text{ GeV}/p_T^{\mu}, 0.3)$ around the transverse momentum direction p_T^{μ} of the muon while the second one uses the sum of all transverse energies of topological clusters within a cone of $\Delta R = 0.2$ around the muon candidate [50].

4.6 Electrons

Electrons are not used explicitly in this thesis, but only indirectly for the overlap removal described in section 4.7. Hence their reconstruction is described only roughly here. More details can be found in [53] where also the information used in this section are taken from.

The reconstruction in the central region ($|\eta| < 2.5$) starts with selecting energy clusters in the ECAL with

a total transverse energy larger than 2.5 GeV. In the next step, tracks with a transverse momentum above 0.5 GeV are extrapolated from the ID to the ECAL and are then assigned to the clusters. Finally, the cluster sizes are corrected to take effects like the deposition of energy in material in front of and behind the ECAL into account.

The reconstruction in the forward region $(2.5 < |\eta| < 4.9)$ is not relevant for this thesis since the used electrons are required to be reconstructed in the central region. Details can again be found in [53].

4.7 Overlap removal

Altough the reconstruction algorithms for the different particle types have mostly high efficiencies, there is still a non-negligible probability that one particle is reconstructed by more than one reconstruction algorithm. This would lead to double-counting of signatures in the detector and thus wrong physics results. To get rid of this feature, an overlap removal is applied after all reconstruction algorithms are done to make sure that each signal is assigned to only one object. The signatures are removed in the following order:

- Rejection of tau candidates in case of an overlap with an electron or a muon within $\Delta R < 0.2$
- Rejection of jets in case of an overlap with a tau candidate or an electron within $\Delta R < 0.2$
- Rejection of muons in case of an overlap with a jet within $\Delta R < 0.4$
- Rejection of electrons in case of an overlap with a jet within $0.2 < \Delta R < 0.4$

The order depends on the reconstruction efficiencies of the different algorithms. For instance, electrons and muons can be reconstructed very efficiently while a tau lepton is usually also reconstructed as a jet.

CHAPTER 5

Event Simulation

An exact simulation of all relevant signal and SM background processes is one of the most important requirements in particle physics experiments. They are used to make predictions about the expected signal and SM events in a certain phase-space region which can be exploited to find new particles or to calculate exclusion limits by comparing the predictions with the data. Moreover the background composition in data can be investigated this way. In particle physics, a common way to simulate events is using Monte Carlo (MC) techniques.

In this chapter the different steps performed in MC simulations are described. Furthermore the tools for the simulation of the electroweak backgrounds used in this thesis are listed. Afterwards two kinds of re-weighting are mentioned which have to be applied to the simulated events in order to get rid of mismodelling effects. The chapter ends with a discussion about why the multijets background for SUSY searches cannot be simulated with MC. The simulation of signal events is not mentioned in this chapter since no signal prediction is used in this thesis.

5.1 Simulation of electroweak backgrounds

All for this thesis relevant backgrounds originating from electroweak processes like W+jets or Z+jets are generated in MC simulations. They correspond to those from the two analyses presented in [29] and [54]. The simulation starts with the matrix element of the hard interaction process, followed by a phase-space integration. For the resulting particles their decay, the hadronization process and the showering are simulated. Additionally the underlying event has to be simulated. For the latter, as well as for hadronization and showering, phenomenologically models are used since perturbation theory is not applicable here. Afterwards the generated events have to go through a detector simulation in order to simulate the signals produced by the particles in the different sub-detectors. The produced signatures finally undergo the same reconstruction procedure as the data.

Two of the most common software packages for the simulation of events which are used in this thesis as well are PYTHIA [55, 56] and SHERPA [57–59]. These frameworks include all relevant steps of event simulations described above. In this thesis, they are used in the simulation of different backgrounds, for example for the event generation as well as the modelling of the showering and the underlying event, respectively. In addition several tools can be used to further improve the matrix element calculation. The relevant program for this thesis is POWHEG [60–62]. It computes the matrix element up to next-to-leading order (NLO) which leads to an overlap between the matrix element and the showering [18]. To get rid of this the showering is ordered by $p_{\rm T}$ and the matrix elements are calculated up to the first

process at NLO level. The following showering is vetoed then [18].

As explained in section 2.2.2, strong production SUSY events with hadronically decaying tau leptons are typically characterized by a tau lepton, numerous jets with high p_T and large missing transverse energy. An overview about the relevant electroweak background processes which can also produce such signatures is given in Tab. 5.1. Furthermore the generator, the PDF set, the parton showering tool and the underlying event model used for the simulation of the electroweak samples is given in this table.

Channel	Generator	PDF set (tune)	Showering (PDFset)	Underlying event		
W+jets	SHERPA 2.1.1	CT10 [63]	-	SHERPA integrated		
Z+jets	SHERPA 2.1.1	CT10	-	SHERPA integrated		
tī	POWHEG-Box v2 [62]	CT10	PYTHIA 6.428 (CTEQ6L1) [64]	Perugia 2012 tune [65]		
Single top						
Wt-channel	POWHEG-Box v2	CT10	PYTHIA 6.428 (CTEQ6L1)	Perugia 2012 tune		
s-channel	POWHEG-Box v2	CT10	PYTHIA 6.428 (CTEQ6L1)	Perugia 2012 tune		
t-channel	POWHEG-Box v1	CT10f4	PYTHIA 6.428 (CTEQ6L1)	Perugia 2012 tune		
Diboson	SHERPA 2.1.1	CT10	-	SHERPA integrated		

Table 5.1: Overview of the MC generators, the PDF sets, the showering tools and the underlying event simulations used for the production of the electroweak background samples. The information are taken from [29].

5.2 Normalization to recorded luminosity

In the sample production with MC the number of generated events is fixed. Afterwards the MC events have to be normalized to the integrated luminosity recorded with ATLAS to ensure that the simulated events reproduce the distributions in data. Therefore a weight is assigned to each event. This *luminosity weight* can be calculated by the ratio of the integrated data luminosity and the luminosity of the simulated sample. The latter can be computed by dividing the number of produced MC events by the production cross-section of the corresponding process.

5.3 Re-weighting

Since the simulations rely on several theoretical predictions and models, they are not perfect. Hence the simulated events do not necessarily represent the data. In order to compensate these differences, some corrections have to be applied to the MC events.

Two of the most important corrections are the *pile-up re-weighting* and the *b-tag re-weighting*. The first one accounts for differences between the simulated pile-up profile and the real one. In this context also the assumed beam spot size differs from the true one which leads to an incorrect reconstruction of the number of primary vertices [18].

The *b*-tag re-weighting accounts for differences between data and MC in the identification efficiency of jets originating from *b*-quarks. This is especially important in $t\bar{t}$ samples since there *b*-jets are used to ensure that the jet was produced in a top quark decay. More detailed information can be found in [18] where also other corrections are discussed.

5.4 Multijets background

Another relevant background arises from multijets events. These are events produced via QCD processes resulting in multiple jets. Such events can also exhibit signal-like signatures with tau leptons, multiple jets and missing transverse energy (c.f. section 2.2.2). The condition is that jets are mis-measured in the detector and at least one jet is falsely reconstructed as a tau lepton. An example of such an event is illustrated in Fig. 5.1. The probability that an event is affected by both of these effects is small, but due to the large production cross-section of multijets events there is a non-negligible number of such events. In contrast to the electroweak events, it is nearly impossible to simulate multijets events in sufficient number and quality in Monte Carlo simulations as will be explained in section 6.1. Instead they are modelled by a data-driven technique called *Jet Smearing* which is presented in chapter 6.



Figure 5.1: Example Feynman diagram of a multijets process faking a characteristical SUSY signature. Particles which are reconstructed as jets are marked in green, the blue gluon illustrates a jet that is reconstructed as a tau lepton and the red jet is mis-measured in the detector resulting in fake- E_T .

CHAPTER 6

Jet Smearing

As mentioned in chapter 5, it is common practice to generate particle physics events in Monte Carlo simulations. However, this is not possible for all backgrounds. As described in section 5.4, multijets events form such a background. The reason is that QCD processes cannot be simulated properly. This is a consequence of the property of gluons to carry colour charge and subsequently interact among each other. Hence all different kinds of high order loop effects contribute to QCD processes which cannot be computed completely. For many processes, a common approach to remedy this issue is perturbation theory. For QCD processes, however, this is not possible due to the running coupling constant of the strong interaction. For low energies, the coupling strength becomes too large to apply perturbation theory. Therefore, it is nearly impossible to achieve a correct modelling of all interactions taking place between the quarks and gluons inside a proton.

In order to get a reliable multijets background estimation for SUSY searches requiring large missing transverse energies, a data-driven technique called *Jet Smearing* can be used. It selects data events fulfilling certain measurement quality conditions and produces multijets events with artificially increased $\not\!\!\!E_T$ by varying the jet four-momenta in the event according to the measurement response of the detector. Another advantage of data-driven methods is the fact that the artificially created events contain information from data which are correct by construction as will be explained in section 6.1.

In this chapter the principle of the Jet Smearing method is decribed which is used to generate multijets events in this thesis. Furthermore the selection of well measured seed events as well as the measurement of the jet response functions and their modifications are explained.

The general information about the relevant measurements and analyses described in this chapter are basically taken from older analyses for 7 TeV and 8 TeV data and can be found in [66–68].

6.1 Motivation for Jet Smearing

As explained in the introduction of this chapter, it is not possible to generate enough multijets events with sufficient reliability in Monte Carlo simulations. This is also illustrated in the example of dijet MC in Fig. 6.1 which shows the m_T (see section 7.1 for definition) distribution of the leading tau lepton in the multijets control region developed later in chapter 8. The black points represent the data, the red distribution the dijet multijets background and the other colours denote the electroweak backgrounds. All backgrounds are stacked on top of each other, thus the whole SM background should reproduce the data. This is obviously not the case. The SM background is underestimated and especially in the peak at approximately 130 GeV, which comes from different reconstruction thresholds of the tau



Figure 6.1: Transverse mass of the leading tau lepton with dijet MC used for the multijets background.

lepton, a large gap between data and MC can be observed. Since the electroweak backgrounds can be reliably simulated using MC and are subsequently normalized to the data luminosity and finally checked for consistency in dedicated control regions, this deviation is mainly caused by an incorrect multijets background description.

The bad multijets modelling in the dijet MC arises from the characteristics of SUSY signatures. As explained in section 2.2.2, SUSY particles as well as neutrinos leave the detector undetected resulting in large missing transverse energy. Thus usually $\not\!\!\!E_T$ triggers are used in SUSY searches whose efficiency curves reach the plateau at high $\not\!\!\!E_T$ values. This requirement on events to have large $\not\!\!\!E_T$ is the key point why dijet MC is not usable in such analyses. Neglecting neutrinos, the only source for missing transverse energy in dijet events are mis-measurements of at least one of the jets in the detector. These mis-measurements, however, are not large enough to cause sufficient $\not\!\!\!\!E_T$. Hence, due to the large $\not\!\!\!\!E_T$ requirement, only events from the far tail of the $\not\!\!\!\!\!E_T$ distribution are taken where the statistics is quite small. Therefore, as a consequence of the large multijets production cross-section at the LHC, it is not possible to generate a sufficient number of multijets events in the tails of the kinematical distributions with MC. This leads to a small statistics of multijets events and finally to the observed discrepancies between data and SM background.

Another motivation for Jet Smearing is the lacking precision in many MC generators for larger values in kinematical distributions. While the lower regions are in general well modelled, in the far tails often deviations can be observed between data and simulation. As an example a comparison of the H_T (see section 7.1 for definition) modelling from different MC generators is shown in Fig. 6.2(b), where, depending on the generator, the prediction does not fit to the data at high values. However, this causes a severe problem since in the most kinematical distributions especially the tails are of interest in SUSY searches. Moreover also the jet multiplicity modelling is difficult in MC due to initial and final state radiation which cannot be simulated properly. This leads to deviations between data and MC particularly for higher jet multiplicities as illustrated in Fig. 6.2(a).

Another important point is the simulation of jets which are falsely reconstructed as tau leptons. Although the modelling of such fake- τ leptons is better in Run-II than in Run-I, it is still not sufficient. Thus it is

convenient to use the fake- τ information directly from data. This can be achieved with Jet Smearing as well since it is a data-driven method where the multijets events are generated from data events.

For all these reasons no reliable multijets background estimate was available for the analysis in [29]. Instead only a rough estimation of the total number of expected multijets events with an assumed uncertainty of 100 % was used. Thus also no shape information were given.



Figure 6.2: Comparison of different MC generators in the number of jets (left) and the H_T (right) distribution. Figures taken from [69].

6.2 Seed event selection

As explained in section 6.1, the number of events with large missing transverse momentum is very limitted in dijet MC. In constrast to this a large abundance of events with low $\not\!\!E_T$ is available. The idea of Jet Smearing is to produce events with large $\not\!\!E_T$ by taking events with low $\not\!\!E_T$ and increase it artificially. A detailed description of this procedure will be given in section 6.5.

The events entering the Jet Smearing procedure are called *seed events*. They have to fulfill the "baseline" conditions for events described in section 7.2 as a basic requirement. They are selected from data by single-jet triggers with different thresholds¹. Since even after the triggers the event rate is too high to record every event, only a certain fraction of them is written to disk. This fraction depends on the trigger threshold since the p_T spectrum of jets has its maximum at lower values and decreases with increasing momentum. Therefore the event rate is higher for lower thresholds. In order to get the "real" number of events the recorded events are scaled afterwards. This *prescaling* depends on the trigger threshold and the run number to take the individual run conditions into account.

An event which passed one of the single-jet triggers has to be well measured to be selected as seed event. That means they need to have only small missing transverse momenta. This is important because the Jet Smearing method is designed such that jet fluctuations are introduced by the smearing itself and the

¹ The thresholds are at transverse jet momenta of 15 GeV, 25 GeV, 55 GeV, 60 GeV, 85 GeV, 100 GeV, 110 GeV, 150 GeV, 175 GeV, 200 GeV, 260 GeV, 300 GeV, 320 GeV, 360 GeV, 380 GeV and 400 GeV.

response function is measured and constrained in a way that it reproduces the jet response in data as will be explained in the following sections. Thus, the by the smearing artificially generated $\not\!\!\!E_T$ is added to that of the seed events and the resulting value would be too large if the seed events have initially large missing transverse energy. Since the latter is the negative sum of the transverse energies of the particles in an event, it is directly correlated to the p_T of the particles. Hence, cutting directly on $\not\!\!\!E_T$ would lead to a bias in different variables like for example the p_T spectrum of the jets. This is a consequence of the p_T -dependence of the detector resolution which is better for particles with higher transverse momenta. This results in larger energy fluctuations for low- p_T jets and consequently also in larger $\not\!\!\!\!E_T$. Thus cutting directly on the missing transverse energy shifts the p_T spectrum of the jets towards lower values. This impact can be avoided by applying a cut on the $\not\!\!\!\!\!\!\!E_T$ -significance [66]

instead. However, also with a cut on *S* a p_{T} -bias can be observed which originates from reconstruction effects. To eliminate nearly any influence on p_{T} , a subtraction of 8 GeV from the missing transverse energy is recommended by the Jet Smearing group. The final cut value is given by the position of the maximum in the *S* distribution. This ensures a sufficient number of seed events on the one hand and events with jet fluctuations in the gaussian core of the response function (c.f. section 6.3) on the other hand what reduces the probability of double-counting events [66]. The jet response functions are determined separately for light and *b*-jets as will be explained in the next section. Therefore the cut on *S* depends on the number of *b*-jets, N_{b-jet} , in the event to avoid biases on the \not{E}_{T} -significance spectrum [68]. Furthermore the Jet Smearing group recommends an additional cut on \not{E}_{T} over the average p_{T} of the two leading jets, which leads to a further improvement. The final cuts for the seed event selection are taken from the official recommendations of the Jet Smearing group. They are namely

$$S = \frac{\not\!\!E_{\rm T} - 8\,{\rm GeV}}{\sqrt{\sum E_{\rm T}}} < \left(0.5 + 0.1 \times N_{b\text{-jet}}\right) {\rm GeV}^{1/2} \qquad \text{and} \qquad \frac{\not\!\!E_{\rm T}}{\left\langle p_{\rm T}^{\rm jet\,1,2} \right\rangle} < 0.2 \; .$$

6.3 Quantification of energy fluctuations in the detector

After the selection of seed events for the smearing, it is necessary to know how strongly these events have to be modified. In order to reproduce the data it is essential to understand the energy mis-measurements in the detector since these are the source of $\not\!\!E_T$ in multijets events. In a perfect detector, the measured transverse momentum of a particle would be equal to the true one. As described already in section 3.2.3, this is not the case in reality where large fluctuations can be observed. Especially for jets there are several sources for missing transverse energy like dead detector material, high-energetic jets which are not fully stopped in the calorimeter or jets containing neutrinos [67] (c.f. section 3.2.3).

