

Probing natural supersymmetry at the CMS experiment based on precise jet measurements

Dissertation
zur Erlangung des Doktorgrades
des Fachbereichs Physik
der Universität Hamburg

vorgelegt von
Kristin Goebel geb. Heine
aus Gifhorn

Hamburg
2014

Gutachter der Dissertation: Prof. Dr. bla
Prof. Dr. bla

Gutachter der Disputation: Prof. Dr. bla
Prof. Dr. bla

Datum der Disputation: ?? . ?? 2014

Vorsitzender des Prüfungsausschusses: Dr. bla

Vorsitzender des Promotionsausschusses: Prof. Dr. bla

Leiterin des Fachbereichs Physik: Prof. Dr. bla

Dekan der Fakultät für Mathematik,
Informatik und Naturwissenschaften: Prof. Dr. bla

Abstract

The search for signals of new physics processes beyond the standard model of particle physics is one of the main goals of the CMS experiment at the CERN Large Hadron Collider. Since many theories like e.g. supersymmetry involve the production of new coloured particles which feature jets as their experimental signature, it is important to have a good understanding of jet properties in order to allow such searches.

In the first part of this thesis, a measurement of the jet transverse momentum resolution is presented based on the analysis of proton-proton collision data recorded at a center of mass energy of 8 TeV by the CMS experiment. The measurement utilizes the transverse momentum balance of dijet events at particle level. The main focus lies on the determination of the data-to-simulation ratio of the resolution which can be used to derive scale factors in order to correct the resolution in simulated events to match the observed resolution in data.

The second part of the thesis focuses on searches for supersymmetry in final states with jets and missing transverse momentum. First, a search performed on collision data recorded at $\sqrt{s} = 8$ TeV is presented which targets mainly signatures sensitive to the production of light squarks and gluinos as well as the production of gluino mediated third generation particles. Here the main focus lies on the estimation of background contributions arising from QCD multijet events. Second, a study based on simulated events at a center of mass energy of 13 TeV is shown investigating different search strategies towards the identification of directly produced top squarks. Utilizing algorithms for the identification of boosted hadronically decaying top quarks arising from the decay of heavy top squarks, a search sensitivity of top squark masses up to 1 TeV is expected for LSP masses less than approximately 200 GeV.

Kurzfassung

Die Suche nach Hinweisen auf Prozesse von neuer Physik jenseits des Standardmodells der Teilchenphysik ist eines der Hauptziele des CMS Experiments am CERN Large Hadron Collider. Da zahlreiche Theorien wie beispielsweise Supersymmetrie die Produktion von neuen farbgeladenen Teilchen, welche als experimentelle Signatur Jets aufweisen, beinhalten, ist es wichtig ein gutes Verständnis dieser Objekte zu erlangen um derartige Suchen zu ermöglichen.

Im ersten Teil dieser Arbeit wird eine Messung der Jet-Transversalimpuls-Auflösung vorgestellt, welche auf der Analyse von Proton-Proton-Kollisionsdaten basiert, die bei einer Schwerpunktsenergie von 8 TeV vom CMS Experiment aufgezeichnet wurden. Die Messung nutzt dabei die Transversalimpulsbalance von Zweijet-Ereignissen auf Teilchenniveau. Der Fokus liegt dabei auf der Bestimmung des Verhältnisses der Auflösung in Daten zu der Auflösung in simulierten Ereignissen, welches verwendet werden kann um Skalierungsfaktoren zu bestimmen, welche die in simulierten Ereignissen gemessene Auflösung an die in Daten beobachtete anpassen.

Der zweite Teil der Arbeit konzentriert sich auf Suchen nach Supersymmetrie in Endzuständen mit Jets und fehlendem Transversalimpuls. Zunächst wird eine Suche unter Verwendung von bei einer Schwerpunktsenergie von $\sqrt{s} = 8$ TeV aufgezeichneten Kollisionsdaten durchgeführt, die auf Signaturen abzielt, welche hauptsächlich sensitiv sind auf die Produktion von leichten Squarks und Gluinos sowie die Produktion von gluinoindizierten Teilchen der dritten Generation. Der Schwerpunkt liegt hier auf der Abschätzung des Untergrundbeitrags durch QCD Mulijet Ereignisse. Als zweites wird eine Studie vorgestellt, welche auf simulierten Ereignissen bei einer Schwerpunktsenergie von 13 TeV basiert und unterschiedliche Analysestrategien zur Identifikation von direkt produzierten Top Squarks untersucht. Unter Verwendung von Algorithmen zur Identifikation von geboosteten hadronisch zerfallenden Top Quarks aus den Zerfällen von Top Squarks, kann eine Sensitivität der Suche für Top Squarks Massen bis 1 TeV für LSP Massen bis circa 200 GeV erwartet werden.

Contents

1. Introduction	1
2. Phenomenological Aspects of the Standard Model and Beyond	3
2.1. The Standard Model of Particle Physics	3
2.1.1. Limitations of the Standard Model	6
2.2. Supersymmetry	7
2.2.1. The MSSM	9
2.2.2. SUSY-Breaking	11
2.3. Searches for Supersymmetry and Current Constraints	12
2.3.1. Indirect Constraints	12
2.3.2. Direct Searches at Collider Experiments	12
3. Experimental Setup	20
3.1. The Large Hadron Collider	20
3.2. The CMS Experiment	22
3.2.1. Coordinate Conventions and Kinematic Variables	23
3.2.2. Superconducting Magnet	24
3.2.3. Inner Tracking System	24
3.2.4. Electromagnetic Calorimeter	26
3.2.5. Hadron Calorimeter	27
3.2.6. Muon System	29
3.2.7. Trigger System	30
3.3. LHC Operation and Data Taking	31
3.4. Event Simulation	32
4. Object Reconstruction and Particle Identification	38
4.1. Global Event Description with the Particle-Flow Algorithm at CMS	38
4.2. Reconstruction of Jets	39
4.2.1. Jet Algorithms	39
4.2.2. Jet Types at CMS	42
4.2.3. Jet Transverse-Momentum Response	43
4.2.4. Jet Energy Calibration	44
4.3. Identification of b-Quark Jets	46
4.4. Identification of Boosted t-Quark Jets	48
5. Measurement of the Jet Transverse-Momentum Resolution	54
5.1. Components of the Jet Response	54
5.2. Basic Concept of the Dijet Asymmetry Method	56
5.3. Application to Realistic Collision Events	56
5.4. Samples and Event Selection	57

5.4.1. Datasets and Triggers	57
5.4.2. Selection Criteria	58
5.5. Definition of the Asymmetry Width	62
5.6. Corrections to the Dijet Asymmetry	63
5.6.1. Correction for Additional Jet Activity	65
5.6.2. Correction for Particle-Level Imbalance	67
5.6.3. Results of the Corrections to the Asymmetry	68
5.7. Determination of the Data-to-Simulation Ratio	68
5.8. Validation of the Method	70
5.8.1. Validation in Simulated Events	70
5.8.2. Validation of the Measured Data-to-Simulation Ratio	71
5.9. Systematic Uncertainties	71
5.10. Extension of the Method to the Forward Detector Region	78
5.11. Measurement for simulated events obtained with the HERWIG++ generator	80
5.12. Results	81
5.12.1. Determination of a combined result	81
5.12.2. Comparison to Other Measurements	84
5.13. Adjustment of the MC-Resolution to Data	86
6. Search for New Physics with Jets and Missing Transverse Momentum	87
6.1. Event Selection	88
6.1.1. Data Samples and Trigger	88
6.1.2. Event Cleaning	90
6.1.3. Baseline Selection	91
6.1.4. Exclusive Search Regions	93
6.2. Estimation of Non-QCD Backgrounds	94
6.2.1. Invisible Z Background	95
6.2.2. Hadronic τ Background	96
6.2.3. Lost-Lepton Background	98
6.3. QCD Background Estimation with the Rebalance-And-Smear Method	100
6.3.1. General concept of the Rebalance-and-Smear method	100
6.3.2. Application to Realistic Collision Events	105
6.3.3. Systematic Uncertainties	110
6.3.4. QCD Background Prediction	112
6.4. Results and Interpretation	112
6.4.1. Comparison to Other Measurements	118
6.5. Status of natural supersymmetry after LHC Run-I	119
7. Prospect Studies for a Search for Top Squarks at $\sqrt{s} = 13 \text{ TeV}$	120
7.1. Data samples	121
7.2. Sensitivity of a Basic Selection using H_T and \cancel{E}_T	122
7.3. Sensitivity Improvement using b Tagging	126
7.4. Sensitivity Improvement using top Tagging	129
7.5. Performance Comparison of Various Kinematic Selections	132
7.6. Performance Comparison to Selection Based on Published Stop Analysis at $\sqrt{s} = 8 \text{ TeV}$	139
7.7. Sensitivity to Gluino-Mediated Stop Production	142

7.8. Results and Discussion	143
8. Conclusions	145
A. Jet Transverse-Momentum Resolution	147
A.1. Primary Vertex Distributions	147
A.2. Extrapolation Graphs	149
A.3. Correlation of Asymmetry Widths	154
B. Search for New Physics with Jets and Missing Transverse Momentum	156
B.1. Data-to-Simulation Ratios: Core	156
B.2. Data-to-Simulation Ratios: Tails	157
C. Project Studies for Stop Quarks at $\sqrt{s} = 13 \text{ TeV}$	158
C.1. Kinematic Distributions	158

1. Introduction

The current knowledge and understanding of the fundamental constituents of matter and occurring interactions between them are summarized in the well-established standard model (SM) of particle physics. The SM which has been introduced in the early 1970's is to date a very successful theory as it was able to predict new particles in the past and is tested to very high precision. Beyond that there are several other fundamental questions unanswered like the origin of dark matter or the unification of coupling constants at a certain scale. One of such theories which goes beyond the standard model and is supposed to provide a solution to such open problems is *supersymmetry* (SUSY). A short introduction to the phenomenology of the standard model as well as supersymmetry can be found in Chapter 1 of this thesis.

In order to address all such questions about tests of the SM and new theories beyond, the CMS experiment located at the Large Hadron Collider (LHC) at CERN¹ has been build. It is designed to analyse particle collisions delivered by the LHC and is the experiment in whichs context this thesis has been performed. An overview of the technical setup of the CMS experiment as well as the current status of its' operation is given in Chapter 2. The general introduction to the CMS experiment is followed in Chapter 3 by an introduction to the reconstruction of certain objects recorded in the particle collisions as well as dedicated algorithms to identify specific decay processes, like e.g. decays of top-quarks.

As the name already indicates, the LHC is a hadron collider and thus the predominant objects to be measured in the detector are expected to be jets – the experimental signature of quarks and gluons. Thus it is very important to have a good understanding of these objects as e.g. a precise knowledge of the jet transverse momentum resolution. This can be measured utilizing events with a momentum balance in the transverse plane, like e.g. $\gamma + \text{jet}$ events, $Z + \text{jet}$ events or dijet events. A measurement of this quantity using dijet event topologies in data recorded by the CMS experiment at a center-of-mass energy of 8 TeV as well as simulated events is a main topic of this thesis and discussed in detail in Chapter 4.

The detailed knowledge about jets and their resolution can afterwards be exploited in a search for new physics mainly targeting signatures of decays involving supersymmetric particles in events with missing transverse energy and several jets. Such searches are especially sensitive to the production of gluinos and first and second generation squarks in supersymmetric models. The main focus regarding this new physics search lies on the estimation of background contributions arising from mismeasured jets and decays of heavy flavour quarks to the final states of interest. An overview of this analysis which is also based on the full dataset recorded by the CMS experiment at a center-of-mass energy of 8 TeV with special emphasis on the QCD multijet background determination is presented in Chapter 5.

Finally, Chapter 6 summarizes a prospect study for searches for supersymmetry at a

¹European Organization for Nuclear Research near Geneva, Suisse

center-of-mass energy of 13 TeV which will be most likely the operation energy of the LHC in 2015 when the next running period starts. In this study, it is investigated how the classical approach of the analysis presented in Chapter 5 can be adapted and improved in order to be sensitive not only to gluino and first and second generation squarks but to the direct production of top squarks. Special emphasis lies on the study of specific kinematic variables and the application of jet substructure tools.

The thesis ends with a short summary and outlook.

2. Phenomenological Aspects of the Standard Model and Beyond

The standard model of particle physics (SM) describes the fundamental particles and interactions between them. It is a theory that successfully predicted the existence of several at that time undiscovered particles. In addition, it has been tested extensively in electroweak precision measurements at LEP.

Although the SM shows so far a remarkably successful performance there are also some open questions which can not be answered within the SM. Thus several theories have been developed to address problems which go beyond the SM. One of such well-motivated extensions is supersymmetry (SUSY) for which however no experimental evidence has been found so far.

After a short introduction to the phenomenology of the standard model including a discussion of specific shortcomings, the basic concepts of supersymmetry are introduced in this chapter. In addition, general concepts of searches for supersymmetry at collider experiments are discussed together with a summary of the current status of the results of such searches which have been performed in the past.

2.1. The Standard Model of Particle Physics

The description of the SM comprises the elementary particles and their interactions [1]. In general, one distinguishes between two types of particles: fermions and bosons. While matter particles are fermions with half-integer spin, the fundamental forces are mediated via bosons carrying integer spin. An overview of the contents of the SM is given in Fig. 2.1 where the particles are denoted together with their interactions ¹.

Mathematically the standard model is a quantum field theory where interactions between particles are described via gauge symmetries. The underlying gauge group of the standard model is

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$$

where $SU(3)$ is the gauge group of the strong force and C indicates that this force acts on the colour charge, $SU(2)$ represents the weak force and L denotes that this force only acts on left-handed fermions, while $U(1)$ represents the electromagnetic force acting on the hypercharge Y .

A brief description of the properties of the particles contained in the SM and the corresponding interactions is given in the following:

Matter Constituents: In the SM, one distinguishes between twelve different fermions being the elementary constituents of matter. For each fermion there exists also an antiparticle which carries the opposite signed quantum numbers.

¹Gravity is not included in the current representation of the standard model and thus it is not discussed in this thesis.

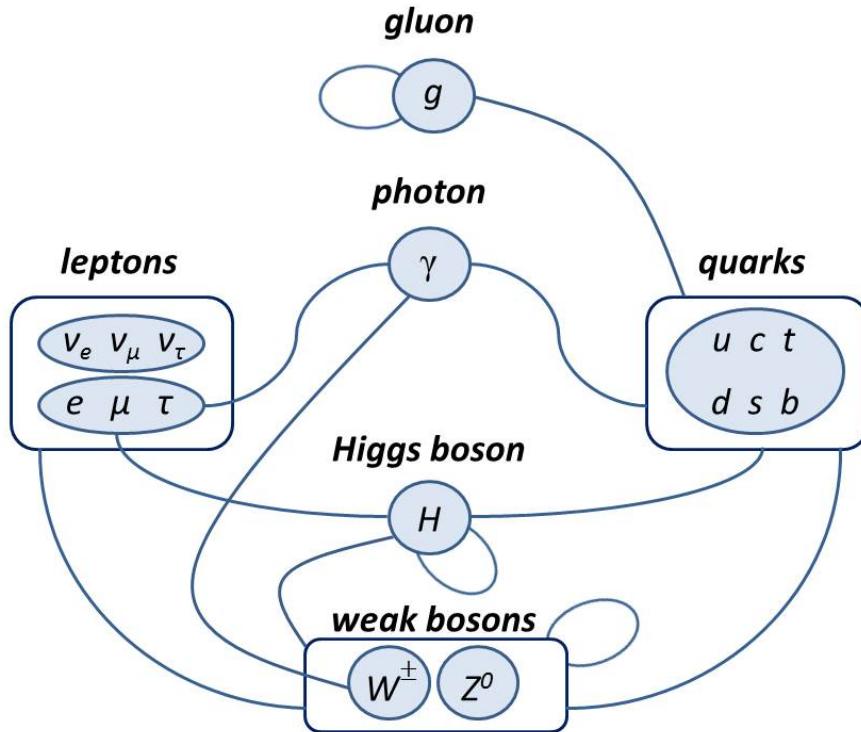


Figure 2.1.: Overview of particles contained in the standard model. Blue lines indicate interactions between different particles.

Leptons: The SM contains in total six leptons which are three negatively charged leptons (e, μ, τ) and three neutral leptons (ν_e, ν_μ, ν_τ), the neutrinos. In addition to the charge, leptons are also distinguished according to the lepton numbers which are electron number $L_e = 1$ for electron and electron-neutrino, muon number $L_\mu = 1$ for muon and muon-neutrino and tauon number $L_\tau = 1$ for tauon and tauon-neutrino. Each pair of lepton and neutrino carrying the same lepton number is forming a so-called *generation* where e and ν_e belong to the first generation, μ and ν_μ to the second and τ and ν_τ to the third, respectively.

Quarks: The remaining six fermions in the SM are quarks and can be grouped into generations analogous to the leptons. The first generation is populated by up- and down-quark (u, d), the second by charm- and strange-quark (c, s) and the third by the top- and bottom-quark (t, b). All quarks carry electrical charge but different than for leptons it is not integer but $+2/3$ for the up-type quarks (u, c, t) and $-1/3$ for down-type quarks (d, s, b). Besides to the electrical charge, quarks also carry color charge which comes in three types.

In addition to the attributes described above, fermions are furthermore characterized by the weak isospin. Left-handed fermions in each generation form an isodoublet with a weak isospin of $\pm 1/2$ while right-handed components are isosinglets with a weak isospin of 0.

Fundamental Forces: Matter particles interact with each other through fundamental forces mediated via gauge bosons. These bosons arise from the principle of local gauge invariance under symmetry transformations.

Electromagnetic Force: The description of the electromagnetic force is based on the theory of *Quantum Electrodynamics* (QED). It is communicated between electrically charged particles, like the charged leptons and quarks, by the exchange of photons. These are massless and electrically neutral resulting in the property that the electromagnetic force is long ranged.

Weak Force: The weak force acts on the weak charge which is carried by all fermions in the SM. It is mediated by three vector bosons – namely the two charged W^\pm bosons and the neutral Z boson. These bosons are unlike the photon massive with masses of $W^\pm = 80.385 \pm 0.015 \text{ GeV}$ and $Z = 91.1876 \pm 0.0021 \text{ GeV}$ [1]. As a result the weak interaction is suppressed with respect to the electromagnetic force.

Weak interactions preferably take place within one generation, e.g. e or μ transform into the respective neutrinos when emitting a W^- . However, since the mass eigenstates in the weak interaction differ from the flavour eigenstates, W^\pm bosons can also couple to fermions between different generations. In the quark-sector typically a representation is chosen where the up-type flavour eigenstates correspond to the mass eigenstates and the down-type quarks mix. This mixing is described by the CKM-matrix [2,3]. This is an unitary matrix, described by three mixing angles and one CP-violating phase, which indicates the relative strength between individual transitions.

Strong Force: The theoretical framework describing the strong force is called Quantum Chromodynamics (QCD). It is mediated via eight massless gluons and acts on the colour charge which is carried for instance by quarks. In contrast to the photon which is electrically neutral and thus can not interact with itself, gluons carry a colour charge and hence are able to couple to themselves. The colour charge exists in three different states commonly denoted as 'red', 'green' and 'blue'. Quarks and gluons are collectively referred to as partons.

Regarding the dependence on the distance the strong force behaves differently than other fundamental forces: the coupling strength increases with rising distance. This is a consequence of the different colour states and the self-coupling property of gluons. It is usually referred to as *confinement* [4] and causes the effect that coloured objects can not exist freely. In fact when separated, coloured objects start to build new coloured particles until only a colour neutral formation is left. Such colourless objects linked by the strong force are named hadrons. On the other hand, particles taking part in the strong interaction start to behave quasi-free when they are brought close together. This feature is known as *asymptotic freedom* [5,6].

First proposed by Salam, Glashow and Weinberg [7,8], the electromagnetic and the weak force could be successfully unified into the electroweak force. As denoted earlier, the weak force acts on the weak isospin T_3 while the electromagnetic force acts on the hypercharge Y . These two quantities are connected via the following

relation to the electric charge Q

$$Q = T_3 + Y/2$$

In the electroweak theory, three gauge bosons $W_\mu^{1,2,3}$ are introduced for $SU(2)_L$ and one gauge boson B_μ for $U(1)_Y$. The physical states photon, W^\pm and Z are formed by mixing of these massless states. While the charged W_μ^\pm bosons are superpositions of W_μ^1 and W_μ^2 , the fields A_μ of the photon and Z_μ of the neutral vector boson are obtained by a mixing of the gauge fields W_μ^3 and B_μ parametrized by the weak mixing angle θ_W .

Formel?

Higgs Boson: The electroweak theory in the current representation requires that fermions and bosons are massless particles as mass terms violate the gauge invariance under $SU(2)_L \otimes U(1)_Y$ transformations. This is in contradiction to experimental observations which have shown that all particles, except for photon and gluon, in fact have mass.

An explanation for the generation of particle masses without violation of the principles of the electroweak theory is provided by the *Higgs-mechanism* [9–11] which is based on the concept of spontaneous symmetry breaking. The main idea behind this mechanism is that in general the principle of local gauge invariance is obeyed, but that it is explicitly broken by the ground state. In the context of the Higgs-mechanism this is realized by the introduction of the Higgs field described by a potential with a non-zero minimum value. This non-vanishing minimum represents the expectation value of the quantum field in the vacuum – the vacuum expectation value. The masses of particles are eventually generated by the couplings to the Higgs field.

Formel?

Furthermore, the Higgs-mechanism also predicts the existence of a new heavy boson, the Higgs boson, which is a quantum excitation of one of the components of the Higgs field. It is supposed to be a scalar and the Higgs mass one of the free parameters of the SM.

The discovery of a new boson at a mass of around 125 GeV has been announced by the ATLAS and CMS collaborations in 2012 [12, 13]. As all properties of this new boson are consistent with SM predictions so far (c.f. e.g. [14–17]), this indicates that the last remaining gap of the SM could finally be closed.

2.1.1. Limitations of the Standard Model

Although the SM has been incredibly successful so far and lead to several discoveries while withstanding numerous precision tests, it is known to be on the other hand also an incomplete theory. Some of known shortcomings of the SM are discussed in some detail in the following.

Gravity: As stated already earlier, the SM contains no description of gravity. In particular, it is currently not possible to unify general relativity and quantum theory in one common concept.

Matter antimatter asymmetry: According to the SM matter and antimatter exist to equal amounts in the universe which is in fact not the case. A theory which would be able to explain such an asymmetry needs some source of CP-violation.

The only source of CP-violation within the SM is arising from the CKM matrix as described in 2.1. However, this is not enough to be able to explain the degree of matter antimatter asymmetry in the universe [18].

Unification of couplings: The unification of the electromagnetic and the weak force leads to the question, if it is even possible to further unify the electroweak force with the strong force in order to build a combined theory usually referred to as Grand Unified Theory (GUT). This would imply that the coupling constants of the SM intersect when extrapolating them from the electroweak to the GUT scale. However, this feature is not observed within the SM.

Origin of dark matter and dark energy: There exist several cosmological observations that indicate that the matter described by the SM makes up only 4.9% of the universe [19]. A by far larger part of 26.8% is assigned to so-called *dark matter* which is presumably neutral and only weakly interacting. The only particles within the SM possessing such attributes are neutrinos. However, they are not able to account for the whole relic density present in the universe [20]. Furthermore, for the major part of the universe making up 68.3% there is no hint at all what its' nature could be. Thus it is typically denoted with *dark energy*.

Hierarchy problem: The observable mass of the Higgs boson is given by the bare mass of the Higgs boson plus contributions arising from higher order corrections caused by each massive SM particle. These higher order corrections are usually dependent on an ultraviolet cut-off scale. Conventionally this UV cut-off scale is chosen to be the Planck scale. This would result in a Higgs mass considerably larger than the observed mass of around 125 GeV and requires an enormous amount of fine tuning in order to address this problem.

Formel?

2.2. Supersymmetry

In order to overcome the weaknesses of the SM and to provide explanations for so far unsolved problems of which some have been discussed in 2.1.1, several theories have been developed which go beyond the SM. Among those, a favoured extension is *supersymmetry* (SUSY) as it is able to provide several benefits at once. In this section a brief introduction to the general concept of supersymmetry is given with focus on the *Minimal Supersymmetric Standard Model* (MSSM). For detailed reviews see e.g. [21, 22].

The basic idea of a supersymmetric theory is that a fermionic state is converted into a bosonic state and vice versa by the generator of a supersymmetry transformation Q according to

$$Q |\text{fermion}\rangle = Q |\text{boson}\rangle, \quad Q |\text{boson}\rangle = Q |\text{fermion}\rangle.$$

The supersymmetric fermionic and bosonic partner particles are called *superpartner* and form together the irreducible representations of the supersymmetry algebra named *supermultiplets* with the same number of fermionic and bosonic degrees of freedom. In case of unbroken supersymmetry, partner particles within one supermultiplet have the same mass as well as the same quantum numbers like electric charge, weak isospin and color

degrees of freedom, except for the spin. Commonly, supersymmetric particles are denoted *sparticles*.

Furthermore, in a general supersymmetric theory fulfilling the criteria of gauge invariance and renormalisability, processes are allowed violating either the lepton or baryon number conservation. However, such a process has not been observed experimentally so far and would imply for instance a rapid decay of protons. The recent lower limit on the proton lifetime is found to be 5.9×10^{33} years at 90% confidence level [23] and indicates that such processes must be suppressed. In order to achieve this, a new quantum number called *R-parity* is introduced according to

$$R = (-1)^{3(B-L)+2S}$$

with baryon number B , lepton number L and spin S . It is a multiplicative quantum number and amounts to $R = +1$ for SM particles while it is $R = -1$ for supersymmetric particles. Assuming R-parity conservation no baryon or lepton number violation processes occur². In addition, the assumption of R-parity conservation leads to further phenomenological implications:

- SUSY particles can only be produced in pairs at collider experiments as only even numbers of supersymmetric particles can occur in an interaction vertex.
- The lightest supersymmetric particle (LSP) is stable and thus any decay chain of a supersymmetric particle finally ends in a state containing an odd number of LSPs.

A R-parity conserving supersymmetric theory as described above provides some elegant solutions to open questions as raised in section 2.1.1. As discussed there, the Higgs mass suffers from quadratically divergent contributions arising from higher order corrections caused by SM particles. However, since in SUSY each SM particle gets a supersymmetric partner these higher order corrections cancel as all supersymmetric particles contribute in an unbroken theory with equal size but opposite sign as their corresponding SM partner. Thus SUSY is able to provide a solution to the hierarchy problem. But the observation of such kind of supersymmetric particles with the exact same masses as their SM correspondents would have happened already. Since this is not the case, one knows from all experimental results that supersymmetry in fact has to be a broken symmetry. In order to still be able to provide a solution to the hierarchy problem, supersymmetric particles are expected to be not heavier than $\mathcal{O}(1\text{ TeV})$. This is the main argument why one would expect supersymmetric particles to be in the TeV range, well within the reach of the LHC. Especially the superpartner of the top quark is expected to be not too heavy in order to be able to cancel the contributions from top loops. This is necessary as these give the largest contributions since the top quark is the heaviest particle in the SM.

In addition to this naturalness aspects, the coupling constants of the forces which do not intersect within the SM actually meet in one point when extrapolating the couplings from the electroweak to the GUT scale considering the existence of supersymmetric particles. This effect is illustrated in Fig. 2.2. It is visible that the evolution of the couplings is modified with respect to the SM at that energy scale where the supersymmetric particles

²There exist also several R-parity violating SUSY models which are not in contradiction to the observed proton lifetime (cf. e.g. [22]) However, these are not subject of this thesis and thus not discussed.