The energy mis-measurements can be quantified in a so-called *jet response function* which is defined by the ratio of the reconstructed p_T and the true p_T of a jet. It is a measure for the probability for a jet to undergo a certain fluctuation in the detector. For the determination of the response function a few assumptions are necessary [66]:

- All sources (true and fake) of jet fluctuations can be combined in one response function which then can be used for all jets
- The $\not\!\!E_T$ in multijets events is mainly caused by jet fluctuations

• The impact of event-wide properties (e.g. the jet multiplicity) on the response function is negligible. Thus Jet Smearing can be applied jet-by-jet.

The response function is measured in several intervals of p_T and η since most of the sources of jet mis-measurements are p_T -dependent and the various detector components have varying resolutions in different pseudorapidity regions. The resulting response functions are merged to a response map in which the jet response is plotted against the p_T intervals. The response map is measured separately for light and b-jets, since the neutrinos in b-jets influence the shape of the response function at lower values.

The jet response is measured in multijets MC events generated with PYTHIA. After the standard object definitions and the overlap removal (for both see chapter 4) also an event and jet cleaning is applied to ensure to use only well resonstructed jets for the response measurement [67]. The response is only calculated for reconstructed jets with a spatial distance of $\Delta R > 1.0$ to other reconstructed jets and $\Delta R < 0.1$ to only one truth jet. Neutrinos in the jet cone are added to the jet as a result of the assumption that all sources of \not{E}_T can be included in one response function. In the end, the jet response

$$R = \frac{p_{\rm T}^{\rm reco}}{p_{\rm T}^{\rm true}}$$

is determined for each jet fulfilling the criteria described above. The response values are stored in different histograms which are binned in intervals of η and the true jet p_T als already mentioned above. The final response map is a two-dimensional histogram with the p_T^{true} intervals on the x-axis and the corresponding response function on the y-axis. As examples the final jet response map for the smearing of light jets within this thesis and the corresponding response function for jets with 100 GeV $< p_T^{jet} < 120$ GeV are shown in Fig. 6.3(a) and Fig. 6.3(b), respectively. Both are extracted from the used version of the Jet Smearing tool². The response function has a gaussian core, coming from statistical fluctuations in the deposited calorimeter energy, and non-gaussian tails due to neutrinos and effects like punch-though, shower leakage and defect detector parts. The binning in intervals of the true jet p_T is done to avoid migration effects caused by the falling p_T spectrum.



Figure 6.3: (a) Jet response map and (b) jet response function for $100 \text{ GeV} < p_T^{\text{jet}} < 120 \text{ GeV}$ used within this thesis for the smearing of light jets.

² The used version is JETSMEARING-01-00-24.

6.4 Modification of the jet response function

The jet response function is measured in MC events. By comparing the final pseudo-data events produced with Jet Smearing with the data, it turns out that the pseudo-data does not reproduce the data perfectly. Therefore the MC jet response function has to be modified to agree with the jet response in data. There are two analyses to modify the jet response [66]: the *dijet analysis*, described in section 6.4.1, to widen the response by applying an additional $p_{\rm T}$ -dependent Gaussian smearing, and the "Mercedes" analysis, explained in section 6.4.2, to modify the non-Gaussian low tails.

It should be noted that neither the measurement nor the modification of the jet response functions are performed for each analysis individually, but the constrained response functions are provided by the Jet Smearing Tool. In this chapter, some parameter values are listed which are used in the Jet Smearing tool version relevant for this thesis. These are explicitly labelled.

6.4.1 Dijet analysis

In dijet events the transverse momenta of the two jets should be perfectly balanced assuming an ideal detector, meaning they are aligned back-to-back and have both the same magnitude. Due to mismeasurements this is not the case in reality. The dijet analysis exploits this asymmetry to quantify statistical jet fluctuations from calorimeter mis-measurements which dominate the gaussian core of the response function. However, the core is broader in data than in MC. To account for this, the response function is widened by an additional smearing with a Gaussian with mean one and a width σ_{corr} (p_T) which is calculated in the dijet analysis. The description of the method is taken from [66].

The events used in the dijet analysis are required to have two jets being almost back-to-back with transverse momenta above certain thresholds for leading and sub-leading jet³. The p_{T} -asymmetry in such events is defined by

$$A = \frac{p_{\mathrm{T},1} - p_{\mathrm{T},2}}{p_{\mathrm{T},1} + p_{\mathrm{T},2}}$$

with $p_{T,1}$ and $p_{T,1}$ being the transverse momentum of the leading and the sub-leading jet, respectively. A has a gaussian shape with a width of [70]

$$\sigma_{A} = \frac{\sqrt{(\sigma(p_{\mathrm{T},1}))^{2} + (\sigma(p_{\mathrm{T},2}))^{2}}}{\langle p_{\mathrm{T},1} + p_{\mathrm{T},2} \rangle} .$$
(6.1)

Under the assumption that both jets have nearly the same rapidity, their $p_{\rm T}$ uncertainties are roughly equal, leading to $\sigma(p_{\rm T,1}) = \sigma(p_{\rm T,2}) = \sigma(p_{\rm T})$. With $\langle p_{\rm T,1} + p_{\rm T,2} \rangle = 2 \langle p_{\rm T,avg} \rangle = p_{\rm T}$, Eq. (6.1) can be approximated by [70]

$$\sigma_A \approx \frac{\sigma(p_{\rm T})}{\sqrt{2}p_{\rm T}}$$
 (6.2)

For both, data and MC, the asymmetry is measured in intervals of p_T . Afterwards, Gaussians with mean zero and widths σ_A are fitted to them. The σ_A are then plotted against the transverse momentum and a

³ No cut values found for the used version of the Jet Smearing tool for a centre-of-mass energy of 13 TeV. For the dijet analyses performed for energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, respectively, values can be found in Table 6.1 in [68].

function of the form

$$\sigma_A = \frac{a}{p_{\rm T}} + \frac{b}{\sqrt{p_{\rm T}}} + c \tag{6.3}$$

is fitted to the distribution [66] where a, b and c are parameters determined in the fit. As mentioned already, the MC jet response function is further smeared with a Gaussian with the width σ_{corr} to reproduce the jet response in data. Hence, using the convolution of two Gaussians, it follows that [66]

$$\left(\frac{\sigma_{A,\text{data}}(p_{\text{T}})}{p_{\text{T}}}\right)^{2} = \left(\frac{\sigma_{A,\text{MC}}(p_{\text{T}})}{p_{\text{T}}}\right)^{2} + \sigma_{\text{corr}}^{2}(p_{\text{T}}) .$$

With Eq. (6.2) this can be rewritten to

$$\left(\sqrt{2} \times \sigma_{A,\text{data}}\right)^2 = \left(\sqrt{2} \times \sigma_{A,\text{MC}}\right)^2 + \sigma_{\text{corr}}^2(p_{\text{T}}).$$

Solving for $\sigma_{\rm corr}(p_{\rm T})$ gives finally

$$\sigma_{\rm corr} \left(p_{\rm T} \right) = \sqrt{2} \times \sqrt{\sigma_{A,\rm data}^2 - \sigma_{A,\rm MC}^2} \,. \tag{6.4}$$

The width σ_{corr} of the correcting Gaussian can thus be determined from the fits of Eq. (6.3) to data and MC, respectively. The calculation of the uncertainty of σ_{corr} is decribed in [66]. The parameter values used within this thesis in Eq. (6.3) to calculate σ_{corr} for the p_{T} smearing can be found in Tab 6.1. They were obtained from fits using 6 fb⁻¹ of 13 TeV data recorded in 2015 and 2016.

Parameter	Value
a _{data}	-4.57202
$b_{\rm data}$	1.05512
Cdata	0.0130301
a _{MC}	-5.13494
$b_{\rm MC}$	1.09213
c _{MC}	0.0137621

Table 6.1: Parameter values for the calculation of σ_{corr} used in the dijet analysis. The values are taken from the used version of the Jet Smearing Tool. They were obtained from fits for 6 fb⁻¹ of 2015 and 2016 13 TeV data.

6.4.2 "Mercedes" analysis

The "Mercedes" analysis uses events with three jets where one jet is aligned parallel or anti-parallel to the missing transverse energy. With these information the tails of the jet response function can be constrained. In contrast to dijet events, in Mercedes events the \not{E}_T can be considered to originate from fluctuations of the associated jet. This also allows a distinction between fluctuations coming from the high and the low tail of the response function. The information used in this section are taken from [66]. Events considered in the Mercedes analysis need to have at least three jets with transverse momenta above certain thresholds for the different jets whereas the leading jet threshold is chosen to be very high. This is necessary to reject other backgrounds and prescaled data events which would influence the tails of the response function. In addition, the events have to fulfill certain requirements on \not{E}_T which can

be found in [66] for the analysis performed in this reference⁴. To ensure that the \not{E}_T is unambiguously assigned to one jet, all jets in the event are sorted by their angular distance $\Delta \phi$ to the \not{E}_T and only those events are used in which the closest jet has a certain maximal angular distance $\Delta \phi^{\text{match}}$ to the \not{E}_T and is isolated in ϕ from other jets by at least $\Delta \phi^{\text{isol}}$. It turned out that in many events the jet with the largest angular distance to the \not{E}_T is nearly back-to-back to the jet being closest to the \not{E}_T . Since in such events a clear assignment of the \not{E}_T to only one jet is not possible, those events are rejected.

All remaining events are finally used to determine the jet reponse in the tails of the response function. The response *R* of the jet associated with the $\not\!\!E_T$ is vectorially defined by [66]

$$R = \frac{\vec{p}_{\rm T}^{\rm reco} \cdot \vec{p}_{\rm T}^{\rm true}}{\left| \vec{p}_{\rm T}^{\rm true} \right|^2}$$

Under the assuption that the true transverse momentum of the respective jet is approximately the sum of the reconstructed p_T and the $\not\!\!E_T$,

$$\vec{p}_{\rm T}^{\rm true} \approx \vec{p}_{\rm T}^{\rm reco} + \vec{E}_{\rm T},$$

the jet response becomes

$$R \approx \frac{\vec{p}_{\rm T}^{\rm reco} \cdot \left(\vec{p}_{\rm T}^{\rm reco} + \vec{E}_{\rm T}\right)}{\left|\vec{p}_{\rm T}^{\rm reco} + \vec{E}_{\rm T}\right|^2}$$

As in the core region also in the tails the jet response does not exactly reproduce the response in data. Hence the response functions are fitted with a functional form. By varying a parameter in the functional form the tails can be modified until they match the ones in data. The best functional form turned out to be a Gaussian [66] which is fitted in a certain range that is given by the interval from zero to the crossing of the core and the lower tail region. The crossing point itself can be determined by fitting a Crystal Ball function [66] to the response function in the range of the lower tail and the core [66]. The tail of *R* can then be changed by multiplying the width σ_{tail} of the fitted Gaussian by a factor $\Delta \sigma_{tail}$ which can be determined by applying χ^2 fits to a series of pseudo-data samples produced with different $\Delta \sigma_{tail}$.

6.4.3 Smearing of the azimutal angle

Besides the momentum of the jets also their azimutal angle ϕ has to be smeared to compensate for biases from the seed event selection and the smearing process. The influence of the smearing itself is caused by the resolution of the inner detector which scales with the momentum of a particle. If the p_T of a jet is smeared down to lower values, the ϕ resolution in data decreases. This leads to differences between the data and the smeared events [68]. Another bias is introduced by the seed selection where only events with low \not{E}_T are selected which requires the jets in a dijet event to be nearly back-to-back. This topology stays unchanged even after the p_T smearing since only the magnitudes of the jet momenta are affected but not their angles. However, after the smearing at least one jet can show large fluctuations which means that the energy in the event is badly measured. This can rotate the jet axis [68].

To remedy these effects also the ϕ of each jet is smeared by adding a random number drawn from a Gaussian with mean zero and width $\sigma_{\Delta\phi}$. The latter can be determined analogously to the σ_{corr} described in section 6.4.1 by fitting Eq. (6.3) to the $|\pi - \Delta\phi(j_1, j_2)|$ distribution for different p_{T} intervals.

⁴ No current values found.

The parameter values used within this thesis in Eq. (6.3) to calculate the σ_{corr} for the p_T smearing can be found in Tab. 6.2. They were obtained from fits using 6 fb⁻¹ of 13 TeV data recorded in 2015 and 2016.

Parameter	Value
a _{data}	6.42374
$b_{\rm data}$	0.177568
c _{data}	0.0373842
a _{MC}	3.53819
$b_{\rm MC}$	0.19432
CMC	0.0367071

Table 6.2: Parameter values for the $\sigma_{\Delta\phi}$ calculation used within this thesis for the jet ϕ smearing. The values are taken from the used version of the Jet Smearing Tool. They were obtained from fits for 6 fb⁻¹ of 2015 and 2016 13 TeV data.

6.5 The Jet Smearing technique

After measuring and constraining the jet response function, the actual smearing of the selected seed events can now be applied. This means that the four-vectors of all jets in an event are modified. This is done by multiplying them by a random number drawn from the constrained jet response function. In addition to the p_T smearing, usually also the azimutal angles ϕ of the jets are modified by a Gaussian with the width $\sigma_{\Delta\phi}$ determined in section 6.4.3. The whole smearing procedure is repeated several times for each seed event resulting in many different events. This is necessary in order to get enough statistics. It should be noted that in principle the number of smearing iterations per seed event can be chosen freely. However, smearing one seed event too often can result in double-counting of events. A maximal number of approximately 10 000 smearing iterations is recommended by the Jet Smearing group. An illustration of the Jet Smearing principle is shown in Fig. 6.4.

For reasons of clarity the whole Jet Smearing procedure is summarized again in the following. It includes four steps:

- 1.) Selection of well measured seed events with low $\not\!\!E_T$ from data. For details see section 6.2.
- 2.) Measurement of the jet response function in MC. For details see section 6.3.
- 3.) Modification of the jet response function. It is constrained with data to achieve a good agreement between data and MC. For details see section 6.4.
- 4.) Smearing of the seed events. Each jet four-vector in the event is multiplied by a random number drawn from the jet response function. Usually also a smearing of the azimutal angle φ is applied. Each seed event is smeared several times.



Figure 6.4: Principle of the Jet Smearing technique. The four-vectors of all jets (blue arrows) in a seed event are multiplied with a random number drawn from the modified jet response function. This results in artificially increased \not{E}_T (red arrows). This procedure is repeated several times per seed event.

CHAPTER 7

Event Selection

One key part of every analysis searching for new physics by investigating a certain model is the knowledge about the predicted signatures and how to search for them. In this context it is also essential to know how to separate this signal from events arising from other, already known background processes. At the beginning of this chapter some important variables are defined. Afterwards a common preselection

is described which is applied to every event in order to ensure to consider only well reconstructed events, recorded under optimal detector conditions. In addition the preselection rejects all events which exhibit basic signatures that are not relevant for the this analysis. In the end the concept of signal, control and validation regions is explained and the relevant regions for this thesis are defined.

7.1 Definitions of important variables

In this thesis, several variables are used to quantify event properties or to cut on in order to apply different selections. In this section the most important variables are defined and explained which are not a priori clearly understandable by their names or symbols.

- Total transverse hadronic activity $H_{\rm T}$ caused by all tau leptons and jets in an event: $H_{\rm T} = \sum_i p_{\rm T}^{\tau_i} + \sum_j p_{\rm T}^{jet_j}$
- "Stransverse mass" m_{T2} of a 2τ system in an event [71]: $m_{T2}^{\tau\tau} = \sqrt{\min_{\vec{p}_T^a + \vec{p}_T^b = \vec{p}_T} \left(\max\left[m_T^2(\tau_1, \vec{p}_T^a), m_T^2(\tau_2, \vec{p}_T^b) \right] \right)}$ \vec{p}_T^a and \vec{p}_T^b denote the momenta of the invisible decay products *a* and *b* of the two tau leptons. m_{T2}
- Effective mass m_{eff} : $m_{\text{eff}} = H_{\text{T}} + \not\!\!\!E_{\text{T}}$
- The sum of the transverse masses of the two leading tau leptons and all jets in an event, m_T^{sum} : $m_T^{\text{sum}} = \sum_{i=1,2} m_T^{\tau_i} + \sum_j m_T^{\text{jet}_j}$

is a measure for the transverse mass of a particle decaying into two invisible particles.

7.2 Baseline selection

All data and background events in this thesis have to fulfill certain quality conditions. These ensure that only events are considered in the analysis which are not affected by any reconstruction or detector issues. This preselection contains the following requirements:

- The event was recorded in a data taking period listed on the *Good Runs List* (GRL). This guarantees that the event was recorded under optimal detector conditions.
- The event is not affected by any known detector issues.
- The event has a reconstructed primary vertex with at least five tracks.
- The event does not contain cosmic muon candidates.
- The event does not contain badly reconstructed muons or jets.

For this analysis also at least one tau lepton and one jet are required in each event. Together with the quality cuts mentioned above, they form the *baseline* selection for this thesis.

7.3 Signal, control and validation regions

In every analysis searching for new physics like SUSY, the data is compared to the expected SM background and the signal prediction. Depending on the SUSY model to be investigated, different signals are expected. Since the number of expected signal events is usually very low in contrast to the background, dedicated signal regions (SRs) are designed by cuts on different variables in the phase-space to increase the signal-to-background ratio. In order to calculate reliable discovery or exclusion limits, the backgrounds in the SRs need to be well modelled. Therefore a *control region* (CR) is developed for each significant background. They are designed in a way that the respective background is maximally enriched with a high purity of the respective background while being orthogonal to each signal region. This means any phase-space overlap between the CRs and SRs has to be avoided. In addition the CRs have to be signal-free. In control regions, the backgrounds are investigated individually and the shapes can be checked for consistence. Since no signal events are expected in control regions, all backgrounds stacked on top of each other should reproduce the data. However, due to mis-modelling effects in Monte Carlo this is usually not the case. To improve the agreement the backgrounds are fitted simultaneously in their corresponding control region. Afterwards the modelling in the control regions is cross-checked in validation regions (VRs) which are usually defined by changing the cut on one variable in the phase-space to the respective cut value in a signal region. If the agreement between the data and the backgrounds is good enough, the latter can finally be extrapolated from the control regions into the signal regions by applying transfer factors.

In this thesis, two different analyses are considered, one using the 3.2 fb⁻¹ of 13 TeV data collected in 2015 [29] and the other one for the full 36 fb⁻¹ of 13 TeV data recorded 2015 and 2016 [54]. Therefore two sets of control, validation and signal regions have been invented, one for each analysis. They are listed in the tables 7.1 to 7.3 for the analysis with the 2015 data and in Tab. 7.4 to Tab. 7.5 for the analysis with the full 2015 and 2016 data. Since at the time of writing this thesis no validation regions are yet defined for the analysis with the 36 fb⁻¹ dataset, only the signal and control regions are given. The final validation regions can be found in [54]. The definitions are taken from the references cited above and correspond to those of the 2τ channel. In these references the motivation for the cuts in the different regions can be found as well.

In this chapter only the control regions for the electroweak backgrounds are listed. The development of a multijets control region is one of the main parts of this thesis and is thus explained in more detail in chapter 8.

	Compressed SR	High-Mass SR	GMSB SR				
Trigger plateau	$E_{\rm T} > 180$)GeV					
Tau leptons	$N_{\tau}^{\text{loose}} \ge 2, \ p_{\mathrm{T}}^{\tau} > 20 \mathrm{GeV}$						
Multijets rejection	$\Delta \phi \left(jet_{1,2}, \vec{p}_{\rm T} \right) \ge 0.4$						
$m_{\rm T}^{ au_1} + m_{\rm T}^{ au_2}$	_	> 350 GeV	> 150 GeV				
$H_{\rm T}$	_	> 800 GeV	> 1 700 GeV				
N _{jet}	≥ 2	≥ 3	≥ 2				
$m_{\mathrm{T2}}^{ au au}$	> 60 GeV	—	—				
m ^{sum} _T	> 1 400 GeV	_	_				

Table 7.1: Signal regions for the 2τ channel of the analysis with 3.2 fb⁻¹ of $\sqrt{s} = 13$ TeV data collected in 2015. $H_{\rm T}$ is the hadronic activity in the transverse plane, $m_{\rm T}$ denotes the transverse mass of a particle, $m_{\rm T2}^{\tau\tau}$ the stransverse mass of the two tau leptons and $m_{\rm T}^{\rm sum}$ is the sum of the transverse masses of all jets and tau leptons in the event. The definitions of the variables can be found in section 7.1. The definitions of these regions are taken from [29].

	kinematic CR		fake tau	I CR	true tau	CR			
	W+jets	top	W+jets	top	W+jets	top	Z+jets		
Trigger plateau			₿ _T)GeV					
Tau leptons	N_{τ}^{loose}	= 0	$N_{\tau}^{\text{loose}} = 1$				$N_{\tau}^{\text{loose}} = 2$, opposite charge		
Light leptons		N_{μ}	= 1		$N_{\mu} =$	0	_		
N _{jet}	≥ 3		≥ 2	r.	≥ 3		≥ 2		
N _{b-jet}	= 0	$= 0 \ge 1 = 0$			= 0	≥ 1	= 0		
Multijets rejection		0.4	$<\Delta\phi(\text{jet}_{1,2})$	$(\underline{p}_{2}, \vec{p}_{T}) <$	< 2.9	$\Delta \phi \left(\text{jet}_{1,2}, \vec{p}_{\text{T}} \right) > 0.4$			
CP solution	$H_{\rm T} < 1200{\rm GeV}$						$H_{\rm T} < 800 {\rm GeV}$		
	_		$m_{\rm T}^{\mu} < 100$	$m_{\rm T}^{ au_1} + m_{\rm T}^{ au_2} < 150 { m GeV}$					

Table 7.2: Control regions for the 2τ channel of the analysis with 3.2 fb⁻¹ of $\sqrt{s} = 13$ TeV data collected in 2015. $H_{\rm T}$ is the hadronic activity in the transverse plane and $m_{\rm T}$ denotes the transverse mass of a particle. The definitions of the variables can be found in section 7.1. The definitions of these regions are taken from [29].

	W+jets VR	top VR	Z+jets VR				
Trigger plateau	$E_{\rm T} > 180 {\rm GeV}, \ p_{\rm T}^{\rm jet_1} > 120 {\rm GeV}$						
Tau leptons		$N_{\tau}^{\text{loose}} \ge 2$					
N _{jet}	≥ 2						
N _{b-jet}	= 0	≥ 1	= 0				
Multijets rejection	$\Delta \phi (j$	≥ 0.4					
H_{T}	< 800 C	> 800 GeV					
$m_{\rm T}^{ au_1} + m_{\rm T}^{ au_2}$	> 150 C	150 GeV < 150 GeV					
$m_{\mathrm{T2}}^{ au au}$	< 60 GeV –						

Table 7.3: Validation regions for the 2τ channel of the analysis with 3.2 fb⁻¹ of $\sqrt{s} = 13$ TeV data collected in 2015. $H_{\rm T}$ is the hadronic activity in the transverse plane, $m_{\rm T}$ denotes the transverse mass of a particle and $m_{\rm T2}^{\tau\tau}$ the stransverse mass of the two tau leptons. The definitions of the variables can be found in section 7.1. The definitions of these regions are taken from [29].

	Compressed SR	High Mass SR				
Trigger plateau	$\not\!$					
	$p_{\rm T}^{\rm jet_1} > 12$	20 GeV				
	$N_{\rm jet}$ >	> 1				
	$p_{\rm T}^{\rm jet_2} > 25 { m GeV}$					
Multijets rejection	$\Delta \phi \left(jet_1, \vec{p}_T \right) > 0.4$					
	$\Delta \phi \left(\text{jet}_2, \vec{p}_T \right) > 0.4$					
Tau leptons	$N_{\tau}^{\text{medium}} \ge 2$					
		$m_{\rm T}^{\tau_1} + m_{\rm T}^{\tau_2} > 350 {\rm GeV}$				
	$m_{\rm T2} > 60 {\rm GeV}$					
General event properties	$H_{\rm T} < 1100{\rm GeV}$	$H_{\rm T} > 1 \ 100 {\rm GeV}$				
	$\sum m_{\rm T}^{\rm taus, jets} > 1600{\rm GeV}$					

Table 7.4: Signal regions for the 2τ channel of the analysis with 36 fb⁻¹ of $\sqrt{s} = 13$ TeV data collected in 2015 and 2016. $H_{\rm T}$ is the hadronic activity in the transverse plane, $m_{\rm T}$ denotes the transverse mass of a particle, $m_{\rm T2}^{\tau\tau}$ the stransverse mass of the two tau leptons and $\sum m_{\rm T}^{\rm taus, \, jets}$ the sum of the transverse masses of all tau leptons and jets. The definitions of the variables can be found in section 7.1. The definitions of these regions are taken from [54].

Z ightarrow au au						$N_{\tau}^{\text{medium}} = 2, \text{ OS}$			= 0					$m_{ m T}^{ au_1} + m_{ m T}^{ au_2} < 150{ m GeV}$	$m_{ m T2} < 70~{ m GeV}$			
Z ightarrow w	.80 GeV 120 GeV	t > 1 25 GeV	$\vec{p}_{\mathrm{T}} > 0.4$	$\vec{p}_{\rm T}$ > 0.4		=]	1	$N_{\mu} = 0$	$N_{b-\mathrm{jet}}$	300 GeV			$100 \mathrm{GeV} \le m_{\mathrm{T}}^{\tau} < 200 \mathrm{GeV}$	•		$\Delta \phi \left(\operatorname{jet}_{1}, \vec{p}_{\mathrm{T}} \right) > 2.0$	$\Delta \phi \left(au_{1}, \vec{p}_{\mathrm{T}} ight) > 1.0$	$I_{\rm T}/m_{\rm eff} > 0.3$
top/W fake tau	$E_{\rm T} > 1$ $p_{\rm T}^{\rm jet_1} >$	$p_{\mathrm{T}}^{\mathrm{jet}_2} > p_{\mathrm{T}}^{\mathrm{jet}_2}$	$\Delta \phi (jet_1,$	$\Delta \phi$ (jet ₂ ,	0.2)	$N_{ au}^{ ext{medium}}$		$N_{\mu} = 1$		$H_{\rm T} < 8$	$t_{\rm T} < 300{ m GeV}$	$m_{\mathrm{T}}^{\mu} < 100 \mathrm{GeV}$						
top/W true tau					$\left(\operatorname{jet}_{1}, \vec{p}_{\mathrm{T}}\right) < (\pi - 0)$		> 2	$N_{\mu} = 0$	$N_{b-\text{jet}} = 0/ \ge 1$		1		$m_{\rm T}^{\tau} < 80 {\rm GeV}$					
top/W kinematic					$\nabla \phi$	$N_{\tau}^{\text{medium}} = 0$	Njet	$N_{\mu} = 1$				$m_{\mathrm{T}}^{\mu} < 100 \mathrm{GeV}$						
	Trigger plateau		Multijets rejection			Tau leptons	Additional jets	Light leptons	W/Top separation	CR cuts								

Table 7.5: Control regions for the 2τ channel of the analysis with 36 fb⁻¹ of $\sqrt{s} = 13$ TeV data collected in 2015 and 2016. H_T is the hadronic activity in the transverse plane, m_T denotes the transverse mass of a particle and m_{eff} is the effective mass. The definitions of the variables can be found in section 7.1. The definitions of these regions are taken from [54].

CHAPTER 8

Development of a Multijets Control Region

The main goal of this thesis is to provide a reliable multijets background estimate for the analysis presented in [54]. Therefore a normalization factor for the multijets background has to be determined since the previous one is no longer correct after applying the Jet Smearing. This has to be done in a dedicated control region in which the multijets background can be investigated in terms of correct shape modelling and be normalized afterwards. The development of such a multijets control region is described in this chapter.

8.1 Basic cuts of the multijets control region

As mentioned already in section 7.3, a control region has to be designed in a way that the corresponding background is enriched compared to the other ones, meaning it constitutes a large fraction of the overall background.

At the beginning of this thesis¹, the dataset used in the analysis described in [29] was also used for the design of the multijets control region in order to be able to make consistency checks. The dataset corresponds to the $3.2 \,\text{fb}^{-1}$ of 13 TeV data collected with the ATLAS detector in 2015. Although the multijets estimate will finally be used for the analysis in [54] with 36 fb⁻¹, there will be no bias by using the data sample from the previous analysis since the higher luminosity in the second analysis is the only difference between both samples. From the $3.2 \,\text{fb}^{-1}$ dataset, a SUSY11 derivation containing events selected by single-jet triggers is used as input for the Jet Smearing in order to produce a multijets sample by smearing each seed event 500 times.

In the later analysis, every event has to pass the trigger cut, in this case the HLT_xE70 trigger item², as well as the trigger plateau cuts listed in the tables 7.1 to 7.3 which ensure a maximum trigger efficiency. With these cuts, the sample is filtered directly after the smearing resulting in a total number of 166 648 events. Moreover also cleaning cuts, namely a tile veto for jets and tau leptons, have to be applied which reject events with badly measured jets or tau leptons due to defect tile calorimeter modules.

The development of a multijets control region starts with the particle contents. These are constrained by the expected signal which are in this case tau leptons and many jets. As explained already, multijets events can produce signal-like signatures if at least one jet is identified as a tau lepton and other ones are mis-measured in the detector. Therefore one tau lepton and a number of jets are required in the CR selection. It should be noted that the tau lepton is always a fake- τ in this case. It has to fulfill the

¹ In April 2016

² HLT_xE70 means that the event needs to have a missing transverse energy of more than 70 GeV.

criteria of the *loose* working point of the tau identification to ensure a sufficient large statistics. In order to use only generic events with one tau lepton and several jets for the determination of the multijets normalization factor, all events containing muons are rejected. The required particle content almost corresponds to that of the true tau CR which is thus taken as basis for the control region design. Hence the number of jets is chosen to be larger than two. However, all cuts of the true tau CR on kinematic variables are omitted for the present.

Multijets events are by far the most abundant background at hadron colliders due to their large production cross section. In order to get a sufficient signal-over-background ratio in the signal regions and a high purity of the other backgrounds in the electroweak (EW) control regions, it is essential to suppress the multijets background in these regions. This is achieved by rejection cuts on the angular separation between the jet axis and the $\not\!\!E_T$ for the two leading jets, namely

$$\Delta \phi \left(\operatorname{jet}_{1,2}, \mathbb{E}_{\mathrm{T}} \right) > 0.4$$

In order to get as many multijets events as possible in the corresponding control region, the rejection cuts are inverted to

$$\Delta \phi (\operatorname{jet}_1, \not\!\!E_T) < 0.4$$
 or $\Delta \phi (\operatorname{jet}_2, \not\!\!E_T) < 0.4$.