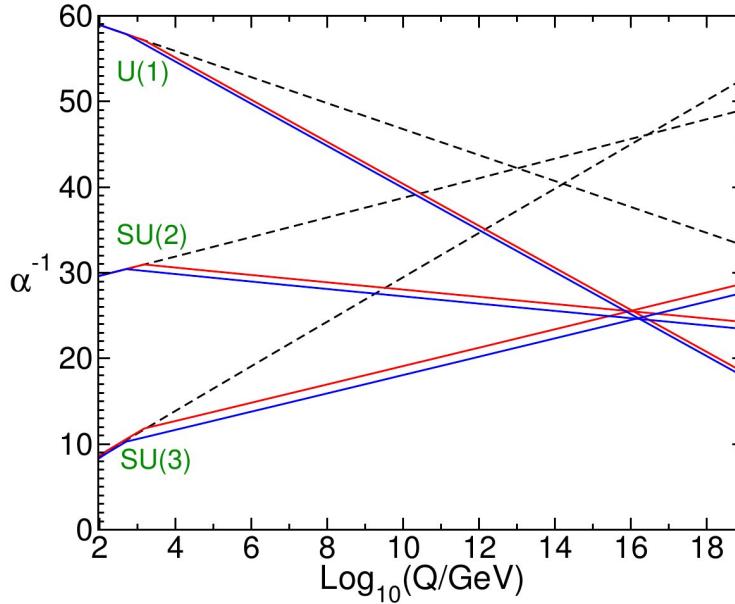


Figure 2.2.: Comparison of the renormalization group evolution of the couplings α_a^{-1} in the SM (dashed lines) and the MSSM (solid lines) including two-loop effects. The masses of the supersymmetric particles in the MSSM are considered as a common threshold changing between 500 GeV and 1.5 TeV while $\alpha_3(m_Z)$ is varied between 0.117 and 0.121. Taken from [22].

enter. In general, this hints to the possibility of a grand unification.

Besides the two advantages discussed above, R-parity conserving SUSY models also provide a suitable dark matter candidate. As discussed previously, each decay of supersymmetric particles finally leads to the existence of an LSP which can not decay further. Thus, it is an adequate DM candidate when it is electrically uncharged and only weakly interacting.

2.2.1. The MSSM

The smallest possible supersymmetric extension of the SM including the full particle spectrum and interactions is comprised in the MSSM. An overview of the respective particle content is given in Tab. 2.2.1. This extension is minimal in that sense that it introduces the least feasible number of additional particles to the existing SM particles. Thus, each SM fermion gets an individual superpartner while the interactions and couplings of the supersymmetric particles are the same as for the SM counterparts. The different transformation of left- and right-handed fermions under the gauge groups implies the necessity to introduce separate superpartners for left- and right-handed states as well. These are arranged with their bosonic (spin 0) superpartner in a *chiral* supermultiplet. The labels indicating the left- and right-handed states refer to the helicity of the respective SM particle. These supersymmetric partners of fermions are named *sfermions* distinguishing between *sleptons* and *squarks* – the supersymmetric partners of leptons and quarks. In a similar manner the SM gauge bosons are arranged in *gauge* supermultiplets together with their

Type	Spin	Gauge eigenstates	Mass eigenstates
Higgs bosons	0	$H_u^0 \ H_d^0 \ H_u^+ \ H_d^-$	$h^0 \ H^0 \ A^0 \ H^\pm$
Squarks	0	$\tilde{u}_L \ \tilde{u}_R \ \tilde{d}_L \ \tilde{d}_R$	see left
		$\tilde{s}_L \ \tilde{s}_R \ \tilde{c}_L \ \tilde{c}_R$	see left
		$\tilde{t}_L \ \tilde{t}_R \ \tilde{b}_L \ \tilde{b}_R$	$\tilde{t}_1 \ \tilde{t}_2 \ \tilde{b}_1 \ \tilde{b}_2$
Sleptons	0	$\tilde{e}_L \ \tilde{e}_R \ \tilde{\nu}_e$	see left
		$\tilde{\mu}_L \ \tilde{\mu}_R \ \tilde{\nu}_\mu$	see left
		$\tilde{\tau}_L \ \tilde{\tau}_R \ \tilde{\nu}_\tau$	$\tilde{\tau}_1 \ \tilde{\tau}_2 \ \tilde{\nu}_\tau$
Neutralinos	1/2	$\tilde{B}^0 \ \tilde{W}^0 \ \tilde{H}_u^0 \ \tilde{H}_d^0$	$\tilde{\chi}_1^0 \ \tilde{\chi}_2^0 \ \tilde{\chi}_3^0 \ \tilde{\chi}_4^0$
Charginos	1/2	$\tilde{W}^\pm \ \tilde{H}_u^\pm \ \tilde{H}_d^\pm$	$\tilde{\chi}_1^\pm \ \tilde{\chi}_2^\pm$
Gluino	1/2	\tilde{g}	see left
Gravitino	3/2	\tilde{G}	see left

Table 2.1.: Supersymmetric particles contained in the MSSM neglecting mixing in the first two sfermion generations. Following [22].

fermionic (spin 1/2) supersymmetric correspondents. The SUSY partners of the gauge bosons are named *gauginos* so that the superpartners in the gauge supermultiplets are the *gluino*, *wino* and *bino*. The corresponding gaugino mixtures of the neutral wino and the bino are the *photino* and the *zino*. Beyond that, the supersymmetric particle spectrum is extended by another supermultiplet containing the graviton (spin 2) and the respective supersymmetric partner – the *gravitino* (spin 3/2).

As described in Sec. 2.1, masses arise in the SM from the concept of spontaneous symmetry breaking implying the existence of the Higgs boson. The supersymmetric partner of the Higgs boson is named *higgsino*. While in the SM one Higgs doublet is sufficient to give mass to all particles, the Higgs sector needs to be extended in the MSSM. Here, two Higgs doublets are introduced where one doublet H_u gives mass to the up-type quarks and the other one H_d to the down-type quarks, respectively. These two doublets have together eight degrees of freedom of which three are needed to give mass to the gauge bosons of the weak interaction as in the SM. This results in five physical Higgs bosons which are the two scalar Higgs particles h^0, H^0 , the pseudoscalar A^0 as well as the charged Higgs bosons H^\pm . As further consequence there are two vacuum expectation values v_u and v_d present – each assigned to one Higgs doublet – whose ratio $\tan\beta = v_u/v_d$ is a free parameter of the model.

As in the SM, the gauge eigenstates of the theory do not necessarily have to be also the mass eigenstates. A mixing occurs especially in the gaugino sector. Here, the neutral components of the bino and wino mix with the neutral higgsinos and form four mass eigenstates – the *neutralinos* $\tilde{\chi}^0$. Similarly, also the charged gauginos and higgsinos mix to the four *charginos* $\tilde{\chi}^\pm$. Furthermore, mixing can also appear in the third squark and slepton generation. The mixing is supposed to be significant only for fermions of the third generation as the off-diagonal elements of the sfermion mass matrix are proportional to the mass of the respective SM partner.

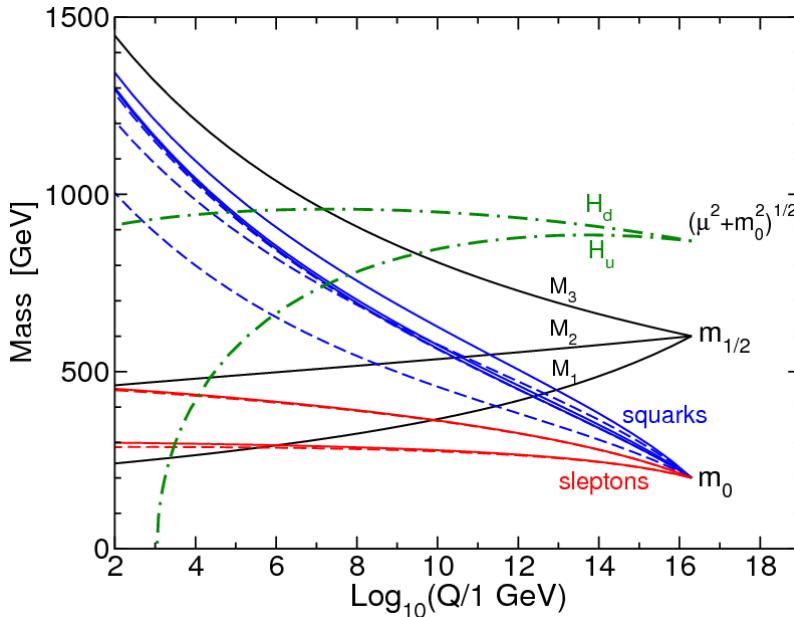


Figure 2.3.: Evolution of scalar and gaugino mass parameters in the MSSM with mSUGRA boundary conditions imposed at $Q_0 = 2 \times 10^{16} \text{ GeV}$. The parameter $\mu^2 + m_{H_u}^2$ runs negative, provoking electroweak symmetry breaking. Taken from [22].

2.2.2. SUSY-Breaking

As discussed above, SUSY particles are expected to have the same mass as their corresponding SM partner particle. However, none of such particles has been observed so far which implies that supersymmetry in fact must be broken and that sparticles are actually heavier than the SM counterpart. If SUSY is expected to provide a solution to the hierarchy problem, the mass difference should consequently not be too large in order to still allow a cancellation of loop contributions from SM particles to the Higgs mass. The characteristic mass scale should be around 1 TeV [22].

Typically, SUSY breaking is introduced such that so-called *soft breaking* terms are added to the theory in addition to the terms determining the gauge and Yukawa couplings. Since the MSSM does not contain such violating terms, it is extended assuming that the SUSY breaking occurs in a hidden sector which does not couple directly to the visible sector represented by the supermultiplets. Thus, a specific breaking mechanism must be assumed in order to describe the mediation of the supersymmetry breaking from the hidden to the visible sector. In total, the MSSM features several new phases, mixing angles and masses which add another 105 free parameters to the already existing parameters of the SM [24].

The two most popular SUSY breaking scenarios are either based on gravity-mediated or gauge-mediated interactions and known as *minimal supergravity* (mSUGRA) [25, 26] or *constrained MSSM* (CMSSM) [27, 28] and *gauge-mediated supersymmetry breaking* (GMSB) [29, 30]. Assuming a specific breaking scenario usually allows to drastically reduce the number of free parameters in the theory and determines the phenomenology of the respective model. In case of mSUGRA/CMSSM the whole model can be described by five

parameters which are the common scalar mass m_0 and the common mass of the gauginos and higgsinos $m_{1/2}$ at the GUT scale, the common trilinear coupling A_0 , $\tan\beta$ and the sign of the higgsino mass parameter μ . In Fig. 2.3 the evolution of the corresponding mass parameters to the electroweak scale are illustrated.

2.3. Searches for Supersymmetry and Current Constraints

The appealing attributes of supersymmetry which have been discussed above initiated a couple of indirect and direct searches looking for hints of supersymmetric particles. Although no sign for SUSY has been observed in nature so far, several results have been used in order to constrain the allowed parameter space. This section gives a short introduction to general search strategies for supersymmetry with main focus on collider experiments and provides a short overview of the status of experimental results. Results of direct searches at the LHC discussed in this section are restricted to results from searches performed previous to the analysis presented in this thesis. Further discussion of the status of supersymmetry after LHC Run-I follows in Sec. 6.5.

2.3.1. Indirect Constraints

Influences from the existence of supersymmetric particles can show up in manifold ways. For instance, higher-order contributions to SM processes could be induced from SUSY. Such contributions might impact for instance electroweak precision data, rare decays of B-mesons or the anomalous magnetic moment of the muon. However, global fits to electroweak precision data using several precision measurements of SM parameters and particle masses, like m_W and m_t , together with theoretical calculations have found no evidence for any inconsistency of the SM only hypothesis so far [31–34]. Furthermore, precise measurements of rare processes in B meson decays, like $B_s^0 \rightarrow \mu^+ \mu^-$, are in good agreement with expectations from the SM [35–37]. The most compelling difference between experimental results and SM prediction, currently is observed for the anomalous magnetic moment of the muon [38–40]. Here, deviations at the level of 3.6σ occur. However, discussions about the accuracy of the SM calculation are ongoing [41].

Further constraints arise from astrophysical and cosmological observations. These occur from direct [42] and indirect [43] dark matter searches. Moreover, several observations suggest that a considerable amount of cold dark matter contributes to the composition of the universe [19, 44]. Good candidates are weakly interacting massive particles (WIMPS), just as the neutralino, in case it is the lightest supersymmetric particle. Consequently, also the observed cold dark relic density can put strong constraints on the MSSM parameter space.

2.3.2. Direct Searches at Collider Experiments

Although the exploitation of indirect searches for supersymmetry is very useful in order to constrain the allowed SUSY parameter space, the most stringent exclusion limits are derived from direct searches at collider experiments. Typically, searches for SUSY at colliders make use of the specific production and decay properties which are caused by the assumption of R-parity conservation. These imply that sparticles are only produced in pairs and decay via cascades into the lightest supersymmetric particle which is often

assumed to be the lightest neutralino. This leaves the experiment undetected and manifests in missing energy or missing momentum. Such missing energy signatures are thus a key-feature of searches for supersymmetry in collider experiments. Depending on the type of supersymmetric particle produced the missing energy can be accompanied by several leptons, photons or jets. Usually, searches are classified according to their targeted final state and are aimed at a specific kind of supersymmetric particle. If searches are designed quite generic in order to be sensitive to various types of particles and models, they are called *inclusive* searches.

Extensive searches resulting in the tightest exclusion limits in the pre-LHC era have been realised by the experiments performed at HERA, LEP and Tevatron:

HERA [45,46]: At HERA searches for R-parity conserving supersymmetric models have been mainly targeting processes involving the production of a selectron and a squark. The dominant MSSM process at HERA is the production of a selectron and a squark via the exchange of a neutralino where \tilde{e} and \tilde{q} subsequently decay into any lighter gaugino and their respective SM partner. Thus, a distinct experimental signature is given by an electron, hadrons and missing energy and momentum. Since no excess over the SM expectation has been observed, exclusion limits for the masses of the selectron and squark in the context of the MSSM have been derived. Depending on the specific choices made to constrain the MSSM parameter space, the excluded region extends to around 60 – 70 GeV for squark masses and around 40 GeV for the LSP mass.

LEP [47–51]: At LEP various searches for supersymmetry have been performed targeting different species of supersymmetric particles. Scalar leptons and quarks are mainly pair-produced in the s-channel via γ or Z exchange. In case of selectrons also the t-channel exchange of neutralinos yields important contributions. The typical energy scale at LEP results in decays into the respective SM partner particles – except for the scalar top – and the lightest neutralino while also cascade decays are possible. Typical final states contain missing energy and a pair of acoplanar leptons or jets where the direction of one of the leptons/jets is not in the plane defined by the direction of the second lepton/jet and the beam direction. Similarly, neutrinos and charginos are expected to be pair-produced via Z and Z/γ s-channel exchange or t-channel selectron or sneutrino exchange, respectively. Typically, the lightest neutralino and fermion pairs are the final decay products of such events. Hence, the final state in case of chargino production is characterized by missing energy accompanied by four jets, two jets and one lepton or only leptons depending on the specific decay mode of the chargino while the most important signature for neutralino production are acoplanar pairs of jets or leptons coming along with large missing momentum. However, the exact decay topologies strongly depend on the particular mass spectrum of the supersymmetric particles so that the above mentioned topologies could also be accompanied by photons or manifest in multiple jets or leptons in cascade decays. Interpretations of combined results from all four experiments in the mSUGRA model lead to exclusion limits showing that $m_{1/2}$ has to be greater than $\approx 100 – 200$ GeV over a wide range of m_0 up to the TeV-region for specific fixed other parameters. The lower limit on the LSP mass is found to be around 50 GeV nearly independent of other model constraints.

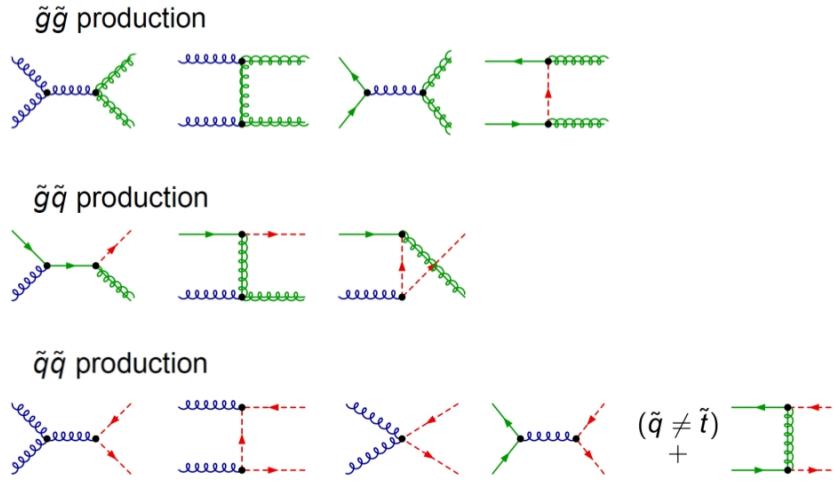


Figure 2.4.: Example diagrams for the production of supersymmetric particles in hadron collisions at parton level.

Tevatron [52–54]: The Tevatron accelerator was the best place to search for supersymmetry before the actual start of the LHC, as the center of mass energy exceeded that of HERA and LEP by several orders of magnitude. Since sparticles are expected to be heavy, thus a rich program of supersymmetric searches was enabled covering various final states of different lepton, photon or jet multiplicities. Of special interest is the search for colored sparticles like squarks and gluinos, as Tevatron is a hadron collider. The expected decay topologies are very similar to those at the LHC and thus discussed below. Results from susy searches at the Tevatron have been interpreted in the context of mSUGRA and extended the LEP results in the parameter region of $m_0 = 70 - 300 \text{ GeV}$ and $m_{1/2} = 125 - 165 \text{ GeV}$. This allows to exclude gluinos below around $280 - 300 \text{ GeV}$ for all squark masses and squarks below 380 GeV independent of the gluino mass.

The absence of SUSY-like signals at any collider experiment performed previous to the start of the LHC and exclusion limits at the order of a few hundred GeV on sparticle masses activated an abundant amount of searches for supersymmetry at the LHC as it opened already with a centre of mass energy of 7 TeV an even larger phase space than the Tevatron before. At the LHC, a variety of searches targeting the various different production and decay modes of supersymmetric particles are performed. In general, they can be classified into searches for electroweakinos, third generation sfermions and searches for squarks and gluinos. Of particular interest are searches for coloured particles, as the LHC is as well as Tevatron a hadron collider.

At leading order sparticles in R-parity conserving models are predominantly produced in processes like [55–59]

$$pp \rightarrow \tilde{g}\tilde{g}, \tilde{g}\tilde{q}, \tilde{q}\tilde{q}. \quad (2.1)$$

Some example diagrams for the production modes of such processes at parton level are shown in Fig. 2.4. Typically, squarks are assumed to be mass degenerate and refer to the

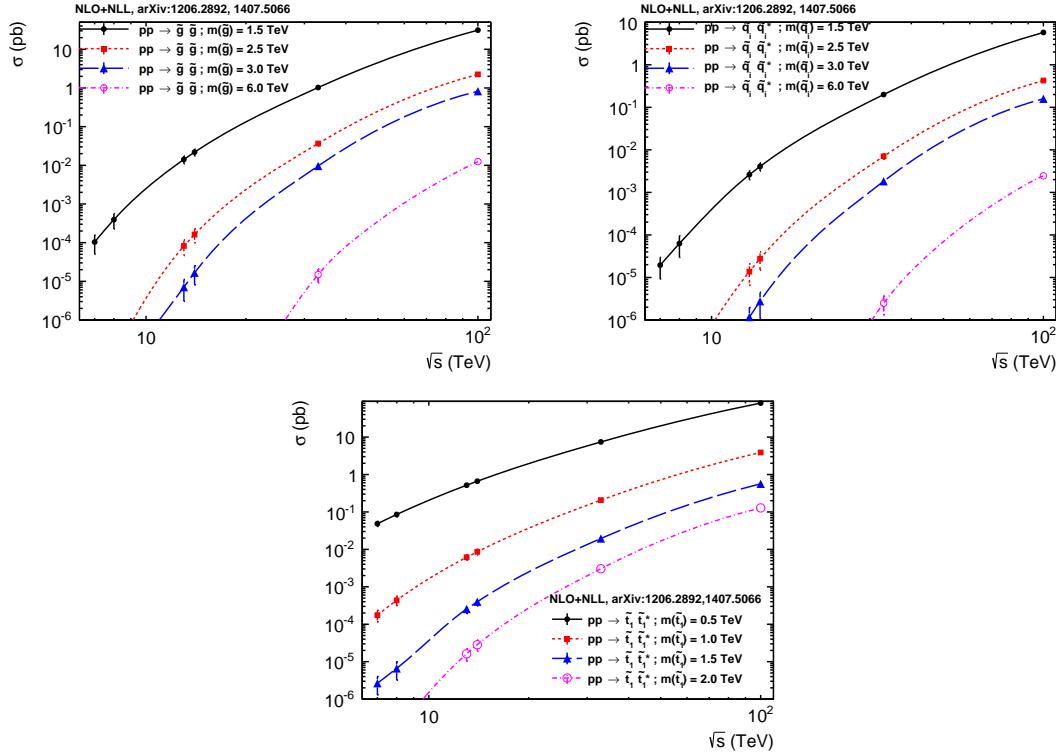


Figure 2.5.: SUSY production cross sections of processes $pp \rightarrow \tilde{g}$ (top left), $pp \rightarrow \tilde{q}$ (top right) and $pp \rightarrow \tilde{t}$ (bottom) displayed for different sparticle masses shown as function of the centre of mass energy [60, 61].

partners of the light-flavour (u, d, s, c) quarks. Supersymmetric partners of the bottom and top quark are considered separately due to a potentially large mixing affecting the mass splitting.

Most recent SUSY cross section calculations consider higher order corrections caused e. g. by quark radiation or gluon loops typically up to next-to-leading order (NLO). Production cross sections for the processes $pp \rightarrow \tilde{g}$, $pp \rightarrow \tilde{q}$ and $pp \rightarrow \tilde{t}$ are illustrated in Fig. 2.5 for different sparticle masses as function of the centre of mass energy. Typically, the relative size of the different channels depends on the respective squark and gluino masses as well as the energy of the collider. While for small masses of SUSY particles or large collider energies the cross sections of gluinos are dominant, associated squark-gluino production and squark-pair production are favoured in case of large SUSY masses and low collider energies. For instance for a gluino with a mass of 1.5 TeV, the production cross section is expected to be at the order of 10^{-4} pb at $\sqrt{s} = 7$ TeV.

In addition to various different production channels, also the decay of supersymmetric particles offers a rich variety of different modes. Depending on the specific mass hierarchy different decay topologies are expected. In Fig. 2.6 some example diagrams for possible decay modes of squarks and gluinos are illustrated. While squarks are expected to decay preferably into a quark and the LSP, gluinos predominantly decay into a quark pair and the LSP. In case of gluino decays, the quark pair can also be for instance a pair of top quarks representing the case of gluino-mediated stop production. In all such cases a mul-

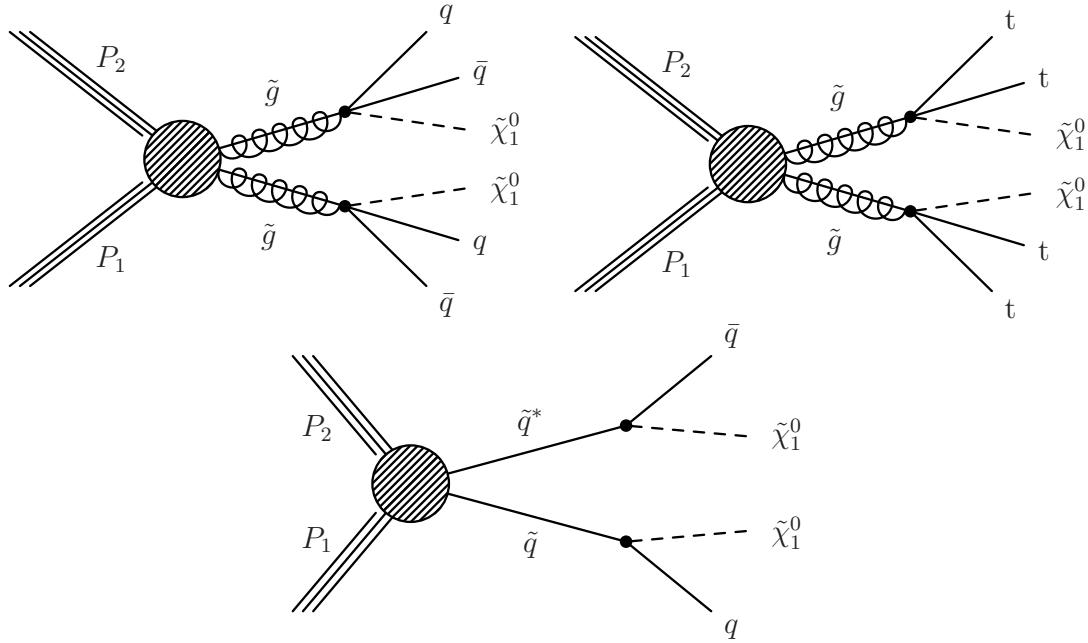


Figure 2.6.: Example diagrams of different SUSY decay channels showing $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_1^0$ (top left), $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ (top right) and $\tilde{q} \rightarrow q\tilde{\chi}_1^0$ (bottom).

tijet final state accompanied by missing transverse momentum with no isolated leptons is expected as experimental signature. These processes and final states are also the main subject of this thesis.

However, final states containing multiple jets accompanied by large values of missing transverse energy do not only arise from SUSY events, but are also realised for several SM processes. For any new physics search such SM processes resulting in the same final state have to be considered as background. In case of multijet + \cancel{E}_T searches these are typically $Z + \text{jets} \rightarrow \nu\bar{\nu} + \text{jets}$ events in which large genuine \cancel{E}_T is caused by the neutrinos. This background is denoted *invisible Z background* in this thesis. Furthermore, events with intrinsic missing energy stem from $W + \text{jets}$ and $t\bar{t}$ events. The top-quark is the heaviest quark in the SM with a mass of 173.34 ± 0.76 GeV [62]. Since its lifetime is smaller than the hadronisation timescale, it decays rapidly and behaves as a bare quark without forming colour-neutral hadrons. The decay takes place via the weak interaction and as denoted in Sec. 2.1, the quark mixing is parametrized by the CKM-matrix. Since the corresponding matrix element $V_{tb} \approx 1$, the top quark decays thus almost exclusively into a W boson and a b-quark. Hence, the decay of the top-quark is experimentally characterized by the decay of the W boson. In general, the W boson can decay into a charged lepton and its corresponding neutrino or a pair of light quarks of the first two generations. Taking into account the three possible colour states for each quark pair, this results in nine different W decay modes. Consequently, two thirds of top quark decays result exclusively in hadrons. However, since the targeted final state is supposed to contain a significant amount of missing energy only the semi-leptonic top quark decays relevantly contribute as background events. Thus, $W + \text{jets}$ and $t\bar{t}$ events featuring a decay containing an electron or muon that is not reconstructed, not isolated or falling out of the detector acceptance have to be

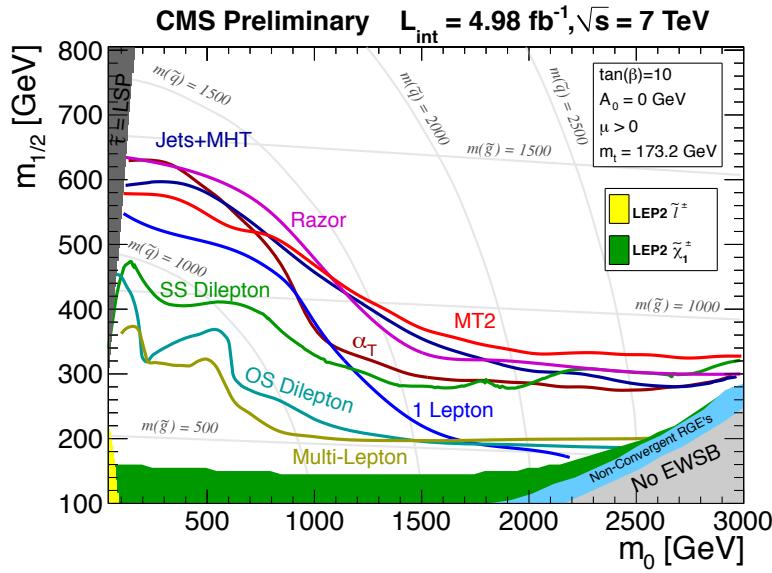


Figure 2.7.: Interpretation of searches for supersymmetry at the CMS experiment within the CMSSM. Shown are the 95% CL exclusion limits in the m_0 and $m_{1/2}$ plane for various searches performed using different final state topologies [65].

considered as possible background. This source of background events is referred to as *lost-lepton background*. Furthermore, also $W + \text{jets}$ and $t\bar{t}$ in which the lepton is a hadronically decaying τ lepton have to be accounted for. This background is known as *hadronic-tau background*. Another source of background events is arising from QCD multijet events. Although these possess no intrinsic missing energy, severely mismeasured jets can give rise to large amounts of \cancel{E}_T for instance because of instrumental effects or semi-leptonically decaying heavy-flavour quarks. This background process is denoted *QCD background* in the following. Contributions from other SM processes are found to be negligible [63, 64].

As soon as first collision data have been obtained at the LHC at a centre of mass energy of $\sqrt{s} = 7 \text{ TeV}$, searches for supersymmetry have been carried out based on various final states. Since however similar to previous collider experiments, no hints of new physics have been found, the results have been interpreted in various SUSY models. In Fig. 2.7 interpretations of CMS searches for supersymmetry based on various different hadronic and leptonic final states in the context of the CMSSM are summarized. For comparison, also the exclusion curves from the LEP experiments are illustrated in the $m_0/m_{1/2}$ -plane. These have been widely exceeded already with those early searches performed at the LHC. In general, the exclusions in the CMSSM $m_0/m_{1/2}$ -plane translate into constraints on the respective sparticle masses of around 1.3 TeV in case of $m_{\tilde{g}} = m_{\tilde{q}}$ obtained from searches based on all-hadronic states as described above.

In addition to such interpretations in full models like the CMSSM in which several experimental observables are predicted based on few theoretical parameters, results of SUSY searches are also often interpreted in the context of *simplified model spectra* (SMS) [66–69]. Since often many SUSY models predict a similar phenomenology, simplified models do not rely on detailed descriptions of specific model parameters but moreover characterize the

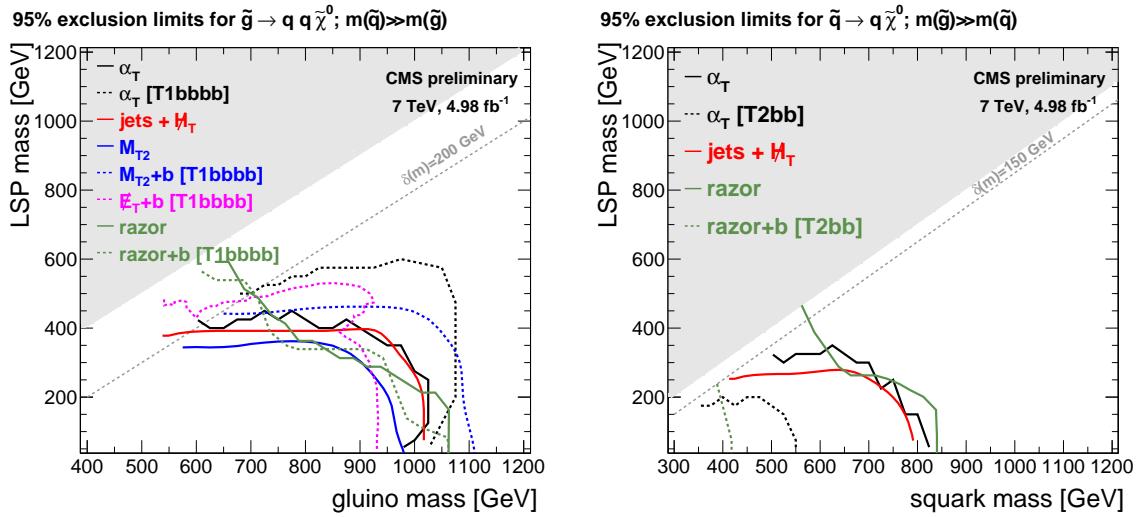


Figure 2.8.: Interpretation of searches for supersymmetry at the CMS experiment with simplified models in $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}^0$ (left) and $\tilde{q} \rightarrow q\tilde{\chi}^0$ (right) topologies. Shown are the 95% CL upper limits on the produced particle and LSP masses. The grey area represents the region where the respective decay mode is forbidden [69].

dominant features of SUSY events that are common for several SUSY and SUSY-like models. Thus, the characterization of basic properties allows a comparison of search results to any (more complex) models. Consequently, a simplified model is described by a set of particles and a certain sequence of their production and decay. Typical benchmark scenarios are for instance those illustrated in Fig. 2.6 in which the only free parameters are the two sparticle masses. The branching ratios of the pair-produced initial particles into the final state particles are assumed to be 100% and all other sparticles are decoupled by assuming their masses to be much larger. Interpretations of SUSY searches at the CMS experiment within the context of these SMS are illustrated in Fig. 2.8 for $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}^0$ and $\tilde{q} \rightarrow q\tilde{\chi}^0$. These illustrate the 95% confidence level upper limit on the product of the cross section and branching fraction as function of the sparticles masses. Hence, the values of cross sections times branching ratio can be compared to any theoretical prediction in order to determine whether the specific model is compatible with data. The exclusion curves shown in Fig. 2.8 indicate that in the context of these specific simplified models gluinos with masses up to around 1 TeV and light-flavour squarks around 800 GeV are excluded in case of LSP masses up to around 100 GeV.

Although the SUSY parameter space has been investigated already intensively with the LHC data obtained at $\sqrt{s} = 7$ TeV and exclusion limits on sparticle masses enter the TeV range already, searches for supersymmetry stay a very vital field within the CMS experiment also for $\sqrt{s} = 8$ TeV. As seen already in Fig. 2.5, production cross sections are expected to largely increase for increasing centre of mass energies. In particular, gluino and light-flavour squark cross sections profit a lot from the collider energy increase. Thus, in particular searches for those sparticles based on final states containing several hard jets

and high values of missing transverse momentum are of major interest for analyses of the LHC $\sqrt{s} = 8 \text{ TeV}$ data.

3. Experimental Setup

In order to probe the standard model or search for hints of new physics beyond the SM, experiments in particle physics often make use of powerful particle accelerators. In these machines, particles are collided to examine the constituents of matter and interactions between them. The analyses presented in this thesis are performed in the context of the CMS experiment located at the Large Hadron Collider at CERN near Geneva, which is the most powerful accelerator to date.

The first section of this chapter provides an introduction to the LHC, which is followed by an overview of the detector system of the CMS experiment. Afterwards, the different periods of collision data taking at the LHC are discussed. This chapter concludes with an introduction to the generation of simulated events which are needed for the analysis of real data events.

3.1. The Large Hadron Collider

The Large Hadron Collider [70, 71] is a ring accelerator designed to provide particle collisions of hadrons. It is built in the tunnel of the former LEP [72] collider 45–170 m below the surface and has a circumference of 26.7 km. The LHC is a particle-particle collider and composed of two rings with counter-rotating beams. It can be operated in different modes with either proton or heavy ion (e.g. lead) beams.¹

In each beam, protons are grouped together in bunches and accelerated in two evacuated beam pipes using superconducting radio-frequency cavities. With a nominal bunch spacing of 25 ns, the collision frequency is 40 MHz. Each of the 2808 individual bunches per beam contains 1.15×10^{11} protons, at design conditions. In order to bend the beams around the LHC ring, superconducting dipole magnets are used at a temperature of 1.9 K. They provide a magnetic field of up to 8.33 T while additional quadrupole and sextupole magnets are utilized to squeeze and focus the beams.

Before the protons are injected into the LHC, they are already pre-accelerated in various smaller accelerators, which are namely: Linac2, the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). With this injector chain, a beam energy of 450 GeV is achieved. An overview of the accelerator complex at CERN is given in Fig. 3.1.

The main goal of the LHC is to provide proton-proton (pp) collisions to the experiments with centre of mass energies up to 14 TeV in order to explore physics processes at novel energy regimes. The expected number of events N for a certain process is given by the product of the cross section σ of that process and the integral $L = \int \mathcal{L} dt$ of the instantaneous luminosity \mathcal{L} over time such that

$$N = \sigma \cdot L. \quad (3.1)$$

¹ All studies presented in this thesis are based on proton-proton collisions. Thus, the operation with heavy ions is not discussed.

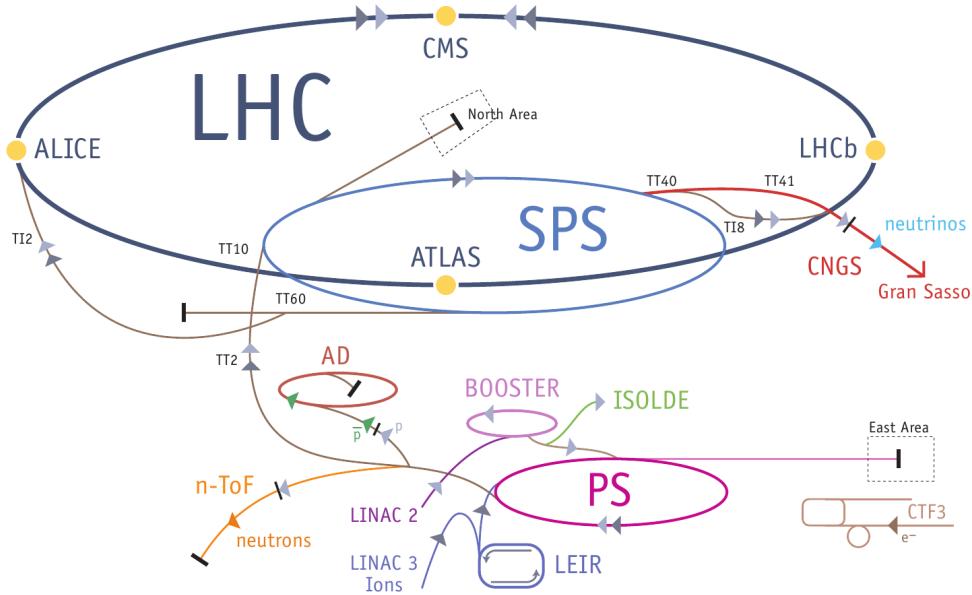


Figure 3.1.: Illustration of the CERN accelerator complex including the injector chain of the LHC ring [73].

The luminosity is a machine parameter and can be expressed for beams with Gaussian-shaped profiles as

$$\mathcal{L} = \frac{fn_1n_2}{4\pi\sigma_x\sigma_y} \cdot F \quad (3.2)$$

with the collision frequency f , the number of particles n_1 and n_2 contained in the two colliding bunches and the transverse beam sizes σ_x (σ_y) in the horizontal (vertical) direction. In order to take the inclination of the two beams into account, the geometrical correction factor F is introduced. With design conditions, the nominal peak luminosity of the LHC is $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. Since the total inelastic proton-proton cross-section at a centre of mass energy of 14 TeV is around 100 mb, as indicated in Fig. 3.2, the expected event rate is approximately 10^9 events per second.

The four main experiments are located at four of eight locations along the LHC ring where the beams cross and can be brought to collision. The two high luminosity experiments ATLAS [75] and CMS [76, 77] are designed for multiple purposes like precision measurements of SM quantities, search for the standard model Higgs Boson and searches for signals indicating new physics. The LHCb detector [78], however, is a specialised experiment and focuses on the measurement of CP violation in the interactions of hadrons containing b quarks. The only experiment especially designed for the analysis of heavy ion collisions is the ALICE [79] detector with the main emphasis on the physics of strongly interacting matter at extreme energy densities, like quark-gluon plasma.

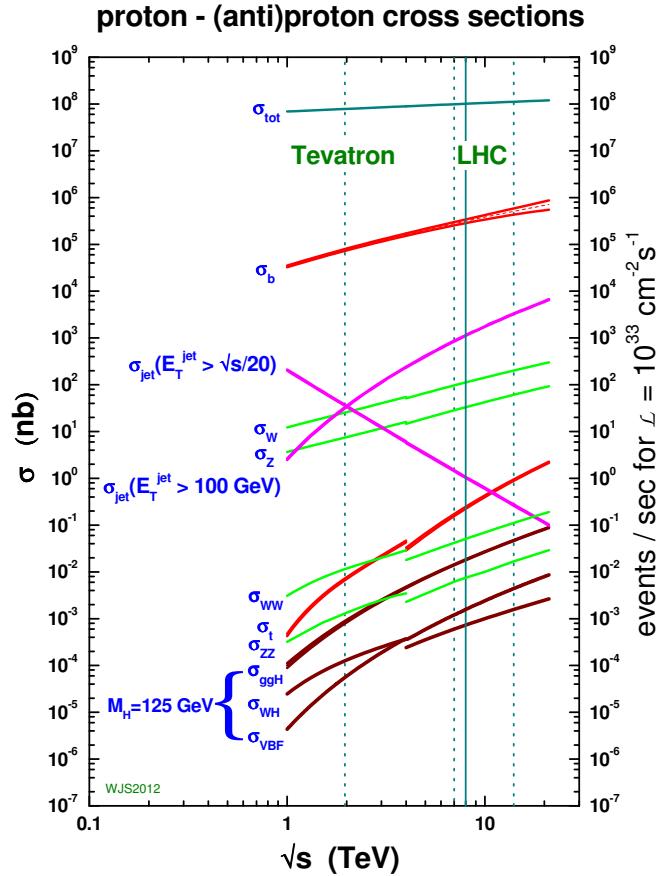


Figure 3.2.: Summary of cross sections for various standard model processes in proton-antiproton ($\sqrt{s} < 4$ TeV) and proton-proton ($\sqrt{s} > 4$ TeV) collisions as a function of the centre of mass energy. The right axis displays the corresponding event rate at a luminosity of $10^{33} \text{ cm}^{-2}\text{s}^{-1}$ [74].

3.2. The CMS Experiment

The CMS detector is one of the two general-purpose experiments at the LHC designed to address several physics questions. In addition to tests of the SM at the TeV scale, studies of the nature of electroweak symmetry breaking which might show up in the presence of a Higgs boson and searches for so far unknown particles pointing to, e.g. new symmetries in nature, are the primary targets of these experiments. These ambitious goals can only be achieved by fully exploiting the provided collision energy and luminosity with a suitable detector concept.

The CMS detector with its typical cylindrical design of different sub-detector components around the beam line is designed to meet these requirements. A sketch of the CMS detector and the different sub-detectors is shown in Fig. 3.3. As a typical high-energy particle experiment, the CMS detector makes mainly use of tracking detectors and calorimeters to

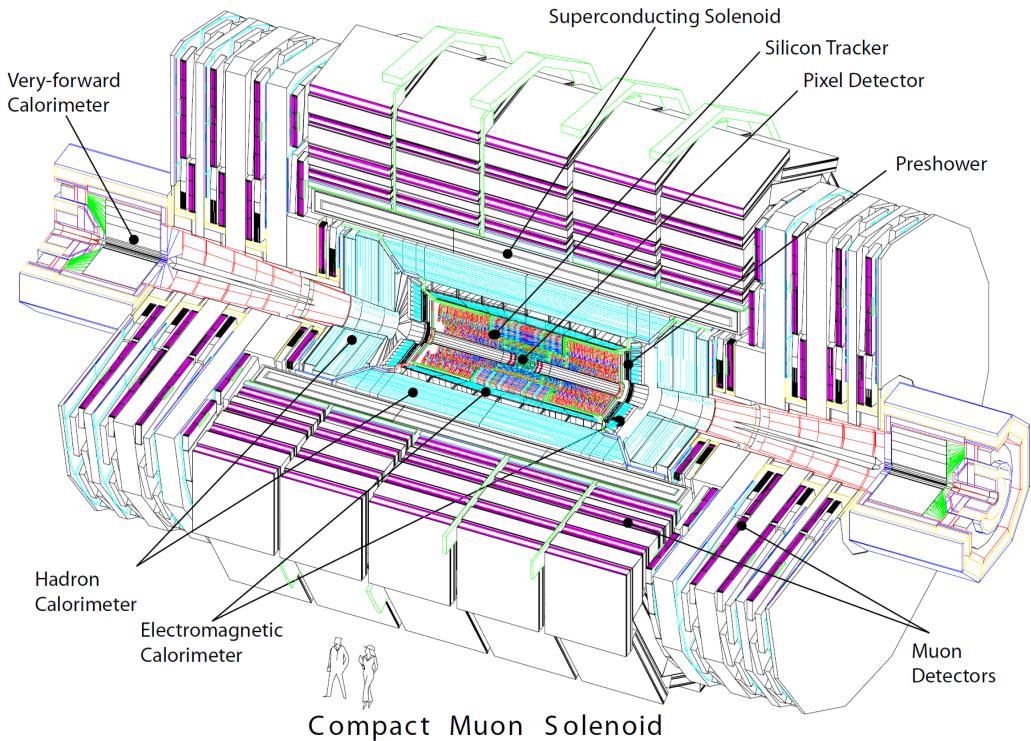


Figure 3.3.: A perspective view of the CMS detector [76].

measure momenta of particles, energy depositions and flight directions in order to identify the objects emerging from the particle collisions. The CMS detector with a length of 21.6 m and a diameter of 14.6 m exhibits a total weight of 12 500 t.

The following sections comprise a description of the CMS detector and individual sub-detectors. A detailed discussion of the detector design can be found in [76, 77]. After the introduction of the CMS coordinate conventions in Sec. 3.2.1, the superconducting magnet is discussed (Sec. 3.2.2). Afterwards, the tracking system (Sec. 3.2.3), the calorimeters (Sec. 3.2.4 and 3.2.5) and the muon system (Sec. 3.2.6) are described moving from the inside of the detector outwards. The detector description ends with a short review of the trigger system in Sec. 3.2.7.

3.2.1. Coordinate Conventions and Kinematic Variables

In order to describe the particle collisions, the CMS experiment makes use of a right-handed coordinate system with its origin at the centre of the detector at the nominal interaction point. While the z -axis is defined along the direction of the anti-clockwise beam, the x -axis points to the center of the LHC ring and the y -axis vertically upwards. In this xy -plane the azimuthal angle ϕ is measured where $\phi = 0$ coincides with the x -axis. The polar angle θ is defined with respect to the positive z -axis. A quantity closely related

to the polar angle is the pseudorapidity η defined as

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (3.3)$$

which is widely used in experimental particle physics. A pseudorapidity $\eta = 0$ corresponds to the direction perpendicular to the beam while $|\eta| \rightarrow \infty$ points along the beams. Based on the pseudorapidity, the distance between two objects ΔR , which is invariant under Lorentz boosts in z -direction, can be written as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (3.4)$$

At the LHC, the initial conditions of the primary collisions are not known, as the specific energy fraction of the proton which each parton carries can not be identified. Thus, conservation of the total momentum can not be utilized directly to describe the momentum balance in the final state. However, it is known that the initial particles have no significant momentum orthogonal to the beam axis which is referred to as *transverse momentum*. Consequently, the transverse momentum is defined as

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (3.5)$$

and momentum conservation in the transverse plane is then used to describe the final state conditions. Hence, the vector momentum imbalance in the transverse plane (\vec{E}_T) is determined as the negative vector sum of the momenta of all particles. The absolute value of the vector momentum imbalance E_T is typically termed *missing transverse momentum* or *missing transverse energy* and often exploited to describe undetected particles.

3.2.2. Superconducting Magnet

The CMS experiment makes use of a large superconducting solenoid magnet which is a crucial component of the whole detector design. This magnet provides a uniform magnetic field in z direction of up to 4 T and allows to precisely determine the momenta and charge of charged particles from the curvature in the (x, y) -plane of the bended tracks, since it surrounds the tracking and calorimeter systems.

With a length of 12.5 m and a diameter of the free bore of 6.3 m, the total cold mass reaches 220 t. The magnet is made up of a niobium-titanium coil which is winded in four layers. This configuration allows a storage of 2.6 GJ energy at full current.

In addition, the muon system located outside of the solenoid is interleaved with a 10 000 t heavy-weight iron yoke which is responsible for the return of the magnetic flux and closes the magnetic field lines. This offers the opportunity to also measure muon momenta precisely.

3.2.3. Inner Tracking System

The tracking system of the CMS experiment is the innermost part of the detector and installed directly around the interaction point completely contained in the bore of the magnet system. Its purpose is to precisely measure the trajectories of charged particles arising from the collisions. Furthermore, it is used to identify secondary vertices. Due

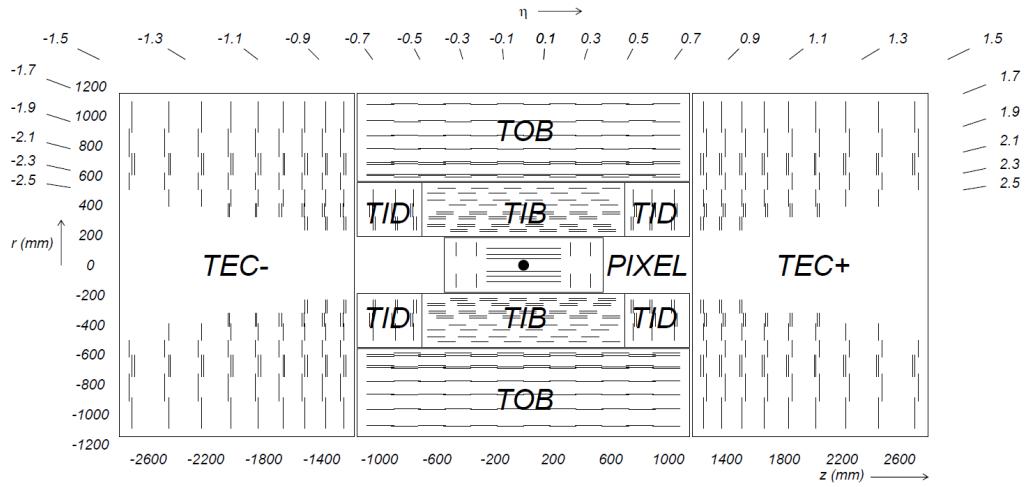


Figure 3.4.: Sketch of the CMS tracking system in a $r\text{-}z$ -view. Each tracker module is represented by one line [76].

to the location close to the interaction point, the tracking system has to cope with a high particle flux. Hence, high requirements on response time and granularity are set to properly identify particle tracks.

In order to fulfill these tasks, the CMS experiment makes use of a tracker design based on silicon detectors. The innermost part is made of silicon pixel detectors. These are surrounded by silicon strip modules. In total, they add up to an active area of 200 m^2 with a length of 5.8 m and a diameter of 2.5 m, covering the detector region up to $|\eta| = 2.5$. A schematic overview of the whole tracking system is shown in Fig. 3.4.

Pixel Detector: The pixel detector consists of three barrel layers extending radially from 4.4 cm to 10.2 cm and two endcap disks on each side. In total, there are 1440 pixel modules installed. The size of one pixel cell is $100 \times 150 \mu\text{m}^2$ providing similar track resolution quality in $r\text{-}\phi$ and z direction. This configuration provides for almost the whole range up to $|\eta| = 2.5$ three precise hits which is especially important for the reconstruction of secondary vertices.

Silicon Strip Tracker: The silicon strip detector which extends to a radius of 1.1 m surrounds the pixel tracker. The more than 15 000 individual strip detector modules are arranged in an inner and an outer detector part. The inner part of the strip tracker is build by the four *Tracker Inner Barrel* (TIB) layers which are accompanied by the three *Tracker Inner Disks* (TID) at the end sides. This inner part provides up to four track measurements in the $r\text{-}\phi$ plane. The TIB/TID system lies within the *Tracker Outer Barrel* (TOB) consisting of another six barrel layers while it is complemented by the *Tracker End Caps* (TEC) which add another nine disks at each side of the tracking system. This layout provides at least around nine hits within the silicon strip system.

The tracking system with the design described above provides a precise impact parameter resolution and tracking efficiency [80]. The impact parameter resolution is of the order of $< 35 \mu\text{m}$ in the plane perpendicular to the beam (for particles with $p_T > 10 \text{ GeV}$)

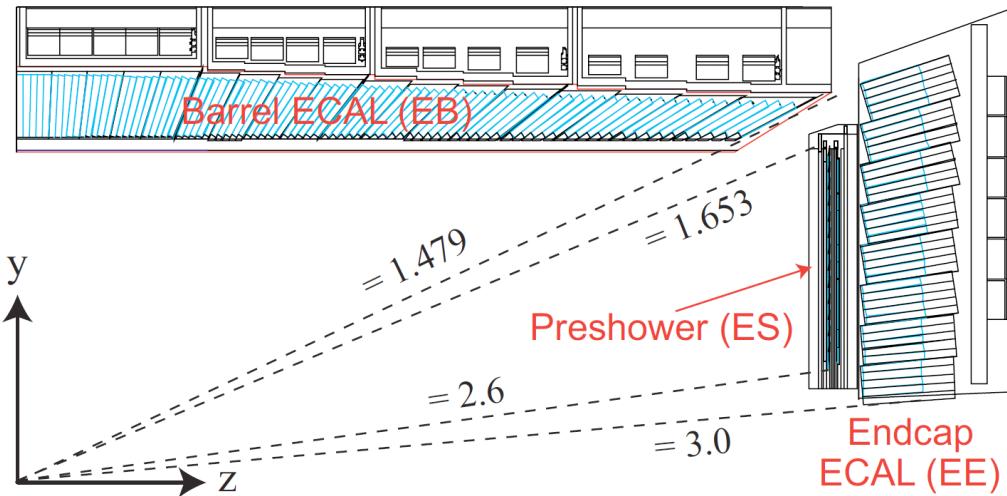


Figure 3.5.: View of one quarter section of the CMS electromagnetic calorimeter in a yz -view [77].

and reaches $75\ \mu\text{m}$ in the longitudinal direction. Furthermore, the track reconstruction efficiency of high energetic electrons is above 90%, that of charged hadrons up to 95% (for $p_T > 10\ \text{GeV}$) and that for muons even better than 98% in the whole covered region up to $|\eta| = 2.5$. This is achieved already for muons with very low transverse momenta around $1\ \text{GeV}$. Altogether, the relative transverse momentum resolution reaches a level of 1–2% for high momentum tracks ($\approx 100\ \text{GeV}$) in the barrel for $|\eta| < 1.6$.

3.2.4. Electromagnetic Calorimeter

The CMS experiment makes use of a homogeneous electromagnetic calorimeter (ECAL), in order to precisely measure the energy deposits of electrons and photons. It is installed around the inner tracking system covering a range up to $|\eta| = 3.0$ and consists of lead tungstate (PbWO_4) crystals. These have been chosen as they provide a high density, short radiation length and a small Molière radius and hence allow to build a compact calorimeter with a fine granularity. As 80% of the scintillation light is emitted within 25 ns, this allows ECAL operation at the bunch crossing rate of the LHC machine. In order to collect the radiated light, photodiodes are glued to the back of each crystal. A typical quantity to describe the extent of high-energy electromagnetic cascades is the radiation length X_0 that describes the characteristic amount of matter traversed by electrons and photons loosing energy by interacting with the material.