Hence the multijets control region is preliminarily defined by the following cuts:

- HLT_xE70 trigger cut
- Trigger plateau cuts
- Cleaning cuts
- $N_{\tau} = 1$
- $N_{\mu} = 0$
- $N_{\text{jet}} > 2$
- $\Delta \phi (\operatorname{jet}_1, \not\!\!E_T) < 0.4$ or $\Delta \phi (\operatorname{jet}_2, \not\!\!E_T) < 0.4$

Applying this selection together with the trigger and cleaning cuts leads to 136 173 remaining multijets events which corresponds to approximately 81.7 % of all events in the sample. This is a sufficient number of events to obtain reliable results. With the nominal QCD rejection cuts instead of the inverted ones, the number of events passing the selection would be 15 286, i.e. inverting the rejection cuts leads to approximately 8.9 times more multijets events. Furthermore it also increases the relative fraction of the multijets background compared to the electroweak ones. It should be noted that in the control regions of the 2015 data analysis with 3.2 fb⁻¹ [29] (c.f. Tab. 7.1), also events with $\Delta\phi$ (jet₁, $\not\!\!E_T$) < 0.4 < 2.9 are rejected. However, this cut is not applied here to avoid any bias since it is also not considered in the signal and control regions (c.f. Tab. 7.4 and 7.5) of the analysis with the full 36 fb⁻¹ dataset. Another important effect of inverting the QCD rejection cuts is that the multijets control region becomes orthogonal to the other control regions and the signal regions.

The multijets sample was produced with seed events selected from single-jet triggers with different thresholds³ and have thus to be prescale-weighted. The weighting increases the number of multijets events to 166 953 for the selection with the inverted QCD rejection cuts, leading to a factor of approximately 1.2

 $^{^{3}}$ The thresholds can be found in section 6.2

times higher statistics than without applying the prescale weight.

In the next step the shapes of the different distributions are checked for consistency between data and multijets for the preliminary selection. Fig. 8.1 shows the distributions for the transverse momentum $p_T^{\tau_1}$ and $p_T^{jet_1}$ of the tau lepton and the leading jet respectively, the missing transverse energy $\not\!\!\!E_T$, the transverse mass of the tau lepton $m_T^{\tau_1}$ and the azimutal angle separations $\Delta \phi$ (jet₁, $\not\!\!\!E_T$) and $\Delta \phi$ (jet₂, $\not\!\!\!E_T$) between the two leading jets and the $\not\!\!\!E_T$. Since both samples have different statistics, the multijets distributions are normalized to the number of data events.



The origin of the technical problems in the Jet Smearing procedure was unknown at the beginning of the control region design, but since the reconstruction effects seem to affect only events in the left peak which have low $m_T^{\tau_1}$, the first five bins are cut away for the present by requiring $m_T^{\tau_1} > 92$ GeV to get rid of this bias. This way the multijets sample can be used further for the control region design.

Due to the normalization of the multijets events to data the shapes are in good accordance with each other after applying the cut on $m_T^{\tau_1}$ as can be seen in Fig. 8.2. However, although the cut improves the agreement the latter is still insufficient in $p_T^{\text{jet}_1}$ and $\Delta \phi$ (jet₂, $\not\!\!\!E_T$). This has two reasons: the first one is that both are jet quantities which are in the case of multijets events affected by the smearing of the jets. The second one is the missing electroweak background which could at least partially balance out the differences between data and multijets.

To be able to compare the SM prediction with the data the electroweak backgrounds are included as well. Since only their sum and not their composition is relevant here they are merged together to one total electroweak background. Both multijets (red) and electroweak background (green) are stacked in the plots to form the whole SM prediction. The multijets background has to be normalized again since numerous events are produced out of one seed event in the smearing. This results in more multijets events than are available in data and the normalization of the former is no longer valid. For a certain selection the normalization factor ω_{QCD} is calculated by

$$\omega_{\rm QCD} = \frac{N_{\rm data} - N_{\rm EW}}{N_{\rm QCD}} \tag{8.1}$$

with N_{data} being the number of data events, N_{EW} the number of electroweak events and N_{QCD} the number of multijets events passing the selection.

Fig. 8.3 shows the same distributions as Fig 8.2 but with the electroweak backgrounds included. Although the data mainly fits the background expectation within one standard deviation, there are some bins in which data and SM prediction deviate significantly from each other, like the last bin in the $\Delta \phi$ (jet₁, \not{E}_T) distribution. In $p_T^{\tau_1}$ the backgrounds are generally overestimated while in $p_T^{jet_1}$ they are underestimated for lower values and overestimated for higher values. This trend seems to be a systematic shift. Besides this disagreement in some quantities, the fluctuations in the leading jet transverse momentum (c.f. Fig. 8.3(b)) in data imply that the statistics is very small. However, at the moment of designing the multijets control region, there was no data sample with higher statistics available.

8.2 Enlargement of the seed event statistics

Besides the fluctuations in data, another reason for the observed deviations between data and SM prediction could be that the multijets events in the final region were produced from a small number of different seed events. As an approach to further enrich the seed event statistics, objects with looser identification criteria are considered. Therefore a second multijets sample with in total 51 614 620 events





Figure 8.3: Comparison of data (black) and the SM backgrounds for the preliminary multijets control region selection after the cut on $m_T^{\tau_1}$. The SM consists of the multijets (red) and the electroweak (green) backgrounds whereas the latter are merged into one total electroweak background. Shown are the transverse momentum $p_T^{\tau_1}$ and $p_T^{\text{jet}_1}$ of (a) the tau lepton and (b) the leading jet respectively, (c) the missing transverse energy $\not{\!\!\!E}_T$, (d) the transverse mass of the tau lepton $m_T^{\tau_1}$ and the azimutal angle separations (e) $\Delta \phi$ (jet₁, $\not{\!\!\!E}_T$) and (f) $\Delta \phi$ (jet₂, $\not{\!\!\!E}_T$) between the two leading jets and the $\not{\!\!\!E}_T$. The multijets distributions are normalized to the difference between the numbers of events in data and in the EW background.

is produced that contains "baseline" muons and tau leptons instead of "signal" ones as before. The definitions of these baseline and signal objects are given in section 4.3 for tau leptons and in section 4.5 for muons and are summarized in Tab. 8.1. The looser criteria of baseline objects ensure a significant increase in events passing the selections and thus also a higher number of different seed events.

However the usage of a baseline sample has one big disadvantage. The multijets events contain baseline

	muons	tau leptons
	$- p_{\rm T} > 10 {\rm GeV}$	$- p_{\rm T} > 20 {\rm GeV}$
baseline	- loose ID	- kinematic selection
		$(\eta < 2.5, 1/3 \text{ tracks})$
aignal	$- p_{\rm T} > 25 {\rm GeV}$	$- p_{\rm T} > 20 {\rm GeV}$
signal	- isolation	- loose ID

Table 8.1: Definitions of baseline and signal objects for tau leptons and muons.

objects, while in the data and the electroweak samples they contain signal objects. Thus due to the looser quality criteria of baseline objects, the number of multijets events passing a certain selection is generally higher while it stays unchanged for data and the EW backgrounds. This leads to the problem that the events are no longer comparable and subsequently the results are not reliable. In order to prevent this issue, the number of multijets events containing baseline objects has to be scaled to the corresponding number with signal objects for the considered phase-space region. Additionally the events have to be normalized to the difference between data and EW background according to Eq. (8.1):

$$\omega_{\rm QCD} = \frac{N_{\rm data} - N_{\rm EW}}{N_{\rm OCD}} = \frac{\Delta N}{N_{\rm OCD}} \,.$$

For the conversion from baseline objects to signal ones the events have to be scaled by

$$a_{\rm QCD} = \frac{N_{\rm QCD}^{\rm signal}}{N_{\rm QCD}^{\rm baseline}} \; .$$

By combining both factors into one scaling factor

$$k_{\rm QCD} = \omega_{\rm QCD} \times a_{\rm QCD}$$

the normalized number of multijets events with signal objects can be computed from the unnormalized number of events containing baseline objects by

$$N_{\text{QCD, scaled}}^{\text{signal}} = k_{\text{QCD}} \times N_{\text{QCD, unscaled}}^{\text{baseline}}$$

In addition to the use of baseline objects, the number of events passing in a phase-space region becomes even higher by demanding that the event was not selected by a specific single-jet trigger but instead only by any one of them. For the preliminary multijets CR this condition increases the number of events significantly from 15 113 to 1 105 028, which is a factor of approximately 73.1. The impact of using baseline objects in the multijets events instead of signal ones is investigated in the preliminary multijets control region. This way also the influence of the tau lepton identification (tau ID) can be studied since the events do not contain any muons due to the muon veto. The sample with the baseline objects is used for both cases. In order to get signal ones the corresponding cuts listed in Tab. 8.1 are applied to the tau leptons. For reasons of comparison the events with baseline objects are not scaled to those with signal

objects. The results are shown in Fig. 8.4. In general, no large differences can be observed between both cases, but the shapes of the multijets background are slightly different. This leads to a minimally better agreement between data and SM with the baseline objects. The overshoot of multijets background in Fig. 8.4(a) in the bin at around 300 GeV is an effect of the Jet Smearing method. It is caused by events which were generated from seed events with large prescale weights.



Figure 8.4: Comparison of data (black) and the SM backgrounds for baseline (left column) and signal (right column) objects. The SM consists of the multijets (red) and the electroweak (green) backgrounds whereas the latter are merged into one total electroweak background. Shown are the transverse momentum $p_T^{\text{jet}_1}$ of the leading jet (top row), the missing transverse energy $\not{\!\!\!E}_T$ (middle row) and the azimutal angle separation $\Delta \phi$ (jet₁, $\not{\!\!\!E}_T$) between the leading jet and the $\not{\!\!\!E}_T$ (bottom row). The multijets distributions are normalized to the difference between the numbers of event in data and in the EW background.

The observed shape differences in the multijets background are an effect of the tau lepton identification since due to the muon veto, the baseline and signal objects differ only in the respective criteria of the tau ID. The deviations are caused by the looser criteria on baseline tau leptons. Since the latter are faked by jets in multijets events the probability to select an event containing a fake- τ becomes higher. This leads to a bias in some variables, among other things, also in jet-related quantities like the p_T of a jet.

As mentioned already, the multijets events have to be scaled from baseline to signal objects in order to be comparable to the data and electroweak events. However, it turned out that the shape of the multijets background for baseline tau leptons does not reproduce the shape for signal ones. Since the tau ID biases the transverse momentum of the tau leptons, a $p_T^{\tau_1}$ -dependent scale factor is tested as well. It turns out that this approach does not lead to satisfying results either. Therefore, and for reasons of having enough seed events in the final sample produced from the whole 36 fb⁻¹ dataset, the approach with baseline objects is discarded. Instead signal objects are used for the further multijets control region development.

8.3 Enrichment of the multijets background fraction

As can be seen in Fig. 8.4, the relative fraction of the multijets background is relatively small and thus has to be enriched. In contrast to this the total number of events has to be as high as possible in order to obtain reliable results. During the work on this topic, it turned out that the technical issues in the Jet Smearing procedure mentioned in section 8.1 were caused by the selection of the seed events. Since the problem was solvable even after the smearing, the multijets sample could be fixed afterwards. Therefore the cut on $m_{\rm T}^{\tau_1}$ is omitted from now on. This leads to a significant gain in statistics in data and also the fraction of multijets events becomes larger as Fig. 8.5 illustrates. In addition the overall agreement between data and SM is better, too. The overestimated multijets background in $p_T^{jet_1}$ in the single bin around 250 GeV is again caused by events produced from seed events with large prescale weights. In E_{T} , significant deviations are only visible in the far tail where almost no multijets events are present. This implies that the disagreement is caused by the electroweak background. For the $\Delta\phi$ distributions data and SM are in good accordance within one standard deviation. Only in N_{jet} and $m_T^{\tau_1}$ the agreement is not sufficient. The peak in $m_T^{\tau_1}$ at around 130 GeV is underestimated in multijets. As mentioned already in section 6.1 and as will be explained in section 9.2 in more detail, this peak is an artefact of different tau lepton reconstruction thresholds and the disagreement is at least partially caused by the missing tau smearing in the Jet smearing procedure. This will be discussed in greater detail in section 10.2.

To enlarge the multijets background fraction in the corresponding control region different variables like $\not\!\!\!E_T$, m_{eff} , H_T or $p_T^{\tau_1}$ are considered and cuts on these quantities as well as on various combinations of them are applied. Some of those combinations are for example

$$\frac{\not\!\!\!E_{\rm T}}{m_{\rm eff}}, \quad \frac{\not\!\!\!\!E_{\rm T}}{H_{\rm T}}, \quad \frac{p_{\rm T}^{\tau_1}}{m_{\rm eff}}, \quad \frac{p_{\rm T}^{\tau_1}}{H_{\rm T}} \quad \text{and} \quad \frac{p_{\rm T}^{\rm Jet_{1,2}}}{N_{\rm jet}}.$$

The shapes of these distributions are depicted in Fig. 8.6. The cut values are chosen such that as much EW background as possible is rejected while keeping almost all QCD events. Following this principle after testing many variable combinations and cuts on more than one variable at the same time, it turns out that the variable $p_T^{\tau_1}/m_{\text{eff}}$ exhibits the best separation power between the two background types. The multijets events are localized only in the first few bins for this variable, while they are distributed over a larger range for the other ones. The cut value is chosen by visual approximation from the plots to be $p_T^{\tau_1}/m_{\text{eff}} < 0.1$. Distributions for other cut combinations can be found in Fig. A.1 to Fig. A.6 in section A.1 in the Appendix.





Figure 8.6: Comparison between data and the SM prediction for the new variable combinations in the preliminary multijets control region. The SM consists of the multijets (red) and the combined electroweak backgrounds (green). Shown are (a) $\not\!\!\!E_T/m_{eff}$, (b) $\not\!\!\!E_T/H_T$, (c) $p_T^{\tau_1}/m_{eff}$, (d) $p_T^{\tau_1}/H_T$, (e) $p_T^{jet_1}/N_{jet}$ and (f) $p_T^{jet_2}/N_{jet}$. For all plots also the tau ID criteria are applied. The multijets distributions are normalized to the difference between the numbers of event in data and in the EW background.



Fig. 8.7 shows some distributions in the preliminary multijets control region with the additional cut on $p_T^{\tau_1}/m_{\text{eff}}$. It is clearly visible that this cut enriches the multijets background fraction. Another positive side effect is the better overall agreement between data and the SM background. Also N_{jet} and $m_T^{\tau_1}$ are well modelled now except for the tails. Therefore this selection is chosen to define the final multijets control region. It is summarized in Tab. 8.2. The multijets control region can now be used to calculate

	Multijets CR
Trigger Plateau	$E_{\rm T} > 180 {\rm GeV}, \ p_{\rm T}^{\rm jet_1} > 120 {\rm GeV}$
Tau leptons	$N_{\tau} = 1$
Light leptons	$N_{\mu} = 0$
Jets	$N_{\rm jet} > 2$
Multijets enrichment	$\Delta\phi\left(\operatorname{jet}_{1},\vec{p}_{\mathrm{T}}\right) < 0.4 \parallel \Delta\phi\left(\operatorname{jet}_{2},\vec{p}_{\mathrm{T}}\right) < 0.4$
CR selection	$p_{\rm T}^{\tau_1}/m_{\rm eff} < 0.1$

Table 8.2: Multijet control region for the 2τ channel determined with the 3.2 fb⁻¹ of $\sqrt{s} = 13$ TeV data collected in 2015. m_{eff} is the effective mass of the event. The definitions can be found in section 7.1.

the multijets normalization factor for the 3.2 fb⁻¹ dataset. This is done using Eq. (8.1) and the number of events in data, multijets background and electroweak background which pass the respective selection. With $N_{\text{data}} = 981$, $N_{\text{EW}} = (686.55 \pm 9.02)$ and $N_{\text{QCD}} = (56455.94 \pm 973.25)$, the normalitazion factor is determined to be

$$\omega_{\rm QCD} = (5.22 \pm 0.18) \times 10^{-3} . \tag{8.2}$$

The reason for the specification of two decimal places will become clear in chapter 9. The number of observed data events and the number of predicted SM events are in agreement by construction. All uncertainties are purely statistical and are calculated from the sum of weights of all events passing the selection. The uncertainty on ω_{QCD} is determined by Gaussian error propagation. After the normalization the number of multijets events is (294.45 ± 11.53). Here the uncertainty is purely statistical again, but also contains the uncertainty on the normalization factor.