An overview of the ECAL layout is shown in Fig. 3.5. The individual sub-components are as follows:

Barrel ECAL (EB): The barrel detector of the ECAL covers the pseudorapidity region up to $|\eta| = 1.479$. Within a radius of about 1.3 m, a total number of 61 200 crystals are installed. Each of them has a length of 230 mm corresponding to a radiation length of $25.8\ X_0$. The crystal cross section in (η, ϕ) is $(0.0174, 0.0174)$. Avalanche photodiodes are used to detect the emitted scintillation light.

Endcap ECAL (EE): The EB is complemented on each side by an endcap which consists of two D-shaped halves. The ECAL endcaps extend from $|\eta| = 1.479$ to $|\eta| = 3.0$. In total, they contain another 14 648 crystals. They have an individual length of 220 mm, which corresponds to 24.7 X_0 . For the collection of scintillation light, vacuum phototriodes are used in the endcaps.

Preshower (ES): In front of the endcap crystals, a preshower detector is placed. It covers the pseudorapidity range of $|\eta| = 1.653\text{--}2.6$. The main purpose is to identify photons emerging from the decay of neutral pions. It is a two-layer sampling calorimeter with lead as absorber material and silicon strip sensors measuring the deposited energy. The total thickness of the preshower is 20 cm (3 X_0).

In order to achieve a stable and uniform energy measurement across the whole ECAL, an accurate calibration is of crucial importance. In addition to the calibration of the absolute energy scale especially channel-to-channel effects referred to as *intercalibration* have to be accounted for. This is mainly done based on physics events. Changes in the transparency of the ECAL crystals during operation caused by irradiation are monitored by a dedicated laser system based on the injection of reference laser pulses into the crystals. The typical relative energy resolution for electrons with a transverse momentum of 120 GeV with this calorimeter configuration is of the order of 0.5%.

3.2.5. Hadron Calorimeter

The calorimetry of the CMS experiment is completed by the hadron calorimeter (HCAL). It is designed to provide an accurate energy measurement of hadron jets and indirectly also of invisible particles, e.g. neutrinos, by the determination of missing transverse energy. In order to measure the missing transverse energy, it is important that the calorimeter is hermetic, meaning that it provides a large geometric coverage to measure all particles emerging from an interaction. Thus, the HCAL is built such that a pseudorapidity range up to $|\eta| = 5.2$ is enclosed.

The hadron calorimeter completely surrounds the inner tracking system and the electromagnetic barrel calorimeter while it is mainly contained within the magnet system. Hence, its radial dimensions are limited on the one hand by the outer circumference of the barrel ECAL and on the other hand by the inner border of the magnet coil. An additional calorimeter component is installed outside the solenoid in the barrel part to ensure that the complete hadronic showers are contained in the HCAL.

An overview of the layout of the CMS hadron calorimeter is shown in Fig. 3.6. It is a typical sampling calorimeter with alternating layers of absorber material and active scintillator layers. The individual sub-components are:

Hadron barrel (HB): The barrel part of the CMS hadron calorimeter covers the pseudorapidity range up to $|\eta| = 1.3$ and is composed of two half barrels each containing 36 identical azimuthal wedges. These wedges hold the absorber plates which are flat brass plates arranged parallel to the axis of the beam. For reasons of stability, the first and last layers are made of stainless steel. The total thickness of the absorber material ranges from 5.82 interaction lengths (λ_I) at $|\eta| = 0.0$ to 10.6 λ_I at $|\eta| = 1.3$. The 17 active plastic scintillator layers alternate with the absorber plates and have a segmentation in $(\Delta\eta, \Delta\phi)$ of (0.087, 0.087).

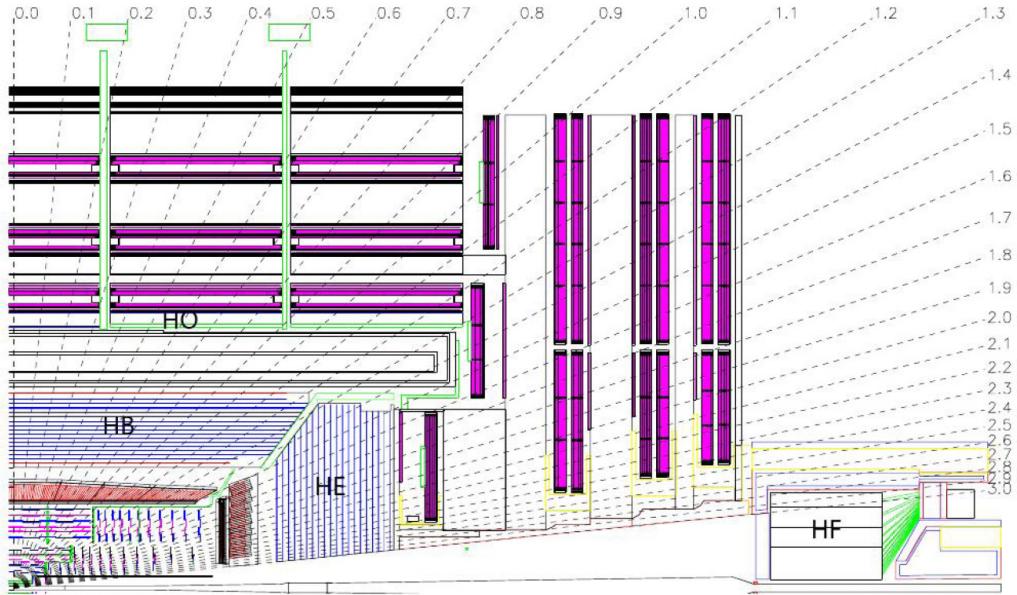


Figure 3.6.: Longitudinal view of one quarter of the CMS detector showing the location of the individual HCAL sub-detector parts [76].

Each half barrel is divided into 16 η -regions for which the individual tiles are optically linked together using wavelength shifting fibres and thus form so-called *HCAL towers*. The read-out of each longitudinal tower is carried out using pixelated hybrid photodiodes.

Hadron outer (HO): The calorimeters in the central pseudorapidity region do not provide a sufficient depth in order to fully contain all hadronic showers. Therefore, the HB is complemented by the outer hadron barrel part which is placed outside the solenoid covering $|\eta| \leq 1.26$. The HO makes use of the solenoid as additional absorber material and adds another one or even two layers in the most central part of scintillators to the barrel region. Thus, the total depth is extended to $11.8\lambda_I$ at $\eta = 0$.

Hadron endcap (HE): The hadron barrel calorimeter is supplemented by the hadron endcap. It is mounted on the endcap iron yoke and covers the pseudorapidity region of $1.3 \leq |\eta| \leq 3.0$ using 18 scintillator layers inserted into brass absorber plates. The granularity of the endcap calorimeter is the same as for the barrel up to $|\eta| = 1.6$ and gets coarser for larger pseudorapidities with $(\Delta\eta, \Delta\phi) \approx (0.17, 0.17)$.

Hadron forward (HF): The forward hadron calorimeter extends the pseudorapidity coverage from $|\eta| = 2.9$ (slightly overlapping with the HE) up to $|\eta| = 5.2$. It is located 11.2 m from the nominal interaction point and has to be radiation hard to cope with the vast particle flux. Thus, the HF is made of steel absorber plates with radiation hard quartz fibres integrated as active material. These fibres are arranged parallel to the beam line and form towers with a size in $(\Delta\eta, \Delta\phi)$ of $\approx (0.175, 0.175)$. The signal is detected as Cerenkov light originating from the quartz fibres.

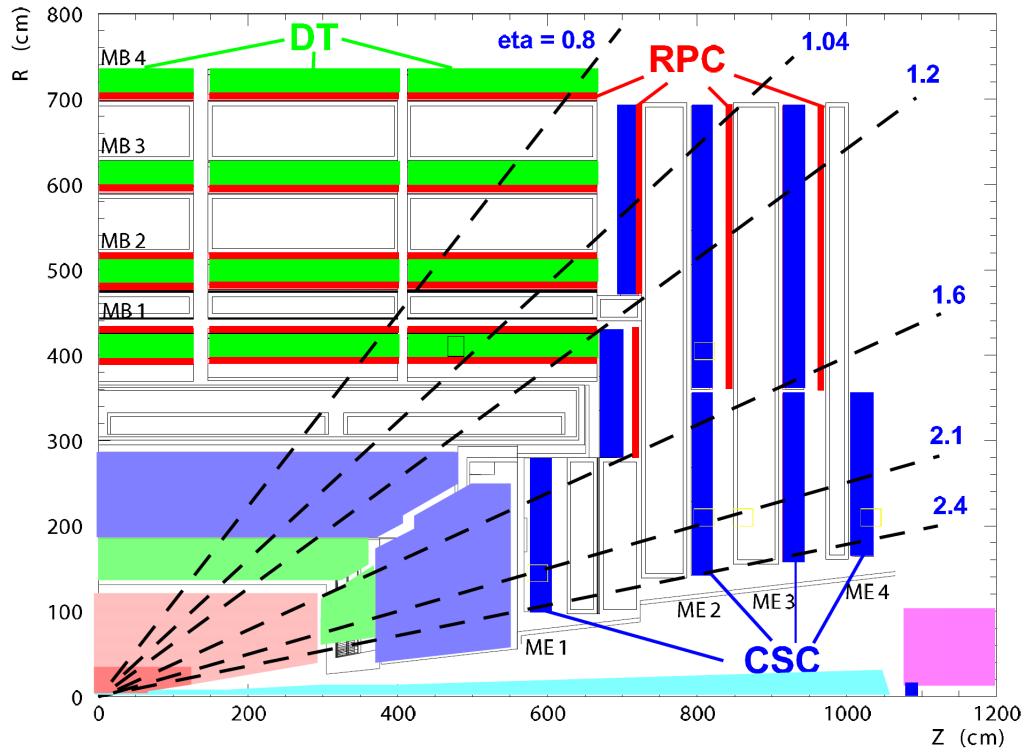


Figure 3.7.: View of one quarter section of the muon system in the $r\zeta$ -plane [77].

Similar to the ECAL, also the performance of the HCAL has to be well monitored during operation. Thus, an initial calibration using a radioactive source is combined with test beam data to derive the absolute energy scale. A continuous update of this calibrations is performed using isolated energetic particles, e. g. from decays of W or Z bosons.

3.2.6. Muon System

The outermost part of the CMS detector, as seen from the interaction point, is made up of the muon system. This component of the detector is assembled in the return yoke of the CMS magnet and consists of a central barrel cylinder complemented by endcap disks in the forward region. This results in a coverage of the pseudorapidity range up to $|\eta| = 2.4$. In the barrel part, four layers of detectors are installed alternating with the iron yoke while the detectors in the endcap are mounted on four discs perpendicular to the beam. In total, about $25\,000\text{ m}^2$ detection planes are employed. The layout of the muon system is illustrated in Fig. 3.7.

To achieve a good muon momentum resolution and consider the different radiation conditions and variations in the homogeneity of the magnetic field, three different types of gaseous detectors are used depending on the pseudorapidity. The different types of tracking chambers are:

Drift tube (DT) chambers: In the barrel region for $|\eta| < 1.2$, the background due to neutrons is low and residual effects from the magnetic field likewise. Here, the muon system is equipped with drift tube chambers. In the barrel, the muon coordinates in

the $r\phi$ -plane are measured in four stations while only the first three layers provide also a measurement of the z -direction. The drift length is restricted to a maximum of 21 mm resulting in a negligible occupancy while keeping the number of channels at an acceptable level. Furthermore, a technology based on tubes is chosen avoiding the issue of possibly broken wires. The resolution in $r\phi$ is designed to reach a precision of 100 μm .

Cathode strip chambers (CSC): The endcap regions are equipped with cathode strip and cover the pseudorapidity range $0.9 < |\eta| < 2.4$. These provide a fast response time and fine segmentation while they are resistant against radiation. Thus, they are well suited for the forward region where the muon and background rates are largely increased and the magnetic field is high and non-uniform. The CSCs, which are multiwire proportional chambers in which anode wires are interlaced with cathode panels, perform a precise position measurement in the $r\phi$ -plane with a spatial resolution of 75–150 μm .

Resistive plate chambers (RPC): Resistive plate chambers are used to complement the drift tube and cathode strip chambers in the range $|\eta| < 1.6$. These are gaseous parallel-plate detectors with a spatial resolution coarser than the DTs and CSCs. However, they provide a very fast response and good time resolution at high particle rates. Thus, they are able to very efficiently detect the bunch-crossing a muon track is associated to.

The global muon reconstruction efficiency is in general about 95–99% and only drops for some $|\eta|$ regions, e.g. in the transition region between the barrel and endcap part around $|\eta| = 1.2$. Since muons reaching the muon system are affected by multiple scattering, the resolution for muons with transverse momenta below $\approx 200 \text{ GeV}$ is in general better based on the inner tracking system than for the muon system alone. For highly energetic muons it is comparable. In general, the muon momentum resolution can be improved by combining the information from the inner tracker and the muon system due to an improved fault finding. This combined approach results in a relative muon momentum resolution for muons with high momenta around 1 TeV of about 5%.

3.2.7. Trigger System

The LHC operating at design conditions performs particle collisions with a bunch crossing rate of 40 MHz. This results in an enormous amount of event data which have to be processed and stored for later analyses. With an approximate event size of 1 MB, it is technically impossible to record all such events. However, as illustrated in Fig. 3.2, the event rate of interesting potentially new physics events is orders of magnitudes smaller than the total inelastic proton-proton cross section. This allows to already perform in the trigger process a suitable event preselection in order to reduce the amount of data to a storable size while still retaining the information of interest. Hence, the trigger system makes up the first step in the physics analysis process.

In order to achieve the necessary rate reduction, the CMS experiment uses a two-stage trigger system:

Level-1 (L1) trigger: The L1 trigger consists of custom-made fast programmable hardware. It makes use of data received from fast detector components which are the

calorimeters and the muon system at reduced granularity. For that purpose, the calorimeter is divided into so-called *trigger towers* which cover an area in (η, ϕ) of $(0.087, 0.087)$ up to $|\eta| = 1.74$ getting even coarser for higher $|\eta|$. At L1, the trigger decision is based on energy deposits in those trigger towers or certain hit patterns in the muon chambers forming trigger primitive objects which are electrons/photons, muons or jets and global quantities like sums of E_T or \cancel{E}_T . Events are accepted, if those trigger objects pass some predefined criteria like for instance certain p_T thresholds. The L1 trigger latency, i. e. the time between the actual bunch crossing and the delivery of the L1 trigger decision to the front-end electronics, is $3.2\text{ }\mu\text{s}$. During this period, the high resolution data is pipelined in readout buffers for further processing. The L1 trigger reduces the event rate to a maximum of 100 kHz .

High-Level trigger (HLT): Events that are accepted by the L1 trigger stage are transferred to the High-Level trigger for further processing. The HLT is a software system running on several thousand commercial processors. It has access to the full information from all sub-detectors and performs an event reconstruction similar to the later event reconstruction performed for recorded data. Thus, it allows a further rate reduction to the final output rate of a few hundred Hz of events that are finally stored for analyses. Since the HLT is software based, it allows to continuously adjust the used algorithms in order to adapt to changing conditions during operation.

In the two trigger stages, various different conditions can be tested in parallel which define the so-called *trigger paths*. For instance, there exist dedicated trigger paths to select events with single electrons or muons above a certain p_T threshold or paths to collect events with a specific amount of missing transverse energy. All of these individual trigger paths form together the *trigger menu* and are operated with a dedicated rate such that the total manageable output rate is not exceeded. Typically, the instantaneous luminosity changes gradually during operation which can result in changing output rates for individual trigger paths. In order to not exceed the allowed total trigger rate, the number of recorded events per trigger path can be adjusted accordingly by requiring that only each n^{th} triggered event is kept. The respective value of n is denoted *prescale factor*.

3.3. LHC Operation and Data Taking

The first operation of the LHC took place in September 2008. After a major cooling incident only a few days later requiring a longer technical stop, beams were circulated again in November 2009. The first collisions at a center of mass energy of 7 TeV finally happened end of March 2010 [82]. In the following running period in 2010, data corresponding to an integrated luminosity of 44.2 pb^{-1} were delivered to the experiments with a maximum peak instantaneous luminosity of $2.05 \times 10^{32}\text{ cm}^{-2}\text{ s}^{-1}$. These data allowed studies of the detector performance and made first searches for new physics possible. The next data taking period performed during 2011 at the same centre of mass energy even lead to much more pp collision data provided to the experiments. A total amount of 6.1 fb^{-1} was delivered reaching a peak instantaneous luminosity of $3.5 \times 10^{33}\text{ cm}^{-2}\text{ s}^{-1}$. Following another technical shutdown during winter, the centre of mass energy was finally increased to 8 TeV for the running period during 2012 and the peak luminosity reached values of up to $7.7 \times 10^{33}\text{ cm}^{-2}\text{ s}^{-1}$, which is already almost design conditions. In total, 23.3 fb^{-1} of

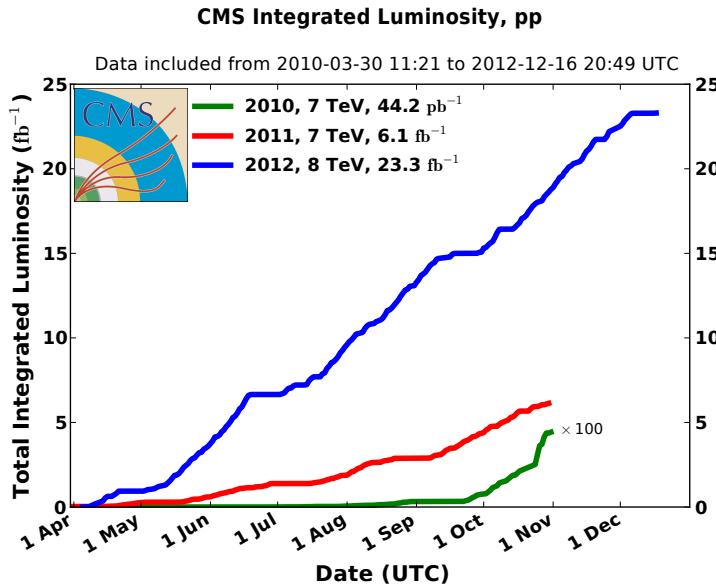


Figure 3.8.: Cumulative integrated luminosity versus day delivered to CMS during stable beams for pp collisions. Different data-taking periods are indicated as follows: green for 2010, red for 2011 and blue for 2012 [81].

integrated luminosity pp collisions were produced by the LHC operating at stable beam conditions. The evolution of the integrated luminosity versus days is illustrated in Fig. 3.8 for data-taking periods in 2010, 2011 and 2012.

During most of the operation in 2012, the LHC was circulating 1380 bunches per beam with a distance of 50 ns. The average bunch intensity, i. e. the number of protons per bunch, was varying from 1.6 to 1.7×10^{11} exceeding even the design value [81, 83]. This was resulting in multiple interactions per bunch crossing known as *pileup* (PU). In general, two sources of pileup are distinguished: *in-time* pileup (IT PU) and *out-of-time* pileup (OOT PU). While IT PU is caused by additional pp collisions occurring within the same bunch-crossing as the primary hard collision and leads to additional tracks in the tracking system and energy deposits in the calorimeters, OOT PU arises from pp collisions in previous and following bunch crossings and contributes further energy deposits in the calorimeters to the hard interaction due to the finite signal decay time in the calorimeters. An overview of the pileup profile in collision data taken in 2012 is given in Fig. 3.9. Here, the mean number of interactions per bunch crossing is illustrated with the mean of this distribution located at 21. The maximum number of mean pileup events even reached a value of ≈ 40 . Dedicated techniques in order to mitigate effects from pileup have been developed of which some are discussed in Chapter 4.

3.4. Event Simulation

An important tool in high energy physics is the use of simulation in order to acquire a good understanding of the behaviour of particle collisions and the collision products as they are observed in the detector. Thus, simulated events are often used as benchmarks

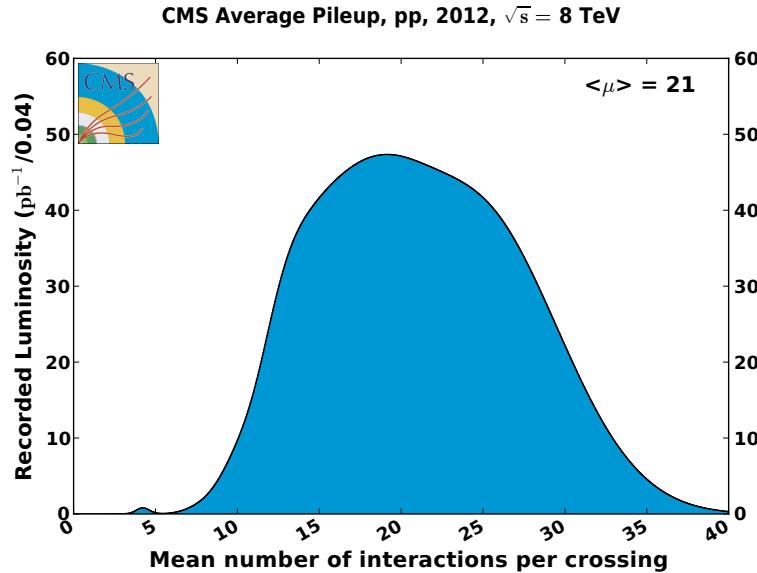


Figure 3.9.: Mean number of interactions per bunch crossing in pp collisions at $\sqrt{s} = 8$ TeV [81].

in the development of new detector concepts. Moreover, they are heavily exploited in the validation and interpretation of the results of actual collision experiments, likewise for the LHC. For instance, simulated events are used to derive expectations for certain physics properties or the detector performance. In particular, they allow to estimate event kinematics of new physics events that can subsequently be used to derive e.g. selection efficiencies.

This section provides a brief introduction to the principles of event simulation in hadron collisions and introduces some event generators including different approaches for the simulation of the detector. A broader overview of event simulation and respective generators for LHC physics can be found e.g. in [84, 85].

The simulation of high energy collisions is a quite challenging task as each collision involves typically several hundred particles with momenta ranging over some orders of magnitude. Furthermore, the collisions being subject to quantum chromodynamics are only calculable within approximation schemes which furthermore often exhibit divergences. Thus, event simulation is primarily utilizing numerical *Monte Carlo* (MC) techniques which rely on the repeated sampling of random numbers, cf. for instance [86]. For convenience, events obtained from simulation are denoted by the label 'MC' in this thesis.

Typically, the generation of an event follows several subsequent steps which are illustrated in Fig. 3.10 and described in the following:

Hard process: The first step in the simulation of collision events is the description of the nominal parton-parton interaction which is typically referred to as *hard process*. The proton itself is not a fundamental particle but exhibits an internal structure. The constituents of which a proton is made off are known as *partons*. Thus, two protons interact when there is a momentum transfer Q taking place between two individual

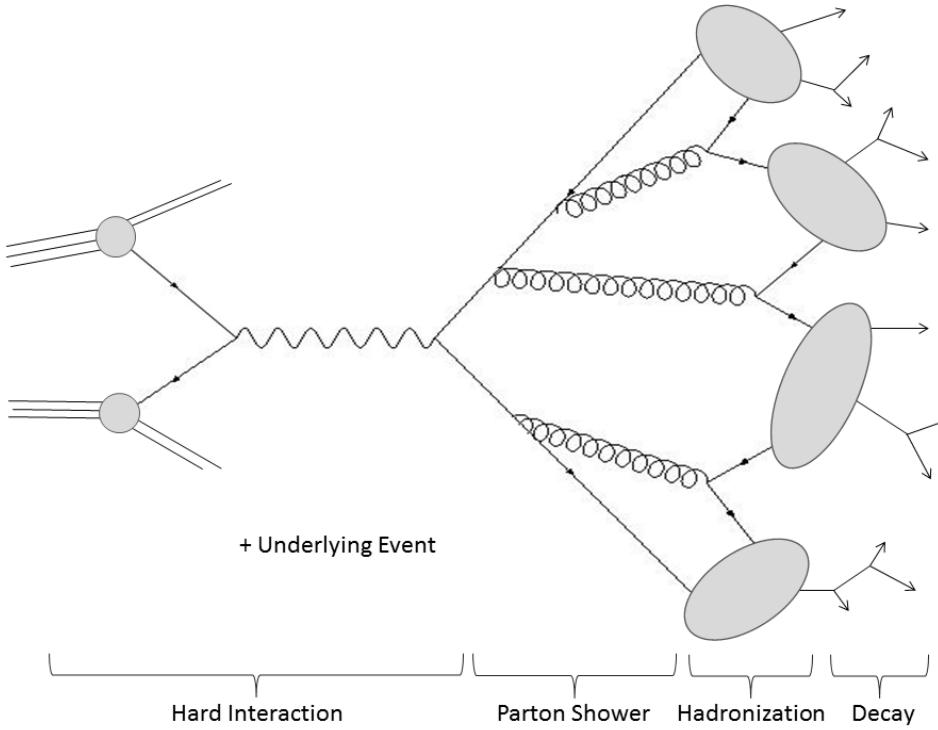


Figure 3.10.: Sketch of individual steps in the generation process of simulated events.
Adapted from [87].

partons. The probability for individual partons to take part in the hard interaction is parametrized by the *parton-distribution functions* (PDFs) which have been determined experimentally for instance in deep-inelastic scattering at the HERA collider [88]. The value of x denotes the fraction of the longitudinal momentum carried by an individual parton. Consequently, the initial state of a parton-parton collision is unknown. Nevertheless, the cross section of a specific process can be calculated following the *factorization theorem* [89, 90]. Here, the hard interaction is described via perturbation theory and low-energy processes are considered in the phenomenological and theoretical models of the respective PDFs. Thus, the cross section for a process $ab \rightarrow n$ is given according to

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int f_a(x_a, \mu_F) f_b(x_b, \mu_F) d\hat{\sigma}_{ab \rightarrow n}(\mu_F, \mu_R). \quad (3.6)$$

Here, $f(x, \mu_F)$ are the PDFs of the interacting protons which depend on the factorization scale μ_F while $\sigma_{ab \rightarrow n}$ indicates the parton-level cross section for the production of a final state n from partons a and b . This parton-level cross section depends on the final-state phase space, the factorization scale and the renormalisation scale μ_R as well as the corresponding matrix element. The factorization and renormalization scales are unphysical and have to be chosen for the generation process. Usually,

the process possesses a typical hard scale Q^2 so that the choice $\mu_F = \mu_R = Q^2$ is made. However, this choice is not fixed by first principles and has to be considered as uncertainty in the calculation.

Parton shower: After the production in the hard interaction, the outgoing partons start to form a *shower*, i. e. cascades of further parton emissions. Typically, this happens with descending amounts of momentum transfer from the high scales down to low scales around 1 GeV. This evolution is typically described by a probabilistic shower algorithm. In addition to a parton shower related to the outgoing partons which is referred to as *final-state radiation* (FSR), also the initial-state partons can radiate off other partons before the actual hard process takes place. This effect is known as *initial-state radiation* and can be described by similar principles as FSR.