After the design of the multijets control region a cut on $p_T^{jet_2}$ was added to the trigger plateau cuts in the other control and signal regions of the analysis in [54]. This cut requires the transverse momentum of the sub-leading jet to be larger than 25 GeV. Hence it is also applied in the multijets control region from now on.
CHAPTER 9

Multijets Background Estimate for the Analysis with the full 36 fb⁻¹ Dataset

This chapter focusses on the results of this thesis. The multijets control region found in chapter 8 is used to determine a multijets normalization factor. The procedure is explained here. In addition also results for the yields of the multijets background in the different control and signal regions are given and discussed. In this context a new approach for the treatment of the statistical uncertainties is presented since the usual calculation method is not valid for events produced via Jet Smearing.

9.1 Treament of the statistical uncertainties

As will be explained in the following, the calculation of the statistical uncertainties is special for events produced by the Jet Smearing technique. Thus a new approach is presented for the treatment of the statistical uncertainties in the analysis with the 36 fb^{-1} dataset.

In general, the total number of events in a bin is given by

$$N = \sum_{i} w_i$$

with w_i being the weight of event *i*. In a histogram with independent events the uncertainty on the bin content can be computed by

$$\sigma_N = \sqrt{\sum_i w_i^2} \,. \tag{9.1}$$

However, this fomula is no longer valid for events produced with Jet Smearing since by the smearing, numerous events are produced from each seed event. Assume two arbitrary bins with different numbers of events coming from j seed events with different weights in both cases, leading to the same total number of N events in each bin. Due to the smearing all events coming from one seed event have the same weight (prescale weight). This situation is depicted in Fig. 9.1.

One issue of the Jet Smearing principle is that events from the same seed event are correlated with each other. Thus Eq. (9.1) is not valid in this case. The statistical uncertainties are not equal for two bins with different numbers of events produced from one seed event or if the seed events have different weights. The degree of correlation, however, cannot be determined exactly since it depends on many different



Figure 9.1: Sketch of the issue of the seed event uncertainty treatment. Both bins contain different numbers of events (red) from the same number of seed events (blue) with different weights leading to an equal total number of events in both cases.

All these influences make a correct determination of the correlations between the events nearly impossible. As an approach to calculate a statistical uncertainty anyway, two different uncertainties are computed, one assuming the events to be uncorrelated and one for a correlation of 100%. The real error has to be somewhere in between then. The uncertainty for uncorrelated events is still given by Eq. (9.1). For 100% correlated events, however, the error of a single seed event *j* can be computed by

$$\sigma_j = \sum_k w_k$$

where k denotes the number of events per seed event j. The bin uncertainty can be obtained by

$$\sigma_N = \sqrt{\sum_j \sigma_j^2} \, .$$

Combining both formulas finally gives

$$\sigma_N = \sqrt{\sum_j \left(\left(\sum_k w_k \right)^2 \right)}.$$
(9.2)

Due to the dependence on the numbers of events and seed events, the uncertainty also depends on the bin widths. Therefore, the statistical uncertainties have to be calculated for each bin of each histogram

individually after filling the histogram. The effect of this error treatment can be nicely seen in Fig. 9.2, where as an example the $\Delta\phi$ (jet₁, $\not\!\!E_T$) distribution in the top/W true CR (c.f. Tab. 7.5) is shown. The plot is built from the full 36 fb⁻¹ dataset, but only with a subsample of the available multijets statistics. The multijets background (red) is normalized in the selected top/W true tau control region. The other coloured backgrounds represent the electroweak ones. The hatched orange band illustrates the statistical uncertainty for uncorrelated events and the turquoise band the uncertainty for a correlation of 100 %. It is clearly visible that data and SM background deviate from each other at approximately $\Delta\phi$ (jet₁, $\not\!\!E_T$) = 2.0, leading to a dip in the data-to-SM ratio. This is caused by many events coming from only a few seed events or events with high weights. Thus the SM background seems to be overestimated at the respective position. However, the disagreement is perfectly covered by the turquoise error band assuming a correlation of 100 % which accounts for the described effects. This shows that the new approach for the treatment of the statistical uncertainty works quite well.

Although this method does not provide an exact error calculation, the statistical uncertainty can be



Figure 9.2: Example for the effect of the new uncertainty treatment. Shown is the $\Delta\phi$ (jet₁, $\not\!\!\!\!/_T$) distribution in the top/W true CR (c.f. Tab. 7.5). The red background represents the multijets background, the other colours denote the electroweak ones. The hatched orange and turquoise bands represent the statistical uncertainties for for the case of uncorrelated and 100 % correlated events, respectively. The plot is built from the full 36 fb⁻¹ dataset but only with a subsample of the available multijets statistics. The multijets background is normalized in the selected region.

narrowed down. Due to the high statistics the difference between both uncertainties becomes negligible for the full 36 fb^{-1} dataset as will become clear in section 9.2. However, although probably not reasonable, all results in this chapter are given with a precision of two decimal places in order to see possible differences between both uncertainties.

9.2 Results

As mentioned in the introduction of this chapter, the goal of this thesis is to provide a reliable multijets background estimate for the different control, validation and signal regions of the analysis using the full 36 fb^{-1} dataset. Thus it is essential to check the modelling of the multijets background again but with the full statistics. As before this is done in the multijets control region found in section 8.3. Afterwards, an

overall multijets background normalization factor is determined which is finally used in all other regions, too. Therefore a new multijets sample is produced by using a SUSY11 derivation of the full available $36 \, \text{fb}^{-1}$ dataset as input for the Jet Smearing.

In Fig. 9.3, some distributions in the multijets control region with the full statistics are shown. The



electroweak backgrounds are split into the different processes from now on. The tau leptons are required to fulfill the criteria of the tau ID working point *loose*. The distributions in Fig. 9.3 as well as all other results in this chapter are pre-fit. This means that in order to obtain the final results all relevant backgrounds are simultaneously scaled in their corresponding control region in a way that the total

SM background fits best to the data. Besides loose tau leptons, also *medium* ones are used in the main analysis. The results for these are shown in Fig. 9.4.

The overall agreement between the data and the SM prediction is quite good for loose tau leptons. For



Figure 9.4: Comparison between data and the SM backgrounds for some distributions in the multijets control region for medium tau leptons. Shown are (a) the transverse momentum $_{T}^{\tau_{1}}$ of the leading tau lepton, (b) the number of jets N_{jet} , (c) the transverse momentum $p_{T}^{jet_{1}}$ of the leading jet, (d) the missing transverse energy $\not{\!\!E}_{T}$, (e) the azimutal angle separations $\Delta \phi$ (jet₁, $\not{\!\!E}_{T}$) between the leading jet and the $\not{\!\!E}_{T}$ and the transverse mass (f) $m_{T}^{\tau_{1}}$ of the tau lepton. The multijets background is normalized to the difference between data and the electroweak backgrounds.

 The yields for data, the expected total nominal EW background and the nominal multijets background in the multijets control region are summarized in Tab. 9.1 for loose and medium tau leptons. Additionally, the number of different seed events. It should be noted that the nominal number of multijets events includes the prescale-weighting. The uncertainties are purely statistical.

By using loose tau leptons instead of medium ones, the number of nominal multijets events is

	data	EW nominal	multijets nominal		seed events
loose taus	15 564	9 265.88 ± 67.35	1 282 194.49	± 30211.69	724 520.00 ± 851.18
				± 30779.93	
medium taus	10 084	7 032.56 ± 58.81	619 122.45	± 18242.56	370 800.00 ± 608.93
				±18290.59	

Table 9.1: Yields for data, the expected total nominal EW background, the nominal multijets background and seed events in the multijets control region for loose and medium tau leptons. In addition, also the number of seed events passing the selection is given. The nominal number of multijets events includes the prescale-weighting. The upper uncertainty on the nominal number of multijets events is the uncertainty for uncorrelated events and the lower one for 100 % correlated events.

approximately twice as large. The same holds for the number of different seed events. This allows a much more reliable statistical treatment. The multijets backgrounds in Fig. 9.3 and Fig. 9.4 are normalized to the difference in the number of data events and the number of total nominal electroweak events by

$$\omega_{\rm QCD} = \frac{N_{\rm data} - N_{\rm EW}}{N_{\rm QCD}}$$

according to Eq. (8.1). The normalization factor is determined to be

$$\omega_{\text{QCD}}^{\text{loose}} = (4.91 \pm 0.13) \times 10^{-3} \tag{9.3}$$



Figure 9.5: Ratio of the number of loose tau leptons to the number of medium ones for different $p_{\rm T}$ intervals.

for loose tau leptons and

$$\omega_{\rm OCD}^{\rm medium} = (4.93 \pm 0.17) \times 10^{-3} \tag{9.4}$$

for medium ones. The uncertainties are calculated with Gaussian error propagation and are purely statistical again. In both cases, the uncertainties for uncorrelated events and for 100 % correlated events are equal in the specified precision since they differ first in the seventh decimal place. The normalization factors for loose and medium tau leptons are also used for all control, validation and signal regions of the analysis in [54], depending on which type of tau leptons is required in the respective region. Normalizing the multijets background leads to 6298.12 ± 220.24 expected events for loose tau leptons assuming uncorrelated events and 6298.12 ± 224.17 events for a correlation of 100 %. With medium tau leptons, the corresponding numbers are 3051.44 ± 140.10 and 3051.44 ± 140.40 , respectively. The uncertainties of these numbers are purely statistical again and contain the uncertainty on the number of events and the one from the normalization factor.

Due to the fact that both types of tau leptons are used in the final analysis, it is convenient to be able to scale between both. Therefore $p_T^{\tau_1}$ -dependent scaling factors are determined by computing the ratio of the two $p_T^{\tau_1}$ distributions. The results can be found in Fig. 9.5. It is clearly visible that for low values of $p_T^{\tau_1}$, the number of medium tau leptons is roughly only half as large as for loose ones. For higher transverse momenta, however, the ratio is approximately one. This behaviour can be explained by the better reconstruction performance for tau leptons with larger $p_T^{\tau_1}$. Hence, there are more tau leptons with low transverse momentum which fulfill the loose tau ID criteria.

The signal and control regions listed in Tab. 7.4 and Tab. 7.5, respectively, require medium tau leptons. With the corresponding normalization factor given in Eq. (11.3), it is now possible to estimate the multijets background in the different regions. Fig. 9.6 shows the $p_T^{jet_1}$ spectrum for the control regions, Fig. 9.7 depicts the same variable for the signal regions. All histograms are pre-fit. Since the signal regions are still blinded in the main analysis, there is no data included in the corresponding plots. From Fig. A.7 to Fig. A.14 in Appendix A.2, further distributions can be found for all control and signal regions. There are no plots for the kinematic control regions available since the events in the used multijets sample are preselected to have at least one tau lepton and thus no events pass the selections of the kinematic CRs.



Figure 9.6: $p_T^{\text{jet}_1}$ for the control regions of the analysis using the full statistics. The control regions are (**a**) the W true tau CR, (**b**) the top true tau CR, (**c**) the W fake tau CR, (**d**) the top fake tau CR, (**e**) the $Z \rightarrow \tau \tau$ CR and (**f**) the $Z \rightarrow \nu \nu$ CR. The multijets background is normalized with the obtained normalization factor for medium tau leptons.

The total background fits the data only in the W true tau CR. In most of the regions, the SM background is understimated and some variables exhibit trends in the data-to-SM ratio. Especially the $Z \rightarrow vv$ CR shows significant deviations. Since this region was quite recently introduced in the main analysis, it has to be further optimized. This could reduce the disagreement at least partially. In addition, the distributions are pre-fit. Although the agreement is presumably not perfect after the fit, it is should be improved. A mismodelling of the multijets background could be another source for the deviations. However, the impact is supposedly not very large since the modelling was quite good in the corresponding control region.



Figure 9.7: $p_T^{\text{jet}_1}$ for the signal regions of the analysis using the full statistics. The signal regions are (**a**) the Compressed SR and (**b**) the High Mass SR. Since the signal regions are still blinded, there are no data points in the plots. The multijets background is normalized with the obtained normalization factor for medium tau leptons.

The yields for data, the expected total nominal EW background, the nominal and normalized multijets background as well as the number of seed events in the different control and signal regions are summarized in Tab. 9.1. The uncertainties on the yields are purely statistical again. For the normalized multijets events, they include two components: the uncertainty on the numbers of events and the one from the normalization factor. Except for the true tau top CR, the uncertainties for uncorrelated and 100 % correlated multijets events are equal since in these regions, every event has a different seed event. Systematic uncertainties on the yields are not available for reasons of time. For the multijets background, however, two sources of systematics are discussed in section 10.3.

In general most events remain in the true tau CRs. This is a result of the muon veto in these regions since for data and the electroweak backgrounds, most of the events contain a true tau lepton. Hence, the number of events in the fake tau CRs is much smaller. The same holds for the multijets events, where only in the true tau CRs a non-negligible number of events is expected. It should be noted that the tau leptons are fake- τ s again. However, the number is very small compared to the data and the electroweak backgrounds. In the two fake tau CRs, no multijets events are left. This can be explained by the fact that jets most likely fake tau leptons and not muons. Therefore, the number of multijets events containing muons is very limited and the remaining events are rejected by the other cuts.

Some control regions exhibit a deviation between the observed number of data events and the total background of more than one standard deviation. This difference would probably be reduced in the fitting procedure.

As expected, the number of background events in the signal regions is very low. Only in the Compressed SR some electroweak events are left. The multijets background is negligible in both signal regions as desired.

Fortunately, in all regions with a non-negligible multijets background, the number of different seed events is high enough to draw reliable conclusions in the main analysis. This is also valid for the signal regions.

	Jata	EW	multijets		and monto
	uala	nominal	nominal	normalized	seed events
W true tau CR	18286	18338.60 ± 126.62	2917.02 ± 637.92	14.32 ± 3.17	190.00 ± 13.78
top true tau CR	13 348	12254.72 ± 75.34	$\begin{array}{r} \pm 6\ 356.56 \\ \pm 6\ 356.57 \\ \pm 6\ 356.57 \end{array}$	171.02 ± 31.77	1291.00 ± 35.93
W fake tau CR	626	592.44 ± 17.38	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
top fake tau CR	712	592.65 ± 15.39	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
$Z \rightarrow \tau \tau CR$	605	572.83 ± 22.15	1.00 ± 1.00	0.01 ± 0.01	1.00 ± 1.00
$Z \rightarrow \nu \nu CR$	5 368	$3.924.05 \pm 81.97$	20257.40 ± 7650.18	99.46 ± 37.72	132.00 ± 11.49
Compressed SR	I	26.57 ± 3.18	32.00 ± 5.66	0.16 ± 0.03	32.00 ± 5.66
High Mass SR	I	5.59 ± 0.88	56.00 ± 7.49	0.28 ± 0.04	56.00 ± 7.48

correlated events. In the signal regions, there are no yields for data since these regions are still blinded in the main analysis. regions with medium tau leptons. In addition, also the number of seed events passing the selection is given. The unnormalized QCD events are prescale-weighted. For the top true tau CR, the upper uncertainty on the nominal number of multijets events is the uncertainty for uncorrelated events and the lower one for 100 % tau CR, the uncertainties for uncorrelated and 100 % correlated multijets events are equal since in these regions, every event originates from a different seed event. The uncertainties are purely statistical. For the normalized multijets events, they contain also the component from the normalization factor. Except for the top true Table 9.2: Yields for data, the expected total nominal EW background, the nominal and normalized multijets background and seed events in the control and signal

CHAPTER 10

Possible Improvements and Outlook

Although section 9.2 points out that the Jet Smearing technique delivers an impressive multijets background estimate the method also has some shortcomings. One of the most severe ones is the limited seed event statistics. This is especially a problem for multijets samples produced with a small dataset since, depending on the considered phase-space region, the multijets events often originate from only a very small number of different seed events. One approach to solve this problem is discussed in section 10.1. Another issue in the context of Jet Smearing is the missing smearing of tau leptons which leads to events in which the jet quantities are varied, but tau leptons stay unchanged for all events produced from the same seed event. Some studies with simple approaches for tau smearing are presented in section 10.2. Furthermore it is studied if the observed discrepancy between data and SM prediction in the peak in $m_{\rm T}^{\tau_1}$ can be explained by this issue.