Hadronization: The evolution of the parton shower continues until low scales in momentum transfer are reached and so the final state partons eventually start to form colour neutral hadrons. In the context of simulation the term *hadronization* describes the particular model which is used in order to specify the transition from the partonic state to the complete hadronic final state. Since the hadronization involves low-energetic processes, it can not be treated within perturbation theory. One distinguishes between *string models* (cf. for instance [91]) and *cluster models* (cf. for instance [92, 93]) in order to characterize the hadronization. While the first model describes directly the transition from the parton to the hadron based on the assumption of linear confinement, i. e. a linearly rising potential, the second model introduces an intermediate step of cluster objects with mass scales around a few GeV. Particles arising from the hadronization are denoted with *generator-level* particles and labelled 'gen' within this thesis.

Decay: After the hadronization process, a couple of unstable hadrons are present in the event whose decay into stable particles must be modelled. Stable in this context means that they do not decay further within the typical collider timescales ($c\tau > 10$ mm). Thus, the observable final-state hadrons result from a convolution of the hadronization process with the decay modelling. Concerning the decay process a choice has to be made regarding the hadrons to be included in the simulation and the respective decay modes which have to be considered. Mainly, these choices are based on experimental results in combination with theoretical assumptions. Typical differences occur for instance in the consideration of excited mesons or heavy baryon multiplets as well as the treatment of matrix elements and spin correlations during the decay.

Underlying event: In addition to hadrons emerging from the process connected directly to the hard interaction, further contributions to the event can be present which is referred to as *underlying event* (UE). The UE activity can be for instance caused by additional interactions occurring for incoming partons, which is known as *multiple parton interaction* (MPI), or by interactions arising from the proton remnants. In general, such effects contribute to the total amount of scattered energy and increase the number of particles appearing in the hadronization process.

The various sub-processes described above are the basis of various event generators which differ in some aspects concerning the treatment of individual sub-processes in the event

simulation. In this thesis, different generators are used:

PYTHIA [94]: PYTHIA is a general purpose event generator which has been used already extensively at previous collider experiments. It is designed to simulate collisions of either hadrons or same-generation leptons. This means that it is well suited to model pp collisions. The highest multiplicity of particles involved in the hard interaction that can be simulated with PYTHIA are $2 \rightarrow 3$ processes. Nonetheless, more than three final state particles can arise for instance from the parton shower. This is traditionally Q^2 ordered. The hadronization model in PYTHIA is based on the aforementioned string model. Besides the simulation of whole physics events including all relevant steps, PYTHIA can also be interfaced to other generator programs in order to carry out the parton shower, hadronization and decay steps only.

HERWIG++ [95]: Like PYTHIA, HERWIG++ is a general purpose MC generator. It provides the possibility to simulate high-energy collisions for lepton-lepton, lepton-hadron and hadron-hadron processes with special emphasis on the modelling of QCD radiation. Thus, the distinct feature of HERWIG++ is that colour coherence effects are taken into account by employing angular ordered parton showers meaning that the coherence of soft radiation is treated correctly. The hadronization process is based on the cluster model.

MADGRAPH [96, 97]: MADGRAPH is a general-purpose matrix-element program and designed to provide accurate descriptions of multiparton processes. It is the only generator that is capable of computing any arbitrary process at tree level or at NLO. In principle, it is possible to generate any $2 \rightarrow n$ process. However, in order to achieve reasonable computing times, the number of partons is typically limited to values < 10 . MADGRAPH can commonly be interfaced to PYTHIA for the realisation of the showering, hadronization and decay.

POWHEG [98]: The POWHEG programme provides accurate QCD computations based on matrix elements up to next-to-leading order following the concept proposed in [99]. It can be interfaced to shower programs like PYTHIA or HERWIG++ in order to perform the subsequent showering process.

In order to evaluate the interaction of the generated particles with the detector material, the passage of the particles through the detector is simulated. This is done based on a detailed model of the CMS apparatus utilizing the GEANT4 programme [100] and referred to as *full simulation*. Full simulation is quite extensive in computing time and thus for several purposes also a simplified scheme of the detector model is used which is known as *fast simulation* [101]. Especially in analyses in which many signal samples are needed to scan a wide parameter range, e.g. in SUSY analyses, the usage of the fast simulation is very beneficial. A comparison of several relevant kinematic distributions in real collision events to simulated events obtained from fast simulation shows a quite good agreement [102]. In particular, also jet related quantities like the sum of calorimetric jet momenta or the missing transverse momentum are modelled with a similar accuracy in full and fast simulation.

Finally, recorded collision data as well as simulated events including the modelling of the passage through the detector are present in the same data format which allows to apply

the same reconstruction algorithms in order to derive physics objects from the low-level event data. These algorithms are discussed int the next chapter.

4. Object Reconstruction and Particle Identification

Particles produced in the pp collisions traverse through the detector and interact with the detector components in a characteristic manner. Thus, it is possible to reconstruct different objects from the detector signals and identify various types of particles which actually emerged from the collision.

The approach for the event reconstruction and identification of specific particles used in CMS is discussed in this chapter. First, the *Particle-Flow (PF) algorithm*, used for a global description of the collision event, is introduced. Moreover, the reconstruction of jets is discussed in Sec. 4.2. Furthermore, the identification of decays from b hadrons and boosted top quarks is reviewed in Sec. 4.3 and 4.4, respectively.

4.1. Global Event Description with the Particle-Flow Algorithm at CMS

The CMS experiment introduced the Particle-Flow algorithm for the reconstruction of collision events and is designed to identify stable particles in an event. Types of particles which are identified by the PF algorithm are electrons and photons, charged and neutral hadrons as well as muons. In order to reconstruct the four-momenta of these particles all sub-components of the detector are utilized.

The CMS detector is very well suited for this task. The silicon tracker enclosed by the uniform magnetic field enables a very efficient track reconstruction yielding only a small track fake rate even down to small transverse momenta of 150 MeV. Furthermore, the strength of the magnetic field together with a high ECAL granularity allows photons to be separated from charged-particle energy deposits. A detailed introduction to the PF algorithm can be found in [103].

The event reconstruction starts with the identification of fundamental objects in the sub-detectors which are charged-particle tracks in the inner tracker, calorimeter clusters and muon tracks in the outer muon system. Tracks emerging from charged particles are formed following an iterative tracking algorithm [104]. Starting from an initial seed trajectory, e. g. pixel hit doublets or triplets, tracks are extrapolated to further tracker layers by taking into account multiple scattering and energy loss in the material. Each iteration proceeds with a removal of unambiguously allocated hits from the previous iteration. With this approach, a high tracking efficiency as well as a low fake track rate can be achieved which is a crucial requirement for a successful particle-flow event reconstruction. Furthermore, calorimeter clusters are formed in each sub-detector separately based on adjacent calorimeter cells. Neighbouring cells are combined to form clusters when their energy exceeds a pre-defined threshold.

A particle traversing the detector gives typically rise to several of such elementary tracks

and clusters so that a dedicated link algorithm is applied in order to connect these elements. Linked elements form blocks and remove a potential double-counting of the same object in different detector parts. First, charged-tracks are associated to calorimeter clusters, if the extrapolated trajectory matches the cluster within the cluster boundaries. This is done considering effects like gaps and cracks between detector components, uncertainty on the shower position or multiple scattering. To account for bremsstrahlung, also tangents of the tracks are extrapolated to the respective energy clusters. In addition, also ECAL and HCAL clusters can be connected to each other by linking clusters in the more granular calorimeter to clusters in the less granular one. Finally, global muons can be defined by associating charged-tracks from the tracker with muon tracks reconstructed in the muon system.

After the identification of such blocks of elements, the PF algorithm creates a list of all particles contained in the event applying dedicated quality criteria interpreting the blocks in terms of particles. The identification of muons and a removal of their tracks from the blocks is followed by an assignment of electrons and associated bremsstrahlung from tracks and linked ECAL clusters. After these have been removed from the list of blocks as well, remaining blocks with a good quality track are considered to be charged hadrons. Their momenta are determined from combining the track momentum and the respective energy in the calorimeter cluster. If the cluster energy largely exceeds the measured momentum from the track beyond the detector resolution, it constitutes a photon and if the excess is larger than the total ECAL energy also a neutral hadron. Finally, remaining ECAL and HCAL clusters not linked to any track give rise to photons and neutral hadrons.

The complete set of particles can then be used to derive further objects and quantities like e.g. jets as discussed in Section 4.2, missing transverse energy \cancel{E}_T or decay products of tau leptons. More detailed information on the specific quality criteria required for the identification of certain particles is given in the Chapters 5, 6 and 7 for each analysis presented in this thesis.

4.2. Reconstruction of Jets

As stated earlier, jets are the experimental signatures of quarks and gluons and constitute of a collimated spray of hadrons. However, in order to identify a particular jet and relate its properties to the original parton, a proper jet definition is needed. Typically, a *jet algorithm* determines how to cluster particles into a jet. Furthermore, it has to be defined how to assign a momentum to the jet. At CMS, the standard procedure is to assign the four-momentum sum of all jet constituents to the jet.

Different jet algorithms are introduced in Sec. 4.2.1 followed by a discussion of different jet types used at CMS in Sec. 4.2.2. Furthermore, the jet transverse-momentum response is defined in Sec. 4.2.3 and the calibration of jet energies is reviewed in Sec. 4.2.4.

4.2.1. Jet Algorithms

A jet algorithm usually provides a prescription how to combine individual particles into a single jet based on some distance criterion. Good jet algorithms though should be able to identify jets that are neither sensitive to the emission of soft particles (*infrared safety*) nor to the collinear splitting of particles (*collinear safety*). These features named *IRC safety* are desirable as otherwise, e. g. cross sections would be dependent on particular effects of

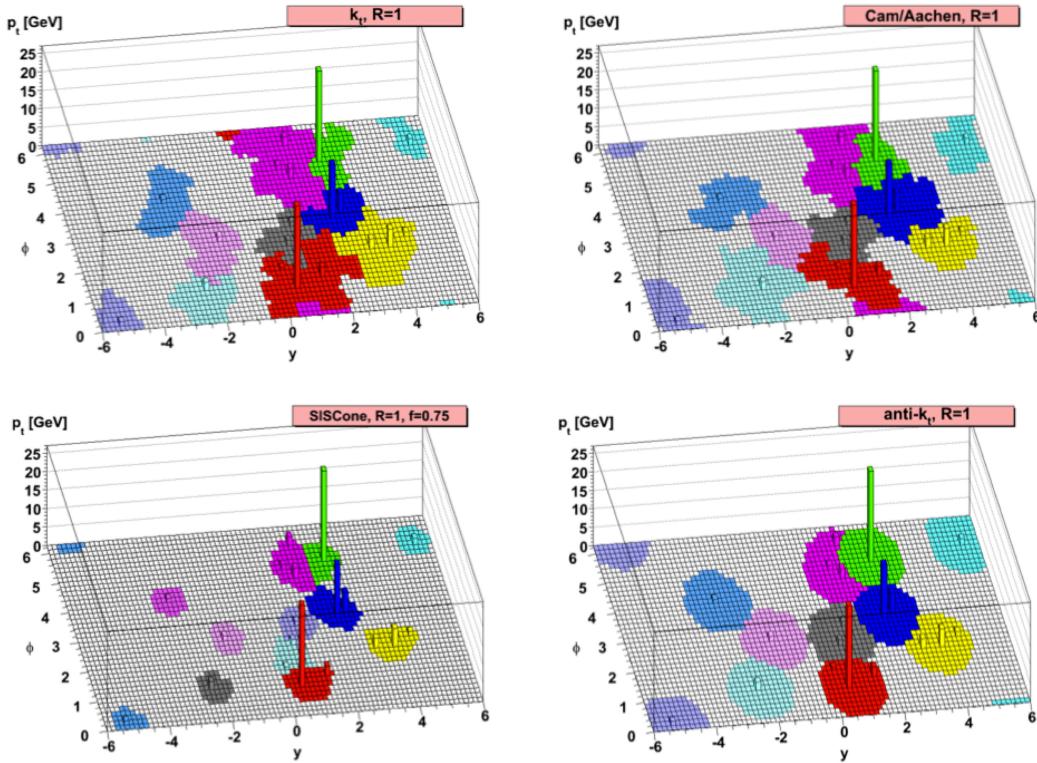


Figure 4.1.: Sample parton-level event generated with HERWIG++, adding many random soft particles, that is clustered with the k_T -algorithm (*top left*), the C/A-algorithm (*top right*), the SiScone-algorithm (*bottom left*) and the anti- k_T -algorithm (*bottom right*). Colours illustrate all particles clustered to the resulting hard jets. Taken from [105].

hadronisation involving random collinear splittings and soft emissions making it difficult to compare the results from experiments to theoretical predictions. The comparability is even more troubled as detectors intrinsically provide regularisation of IRC effects to some extent due to a finite resolution. Furthermore, calculations in QCD perturbation theory rely on the cancellation of divergences related to IRC processes. Thus, if jets are sensitive to such effects cancellations are not ensured and can lead to infinite cross sections. An introduction to the most commonly used jet algorithms known as *cone algorithms* and *sequential recombination algorithms* is given. A comprehensive overview of jet algorithms and properties can be found for instance in [105].

Cone Algorithms: Cone algorithms were the first developed jet algorithms. They are based on the general idea that the main kinematics in an event are not changed by specific effects from hadronisation and thus a jet is defined by a set of particles within a stable cone around their centre of mass. Typically, separate angular or energy parameters are used to perform the jet finding. A very common approach is implemented in *iterative cone* (IC) algorithms. Here, a seed constituent i , which is for instance the constituent with the highest transverse momentum, defines the

initial direction and momenta of all constituents j within a cone defined by

$$\Delta R_{i,j}^2 < R^2 \quad (4.1)$$

as introduced in Sec. 3.2.1, are added to the momentum of the seed. The resulting direction is used as new seed direction and the whole procedure is repeated until a stable cone is achieved. The dimensionless parameter R hence defines the jet radius. After finding such a jet, all constituents are removed from the input list and further jets are clustered from the remaining objects. This progressive removal approach avoids to form jets with overlapping cones. However, such a procedure is not IRC safe as collinear splittings can lead to varying seeds in the event and thus to different final ensembles of jets.

However, cone algorithms are IRC safe when instead of iteratively forming stable cones all stable cone solutions are identified at once. This procedure is denoted *seedless cone* (SC) algorithm. The usage of such algorithms though is typically impractical as the computation time increases exponentially with the number of particles to be considered so that even for 100 particles it is not solveable at any reasonable timescale. A feasible implementation of a seedless cone algorithm featuring a $\mathcal{O}(N \log(N))$ time-dependence is given by the SISCone algorithm [106]. As it is usually nonetheless still more time consuming than sequential algorithms, as described in the next part, the SISCone algorithm is not used by CMS.

Sequential Recombination Algorithms: The basic concept of sequential clustering algorithms is to iteratively group pairs of particles together based on some distance measure and thus, to some extent, reconstruct the evolution of a parton shower. At hadron colliders where the total energy of a collision is unknown, a suitable metric to be used based on variables invariant under longitudinal boosts is

$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}$$

$$d_{iB} = k_{T,i}^{2p}$$

with the distance d_{ij} between final state objects i and j carrying transverse momentum k_T and the distance of the object to the beam d_{iB} . While ΔR_{ij} denotes the spatial separation in the (η, ϕ) -plane, R and p are free parameters of the algorithm. The recombination is done by first calculating d_{ij} and d_{iB} for all objects in the final state and then identifying the minimum value. If the minimum is d_{ij} , the two objects i and j are combined meaning that their four-momenta are added to form one new object. Afterwards distances are computed again. However, if the minimum is d_{iB} , object i is declared a jet and removed from the input list. This procedure is repeated until all objects are assigned. In this context, the parameter R acts as an angular cut-off and thus has a similar role as the jet radius in cone algorithms.

Depending on the choice of the parameter p , different types of algorithms are distinguished which are all IRC safe. An illustration of an arbitrary example event where jets are clustered with the different jet algorithms is shown in Fig 4.1. The k_T -algorithm uses $p = 1$ [107] and thus clusters soft particles first. This results mainly in irregularly shaped jets (cf. Fig 4.1) and makes them sensitive to radiation

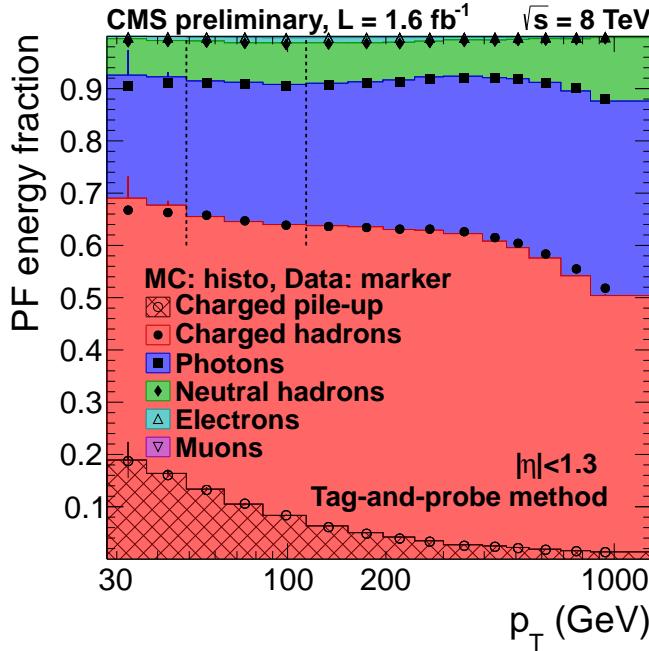


Figure 4.2.: Composition of the PF jet energy versus jet p_T in the barrel detector region $|\eta| < 1.3$ in simulated events (coloured histograms) and data (solid markers) [111].

in the event. Consequently, they are difficult to calibrate and therefore impractical to use at hadron colliders. The *Cambridge-Aachen-* or short *CA*-algorithm [108, 109] utilizes $p = 0$. Thus, it does not rely on momentum information, but clusters jets based on the angular separation of input objects. This allows a direct geometric interpretation of jets and has been proven that to be a jet algorithm especially suited for analyses of jet substructure, as will be discussed later in Sec. 4.4. These jets do still reflect the structure of the parton shower, but are less affected by soft radiation than the k_T -algorithm. Finally, the *anti- k_T* -algorithm uses $p = -1$ [110] and starts the jet clustering beginning with the hardest objects in the event. Hence, the evolution of the parton shower is not reproduced. The *anti- k_T* algorithm tends to form very regular shaped jets, as the four-momentum of the core of the jet is not much affected by the softer components which are clustered later in the process. Typically, the regular shape allows an easy calibration of *anti- k_T* jets.

4.2.2. Jet Types at CMS

The default jets used at the CMS experiment are jets clustered by the *anti- k_T* algorithm using a distance parameter of $R = 0.5$ (AK5 jets). This can be applied to reconstructed detector signals resulting in *detector-level* jets or to final-state particles after hadronisation and decay, in simulated events, giving rise to *particle-level* jets. Implementations of various jet algorithms are provided by the *FastJet* package [112, 113].

Based on the information used from the various CMS subdetectors for the jet clustering, different types of detector-level jets are distinguished. Some of them are [114]:

Calorimeter (Calo) jets: Calo jets are clustered from energy deposits in the calorimeters. For this purpose, calorimeter towers are defined which consist of at least one HCAL cell and the geometrically associated ECAL cells. For instance in case of the barrel detector region, a calorimeter tower consists of one HCAL cell and 5×5 ECAL cells. The four-momentum of each tower is defined by the tower position as seen from the primary interaction vertex and the energy deposit in the tower, above a certain threshold assuming a mass of zero.

Jet-Plus-Track (JPT) jets: JPT jets are reconstructed from calorimeter jets complemented by tracking information [115]. Tracks of charged particles can be associated to calo jets based on the separation of the jet axis and the momentum vector of the track in the (η, ϕ) -plane. Associated tracks are projected to the jet cone and are exploited to correct the jet energy and direction.

Particle-Flow (PF) jets: PF jets are clustered from the four-momentum vectors of particle-flow candidates as identified by the PF algorithm described in Section 4.1. Typically, these types of jets show the best performance as the excellent resolution of the tracking system and the ECAL are utilized. Only neutral hadrons which constitute around 15% of the jet energy (cf. Fig. 4.2) rely on the energy measurement of the HCAL with its relatively poor resolution. Thus, PF jets are the default jets to be used at the CMS experiment, as it is done within this thesis. In order to mitigate influences from pile-up, charged hadrons unambiguously associated to vertices other than the primary vertex can be removed from the jet algorithm input list before the actual jet clustering is performed. This technique is referred to as *charged hadron subtraction* (CHS) and used as default throughout this thesis.

4.2.3. Jet Transverse-Momentum Response

In general, the jet transverse momentum as measured at detector level is not necessarily equal to the energy of the original particle due to for instance the limited detector resolution. This effect is quantified by the *jet transverse-momentum response* \mathcal{R} which is defined as

$$\mathcal{R} = \frac{p_T}{p_T^{\text{particle}}} \quad (4.2)$$

where p_T denotes the transverse momentum of the jet measured at detector level and p_T^{particle} is the transverse momentum of the original particle-level jet. Thus, the jet response provides a measure of the jet momentum visible in the detector compared to the actual momentum of the particle after hadronisation and decay.

The average response $\langle \mathcal{R} \rangle$ is referred to as *jet energy scale* (JES) and has to be calibrated such that $\langle \mathcal{R} \rangle = 1$ for fixed p_T^{particle} . The response usually depends on the jet momentum as well as on the pseudorapidity. This is expected since the precision of the jet measurement is directly related to the energy of the particles and the resolution of the detector sub-components. This is for instance caused by the specific track-reconstruction efficiency, the individual amount of detector material, cracks in the detector layout or different instrumentation.

Furthermore, the width of the response distribution corresponds to the relative *jet transverse-momentum resolution* (JER) and consequently is a function of p_T and η as well.

4.2.4. Jet Energy Calibration

In order to relate the measured jet momentum on average to the momentum of the corresponding particle-level jet, a jet calibration procedure is applied. Within the CMS experiment, a factorized approach is utilized which is described in detail in [114]. The actual set of correction factors used in this thesis, if not stated otherwise, is documented in [116].

The calibrated jet four-momentum vector p^{cor} is obtained from the raw jet momentum four-vector p^{raw} by scaling the raw momentum with a correction factor C according to

$$p^{\text{cor}} = C \cdot p^{\text{raw}} = C_{\text{offset}}(p_T^{\text{raw}}, \eta) \cdot C_{\text{rel}}(\eta) \cdot C_{\text{abs}}(p'_T) \cdot C_{\text{res}}(p''_T, \eta) \cdot p^{\text{raw}} \quad (4.3)$$

where C is composed of the offset correction C_{offset} , the calibration factors C_{rel} and C_{abs} as well as residual correction factors C_{res} applied to data. Each correction factor is applied sequentially after the other in a fixed order such that p'_T denotes the transverse momentum after the application of the offset correction and p''_T is the transverse momentum after applying the respective previous corrections. Some details for each individual correction are given in the following:

L1 Offset: The L1 offset correction is designed to compensate for additional energy contributions arising from instrumental noise or pile-up events. The p_T offset is estimated in dependence of η , the effective jet area A_j and the p_T -density ρ (*hybrid jet area method* [117]). The jet area is determined by adding a large number of infinitely soft four-momentum vectors to the event. The active jet area is then defined as the fraction of soft particles clustered together with the true hard jet components. The p_T -density ρ is defined on an event-by-event basis as the median of the distribution p_T^j/A_j where j denotes all reconstructed jets in the event. The estimated offset in simulated QCD multijet events and data is illustrated in Fig. 4.3.

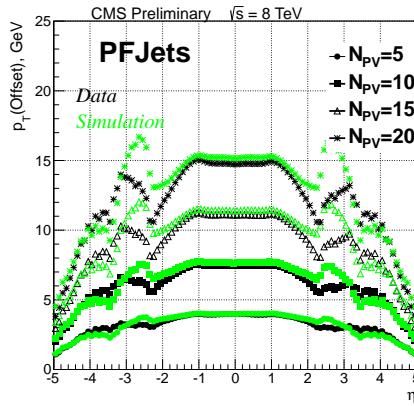


Figure 4.3.: L1 offset transverse momentum correction for AK5 PF jets as a function of jet pseudorapidity in data (black) and simulation (green). Different intervals of reconstructed primary vertices (N_{PV}) are shown with different markers [116].

L2 Relative + L3 Absolute Correction: The L2 relative correction is designed to make the jet energy scale uniform with respect to η while the L3 absolute correction ensures a uniform response versus p_T . Both corrections are entirely estimated from simulated QCD multijet events. The correction is defined as the inverse of the average response $1/\langle R \rangle$ at fixed p_T^{gen} and illustrated in Fig. 4.4.

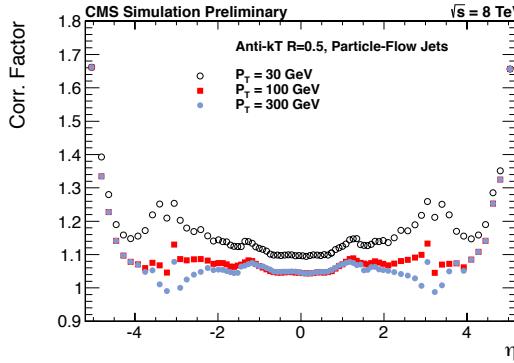


Figure 4.4.: MC-Truth L2 relative corrections for AK5 PF jets as a function of jet pseudorapidity for three reference transverse momentum values: 30 GeV (white hollow circles), 100 GeV (red squares) and 300 GeV (blue circles) [116].

L2L3 Residual Correction: In order to compensate for remaining response differences between simulated events and data, residual correction factors are derived. These are applied to data only and correct for remaining differences in the data-to-simulation ratio of the relative jet energy scale. Such corrections can be derived from events featuring a momentum balance in the transverse plane like dijet events (used for the L2 residual determination) or $Z/\gamma + \text{jet}$ events (used for the measurement of the L3 residual correction). The L2 residual correction is illustrated in Fig. 4.5.

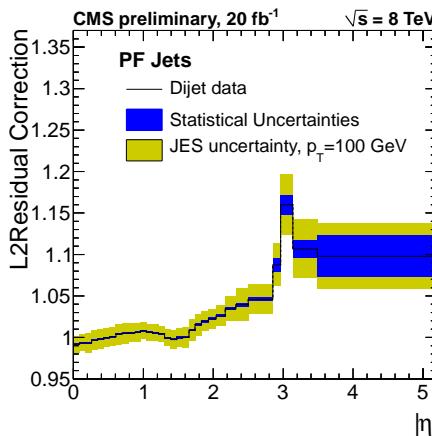


Figure 4.5.: L2 residual corrections for AK5 PF jets as a function of jet pseudorapidity, obtained from dijet events, shown with JES uncertainty (yellow band) and the statistical uncertainty (blue band) [116].

The calibration factors are obtained with respect to the average flavour composition of a QCD multijet sample. Thus, further steps of correction factors can be applied for specific analysis purposes, e. g. correcting the different response for various jet flavours. However, such higher order corrections are not used in this thesis.

4.3. Identification of b-Quark Jets

Jets arising from the hadronization of bottom quarks are usually referred to as *b jets*. As these are present in many physics processes, e.g. the decay of top quarks, it is crucial to identify b jets, i. e. distinguish them from jets initiated by gluons or light-flavour quarks. Typically, the identification of b jets is denoted *b tagging* which exploits the distinct properties of b quarks for the identification of the respective jets. In general, b hadrons possess a lifetime of around $c\tau \approx 500 \mu\text{m}$ that makes them travel in the detector before they actually decay. This results typically in a measurable secondary vertex that is displaced with respect to the primary interaction. Furthermore, b-quark jets feature a high number of charged particles per decay resulting in jets with several tracks and often exhibit soft leptons emerging from semi-leptonic decays of b quarks. The CMS experiment with the ability to perform a precise charged-particle tracking and lepton identification exploits the specific b jet properties in dedicated b-tagging algorithms for an efficient b-jet identification [118].