Besides the physics of tau leptons, that of electrons and muons is equal in all events originating from one seed event as well. This introduces a bias in the sense that the events do not represent the "real" physics. In the future, it is necessary to determine a systematic uncertainty on this feature. Moreover there are also other sources of systematic uncertainties on the multijets background. Two of them are discussed in section 10.3.

10.1 Rebalance-and-smear method

One approach to increase the seed event statistics is the so-called "rebalance-and-smear" method which is described in detail in [72]. In this method seed events for the smearing are constructed by balancing the jet four-momenta in an event in a way that the missing transverse energy is minimal afterwards. This rebalancing is done by maximizing the likelihood that the considered jet with a reconstructed transverse momentum p_T^{reco} has a certain true transverse momentum p_T^{true} . In addition, the likelihood is constrained by the condition that the sum of the true transverse momenta of all particles in the event is zero. The advantage of this technique is that also events with a large initial $\not{\!\!\!E}_T$ can be used as seed events for the smearing. This increases the number of different seed events in a certain phase-space region.

10.2 Tau lepton smearing

The missing smearing of tau leptons affects the modelling of all tau-related quantities. As an example, in Fig. 10.1 depicts scatter plots for data (left) and multijets events (right) of events containing one loose tau

lepton. In these plots the transverse mass of the tau lepton is plotted against its transverse momentum. While in data the events are continuously distributed over $p_T^{\tau_1}$, bands of events with the same $p_T^{\tau_1}$ can be observed for the multijets background. All these events in a band are generated from the same seed event. This illustrates nicely one effect of the missing tau smearing. Also the in Fig. 9.4(f) observed peak in $m_T^{\tau_1}$, which is narrower in the multijets background than in data, is probably caused by this issue.

In this section some simple studies with different approaches are presented to improve the tau lepton



Figure 10.1: Scatter plots of the transverse mass $m_T^{\tau_1}$ of the leading tau lepton versus its transverse momentum $p_T^{\tau_1}$ for (a) data and (b) multijets background for loose tau leptons in the multijets control region.

modelling in multijets events produced with the Jet Smearing technique. All these approaches use parameters and functions derived from jets. For tau leptons those quantities are generally different, but since tau leptons and jets originate from the same objects on reconstruction level, and due to the fact that something like a tau response function is not yet available, jet quantities are considered to be sufficient for the first studies explained in this section.

The first approach is to smear the magnitude of the transverse momentum of the tau lepton by adding a random number drawn from a Gaussian with mean zero and a width corresponding to σ_{corr} used for the correction of the jet response function in the dijet analysis (c.f. Eq. (6.4)):

$$\sigma_{\rm corr}(p_{\rm T}) = \sqrt{2} \times \sqrt{\sigma_{A,{\rm data}}^2 - \sigma_{A,{\rm MC}}^2} \,.$$

The widths $\sigma_{A,data}$ and $\sigma_{A,MC}$ of the p_T -asymmetry in dijet events in data and MC can be determined by (c.f. Eq. (6.3) in section 6.4.1)

$$\sigma_A = \frac{a}{p_{\rm T}} + \frac{b}{\sqrt{p_{\rm T}}} + c$$

whereat for *a*, *b* and *c* the parameters determined in the 8 TeV analysis presented in [68] are used as first attempt for reasons of having no reference points. The values are given in Tab. 10.1. The tau leptons are then smeared by the obtained Gaussian. Afterwards all quantities affected by the smearing need to be recalculated, namely $\not{\!\!\!E}_T$, $\phi_{\not{\!\!\!E}_T}$, $m_T^{\tau_1}$ and $\Delta \phi (\tau_1, \not{\!\!\!E}_T)$. Fig. 10.2 illustrates the effect of smearing the tau leptons with the Gaussian calculated from the values in

Fig. 10.2 illustrates the effect of smearing the tau leptons with the Gaussian calculated from the values in Tab. 10.1 on different tau-related quantities. The selection corresponds to the multijets control region and the dataset to the 3.2 fb^{-1} of 13 TeV data collected in 2015 since the full statistics was not yet available at the time of performing this study. However, this is unproblematic since this study is not aiming for any



Figure 10.2: Effect of a tau lepton smearing with a Gaussian on tau-related quantities in the multijets control region. The luminosity corresponds to the 3.2 fb⁻¹ of 13 TeV data collected in 2015. Shown are the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton (upper row), the azimutal angle separations $\Delta \phi(\tau_1, \not\!\!E_T)$ between the leading tau lepton and the $\not\!\!E_T$ (middle row) and the transverse tau lepton mass $m_T^{\tau_1}$ (lower row) for unsmeared tau leptons (left column) and those smeared by a Gaussian with a width calculated from the parameters found in [68] (right column). The multijets background is normalized to the difference between the numbers of events in data and in the electroweak background, respectively.

Parameter	Value
a _{data} [GeV]	3.856
$b_{ m data}$ [$\sqrt{ m GeV}$]	0.652
<i>c</i> _{data}	0.003
$a_{\rm MC}$ [GeV]	2.833
$b_{ m MC}$ [$\sqrt{ m GeV}$]	0.616
$c_{\rm MC}$	0.002

Table 10.1: Parameter values for a, b and c for data and MC used for the calculation of the width of the Gaussian taken for the tau lepton smearing. The values are taken from [68].

quantitative results. Also all electroweak backgrounds are merged to one total EW background again. In general the modelling of the shown quantities is already feasible even without tau smearing, but especially in $m_T^{\tau_1}$ an improvement is necessary for higher values. The background is slightly underestimated in the peak at about 130 GeV while in the tail it is the other way around. Furthermore, the peak is somewhat broader in data than in the multijets background. While the smearing has almost no visible effect on $p_{T}^{\tau_{1}}$ and $\Delta \phi(\tau_1, \not\!\!\!E_T)$ the peak in $m_T^{\tau_1}$ is slightly broader in the multijets background when smearing the tau leptons. The agreement between data and SM prediction, however, becomes worse in the peak due to the normalization. This shows that the smearing of tau leptons indeed affects the modelling of $m_{\rm T}^{\tau_1}$ and the difference of the peak widths in data and SM background can at least be partially compensated. It should be noted that for different reasons it is very challenging to determine a quantitative measure for the influence of tau smearing on the peak width. The most important one is that the smearing has to be done before the tau ID is applied since due to the modification of their $p_{\rm T}$, some tau leptons will not pass anymore the $p_{\rm T} > 20 \,\text{GeV}$ condition for "baseline" tau leptons (c.f. section 4.3) after the smearing while other ones are above this threshold. In addition the transverse mass of the tau lepton is calculated from three different quantities which are all affected in the smearing. Thus it is not trivial to conclude in which direction $m_{\rm T}^{\tau_1}$ is shifted by the smearing.

In order to investigate the impact of the width of the Gaussian on the tau lepton smearing, higher and lower parameter values for the calculation of the width are tested which are listed in Tab. 10.2. The

Parameter	Value		
	smaller σ	larger σ	
a _{data} [GeV]	2.0	3.856	
$b_{ m data}$ [$\sqrt{ m GeV}$]	0.552	0.652	
C _{data}	0.008	0.005	
a _{MC} [GeV]	1.0	2.833	
$b_{ m MC}$ [$\sqrt{ m GeV}$]	0.316	0.616	
$c_{\rm MC}$	0.005	0.004	

Table 10.2: Parameters a, b and c for data and MC used for the calculation of the width of the Gaussian taken for the tau lepton smearing. The left values are used to reduce the width and the right ones to increase it.

results are shown in Fig. 10.3, where the same distributions as before are depicted for a smaller width (left side) and a larger width (right side) of the Gaussian. The chosen variations of σ have obviously no notable influence on the tau smearing. Only in the peak of $m_T^{\tau_1}$ a minimal difference can be observed. The result seems to be better for a smaller σ since it leads to a broader peak in the multijets background compared to the previous case.



Figure 10.3: Effect of a variation of the width of the Gaussian on the tau lepton smearing in the multijets control region. The luminosity corresponds to the 3.2 fb⁻¹ of 13 TeV data collected in 2015. Shown are the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton (upper row), the azimutal angle separations $\Delta \phi(\tau_1, \not{\!\!\!\! E_T})$ between the leading tau lepton and the $\not{\!\!\!\! E_T}$ (middle row) and the transverse tau lepton mass $m_T^{\tau_1}$ (lower row) for a smaller width (left column) and a larger width (right column) of the Gaussian. The multijets background is normalized to the difference between the numbers of events in data and in the electroweak background, respectively.

Besides the smearing with a Gaussian, the smearing of tau leptons with the jet response function is investigated as well. Therefore the jet response map from the used version Jet Smearing tool (c.f. Fig. 6.3(a)) is considered and the tau leptons are smeared with the response function corresponding to its $p_{\rm T}$. Afterwards the affected quantities have to be recalculated again. The result is presented in the right column in Fig. 10.4. For comparison in the left column the same variables are plotted for unsmeared tau leptons as well. As can be seen also the smearing by the jet response function broadenes the peak in the multijets background. Although still not sufficient, the smearing by the jet response function leads to better results than the smearing with a Gaussian.



Figure 10.4: Effect of a tau lepton smearing with the jet response function from the used version of the Jet Smearing tool. The study is performed in the multijets control region. The luminosity corresponds to the 3.2 fb⁻¹ of 13 TeV data collected in 2015. Shown are the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton (upper row), the azimutal angle separations $\Delta \phi(\tau_1, \not\!\!E_T)$ between the tau and the $\not\!\!E_T$ (middle row) and the transverse tau lepton mass $m_T^{\tau_1}$ (lower row) for unsmeared tau leptons (left column) and those smeared with the jet response function (right column). The multijets background is normalized to the difference between the numbers of events in data and in the electroweak background, respectively.

All described approaches did not lead to satisfying results for the tau lepton modelling in events generated with Jet Smearing. Nevertheless, they have shown that smearing tau leptons as well can reduce the discrepancy between data and SM background. The still visible deviation implies that the use of

parameter values and functions derived from jets is not justified for this purpose. In order to achieve reasonable results, a much more dedicated smearing function has to be determined.

10.3 Determination of systematic uncertainties

Besides the derivation of an appropriate tau lepton smearing function, another important task in the future is the determination of systematic uncertainties resulting from the Jet Smearing technique. For reasons of time this was not possible within this thesis. However, two sources of them are discussed in the following.

The first one is caused by the uncertainties of the jet reponse function. Since these affect the jet quantities in the smeared events the output of the Jet Smearing procedure is different compared to the use of the nominal response function. The strength of this effect can be studied by changing the response function within its uncertainties. From the variations of the resulting events compared to output for the nominal response function, a systematic uncertainty for the impact of a response function variation on the final result can be calculated.

A second systematic uncertainty is introduced by smearing each seed event several times. As a consequence it is generally possible that many events generated from the same seed event are stored in the same bin in the final histogram. This causes correlations between different events and thus makes the treatment of uncertainties very challenging as mentioned already in section 9.1. The impact of multiple smearing on bin migration effects can be investigated using toy MC. This can be produced by generating many multijets samples by running the smearing several times on the same input data. With the different samples it can be studied how events from the same seed event are distributed over the single bins for the different variables. This knowlegde can finally be used to compute a systematic uncertainty on the bin migration effects caused by multiple smearing. In addition, with this method also a better measure for the correlations between the events can be obtained. However, the disadvantage of this approach is that it is very computing-intensive and needs a lot of disk space.

CHAPTER 11

Summary and Conclusion

The correct modelling of all relevant SM backgrounds is an essential ingredient for every analysis. One dominant background in proton-proton collisions is the multijets background. It constists of many jets originating from hadronization processes of quarks and gluons due to the confinement principle, a consequence of the characteristics of the strong interaction. This background cannot be simulated with Monte Carlo techniques with sufficient reliability and statistics, since it is impossible to account for all higher order couplings in strong interaction processes and hence for the large production cross-section of multijets events. Although there exists dijet MC it is especially for SUSY searches not usable due to the large $\not{\!\!E}_T$ requirement in SUSY events. In dijet events missing energy is almost exclusively caused by mis-measurements of at least one of the jets. However, these mis-measurements are mostly too small to generate events with enough $\not{\!\!E}_T$. Hence the relevant events are in the far tail of the $\not{\!\!E}_T$ distribution. This results in too low statistics for a SUSY analysis.

$$R = \frac{p_{\rm T}^{\rm reco}}{p_{\rm T}^{\rm true}}$$

with p_T^{reco} being the reconstructed transverse momentum of a jet and p_T^{true} its true one. The jet response quantifies the energy mis-measurements of jets in the detector. By measuring *R* for many jets, a jet response function can be determined. In order to reflect the jet response in data, the response function is constrained in two dedicated analysis, the dijet analysis and the "Mercedes" analysis. Since the jet response depends on the momentum and the pseudorapidity of the jet, there exists one response function for every p_T - and every η -interval.

The well measured data events for the smearing were selected by single-jet triggers with different thresholds. They had to fulfill criteria on the $\not\!\!E_T$ -significance and the $\not\!\!E_T$ over the average transverse

momenta of the two leading jets:

$$S = \frac{\not\!\!E_{\rm T} - 8\,{\rm GeV}}{\sqrt{\sum E_{\rm T}}} < \left(0.5 + 0.1 \times N_{b\text{-jet}}\right) {\rm GeV}^{1/2} \qquad \text{and} \qquad \frac{\not\!\!E_{\rm T}}{\left\langle p_{\rm T}^{\rm jet\,1,2} \right\rangle} < 0.2 \; .$$

These "seed events" were then smeared after the modified jet response function. This procedure was repeated several times resulting in a large multijets sample. The latter was afterwards used for the development of a multijets control region which was needed to normalize the multijets events since their natural normalization is lost due to the numerously applied smearing of every seed event. In addition the control region was also used to check the modelling.

The multiple smearing of each seed event makes the calculation of the statistical uncertainties more challenging, since the resulting events are correlated with each other. The correlation depends on many factors and cannot be calculated properly. A new approach was presented in this thesis to solve this issue. Two uncertainties were calculated, one assuming a correlation of 0% and one for a correlation of 100%. The real error was then somewhere in between. For the full statistics of 2015 and 2016, the difference between them was negligible.

The control region design was performed with the 3.2 fb^{-1} dataset of 13 TeV data recorded in 2015. It started with a trigger cut, trigger plateau cuts and cleaning cuts. Furthermore the particle content was adapted to the expected one, i.e. and at least three jets and one tau lepton. In addition, a muon veto was applied to use only generic events for the calculation of the normalization factor. The most important cut, however, was the inverted QCD rejection,

$$\Delta \phi (\operatorname{jet}_1, \not\!\!E_{\mathrm{T}}) < 0.4 \quad \text{or} \quad \Delta \phi (\operatorname{jet}_2, \not\!\!E_{\mathrm{T}}) < 0.4 ,$$

which enriched the multijets statistics significantly and made the multijets control region orthogonal to the other control regions and the signal regions. In order to enlarge the relative fraction of the multijets background compared to the other ones, cuts on different variables and combinations of them were applied. It turned out that the best result can be achieved for requiring $p_T^{\tau_1}/m_{\text{eff}} < 0.1$.