Each algorithm determines a discriminator value per jet indicating how b-jet-like a jet behaves. Based on that, working points are defined corresponding to a specific minimum threshold of the discriminator value. These working points are named *loose*, *medium* and *tight* and correspond to a misidentification probability, i. e. the probability to identify a non-b jet as b jet, of 10%, 1% and 0.1% for an average jet transverse momentum of 80 GeV, respectively. B-tagging algorithms commissioned within the CMS experiment are:

Track Counting (TC) algorithm: A powerful discriminator for the decay products of a b hadron from prompt tracks is the *impact parameter* (IP) of a track with respect to the primary vertex. Its significance can be computed by taking the ratio of the IP to its respective uncertainty. Tracks in a jet are sorted by decreasing values of the IP significance by the TC algorithm. Depending on whether the IP significance of the second or the third ranked track is chosen as discriminator the algorithm is denoted *Track Counting High Efficiency* (TCHE) or *Track Counting High Purity* (TCHP) algorithm.

Jet Probability (JP) algorithm: The JP algorithm extends the simple TC algorithm by connecting the information about the IP from a couple of tracks in the jet. A likelihood is calculated that all tracks of the jet stem actually from the primary vertex. This approach can be varied by giving more weight to tracks with the highest IP significance. The maximum of such tracks is four and matches the average number of reconstructed charged particles from the decay of b hadrons. This version is called *Jet B Probability* (JBP) algorithm.

Simple Secondary Vertex (SSV) algorithm: A further useful discriminating feature for b tagging is the presence of a secondary vertex and related kinematic variables like the flight distance and direction which can be determined from the vector between the primary and secondary vertex. The SSV exploits the significance of the flight

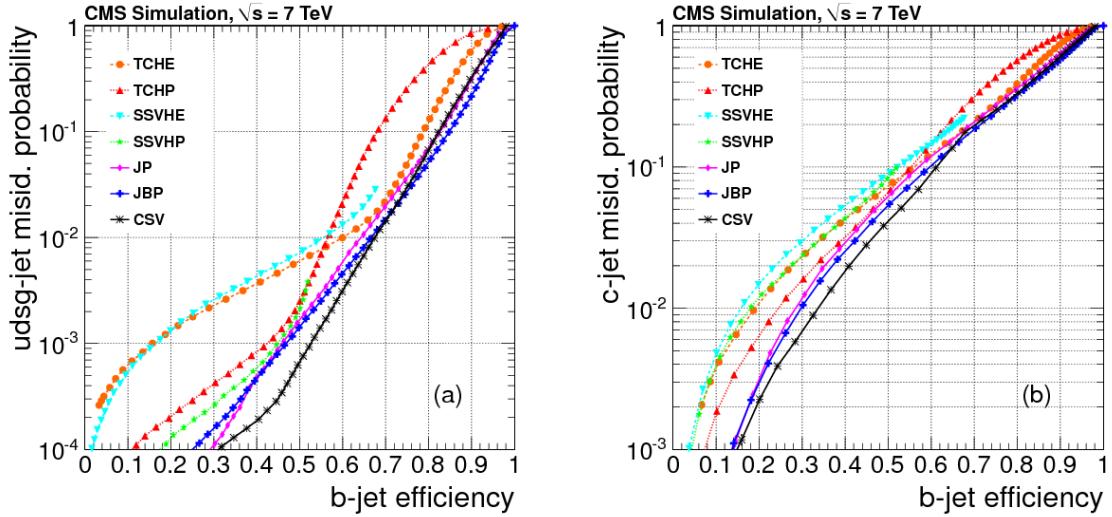


Figure 4.6.: Performance curves obtained from simulation for the algorithms described in the text. (a) light-parton- and (b) c-jet misidentification probabilities as a function of the b-jet efficiency. Taken from [118].

distance which is given by the flight distance divided by the associated uncertainty. Two different versions of this algorithm exist targeting on the one hand a *High Efficiency* (SSVHE) and on the other hand a *High Purity* (SSVHP). While the SSVHE is based on vertices with at least two associated tracks, the SSVHP uses vertices with at least three tracks. Typically, the efficiency of the algorithm is limited by the reconstruction efficiency of secondary vertices which is at the order of 65%.

Combined Secondary Vertex (CSV) algorithm: The CSV algorithm provides utilizes an approach combining information from secondary vertices as well as track-based lifetime information and thus is able to exceed the efficiency of SSV algorithms. It allows even an efficient identification of b jets also in cases when no secondary vertex could be reconstructed. Often, pseudo-vertices can be formed from tracks even when failing the reconstruction of an actual secondary vertex which allows to derive some secondary vertex related quantities. Variables used in the CSV algorithm are, e. g. flight distance significance, vertex mass, number of tracks at the vertex, number of tracks in the jet and the IP significances for the tracks in the jet. These variables are used to compute two likelihood ratios which can be used to distinguish either c and b jets or light-parton and b jets.

In order to determine the quality of a particular b-tagging algorithm, typically the misidentification probability as a function of the b-jet efficiency is compared for various taggers. Such a performance comparison is illustrated in Fig 4.6 for the tagging algorithms described above. The misidentification probability is derived separately for light-flavour and gluon initiated jets as well as c jets. The curves are derived from simulated multijet events using jets with $p_T > 60$ GeV. For loose working points, the b-jet efficiency is around $\approx 80 - 85\%$, and the JBP algorithm shows the best performance. In case of medium and

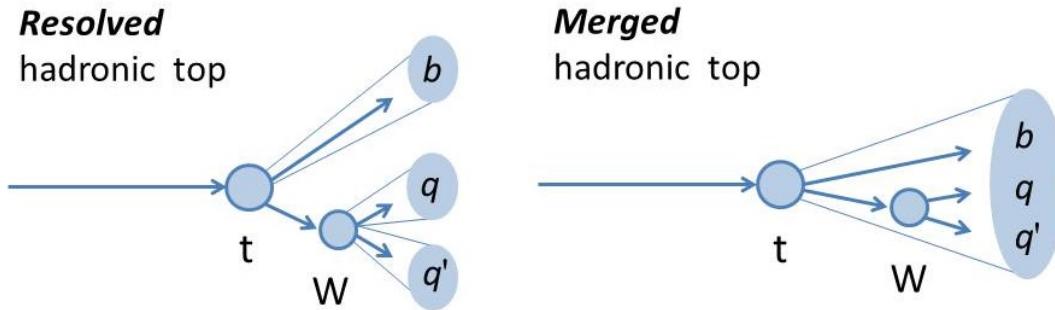


Figure 4.7.: Schematic diagrams of the decay of a top-quark in the resolved case (*left*) and the boosted scenario (*right*).

tight working points, the b-tag efficiency drops to $\approx 45 - 55\%$ and the CSV algorithm performs best. B-tagging algorithms used for analyses of data obtained at $\sqrt{s} = 8 \text{ TeV}$ in 2012 where the TCHP, JP and CSV algorithms [119].

4.4. Identification of Boosted t-Quark Jets

Supersymmetric models or other scenarios describing physics beyond the Standard Model predict the existence of new massive particles. Often, the coupling of these particles especially to quarks belonging to the third generation is sizeable, e.g. in decays of top squarks which predominantly are expected to decay into top quarks. Consequently, such processes lead to highly-energetic top-quarks in the final state which can be identified exploiting the specific properties of the top-quark.

As discussed in Sec. 2.3.2, the decay of the top-quark is experimentally characterized by the decay of the W boson which makes two thirds of top quark decays resulting exclusively in hadrons. This decay mode is here referred to as *hadronic top*. If top quarks are produced with $p_T \ll m_t$, the top quark decay products show up as three distinct objects in the detector. In the case of hadronic tops, these are three well separated jets. However, if the top transverse momentum is high, the decay products are *boosted* and thus collimated in the forward direction. Consequently, they might overlap and merge into a single large jet (*fat jet*). The opening angle of the decay products ΔR is expected to scale as

$$\Delta R \approx 2m/p_T \quad (4.4)$$

where m and p_T are the mass and the transverse momentum of the decaying particle, respectively. Schematic diagrams of resolved and boosted top quark decays are illustrated in Fig. 4.7.

The identification of boosted top-quark decays – here restricted to hadronic tops – is typically known as *top tagging* and aims at the determination of the decay

$$t \rightarrow W + b \rightarrow q\bar{q}' + b \quad (4.5)$$

by analysing the substructure of fat jets and applying kinematic selections. Typically, fat jets are clustered by the Cambridge-Aachen algorithm with distance parameters of $R = 0.8$ (CA8 jet) or $R = 1.5$ (CA15 jets). Such large jets are often more prone to radiation in the event that is wrongly associated to the hard jet than jets with smaller radii. Such unwanted contributions often result in a degraded resolution. In order to mitigate such effects, *jet grooming* techniques can be employed:

Filtering [120]: Jet constituents are recombined by rerunning the Cambridge-Aachen algorithm with a smaller distance parameter, like 0.3. With this procedure, subjets within the fat jet are identified: The filtering is performed by keeping only n of the identified subjets. Thus, the dominant contribution from the hard process is captured while contamination from the underlying event or pileup is removed.

Pruning [121]: In contrast to filtering, pruning does not attempt to find a particular number of subjets. Rather, the idea is to systematically remove soft and large angle recombinations in the jet clustering process. For the recombination of objects i and j , two conditions are defined:

$$z = \frac{\min(p_{T,i}, p_{T,j})}{|\vec{p}_{T,i} + \vec{p}_{T,j}|} < z_{\text{cut}}$$

$$\Delta R_{i,j} > D_{\text{cut}} = \frac{m_J}{p_{T,J}}$$

where m_J and $p_{T,J}$ are the mass and the transverse momentum of the original fat jet, respectively. Whenever these two conditions are satisfied, the recombination of objects i and j is not performed. Thus, the jet clustering is rerun using the initial jet constituents, but in addition employing a veto rule. For the Cambridge-Aachen algorithm z_{cut} is typically chosen to be 0.1.

Trimming [122]: Similarly to the filtering procedure, the jet trimming is based on the concept of identifying subjets within an initially clustered large jet, here called *seed jet*. The constituents of this seed jet are reclustered using a jet clustering algorithm with a smaller distance parameter than for the seed jet. If

$$p_{T,i} < f_{\text{cut}} \cdot \Lambda_{\text{hard}},$$

the contributions from subjet i to the seed jet are removed. Here, f_{cut} is a fixed dimensionless parameter and Λ_{hard} a scale chosen with respect to the specific kinematics of the event. While f_{cut} quantifies the expected hierarchy between FSR and ISR or pileup, the hard scale Λ_{hard} usually represents the transverse momentum of the seed jet or the effective mass of the event.

In order to identify decays of boosted hadronic top quarks, as discussed above, several top tagging algorithms (or short *taggers*) are commissioned within the CMS experiment [123, 124]. Two of them that are used within this thesis are:

CMS Top Tagger: The CMS Top Tagger [125] is based on the Top Tagger developed by Kaplan et al. [126] and acts on jets clustered by the Cambridge-Aachen algorithm with distance parameter $R = 0.8$. These jets, used as input for the algorithm, are

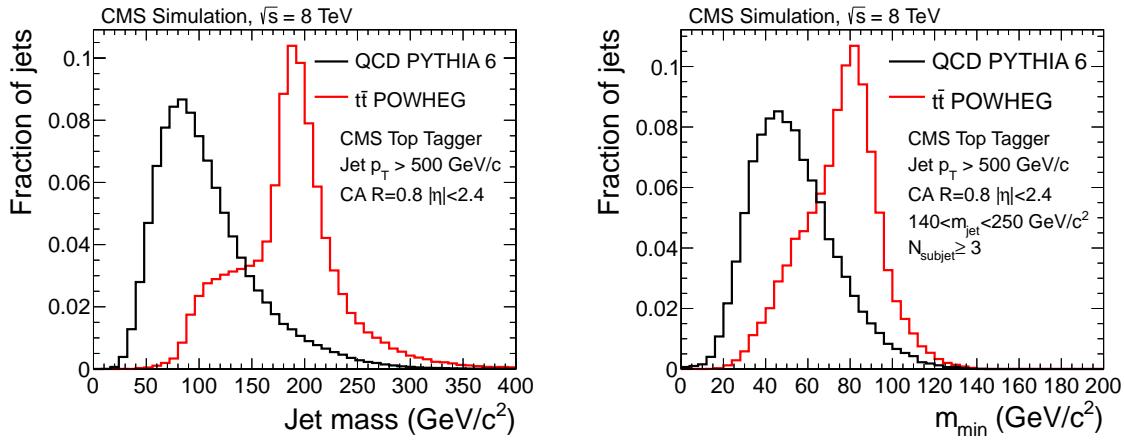


Figure 4.8.: Jet mass (*left*) and minimum pairwise subjet mass (*right*) for CA8 jets with $p_T > 500 \text{ GeV}$ from a simulated $t\bar{t}$ POWHEG sample (red) and from a simulated QCD PYTHIA sample (black). Taken from [123].

denoted as *hard jets*. Since the decay products of the hadronic top are not expected to be all contained within one jet with $R = 0.8$ for low transverse momenta of the top quark, only jets with $p_T^{\text{jet}} > 350 \text{ GeV}$ are considered. In order to identify subjets within the hard jets, a two-stage decomposition procedure is applied which performs the pairwise clustering sequence used to form the hard jet in reverse order. First, the algorithm aims at splitting the hard jet into two subclusters (*primary decomposition*). Second, it is attempted to further split the clusters emerging from the first step (*secondary decomposition*). Typically, subjets are found when they are spatially well separated and carry a significant fraction of the momentum of the hard jet. Details on the actual splitting criteria can be found in [123]. With this approach, up to four subjets are identified within the hard jet. After a successful decomposition procedure, kinematic criteria can be applied to the identified subjets in order to tag top jets. In the CMS top-tagging algorithm, as employed in this thesis, the following criteria are used:

- Number of subjets ≥ 3
- The jet mass m_{jet} , i. e. the mass of the four-vector sum of the constituents of the hard jet, has to be close to the top-mass with $140 < m_{\text{jet}} < 250 \text{ GeV}$.
- The invariant mass of each pair of the three subjets highest in p_T is calculated. The minimum of the pairwise masses m_{min} has to be $> 50 \text{ GeV}$.

In Fig. 4.8, the distributions for jet mass and minimum pairwise mass are illustrated for a simulated $t\bar{t}$ (red) and QCD multijet (black) sample at $\sqrt{s} = 8 \text{ TeV}$.

HEP Top Tagger: The HEP Top Tagger [127] uses jets with a larger distance parameter than the CMS Top Tagger of $R = 1.5$. This makes the HEP top-tagging algorithm especially suited for tops with moderate boost and thus, uses fat jets with a transverse momentum greater than 200 GeV as input. Since a larger jet size

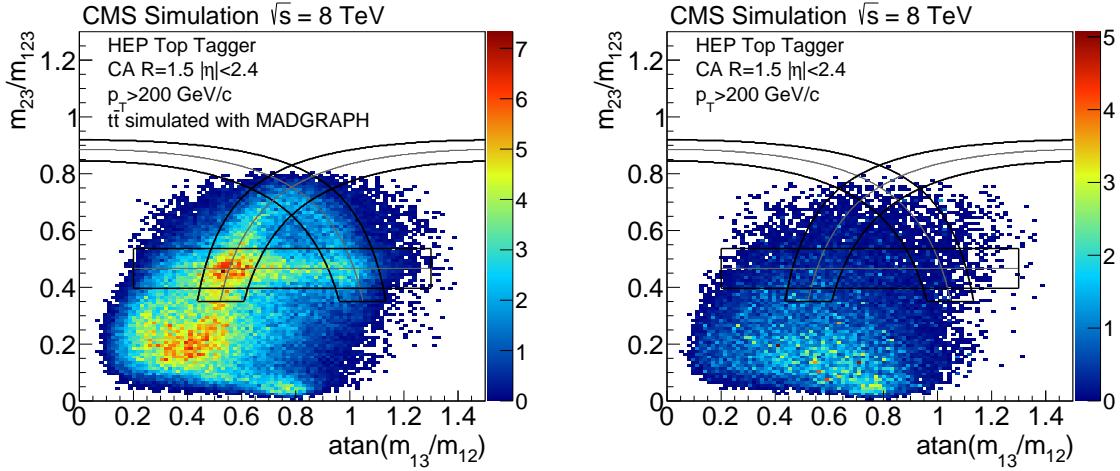


Figure 4.9.: Two-dimensional distributions of m_{23}/m_{123} versus $\text{atan}(m_{13}/m_{12})$ for HEP Top Tagger subjets from CA15 jets with $p_T > 200 \text{ GeV}$ for a simulated $t\bar{t}$ MADGRAPH sample (*left*) and for a simulated background sample composed of cross-section weighted boson + jets, diboson, single-top, $t\bar{t}$ all-hadronic and $t\bar{t}$ leptonic events (*right*). The A-shaped region indicates the selected region by the HEP Top Tagger. Taken from [123].

is in principle more prone to disturbing effects from the underlying event or pile-up, a sophisticated decomposition procedure is applied to distinguish hard subjets from soft components. Similar to the decomposition done for the CMS tagger, the identification of subjets is based on going through the cluster history of the jet in reversed order. First, the jet is decomposed into subclusters applying a *mass drop condition* discarding too soft components. Afterwards, subclusters resulting from the mass drop decomposition are reclustered into subjets and filtering criteria are applied. Details on the mass drop decomposition and reclustering criteria can be found in [123]. Finally, the combination with the mass determined from the filtered subjets closest to the top mass is kept and reclustered to force three subjets. Kinematic selections are applied to these three final subjets in order to identify the top jets. The following quantities are used based on the invariant mass of combinations of subjets:

- The invariant mass of the sum of the four-vectors of the three subjets is required to be in the top mass window: $140 < m_{123} < 250 \text{ GeV}$
- In order to select the W mass, the jet has to satisfy at least one of the following conditions based on the subjet pairwise masses

$$0.2 < \text{atan} \frac{m_{13}}{m_{12}} < 1.3 \text{ and } R_{\min} < \frac{m_{23}}{m_{123}} < R_{\max} \quad (4.6)$$

$$R_{\min}^2 \left(1 + \left(\frac{m_{13}}{m_{12}} \right)^2 \right) < 1 - \left(\frac{m_{23}}{m_{123}} \right)^2 < R_{\max}^2 \left(1 + \left(\frac{m_{13}}{m_{12}} \right)^2 \right) \text{ and } \frac{m_{23}}{m_{123}} > 0.35 \quad (4.7)$$

$$R_{\min}^2 \left(1 + \left(\frac{m_{12}}{m_{13}} \right)^2 \right) < 1 - \left(\frac{m_{23}}{m_{123}} \right)^2 < R_{\max}^2 \left(1 + \left(\frac{m_{12}}{m_{13}} \right)^2 \right) \text{ and } \frac{m_{23}}{m_{123}} > 0.35 \quad (4.8)$$

where the indices of m indicate the rank of the considered subjets with respect to the transverse momentum, $R_{\min} = (1 - f_W) \times m_W/m_t$ and $R_{\max} = (1 + f_W) \times m_W/m_t$ for the W mass width chosen to be $f_W = 0.495$.

The criteria related to the selection of the W mass are illustrated in Fig. 4.9 for signal events as well as background events and indicated by the black solid lines resulting in an A-shaped region.

In analogy to the b-tagging algorithms, different working points for each top tagger can be defined characterized by a specific top-tag efficiency and misidentification rate. Typically, the working points are chosen such that they have a minimum mistag rate for a given signal efficiency. In addition, to the taggers as described above further selection criteria can be added to achieve a potential performance improvement. These criteria are subjet b-tagging, as introduced in [119], and selections on the n -subjettiness [128, 129] ratio τ_3/τ_2 that denotes how likely a fat jet is composed of three subjets rather than two.

The performance of various taggers is compared by determining the top-tag efficiency versus the misidentification rate. Commonly, the top-tag efficiency for simulated events is defined as the number of jets passing the top-tagging selection divided by the number of jets associated to a simulated hadronic top or anti-top passing a certain p_T selection. Similarly, the mistag rate is defined as the number of jets passing the top-tagging selection divided by the number of jets associated to a simulated quark or gluon from the hard process passing the p_T selection. In Fig. 4.10 the mistag rate versus the top-tagging efficiency is illustrated for various taggers and parton momenta. The top-tag efficiency determined from simulated $t\bar{t}$ events for the selection criteria described above amounts to 35.3% for a matched parton- p_T of > 200 GeV for the HEP Top Tagger (HEP WP0) and 38.3% for a matched parton- p_T of > 400 GeV for the CMS Top Tagger (CMS WP0) while the mistag rates as determined from simulated QCD multijet events are 2.6% and 2.5%, respectively. However, the tagging performance shows a moderate dependence on pile-up such that in high pile-up environments a performance degradation is expected. For instance, in the case of the CMS Top Tagger the efficiency stays rather stable as a function of the number of primary vertices with a slope of $0.031\% \pm 0.034\%$. However, the mistag rate increases with a slope of $0.095\% \pm 0.006\%$ as a function of the number of primary vertices.

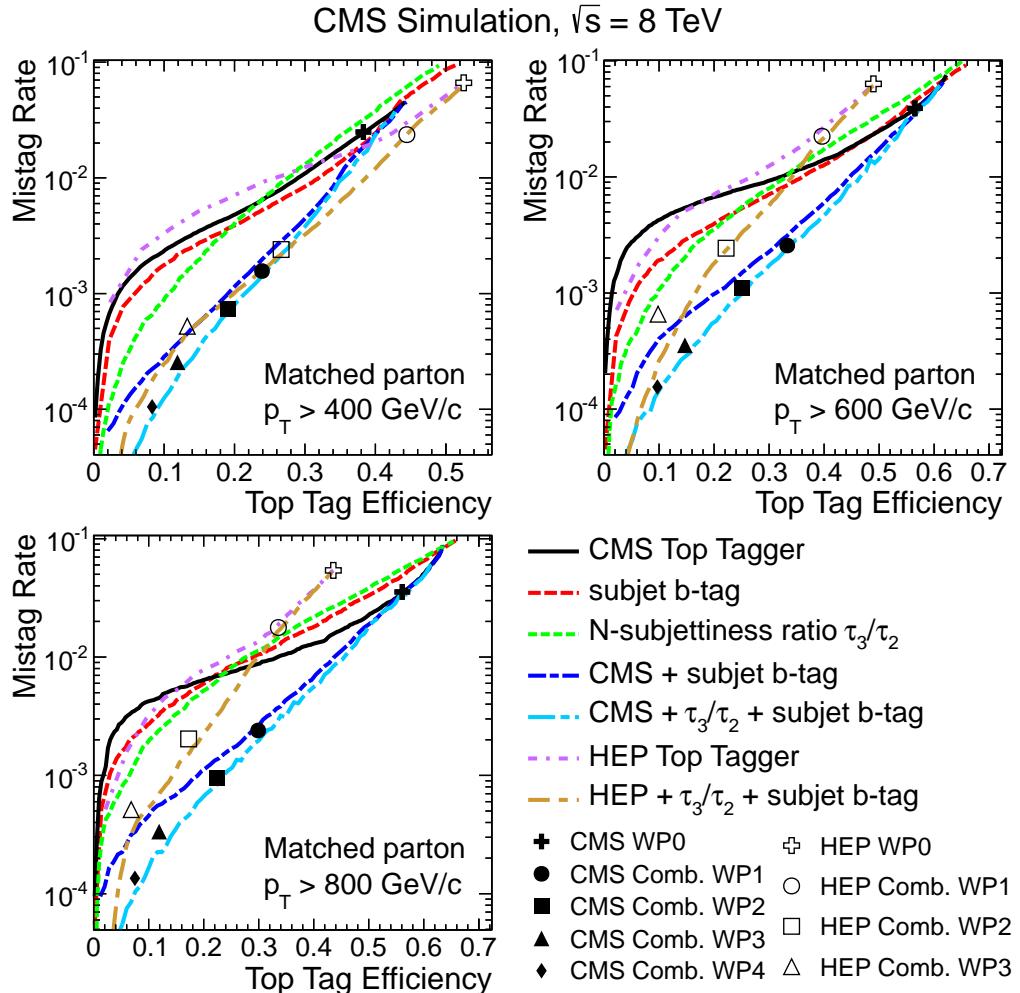


Figure 4.10.: Mistag rate versus top-tagging efficiency as measured from QCD PYTHIA 6 and POWHEG $t\bar{t}$ simulated events, respectively. In the cases where a jet mass cut is applied, the cut is not varied but fixed at $140 < m_{\text{jet}} < 250 \text{ GeV}$. N-subjettiness is calculated using $R = 0.8$ jets except when used in combination with the HEP Top Tagger in which case $R = 1.5$ jets are used. Signal jets are matched to simulated all-hadronic generated top quarks, while background jets are matched to simulated partons from the hard scatter. Distributions are shown for three p_T selections, where the p_T cut is applied to the matched generated parton [123].

5. Measurement of the Jet Transverse-Momentum Resolution

Many measurements of standard model properties or searches for new physics beyond the standard model performed within the CMS experiment rely on events with jets in the final state. Hence a good understanding of jet properties, like for instance the resolution of the jet transverse momentum is of major importance and a crucial ingredient for such kind of analyses. For example, many new physics searches are carried out based on jet final states. Here, QCD multijet events can fake the signature of possible new physics events and constitute a background process, since a mismeasurement of the jet momenta due to the limited detector resolution or the decay of heavy flavour quarks leads to a momentum imbalance in the event and consequently to measurable missing energy. The knowledge of the jet resolution is hence keypoint in the prediction of such background contributions, as discussed in Chapter 6.

In this chapter, an analysis is presented where the jet- p_T resolution in data and in simulated events is derived. The method is based on momentum conservation in the transverse plane of dijet events and offers the possibility to cover a large phase space in p_T and η . A similar approach was already used in previous studies at $\sqrt{s} = 7\text{ TeV}$ [114, 130] while a complementary approach utilizes $\gamma + \text{jet}$ events [131–133]. The measurement shown here is based on collision data corresponding to an integrated luminosity of 19.7 fb^{-1} recorded at $\sqrt{s} = 8\text{ TeV}$ in 2012. Parts of this Chapter are taken from [134], having been written by the author.

5.1. Components of the Jet Response

As introduced in Section 4.2.3 the transverse momentum of a jet at detector level does not equal the momentum of the corresponding particle-level jet which is typically expressed in the jet response. In simulated events the transverse momentum of the particle-level jet is given by the p_T of the generated jet which is clustered from all stable particles after hadronisation including neutrinos. Thus, the *MC-truth response* can be determined as

$$R = \frac{p_T^{\text{gen}}}{p_T^{\text{gen}}} . \quad (5.1)$$

In Fig. 5.1 an example for a jet response distribution derived from simulation is shown. It is obtained from a QCD multijet sample generated with PYTHIA 6 tune Z2 [135] using the CTEQ6L1 PDF set [136] and processed with the full detector simulation. Since the cross-section of the process has been scaled by $\hat{p_T}^{4.5}$ with the scale parameter $\hat{p_T}$ describing the momentum transfer in the hard process, the sample is reweighted with the inverse in order to regain the physical spectrum. For the calculation of the truth response, the two generated jets in the event with highest transverse momentum are selected. The event is rejected, if one of these two generator jets does not have a corresponding reconstructed

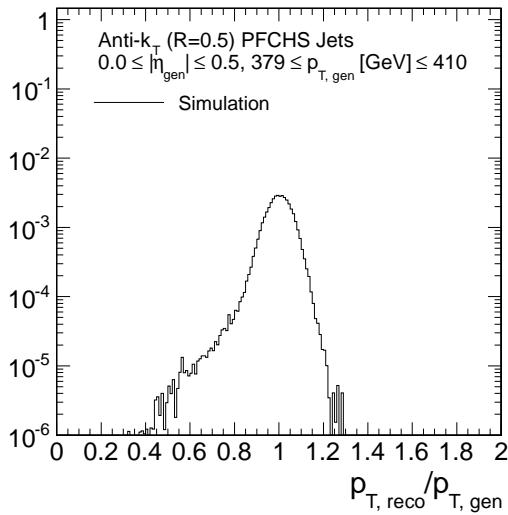


Figure 5.1.: Jet response function for one example $|\eta_{\text{gen}}|$ and $p_{T,\text{gen}}$ interval derived from simulation.

jet within a distance of $\Delta R < 0.25$. Otherwise the response is calculated in intervals of p_T^{gen} and $|\eta^{\text{gen}}|$ since a dependence on the momentum and the respective detector region is expected, as discussed in Section 4.2.3. This is done after applying the jet energy corrections discussed in Section 4.2.4 to the detector-level jet momenta.