The multijets control region was then used to determine multijets normalization factors. They were computed by

$$\omega_{\rm QCD} = \frac{N_{\rm data} - N_{\rm EW}}{N_{\rm QCD}}$$

from the numbers of data, electroweak and multijets events in the control region. For the 3.2 fb^{-1} dataset, the normalization was determined to be

$$\omega_{\rm QCD} = (5.22 \pm 0.18) \times 10^{-3} . \tag{11.1}$$

For the full available statistics of 36 fb^{-1} , a cut on the transverse momentum of the sub-leading jet was added to the control region selection. Then, the normalization was individually determined for events containing tau leptons fulfilling criteria of the tau ID working points *loose* and *medium* since both types are used in the main analysis. The obtained normalization factors are

$$\omega_{\text{QCD}}^{\text{loose}} = (4.91 \pm 0.13) \times 10^{-3} \tag{11.2}$$

for loose tau leptons and

$$\omega_{\rm OCD}^{\rm medium} = (4.93 \pm 0.17) \times 10^{-3} \tag{11.3}$$

for medium ones. The uncertainties are purely statistical.

Using these factors, the expected number of multijets events in the corresponding control region is 6298.12 ± 220.24 with loose tau leptons assuming uncorrelated events and 6298.12 ± 224.17 for a correlation of 100 %. With medium tau leptons, the corresponding numbers are 3051.44 ± 140.10 and 3051.44 ± 140.40 , respectively.

The key results of this thesis are the yields of the multijets background in the other control and signal regions since these are the values of interest for the analysis presented in [54]. It turned out that after the normalization 0.00 ± 0.00 multijets events are expected in the top and W fake tau CRs, 0.01 ± 0.01 in the $Z \rightarrow \tau\tau$ CR, 0.28 ± 0.04 in the High Mass SR and 0.16 ± 0.03 events in the Compressed SR. This is as desired since the signal regions are designed in a way such that they contain as little background as possible. Only in the top and W true tau CR and the $Z \rightarrow \nu\nu$ CR a non-gegligible number of multijets events is found. The top true tau CR contains 171.02 ± 31.77 events, the W true tau CR 14.32 ± 3.17 and the $Z \rightarrow \nu\nu$ CR 99.46 ± 37.72 events. Fortunately, the number of different seed events in these control regions was found to be between 132.00 ± 11.49 and $1.291.00 \pm 35.93$ and is thus high enough to allow for reliable conclusions. In the signal regions, 32.00 ± 5.66 seed events were obtained for the Compressed SR and 56.00 ± 7.48 for the High Mass SR. All these uncertainties are purely statistical as well. For the normalized multijets events, they contain also the uncertainty on the normalization factor. The uncertainties for uncorrelated multijets events and for 100 % correlated ones are equal in the specified precision.

Another important part of this thesis was to check the modelling of the multijets background. This was also done in the multijets control region. In general an overall good agreement between the data and the SM prediction could be observed. In this context the $m_T^{\tau_1}$ distribution played a key role since this variable is calculated out of three different quantities. This means that a good agreement between the data and the whole SM background implies a good multijets modelling of all $m_T^{\tau_1}$ input variables. Indeed, the overall agreement was satisfying. However, the peak around 130 GeV, coming from different reconstruction thresholds of the tau leptons, was broader in data than in the multijets background. This can be partially explained by the missing tau lepton smearing in the Jet Smearing procedure as some studies with different approaches have shown in this thesis.

Last but not least, some possible improvements have been discussed, in particular the determination of a suitable tau smearing function and the calculation of systematic uncertainties.

The produced multijets sample will be used for the analysis in [54] in the future.

Bibliography

- P. W. Higgs, Broken Symmetries and the Masses of Gauge Bosons, Phys. Rev. Lett. 13 (16 1964) 508, URL: http://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on p. 1).
- F. Englert and R. Brout, Broken Symmetry and the Mass of Gauge Vector Mesons, Phys. Rev. Lett. 13 (9 1964) 321, URL: http://link.aps.org/doi/10.1103/PhysRevLett.13.321 (cit. on p. 1).
- G. A. et al., Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Physics Letters B 716 (2012) 1, ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/S037026931200857X (cit. on pp. 1, 12).
- [4] S. Chatrchyan et al.,
 Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,
 Phys. Lett. B716 (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on pp. 1, 12).
- [5] E. Corbelli and P. Salucci, *The Extended Rotation Curve and the Dark Matter Halo of M33*, Mon. Not. Roy. Astron. Soc. **311** (2000) 441, arXiv: astro-ph/9909252 [astro-ph] (cit. on pp. 1, 12).
- [6] J. Wess and B. Zumino, Supergauge transformations in four dimensions, Nuclear Physics B 70 (1974) 39, ISSN: 0550-3213, URL: http://www.sciencedirect.com/science/article/pii/0550321374903551 (cit. on pp. 1, 14).
- S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics 22 (1961) 579, ISSN: 0029-5582,
 URL: http://www.sciencedirect.com/science/article/pii/0029558261904692 (cit. on pp. 3, 6).
- [8] S. Weinberg, Elementary Particle Theory of Composite Particles, Phys. Rev. 130 (2 1963) 776, URL: http://link.aps.org/doi/10.1103/PhysRev.130.776 (cit. on pp. 3, 6).
- [9] S. Weinberg, A Model of Leptons, Phys. Rev. Lett. 19 (21 1967) 1264,
 URL: http://link.aps.org/doi/10.1103/PhysRevLett.19.1264 (cit. on pp. 3, 6).
- H. Fritzsch, M. Gell-Mann and H. Leutwyler, Advantages of the color octet gluon picture, Physics Letters B 47 (1973) 365, ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/0370269373906254 (cit. on pp. 3, 7).
- [11] A. Salam, Weak and Electromagnetic Interactions, Conf. Proc. C680519 (1968) 367 (cit. on pp. 3, 6).

- [12] M. el al., Standard Model of Elementary Particles, modified version, 2014, URL: https://commons.wikimedia.org/wiki/File: Standard_Model_of_Elementary_Particles_modified_version.svg (cit. on p. 4).
- [13] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013 (cit. on pp. 4, 22).
- [14] C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. C 40 (2016) (cit. on pp. 6, 7, 9, 10, 14).
- [15] B. Povh et al., *Teilchen und Kerne*, 8th ed., Springer-Verlag, 2008 (cit. on p. 8).
- [16] A. Harmon, Sherpa and Open Science Grid: Predicting the emergence of jets, URL: https://sciencenode.org/feature/sherpa-and-open-science-gridpredicting-emergence-jets.php (cit. on p. 8).
- [17] JabberWok and Time3000, Feynman Diagram of Decay of tau lepton, modified version, 2008, URL: https://commons.wikimedia.org/wiki/File: Feynman_diagram_of_decay_of_tau_lepton.svg (cit. on p. 9).
- [18] S. Schaepe, *Search for Supersymmetry in Tau Lepton Final States with the ATLAS Detector*, PhD thesis: University of Bonn, 2015 (cit. on pp. 10, 33, 34).
- [19] Measurement of the Mis-identification Probability of Tau Leptons from Hadronic Jets and from Electrons, tech. rep. ATLAS-CONF-2011-113, CERN, 2011, URL: http://cds.cern.ch/record/1375550 (cit. on pp. 11, 29).
- [20] A. Collaboration, Summary plots from the ATLAS Standard Model physics group, 2017, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/ SM/index.html (cit. on p. 11).
- [21] S. P. Martin, A Supersymmetry primer, (1997), [Adv. Ser. Direct. High Energy Phys.18,1(1998)], arXiv: hep-ph/9709356 [hep-ph] (cit. on pp. 12–15).
- [22] A. Refregier, Weak gravitational lensing by large scale structure, Ann. Rev. Astron. Astrophys. 41 (2003) 645, arXiv: astro-ph/0307212 [astro-ph] (cit. on p. 12).
- [23] A. N. Taylor et al., *Gravitational lens magnification and the mass of abell 1689*, Astrophys. J. 501 (1998) 539, arXiv: astro-ph/9801158 [astro-ph] (cit. on p. 12).
- [24] X.-P. Wu et al., A comparison of different cluster mass estimates: consistency or discrepancy ?, Mon. Not. Roy. Astron. Soc. 301 (1998) 861, arXiv: astro-ph/9808179 [astro-ph] (cit. on p. 12).
- [25] G. Hinshaw et al., Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results, Astrophys. J. Suppl. 180 (2009) 225, arXiv: 0803.0732 [astro-ph] (cit. on p. 12).
- P. A. R. Ade et al., *Planck 2013 results. XVI. Cosmological parameters*, Astron. Astrophys. 571 (2014) A16, arXiv: 1303.5076 [astro-ph.CO] (cit. on pp. 12, 13).
- [27] P. Fayet, Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions, Physics Letters B 69 (1977) 489, ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/0370269377908528 (cit. on p. 14).

- [28] S. Dimopoulos and H. Georgi, Softly broken supersymmetry and SU(5), Nuclear Physics B 193 (1981) 150, ISSN: 0550-3213, URL: http://www.sciencedirect.com/science/article/pii/0550321381905228 (cit. on p. 14).
- [29] M. Aaboud et al., Search for squarks and gluinos in events with hadronically decaying tau leptons, jets and missing transverse momentum in proton–proton collisions at $\sqrt{s} = 13$ TeV recorded with the ATLAS detector, Eur. Phys. J. C76 (2016) 683, arXiv: 1607.05979 [hep-ex] (cit. on pp. 16, 33, 34, 39, 48–50, 53, 54).
- [30] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation 3 (2008) S08001, URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08001 (cit. on pp. 17, 18).
- [31] W. der Physik, LHC und Experimente, CERN license, 1997, URL: http: //www.weltderphysik.de/uploads/tx_wdpmedia/1998_LHCUeberblick_CERN_02.jpg (cit. on p. 18).
- [32] T. A. Collaboration, ATLAS Public Results, URL: https: //twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (cit. on p. 19).
- [33] J. Pequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: http://cds.cern.ch/record/1095924 (cit. on p. 20).
- [34] G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 19–24).
- [35] J. Pequenao, "Computer generated image of the ATLAS inner detector", 2008, URL: http://cds.cern.ch/record/1095926 (cit. on p. 21).
- [36] K. Potamianos, "The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider", *Proceedings, 2015 European Physical Society Conference on High Energy Physics (EPS-HEP 2015): Vienna, Austria, July 22-29, 2015, 2016, arXiv: 1608.07850 [physics.ins-det], URL: http://inspirehep.net/record/1484295/files/arXiv:1608.07850.pdf (cit. on pp. 21, 28).*
- [37] J. Jentzsch, C. Goessling and H. Pernegger, *Pixel detector modules performance for ATLAS IBL and future pixel detectors*, Presented 19 Jun 2015, PhD thesis: Tech. U., Dortmund (main), 2015, URL: https://cds.cern.ch/record/2197808 (cit. on p. 21).
- [38] J. Pequenao, "Computer Generated image of the ATLAS calorimeter", 2008, URL: http://cds.cern.ch/record/1095927 (cit. on p. 22).
- [39] J. Pequenao, "Computer generated image of the ATLAS Muons subsystem", 2008, URL: http://cds.cern.ch/record/1095929 (cit. on p. 24).
- [40] M. Aaboud et al., Performance of the ATLAS Trigger System in 2015, (2016), arXiv: 1611.09661 [hep-ex] (cit. on p. 24).
- [41] G. Aad et al., Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, Eur. Phys. J. C75 (2015) 17, arXiv: 1406.0076 [hep-ex] (cit. on p. 27).
- [42] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: **0802.1189** [hep-ph] (cit. on p. 27).

- [43] W. Lampl et al., Calorimeter Clustering Algorithms: Description and Performance, tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, 2008, URL: http://cds.cern.ch/record/1099735 (cit. on p. 27).
- [44] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. B659 (2008) 119, arXiv: 0707.1378 [hep-ph] (cit. on p. 27).
- [45] G. Aad et al., *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex] (cit. on p. 28).
- [46] T. A. collaboration, Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC, (2015), URL: https: //atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-045/ (cit. on pp. 28, 29).
- [47] G. Aad et al., *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at* \sqrt{s} =8 *TeV*, Eur. Phys. J. **C75** (2015) 303, arXiv: 1412.7086 [hep-ex] (cit. on pp. 28–30).
- [48] B. P. Roe et al., *Boosted decision trees, an alternative to artificial neural networks*, Nucl. Instrum. Meth. A**543** (2005) 577, arXiv: physics/0408124 [physics] (cit. on p. 28).
- [49] G. Aad et al., *Performance of algorithms that reconstruct missing transverse momentum in* $\sqrt{s} = 8$ *TeV proton-proton collisions in the ATLAS detector*, (2016), arXiv: 1609.09324 [hep-ex] (cit. on pp. 30, 31).
- [50] G. Aad et al., Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C76 (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 31).
- [51] T. Cornelissen et al., Concepts, Design and Implementation of the ATLAS New Tracking (NEWT), tech. rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, 2007, URL: http://cds.cern.ch/record/1020106 (cit. on p. 31).
- [52] Performance of the ATLAS Silicon Pattern Recognition Algorithm in Data and Simulation at $\sqrt{s} = 7$ TeV, tech. rep. ATLAS-CONF-2010-072, CERN, 2010, URL: http://cds.cern.ch/record/1281363 (cit. on p. 31).
- [53] G. Aad et al., Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data, Eur. Phys. J. C74 (2014) 2941, arXiv: 1404.2240 [hep-ex] (cit. on pp. 31, 32).
- [54] T. A. collaboration, Search for squarks and gluinos in events with hadronically decaying tau leptons, jets and missing transverse momentum in $36fb^{-1}$ of proton-proton collisions at $\sqrt{s} = 13$ TeV recorded with the ATLAS detector, (to be published) (cit. on pp. 33, 48, 50, 51, 53, 65, 73, 87).
- [55] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP 05 (2006) 026, arXiv: hep-ph/0603175 [hep-ph] (cit. on p. 33).
- [56] T. Sjostrand, S. Mrenna and P. Z. Skands, A Brief Introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 33).
- [57] T. Gleisberg et al., Event generation with SHERPA 1.1, JHEP 02 (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on p. 33).
- [58] S. Hoeche et al., QCD matrix elements and truncated showers, JHEP 05 (2009) 053, arXiv: 0903.1219 [hep-ph] (cit. on p. 33).