Apparently the jet response consists generally of two components – the Gaussian-shaped core around the mean referred to as jet resolution and non-Gaussian components referred to as tails.

The core of the response is mainly caused by the intrinsic resolution of the various sub-detector components and the precision of the PF and jet clustering algorithms. Since the jet momenta obtained from the PF constituents are based on tracking- as well as calorimeter-based measurements, the jet resolution is hence closely connected to the evolution of the tracking and calorimeter response with p_T . In the tracking system, the intrinsic resolution is mainly caused by the uncertainty on the track curvature. This is influenced by multiple scattering and the finite hit-position resolution. Typically, these effects cause a degradation of the track- p_T resolution proportional to the transverse momentum. However, the resolution in the calorimeters behaves the other way round and the resolution increases with increasing momentum. At low momenta, the resolution is mainly dominated by electronic noise and pile-up while at medium momenta the resolution is driven by fluctuations of the shower development. At high energies however, the resolution is eventually limited by calorimeter miscalibration and non-uniformities. Furthermore, the response also depends on the jet flavour. Typically, quark jets consist of less and harder particles than gluon jets and thus have a different response due to the non-linearity of the calorimeters.

The response tails are predominantly caused by severe jet-momentum mismeasurements. These can for instance be due to detector effects. Shower leakage can occur, if not the whole shower is deposited in the calorimeters. Due to the limited length of the calorimeters some particles might even cross the whole apparatus (*punch-through*). Also malfunction-

ing detector components can lead to high or low response tails by creating artificial signals. Beyond that also real physics processes lead to response tails. For instance the lower response tails get populated by semi-leptonic decays of heavy-flavour quarks. These contain neutrinos that carry a certain amount of the momentum and leave the detector unnoticed resulting in an overall reduced response.

5.2. Basic Concept of the Dijet Asymmetry Method

As introduced in Section 4.2.3 and 5.1 the jet transverse momentum resolution corresponds to the Gaussian-shaped core part of the jet response. In simulated events the particle-level jet momenta are given by the generator-level jet momenta. In data events however, no such equivalent is present so that the jet resolution is not accessible directly.

One possibility to measure the resolution of the jet transverse momenta in data as well as in simulated events is to utilize the dijet asymmetry A . For events with at least two jets it is defined as

$$A = \frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}}. \quad (5.2)$$

In this equation $p_{T,1}$ and $p_{T,2}$ correspond to the randomly ordered transverse momenta of the two leading jets.

For a sufficient number of events the asymmetry is approximately normally distributed and the standard deviation is given as

$$\sigma_A = \left| \frac{\partial A}{\partial p_{T,1}} \right| \cdot \sigma(p_{T,1}) \oplus \left| \frac{\partial A}{\partial p_{T,2}} \right| \cdot \sigma(p_{T,2}) \quad (5.3)$$

In an ideal dijet topology the two jets are exactly balanced at particle level. If they belong in addition to the same η region, then $\langle p_{T,1} \rangle = \langle p_{T,2} \rangle = \langle p_T \rangle$ and $\sigma(p_{T,1}) = \sigma(p_{T,2}) = \sigma(p_T)$. This allows the simplification of Eq. 5.3 and provides the following important relation between the width of the asymmetry σ_A and the jet- p_T resolution $\sigma(p_T)$

$$\frac{\sigma(p_T)}{\langle p_T \rangle} = \sqrt{2} \cdot \sigma_A \quad (5.4)$$

Already at the Tevatron experiments [137,138], the ATLAS experiment [139] or in previous CMS analyses as stated above this relationship was utilized to measure the jet resolution from dijet events.

5.3. Application to Realistic Collision Events

The measurement of the jet transverse momentum resolution in collision events is based on Eq. 5.4. As discussed in Section 4.2.3 the resolution is a function of p_T and η . Consequently, the asymmetry distribution is a function of p_T and η as well. Typically, the jet- p_T spectrum is affected by migration effects. Due to the limited resolution a particular interval of reconstructed jet momenta is populated not only by jets whose true momentum belongs to that bin, but also from jets outside. In case of a steeply falling spectrum, as this is the case for the jet momenta, more jets are supposed to migrate into a specific interval than out. Consequently, the measured response is systematically higher and the

measured relative response gets biased in favor of the object with the worse resolution. In order to reduce this resolution bias in the analysis, the measurement is performed in intervals of the average momentum of the two leading jets in the event

$$p_T^{\text{ave}} = \frac{1}{2}(p_{T,1} + p_{T,2}) . \quad (5.5)$$

Beyond that the ideal dijet topology with exactly two jets being perfectly balanced is interferred with additional effects in realistic collision events. Very often further jet activity is occuring as momentum of the hard scattering process is transferred to soft-particles or jets arising from initial or final state radiation leading to momentum imbalance in the event. This additional jet activity can in good approximation be described by the variable α which is defined as the ratio of the transverse momentum of the third jet to the average momentum according to

$$\alpha = \frac{p_{T,3}}{p_T^{\text{ave}}} . \quad (5.6)$$

The presence of further jets is neglected in the parametrization of the additional activity in the event, as these have consecutively less and less high momentum due to the strongly decreasing jet production cross section versus jet- p_T [140]. The presence of additional jets and the thereby introduced imbalance leads to a broadening of the observed asymmetry distribution. This effect is also illustrated later in Section 5.4.2. In order to determine the intrinsic resolution from such events the measured resolution has to be corrected for this bias.

A further source of momentum differences between the particle-level and the detector-level jet leading to an overall momentum imbalance in an event is arising from out-of-cone showering effects in the jet-clustering procedure. Typically, some particles might be too soft to be included in the clustered jet. Furthermore, additional contributions from pile-up or the underlying event might be wrongly associated to a jet. Such effects require a correction to the resolution measurement as well. The actual procedure how to determine and apply the required corrections is discussed in Section 5.6.

5.4. Samples and Event Selection

5.4.1. Datasets and Triggers

In this analysis, multijet events from pp collisions are considered which have been recorded in 2012 with the CMS detector at $\sqrt{s} = 8$ TeV. From these data only those are considered where all subdetectors have been reliably operating. The collected data sample used in this analysis corresponds to an integrated luminosity of 19.7 fb^{-1} with an uncertainty of 2.5% (syst.) + 0.5% (stat.) [141].

Multijet events are pre-selected by a set of triggers based on the average transverse momentum of the two leading jets in the event. In order to obtain a good coverage of the p_T^{ave} spectrum, different trigger paths are combined. Since they have different minimum p_T^{ave} thresholds, a broad range in p_T^{ave} is considered. In Tab. 5.1 the different trigger paths used for this analysis are listed and the particular offline p_T^{ave} value for which the respective trigger reaches 99% of the efficiency plateau is given (taken from [142]).

The simulated event sample is generated with PYTHIA, as discussed in Section 5.1 .

details
trigger selection
+ eff

Table 5.1.: Trigger paths with offline p_T^{ave} thresholds at which the trigger efficiency reaches the 99% efficiency plateau. Thresholds are given for PFCHS jets.

Trigger	p_T^{ave} threshold [GeV]
HltDiPFJetAve40	62
HltDiPFJetAve80	107
HltDiPFJetAve140	175
HltDiPFJetAve200	242
HltDiPFJetAve260	310
HltDiPFJetAve320	379
HltDiPFJetAve400	467

5.4.2. Selection Criteria

The physics objects used in the analysis are reconstructed with the PF algorithm including charged-hadron subtraction, as described in Sec. 4.1. Jets are clustered with the anti- k_T algorithm using a distance parameter of $R = 0.5$. They are calibrated up to residual correction factors for data.

The event selection described in the following is designed to enhance the dataset with events featuring a typical dijet-like event topology. This is characterized by two hard jets being back-to-back in the ϕ -plane. Thus, only events with at least two jets are considered for the analysis. These two leading jets in the event have to fulfill loose jet identification (*jet-id*) criteria which remove fake jets originating from detector noise while maintaining an efficiency of more than 99% for real jets [143, 144]. In order to mitigate effects from pile-up only jets with $p_T > 10 \text{ GeV}$ are considered for the analysis. As discussed in Section 5.3 events with a topology of exactly two high- p_T jets and no further jet activity are almost not existent in realistic collisions. Thus, events with additional jets have to be selected to perform the measurement. Since very soft jets do not necessarily have to belong to the hard interaction, but could arise from pile-up or the underlying event activity, it is required that each event has a third jet passing the p_T threshold of 01 GeV while fulfilling also loose jet identification criteria. Furthermore, the additional jet activity has to be restricted to a maximal amount in order to maintain a dijet-like structure of the event. Thus, a maximum threshold for the relative third jet momentum of

$$\alpha_{\max} < 0.25$$

is introduced.

In order to enrich the sample with events close to the ideal dijet topology where two jets point into opposite directions in the transverse plane and to reject events with very large

asymmetries the two leading jets have to fulfill

$$|\Delta\phi| > 2.7 \text{ with } \Delta\phi = \Delta\phi(p_{T,1}, p_{T,2}). \quad (5.7)$$

This requirement keeps most of the well-balanced events, but suppresses those which feature a strong imbalance.

The selection criteria described above are applied to data and simulation in an identical manner. However, a further adjustment is necessary for the simulated sample. Typically, the simulation is performed before the actual data-taking takes place. Thus, it is unknown which specific pile-up conditions will be present in data and the simulation is performed with an estimated pile-up scenario. Hence, it is necessary to adapt the observed pile-up scenario in data for the simulated events. This is done by reweighting the simulated events to match the pileup conditions in data. However, the pileup distribution in data differs for each individual trigger. Since the running conditions changed throughout the data taking in 2012, the pile-up conditions changed accordingly. The trigger paths utilized in this analysis have been mainly operated with pre-scale factors applied. In order to meet the changing running conditions over time, the pre-scales changed accordingly. Thus, each trigger path collected different amounts of data and consequently also the observed amount of pile-up differs per trigger. Hence, the reweighting of the pile-up scheme is done for each trigger path individually. Depending on the offline p_T^{ave} , a simulated event can be unambiguously assigned to the corresponding trigger path, which is fully efficient for that particular p_T range, and reweighted to that particular pile-up scenario. Since the number of primary vertices in the event is a measure for the pile-up activity, the success of the reweighting procedure can be checked by comparing the distribution of the number of primary vertices in data and simulation before and after reweighting, respectively. Such a comparison is shown in Fig. 5.2 for the trigger path with the highest p_T^{ave} threshold. The primary vertex distributions show a good agreement after the application of the reweighting procedure; especially in the bulk of the distribution. Corresponding distributions of other trigger paths used in the analysis are shown in Appendix A.1. The pileup weight is considered as a multiplicative factor for each simulated event.

In order to account for the dependence of the resolution on the transverse momentum and η , the asymmetry distributions are derived for various intervals of $|\eta|$ and p_T^{ave} . These are summarized in Tab. 5.2. The p_T^{ave} intervals are chosen in correspondence to the p_T^{ave} values for which the different triggers become fully efficient. If a trigger path provides enough statistical precision, the interval is separated into two. This definition of p_T^{ave} intervals ensures that all events in one p_T^{ave} interval are selected by the same trigger. The same interval boundaries are chosen for simulated events as well. The $|\eta|$ intervals are chosen to reflect the actual detector geometry. The most central part of the detector is covered by $|\eta| = 0.0 - 0.5, 0.5 - 1.1$ while the transition region, where the ECAL ends, is separated into an individual bin $|\eta| = 2.8 - 3.2$. The forward detector is summarized into one large bin $|\eta| = 3.2 - 5.0$ mainly due to the small number of available events. In order to allow the application of Eq. 5.4 for the determination of the resolution, both leading jets are required to belong to the same $|\eta|$ interval

$$|\eta_{\text{jet},1}| = |\eta_{\text{jet},2}|. \quad (5.8)$$

To account for the α -dependence of the measured asymmetry distributions, the p_T^{ave} and

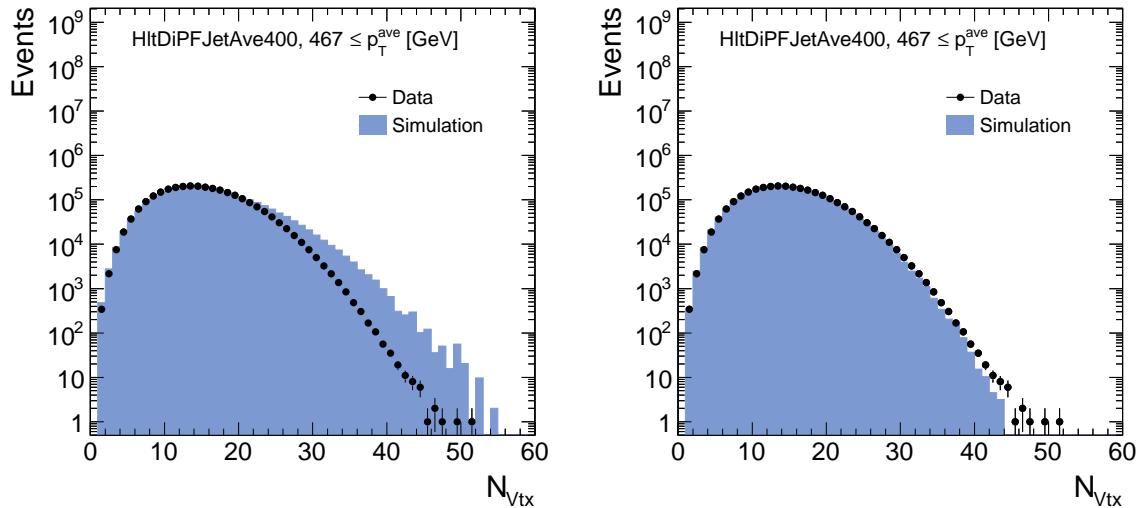


Figure 5.2.: Distribution of number of primary vertices in data (black dots) and simulation (blue histogram) before (*left*) and after (*right*) reweighting of the pile-up scenario in simulation for trigger path HltDiPFJetAve400.

Table 5.2.: Overview of the $|\eta|$ and p_T^{ave} interval boundaries used for the resolution measurement.

$ \eta $
0.0, 0.5, 1.1, 1.7, 2.3, 2.8, 3.2, 5.0
p_T^{ave} [GeV]
62, 107, 175, 205, 242, 270, 310, 335, 379, 410, 467, 600, 1000, 2000

$|\eta|$ intervals are further subdivided in various α intervals. These are chosen such that each α interval ranges from $\alpha = 0.0$ to a particular upper boundary α_{max} . The respective upper boundaries of the α_{max} intervals are 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25. This inclusive definition of the α intervals implicates that one specific event might be assigned to more than one α bin. It is discussed in Section 5.6.1 how this effect is treated during the analysis. The resulting inclusive p_T^{ave} spectrum after applying the the described selection is shown for data and simulationin Fig. 5.3. The p_T spectra of the first three leading jets are shown as well. The number of simulated events in each trigger p_T^{ave} interval is normalized to the respective integral in data. It is visible that the shapes of the spectra in data and simulation agree well. Furthermore, the effect of the pre-scales applied to the low- p_T^{ave} -threshold triggers is nicely visible resulting in the sawtooth-shaped in the p_T^{ave} -spectrum.

Finally, some exemplary asymmetry distributions obtained after the described selection are depicted in Fig. 5.4 for data and simulated events. They have been obtained using Eq. 5.2. Instead of randomly assigning the first and second jet, the absolute value of the asymmetry has been taken. For corrected jet momenta both definitions are equivalent and

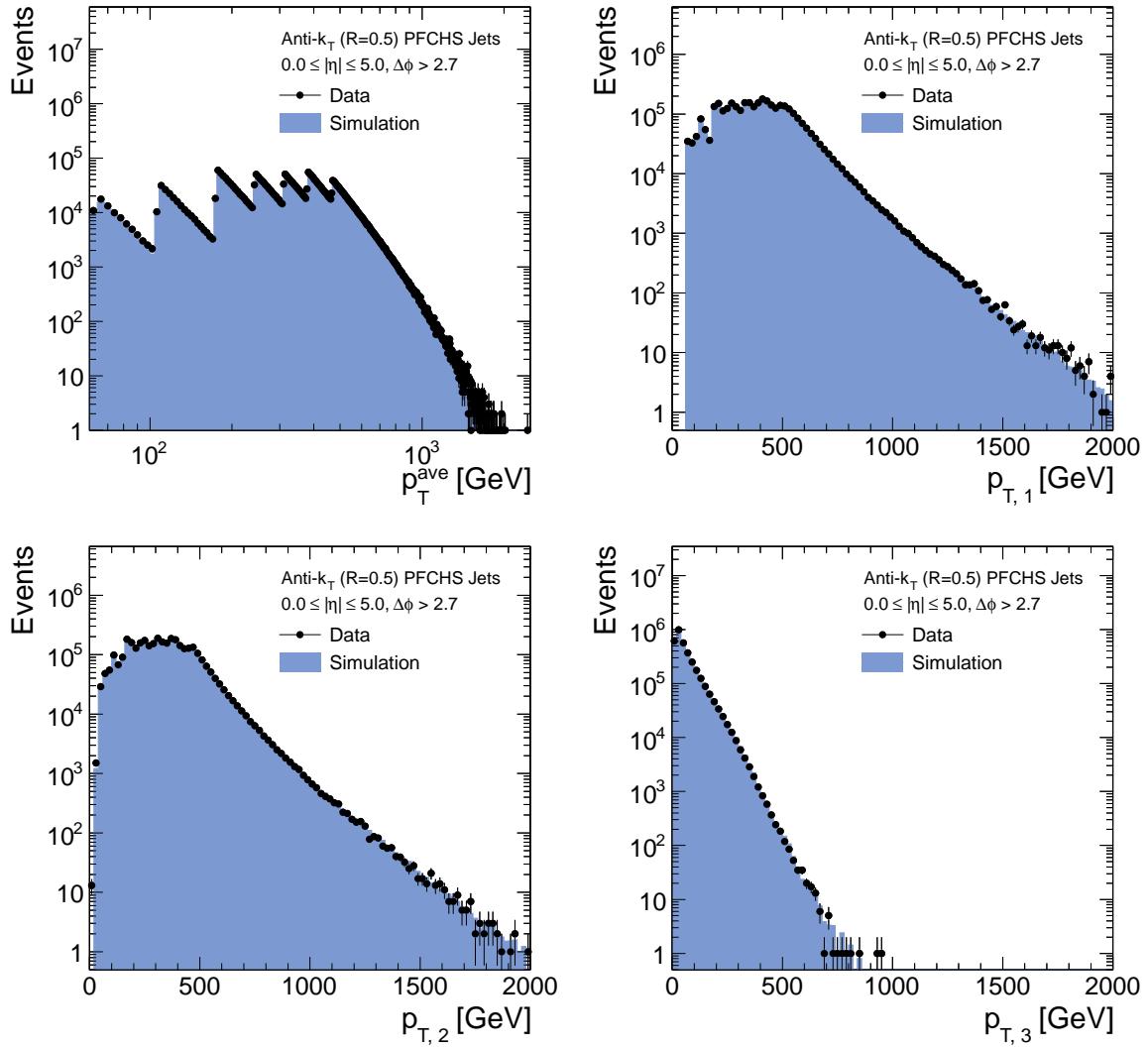


Figure 5.3.: Inclusive p_T^{ave} spectrum of events after the total selection in data and in simulation (top left), leading jet- p_T (top right), subleading jet- p_T (bottom left) and third jet- p_T (bottom right).

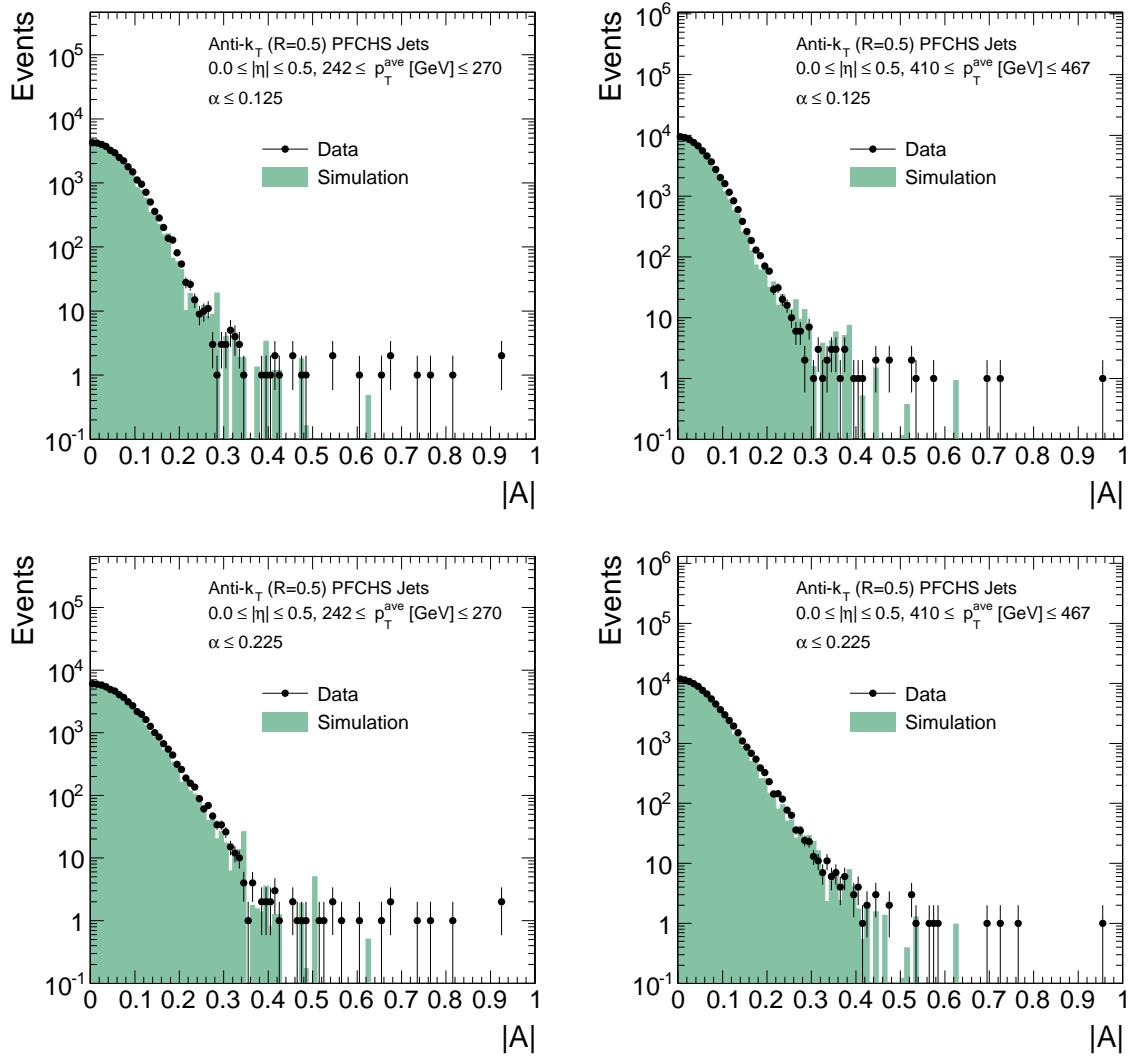


Figure 5.4.: Some example asymmetry distributions for the lowest $|\eta|$ region, two medium p_T^{ave} intervals and for a low (*top*) and a high (*bottom*) α interval.

the mean of the distribution is located at zero. The effect that the asymmetry distribution gets broadened for larger values of α is nicely visible when comparing the distributions in the top and the bottom row. Since the relation between resolution and width of the asymmetry distribution, as expressed in Eq. 5.4, is based on the assumption that the asymmetry is normally distributed, only asymmetry distributions with at least 100 events in the specified p_T^{ave}, η and α intervals are considered for the analysis.

5.5. Definition of the Asymmetry Width

As the measurement of the resolution in collision data is based on the width of the dijet asymmetry distribution, a proper definition of the asymmetry width is of major impor-

tance.

From the tails of the detector response events with large asymmetries can emerge leading to non-gaussian tails in the asymmetry distributions. These tails have to be rejected in the calculation of the asymmetry width in order to avoid a bias of the measurement. Hence, the asymmetry width has to be defined such that the central part of the distribution is described reasonably well by separating the core part from the tails. Since the core of the asymmetry distribution is expected to be Gaussian-shaped, a proper description of the asymmetry core can be tested by comparing the actual asymmetry histogram to a Gaussian function. The standard deviation of the Gaussian function can be chosen according to different definitions of the width of the asymmetry distribution and so the definition of the asymmetry width can be identified which features a good description of the asymmetry distribution by a corresponding Gaussian function.

The width of the asymmetry is determined by taking the whole distribution into account or by truncating a certain percentage of the tail regions. Thus, the asymmetry width is calculated as truncated root-mean-square

$$\sigma_A = \text{RMS}_{t\%} = \sqrt{\frac{1}{\sum_i y_i} \sum_i y_i \cdot A_i^2} \quad (5.9)$$

where y_i denotes the frequency of the asymmetry value A_i and the sum over i includes all values such that a percentage t of the total asymmetry distribution is covered. The statistical uncertainty is hence given by

$$\Delta\sigma_A = \Delta\text{RMS}_{t\%} = \frac{\text{RMS}_{t\%}}{\sqrt{2 \cdot n_{\text{eff}}}} \quad (5.10)$$

with the number of effective entries n_{eff} in the specified $(p_T^{\text{ave}}, |\eta|, \alpha)$ -interval. The truncation is chosen such that the whole distribution, 98.5% or only 95% are considered. A comparison of the asymmetry distributions and Gaussian functions where the standard deviation has been set to the value of the determined asymmetry width is illustrated in Fig. 5.5 for a certain $|\eta|$, p_T^{ave} and α interval. It is visible that the core part of the asymmetry distribution can in good approximation be described by a Gaussian function when choosing $\text{RMS}_{98.5\%}$ as the asymmetry width (cf. middle row in Fig. 5.5). Hence, this is the default definition of the asymmetry width used for this measurement. It is chosen to be the same for data events as well as for simulated events.

5.6. Corrections to the Dijet Asymmetry

The fundamental relation between the width of the asymmetry distribution and the jet energy resolution as expressed in Eq. 5.4 holds in this form only for the case of an ideal dijet topology. In real collision events various effects can occur that disturb the exact balance of the two jets as discussed in Section 5.3. Such effects can be soft radiation or additional jets originating from the hard scattering. They lead to momentum imbalance in the event and hence broaden the observed asymmetry distribution resulting in an increased measured resolution. In order to determine the intrinsic resolution, the measured asymmetry width has to be corrected for such effects. These corrections are explained in the following Sections 5.6.1 and 5.6.2.