- [59] T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 33).
- [60] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 [hep-ph] (cit. on p. 33).
- [61] S. Frixione, P. Nason and C. Oleari, Matching NLO QCD computations with Parton Shower simulations: the POWHEG method, JHEP 11 (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on p. 33).
- [62] S. Alioli et al., A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX, JHEP 06 (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on pp. 33, 34).
- [63] H.-L. Lai et al., New parton distributions for collider physics, Phys. Rev. D82 (2010) 074024, arXiv: 1007.2241 [hep-ph] (cit. on p. 34).
- [64] J. Pumplin et al.,
 New generation of parton distributions with uncertainties from global QCD analysis,
 JHEP 07 (2002) 012, arXiv: hep-ph/0201195 [hep-ph] (cit. on p. 34).
- [65] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. D82 (2010) 074018, arXiv: 1005.3457 [hep-ph] (cit. on p. 34).
- [66] S. Owen, Data-driven estimation of the multijet background to SUSY searches with jets and missing transverse momentum at ATLAS using jet smearing, ATL-PHYS-INT-2012-008, 2012 (cit. on pp. 37, 40, 42–44).
- [67] S. Owen, *Multijet Background Estimation for Supersymmetry Searches Using the ATLAS Detector at the Large Hadron Collider*, PhD thesis: University of Sheffield, 2012 (cit. on pp. 37, 40, 41).
- [68] G. T. Fletcher, Multijet Background Estimation for SUSY Searches And Particle Flow Offline Reconstruction Using The ATLAS Detector At The LHC, PhD thesis: University of Sheffield, 2015 (cit. on pp. 37, 40, 42, 44, 78–80).
- [69] Monte Carlo Generators for the Production of a W or Z/γ* Boson in Association with Jets at ATLAS in Run 2, tech. rep. ATL-PHYS-PUB-2016-003, CERN, 2016, URL: http://cds.cern.ch/record/2120133 (cit. on p. 39).
- [70] Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy sqrts = 7 TeV, tech. rep. ATLAS-CONF-2010-054, CERN, 2010, URL: http://cds.cern.ch/record/1281311 (cit. on p. 42).
- [71] A. Barr, C. Lester and P. Stephens, *m*(*T*2): *The Truth behind the glamour*,
 J. Phys. **G29** (2003) 2343, arXiv: hep-ph/0304226 [hep-ph] (cit. on p. 47).
- [72] S. Chatrchyan et al., Search for New Physics with Jets and Missing Transverse Momentum in pp collisions at $\sqrt{s} = 7$ TeV, JHEP 08 (2011) 155, arXiv: 1106.4503 [hep-ex] (cit. on p. 77).

APPENDIX A

Additional Figures

A.1 Distributions after different combinations of cuts on new variables



Figure A.1: Comparison between data and SM prediction for a selection with one tau lepton, zero muons, at least three jets, inverted QCD rejection cuts and a cut on H_T over m_{eff} . The SM consists of the multijets (red) and the combined electroweak backgrounds (green). Shown are (a) N_{jet} , (b) $p_T^{\text{jet}_1}$, (c) $\not{\!\!E}_T$, (d) $m_T^{\tau_1}$ and (e) $\Delta \phi$ (jet₁, $\not{\!\!E}_T$) after an additional cut on $H_T/m_{\text{eff}} > 0.65$. For all plots also the tau ID criteria are applied. The multijets distributions are normalized to the difference between the numbers of events in data and in the electroweak background.





Figure A.3: Comparison between data and SM prediction for a selection with one tau lepton, zero muons, at least three jets, inverted QCD rejection cuts and and a cut on $p_T^{\tau_1}$ over m_{eff} . The SM consists of the multijets (red) and the combined electroweak backgrounds (green). Shown are (a) N_{jet} , (b) $p_T^{\text{jet}_1}$, (c) $\not{\!\!E}_T$, (d) $m_T^{\tau_1}$ and (e) $\Delta \phi$ (jet₁, $\not{\!\!E}_T$) after an additional cut on $p_T^{\tau_1}/m_{\text{eff}} < 0.1$. For all plots also the tau ID criteria are applied. The multijets distributions are normalized to the difference between the numbers of events in data and in the electroweak background.



Figure A.4: Comparison between data and SM prediction for a selection with one tau lepton, zero muons, at least three jets, inverted QCD rejection cuts and cuts on $p_T^{\tau_1}$ over m_{eff} and H_T over m_{eff} . The SM consists of the multijets (red) and the combined electroweak backgrounds (green). Shown are (a) N_{jet} , (b) $p_T^{\text{jet}_1}$, (c) $\not{\!\!E}_T$, (d) $m_T^{\tau_1}$ and (e) $\Delta \phi$ (jet₁, $\not{\!\!E}_T$) after additional cuts on $p_T^{\tau_1}/m_{\text{eff}} < 0.1$ and $H_T/m_{\text{eff}} > 0.65$. For all plots also the tau ID criteria are applied. The multijets distributions are normalized to the difference between the numbers of events in data and in the electroweak background.







A.2 Pre-fit distributions in the control and signal regions of the analysis with the full 36 fb⁻¹ dataset

Figure A.7: Pre-fit distributions in the W true tau CR of the analysis using the full statistics. Shown are (a) the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton, (b) the number of jets N_{jet} , (c) the missing transverse energy $\not\!\!\!E_T$, the azimutal angle separations (d) $\Delta\phi$ (jet₁, $\not\!\!\!E_T$) and (e) $\Delta\phi$ (jet₁, $\not\!\!\!E_T$) between the leading and sub-leading jet, respectively, and the $\not\!\!\!E_T$ and the transverse mass (f) $m_T^{\tau_1}$ of the tau lepton. The multijets background is normalized with the obtained normalization factor for medium tau leptons.


Figure A.8: Pre-fit distributions in the top true tau CR of the analysis using the full statistics. Shown are (a) the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton, (b) the number of jets N_{jet} , (c) the missing transverse energy $\not\!\!E_T$, the azimutal angle separations (d) $\Delta\phi$ (jet₁, $\not\!\!E_T$) and (e) $\Delta\phi$ (jet₁, $\not\!\!E_T$) between the leading and sub-leading jet, respectively, and the $\not\!\!E_T$ and the transverse mass (f) $m_T^{\tau_1}$ of the tau lepton. The multijets background is normalized with the obtained normalization factor for medium tau leptons.



Figure A.9: Pre-fit distributions in the W fake tau CR of the analysis using the full statistics. Shown are (a) the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton, (b) the number of jets N_{jet} , (c) the missing transverse energy $\not\!\!\!E_T$, the azimutal angle separations (d) $\Delta\phi$ (jet₁, $\not\!\!\!E_T$) and (e) $\Delta\phi$ (jet₁, $\not\!\!\!E_T$) between the leading and sub-leading jet, respectively, and the $\not\!\!\!E_T$ and the transverse mass (f) $m_T^{\tau_1}$ of the tau lepton. The multijets background is normalized with the obtained normalization factor for medium tau leptons.



Figure A.10: Pre-fit distributions in the top fake tau CR of the analysis using the full statistics. Shown are (a) the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton, (b) the number of jets N_{jet} , (c) the missing transverse energy $\not\!\!E_T$, the azimutal angle separations (d) $\Delta\phi$ (jet₁, $\not\!\!E_T$) and (e) $\Delta\phi$ (jet₁, $\not\!\!E_T$) between the leading and sub-leading jet, respectively, and the $\not\!\!E_T$ and the transverse mass (f) $m_T^{\tau_1}$ of the tau lepton. The multijets background is normalized with the obtained normalization factor for medium tau leptons.



Figure A.11: Pre-fit distributions in the $Z \to \tau \tau$ CR of the analysis using the full statistics. Shown are (a) the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton, (b) the number of jets N_{jet} , (c) the missing transverse energy $\not\!\!\!E_T$, the azimutal angle separations (d) $\Delta \phi$ (jet₁, $\not\!\!\!E_T$) and (e) $\Delta \phi$ (jet₁, $\not\!\!\!E_T$) between the leading and sub-leading jet, respectively, and the $\not\!\!\!E_T$ and the transverse mass (f) $m_T^{\tau_1}$ of the tau lepton. The multijets background is normalized with the obtained normalization factor for medium tau leptons.





Figure A.13: Pre-fit distributions in the Compressed SR of the analysis using the full statistics. Shown are (a) the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton, (b) the number of jets N_{jet} , (c) the missing transverse energy $\not\!\!E_T$, the azimutal angle separations (d) $\Delta \phi$ (jet₁, $\not\!\!E_T$) and (e) $\Delta \phi$ (jet₁, $\not\!\!E_T$) between the leading and sub-leading jet, respectively, and the $\not\!\!E_T$ and the transverse mass (f) $m_T^{\tau_1}$ of the tau lepton. There are no data point included in the plots since the signal regions are still blinded in the main analysis. The multijets background is normalized with the obtained normalization factor for medium tau leptons.



Figure A.14: Pre-fit distributions in the High Mass SR of the analysis using the full statistics. Shown are (a) the transverse momentum $p_T^{\tau_1}$ of the leading tau lepton, (b) the number of jets N_{jet} , (c) the missing transverse energy $\not\!\!E_T$, the azimutal angle separations (d) $\Delta\phi$ (jet₁, $\not\!\!E_T$) and (e) $\Delta\phi$ (jet₁, $\not\!\!E_T$) between the leading and sub-leading jet, respectively, and the $\not\!\!E_T$ and the transverse mass (f) $m_T^{\tau_1}$ of the tau lepton. There are no data point included in the plots since the signal regions are still blinded in the main analysis. The multijets background is normalized with the obtained normalization factor for medium tau leptons.

List of Figures

2.1	Overview of the elementary particles included in the standard model	4
2.2	Evolution of the strong interaction coupling constant α_S with energy	7
2.3	Schematic illustration of a proton-proton collision.	8
2.4	Feynman diagram of the decay of a tau lepton	9
2.5	Branching fractions for the main decay channels of tau leptons	10
2.6	Difference between the spread of jets and of tau leptons.	10
2.7	Summary of the total and fiducial production cross-sections for several standard model	
	processes measured with ATLAS.	11
2.8	Loop corrections to the Higgs mass caused by the coupling to a fermion and a scalar.	12
2.9	Fractions of ordinary matter, dark matter and dark energy in the universe	13
2.10	Evolution of the coupling constants of the three SM forces with energy	13
2.11	The for this thesis relevant simplified model	16
3.1	The LHC accelerator and its experiments	18
3.2	Peak luminosities by fill and interactions per bunch crossing (pile-up)	19
3.3	Overview of the ATLAS detector	20
3.4	Setup of the ATLAS inner detector	21
3.5	The ATLAS calorimeter	22
3.6	Overview of the ATLAS muon system and the toroid magnet system	24
4.1	Distributions of some BDT input variables for the signal and the background sample used for the BDT training for the tau lepton identification.	30
5.1	Example Feynman diagram of a multijets event faking a characteristical SUSY signature.	35
6.1	Transverse mass of the leading tau lepton with dijet MC used for the multijets background.	38
6.2	Comparison of different MC generators	39
6.3	Light jet response map and response function for $100 \text{ GeV} < p_T^{\text{jet}} < 120 \text{ GeV} \dots$	41
6.4	Principle of the Jet Smearing technique.	46
8.1	Shape comparison between data and multijets background for different distributions for	
	the preliminary multijets control region selection.	55
8.2	Shape comparison between data and multijets background for different distributions for	
	the preliminary multijets control region selection after the cut on $m_{\rm T}^{\tau_1}$	57
8.3	Shape comparison between data and SM background for different distributions for the	
	preliminary multijets control region selection after the cut on $m_T^{\tau_1}$	58
8.4	Comparison of data and the SM backgrounds for baseline and signal objects	60

8.5	Comparison between data and the SM prediction for the preliminary multijets control region without the cut on $m_{T}^{\tau_1}$.	62
8.6	Comparison between data and the SM prediction for the new variable combinations in the preliminary multijets control region.	63
8.7	Comparison between data and the SM prediction for 3.2 fb^{-1} in the final QCD control region.	64
9.1	Sketch of the issue of the seed event uncertainty treatment.	68
9.2	Example for the effect of the new uncertainty treatment.	69
9.3	Comparison between data and the SM backgrounds for some distributions in the multijets control region using loose tau leptons.	70
9.4	Comparison between data and the SM backgrounds for some distributions in the multijet	
9.5	control region for medium tau leptons	71
0.6	intervals	73
9.6	$p_{\rm T}^{\rm iet}$ for the control regions of the analysis using the full statistics.	74
9.7	$p_{\rm T}^{\rm per}$ for the signal regions of the analysis using the full statistics	75
10.1	Scatter plots of the transverse mass $m_T^{\tau_1}$ of the leading tau lepton versus its transverse momentum $p_T^{\tau_1}$ for data and multijets background for loose tau leptons in the multijets	
	control region.	78
10.2	Effect of a tau lepton smearing with a Gaussian on tau-related quantities	79
10.3	Effect of a variation of the Gaussian width on the tau lepton smearing.	81 82
10.4	Effect of a fau lepton smearing with the jet response function	82
A.1	Comparison between data and SM prediction for a selection with one tau lepton, zero	
A.2	muons, at least three jets, inverted QCD rejection cuts and a cut on H_T over m_{eff} Comparison between data and SM prediction for a selection with one tau lepton, zero	96
	muons, at least three jets, inverted QCD rejection cuts and a cut on the missing transverse energy over m_{eff} .	97
A.3	Comparison between data and SM prediction for a selection with one tau lepton, zero	
	muons, at least three jets, inverted QCD rejection cuts and a cut on $p_T^{\tau_1}$ over m_{eff}	98
A.4	Comparison between data and SM prediction for a selection with one tau lepton, zero muons at least three jets inverted QCD rejection outs and outs on r^{τ_1} over m_{τ_2} and H_{τ_2}	
	muons, at least three jets, inverted QCD rejection cuts and cuts on $p_{\rm T}$ over $m_{\rm eff}$ and $H_{\rm T}$	99
A.5	Comparison between data and SM prediction for a selection with one tau lepton, zero	
	muons, at least three jets, inverted QCD rejection cuts and cuts on $p_{\rm T}^{\tau_1}$ over $m_{\rm eff}$ and the	
	missing transverse energy over m_{eff} .	100
A.6	Comparison between data and SM prediction for a selection with one tau lepton, zero muons, at least three jets, inverted QCD rejection cuts and cuts on $p_T^{\tau_1}$ over m_{eff} , H_T over	
<u>۸</u> 7	$m_{\rm eff}$ and the missing transverse energy over $m_{\rm eff}$.	101
A./	Pre-fit distributions in the top true tau CR of the analysis using the full statistics	102
A.9	Pre-fit distributions in the W fake tau CR of the analysis using the full statistics.	104
A.10	Pre-fit distributions in the top fake tau CR of the analysis using the full statistics	105
A.11	Pre-fit distributions in the $Z \rightarrow \tau \tau$ CR of the analysis using the full statistics	106
A.12	Pre-fit distributions in the $Z \rightarrow \nu\nu$ CR of the analysis using the full statistics	107

A.13 Pre-fit distributions in the Compressed SR of the analysis using the full statistics.	 108
A.14 Pre-fit distributions in the High Mass SR of the analysis using the full statistics	 109

List of Tables

Overview of the tools used for the simulation of the electroweak background samples	34
Parameter values for the calculation of σ_{corr} used in the dijet analysis Parameter values for the $\sigma_{\Delta\phi}$ calculation used within this thesis for the jet ϕ smearing	43 45
Signal regions for the 2τ channel of the analysis with 3.2 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ data collected in 2015.	49
Control regions for the 2τ channel of the analysis with 3.2 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ data collected in 2015.	49
Validation regions for the 2τ channel of the analysis with 3.2 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ data collected in 2015.	50
Signal regions for the 2τ channel of the analysis with 36 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ data collected in 2015 and 2016.	50
Control regions for the 2τ channel of the analysis with 36 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ data collected in 2015 and 2016.	51
Definitions of baseline and signal objects for tau leptons and muons.	59
Multijets control region for the 2τ channel determined with the 3.2 fb ⁻¹ of $\sqrt{s} = 13$ feV data collected in 2015.	65
Yields for data, the expected total nominal EW background, the nominal multijets background and seed events in the multijets control region for loose and medium tau	
leptons	72
multijets background and seed events in the control and signal regions with medium tau leptons.	76
Parameter values for a, b and c for data and MC used for the calculation of the width of the Gaussian taken for the tau lepton smearing.	80
Parameters a, b and c for data and MC used for the calculation of the width of the Gaussian taken for the tau lepton smearing.	80
	Overview of the tools used for the simulation of the electroweak background samples Parameter values for the calculation of σ_{corr} used in the dijet analysis Parameter values for the $\sigma_{\Delta\phi}$ calculation used within this thesis for the jet ϕ smearing Signal regions for the 2τ channel of the analysis with 3.2 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ data collected in 2015