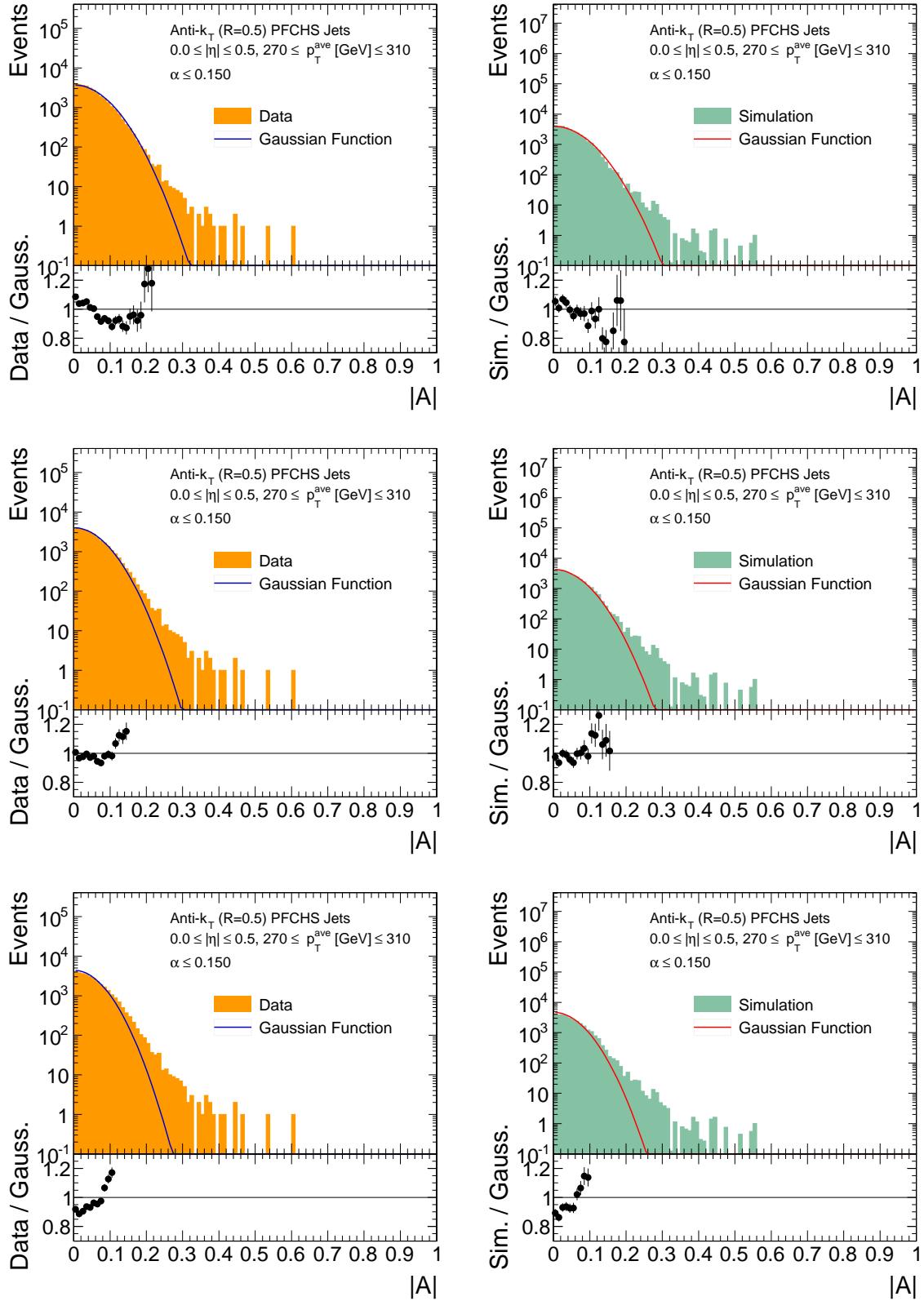


Figure 5.5.: Asymmetry distributions for data (*left*) and simulation (*right*) compared to gaussian distributions obtained with standard deviations corresponding to 0% truncation (*top*), 1.5% truncation (*middle*) and 5% truncation (*bottom*) of the tail regions.

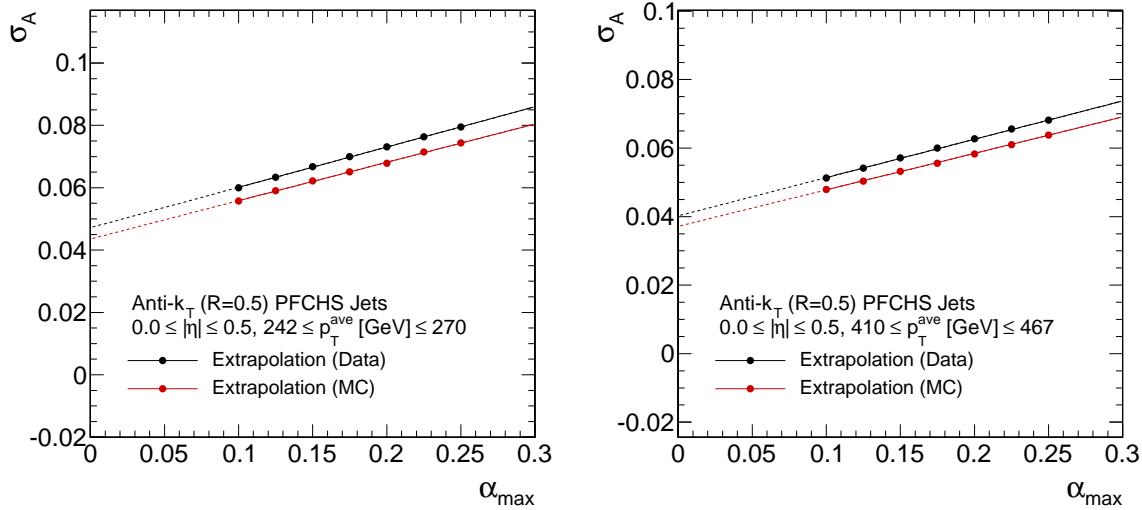


Figure 5.6.: Two exemplary extrapolations of measured values for σ_A in data (black) and simulation (red) to obtain the asymmetry for zero additional jet activity.

5.6.1. Correction for Additional Jet Activity

In addition to the two jets originating from the hard interaction, further jets can occur in an event from the emission of additional partons. The additional jet activity in the event is quantified by α introduced in Eq. 5.6 which is the fractional p_T of the third jet. The impact from jets beyond the third one is neglected since their impact is expected to be small due to the strongly decreasing jet-production cross-section.

The imbalance contribution arising from additional jet activity is taken care of by an extrapolation procedure. As described in Section 5.4.2 the asymmetry distribution is calculated for several intervals in $|\eta|$ and p_T^{ave} with different selections on the maximum value of α ($= \alpha_{max}$). For each of these individual selections the width of the asymmetry is determined as stated in Section 5.5. The measured values of $\sigma_A(\alpha_{max})$ are extrapolated to $\alpha_{max} \rightarrow 0$ assuming a linear behaviour. Thus, the measured asymmetry widths in one particular (p_T^{ave} , $|\eta|$, α)-interval are fitted with a linear function. The y-intercept of the fitted linear function represents the resolution without further jet activity in the event while the statistical uncertainty is given by the respective fit uncertainty for the intercept. The extrapolation procedure for two exemplary $|\eta|$ and p_T^{ave} intervals is illustrated in Fig. 5.6. The performed extrapolation fits for all other non-empty intervals are shown in Appendix A.2.

As stated in Section 5.4.2, the selection is performed in inclusive intervals of $\alpha < \alpha_{max}$. This results in a correlation of the measured values of σ_A for a particular (p_T^{ave} , $|\eta|$)-interval. In order to obtain a realistic estimate of the statistical uncertainty the correlation is propagated to the extrapolation fit. This approach is new with respect to previous analyses where such correlations have not been considered in the extrapolation [114, 130, 139].

In some more detail this means that the measured data points are described by a linear function

$$a \cdot \alpha_{max} + b = \sigma_A(\alpha_{max}) \quad (5.11)$$

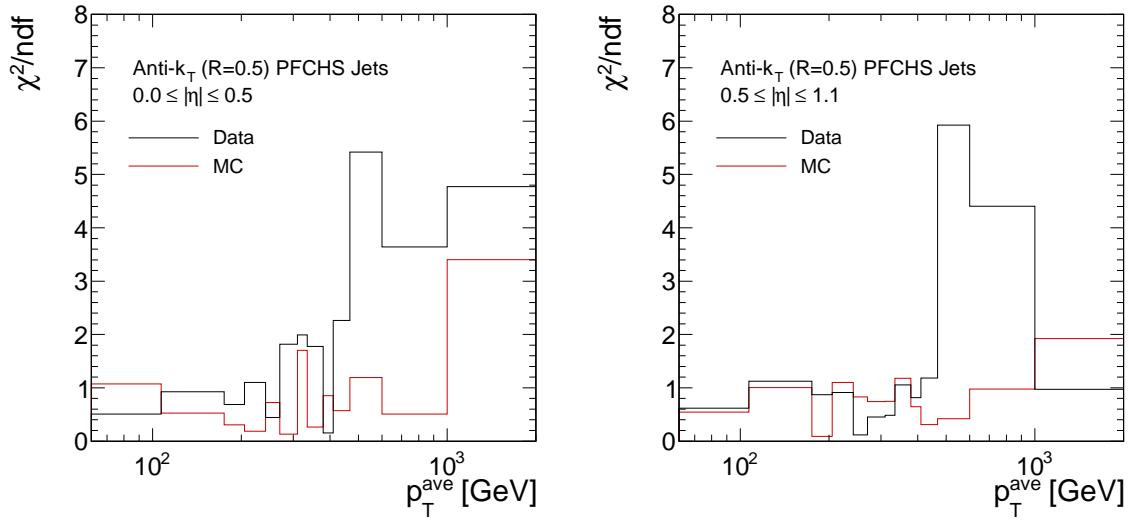


Figure 5.7.: Goodness-of-fit test for the extrapolation fits in data and simulation in the two central $|\eta|$ regions as function of p_T^{ave} .

by determining the parameters a and b . This is done for known α_{max} and $\sigma_A(\alpha_{max})$ by minimizing

$$\chi^2 = dy^T \cdot C^{-1} \cdot dy \quad (5.12)$$

where $dy = y_{\text{measured}} - y_{\text{predicted}}$ and C denotes the covariance matrix. Consequently, the asymmetry width for vanishing additional jet activity is given by $b = \sigma_A(\alpha_{max} \rightarrow 0)$. Under the assumption that all events that belong to α interval i are also completely included in the next higher α interval j ¹ the covariance matrix is given by

$$C_{i,j}(\sigma_{A_i}, \sigma_{A_j}) = (\Delta\sigma_{A_i})^2 \cdot \frac{\sigma_{A_i}}{\sigma_{A_j}} \cdot \frac{n_i}{n_j} \quad (5.13)$$

where n is the number of events in that particular α interval. A discussion regarding the derivation of that expression is summarized in Appendix A.3 [145]. The function minimization itself is performed with the Minuit package [146].

In Fig. 5.7 an overview of a goodness-of-fit test is shown for the extrapolation fits in data and simulation. The resulting χ^2 distributions over the number of degrees of freedom are shown as function of p_T^{ave} exemplary for the two central $|\eta|$ intervals. This is expected to be distributed around 1. Consequently, the fit quality is in general quite good but worsens for larger p_T^{ave} values. Since especially for higher p_T^{ave} intervals in the central $|\eta|$ regions the statistical uncertainty is rather low, the fit is very sensitive to even small deviations from a linear behaviour which results in a minor fit quality. However, a possible non-linearity is considered in the systematic uncertainties of the measurement as discussed in Section 5.9.

¹This assumption is only almost true since the asymmetry distributions are truncated to reject non-Gaussian components and some events might not fulfill this criterium.

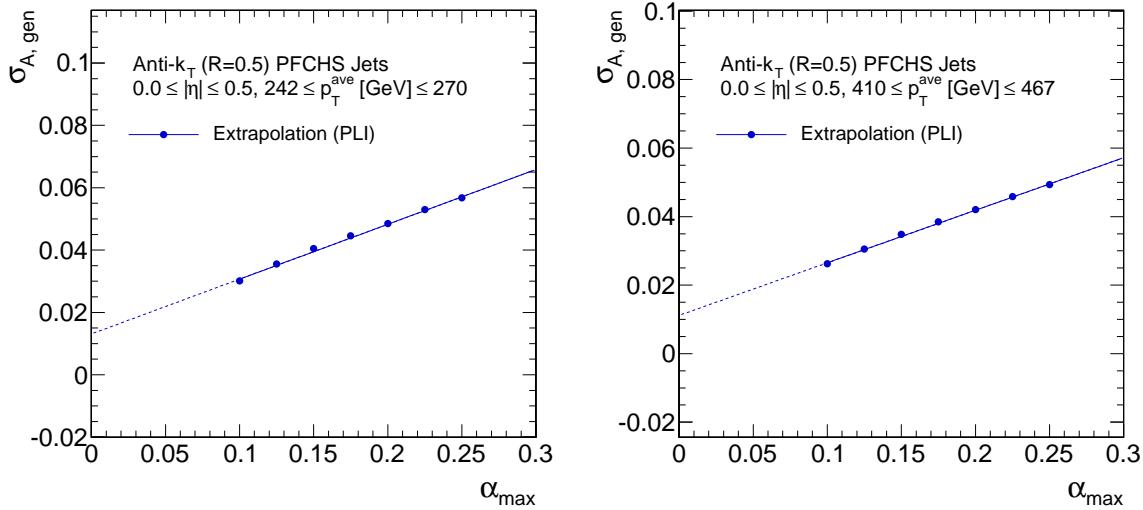


Figure 5.8.: Two exemplary extrapolations of measured values for $\sigma_{A,\text{gen}}$ in simulation to determine the contribution from PLI.

5.6.2. Correction for Particle-Level Imbalance

In addition to an imbalance in dijet events caused by the presence of additional jets, an imbalance in the dijet system at particle level can also arise from the jet clustering process through out-of-cone showering. This additional imbalance contribution is referred to as *particle-level imbalance* (PLI) and estimated from simulation.

The contribution arising from the particle-level imbalance is estimated from the dijet asymmetry defined at generator level. This is defined equivalently to the asymmetry at detector level but based on generator-level jet quantities as

$$A_{\text{gen}} = \frac{p_{T,1}^{\text{gen}} - p_{T,2}^{\text{gen}}}{p_{T,1}^{\text{gen}} + p_{T,2}^{\text{gen}}} \quad (5.14)$$

where $p_{T,1}^{\text{gen}}$ and $p_{T,2}^{\text{gen}}$ refer to the momenta of the two leading generated jets. This distribution is affected by additional parton radiation as well. Thus, the procedure to obtain $\sigma_{A,\text{gen}}(\alpha_{\text{max}} \rightarrow 0)$ is the same as for the asymmetry at detector level. The generator asymmetry is calculated in the same $(p_T^{\text{ave}}, |\eta|, \alpha)$ -intervals as the detector-level asymmetry in order to derive the size of the particle-level imbalance for exactly the same events. The asymmetry width of the generator asymmetry is again calculated as RMS_{98.5%}. In order to obtain the values of $\sigma_{A,\text{gen}}$ for zero additional jet activity, the analogue extrapolation procedure is performed, as described in Section 5.6.1. Some exemplary extrapolations are illustrated in Fig. 5.8.

Finally, the results of the extrapolated values for $\sigma_{A,\text{gen}}$ can be used to quantify the size of the particle-level imbalance. This is given by

$$\sigma_{\text{PLI}} = \sqrt{2} \cdot \sigma_{A_{\text{gen}}}(\alpha_{\text{max}} \rightarrow 0). \quad (5.15)$$

5.6.3. Results of the Corrections to the Asymmetry

The results of the various extrapolation fits to determine $\sigma_A(\alpha_{max} \rightarrow 0)$ in data and simulation as well as for the particle-level imbalance are shown together in Fig. 5.9 as function of p_T^{ave} for the different $|\eta|$ regions.

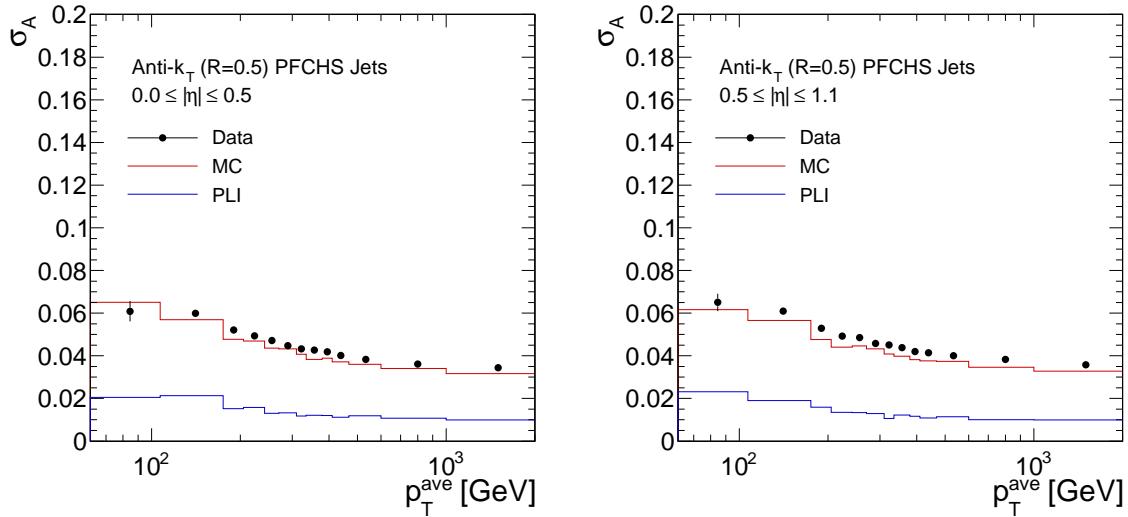


Figure 5.9.: Results of extrapolation fits in the two central $|\eta|$ regions as function of p_T^{ave} for data, simulation and particle-level imbalance.

In order to obtain the jet energy resolution the results of the extrapolated detector-level asymmetry widths are corrected for the effects from particle-level imbalance. This is done by subtracting the PLI correction in quadrature from the measured σ_A after the correction for additional jet activity

$$\sigma_{JER} = \sqrt{2} \cdot \sigma_A(\alpha_{max} \rightarrow 0) \ominus \sigma_{PLI}. \quad (5.16)$$

The same correction for PLI is subtracted both from data as well as from simulation results. As can be seen in Fig. 5.9 the correction due to particle-level imbalance is small compared to the measured asymmetry widths. This justifies to take this correction from simulation also for the data resolution. However, in order to account for a possible imprecise modelling of the particle-level imbalance contribution a systematic uncertainty will be considered, as discussed in Section 5.9.

5.7. Determination of the Data-to-Simulation Ratio of the Jet Transverse Momentum Resolution

The measured resolution in data and simulation can be compared after applying all corrections described in Section 5.6 by calculating the data-to-simulation ratio

$$c(\text{Data/MC}) = \frac{\sigma_{JER, \text{Data}}}{\sigma_{JER, \text{MC}}}. \quad (5.17)$$

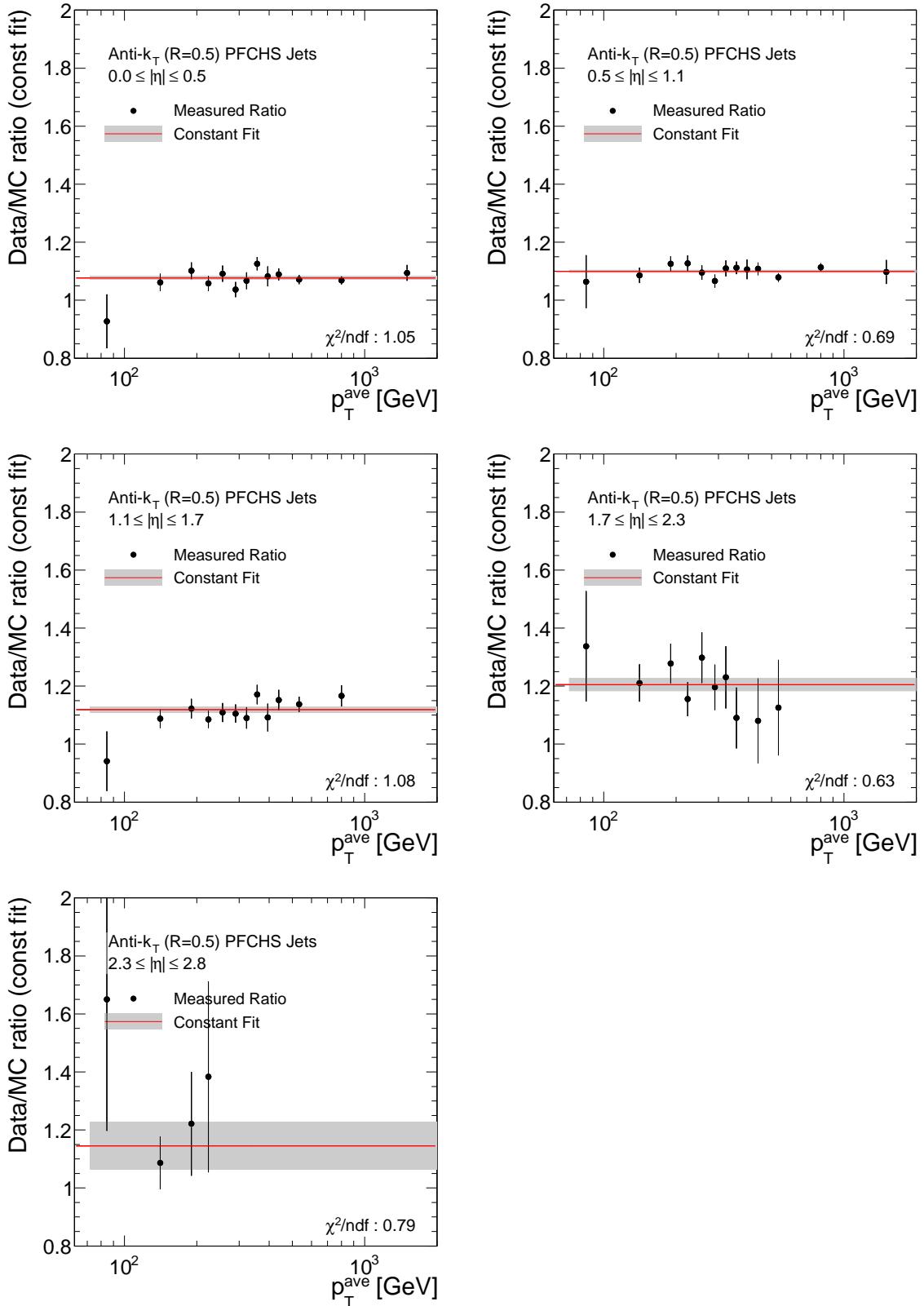


Figure 5.10.: Ratio of the resolution in data to the resolution in the QCD sample in non-empty $|\eta|$ regions as function of p_T^{ave} shown together with a constant fit (red line) and corresponding statistical uncertainty (grey shaded area).

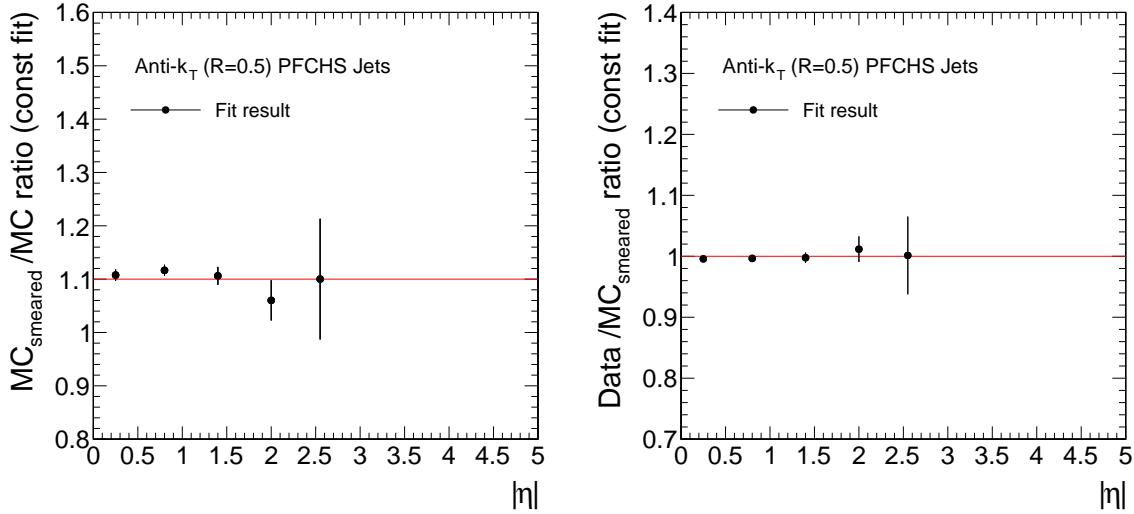


Figure 5.11.: Ratio $c(\text{MC}_{\text{smeared}}/\text{MC})$ of the resolution in the smeared QCD sample to the resolution of the unsmeared QCD sample as function of $|\eta|$ with statistical uncertainty (*left*) and ratio $c(\text{Data}/\text{MC}_{\text{smeared}})$ of the resolution in data to the resolution of the QCD sample smeared with the measured data-to-simulation ratio as function of $|\eta|$ with statistical uncertainty (*right*). The red line indicates the target value of each validation test.

The resulting distributions as function of $p_{\text{T}}^{\text{ave}}$ for the different η regions are shown in Fig. 5.10. As no significant p_{T} dependence is observed, the ratio is parametrized by a constant fit. The fit result shows in each $|\eta|$ region a value larger than one meaning that the resolution in data is in general worse than in simulation. The constant fit is also visualized in Fig. 5.10 (red line) together with the statistical uncertainty resulting from the fit (grey shaded area). Since the χ^2/ndf values resulting from the fit are nicely in agreement with one, the description with a constant is justified. However, a potential hidden trend as function of $p_{\text{T}}^{\text{ave}}$ is considered in the systematic uncertainties of the measurement. Results are summarized later showing also systematic uncertainties in Tab 5.3.

Due to statistical limitations no data-to-simulation ratio could be determined for the two highest $|\eta|$ intervals. As jets in the forward region of the detector have essentially low transverse momentum, these $|\eta|$ intervals are mainly affected by the high pre-scales of the triggers with low momentum thresholds.

5.8. Validation of the Method

5.8.1. Validation in Simulated Events

In order to test the quality of the method to predict the correct data-to-simulation ratio a validation test on simulated events is performed.

In this test, the simulated PYTHIA QCD sample is divided into two sub-samples with

equal number of events. In one of the sub-samples the leading corrected jets² in p_T in each event are smeared with a smearing factor c which increases the p_T resolution. This is done for jets which pass a minimum p_T threshold of 10 GeV. More precisely this means that for each reconstructed jet that has a corresponding generated jet within a $\Delta R < 0.25$ cone, the transverse momentum is scaled according to

$$p'_T = p_T^{gen} + c \cdot (p_T - p_T^{gen}). \quad (5.18)$$

resulting in the smeared momentum p'_T . After the smearing procedure the jet momenta are re-ordered again descendent in p_T .

For this test, the smearing factor has been chosen to be $c = 1.1$ for each $|\eta|$ interval. This is expected to be the resulting ratio when determining the ratio of the resolution in the smeared QCD sub-sample to the resolution in the unsmeared sub-sample derived with the asymmetry method including all corrections as described in Section 5.6. Consequently, the test is passed when the measured ratio recovers the input value according to $c(\text{MC}_{\text{smeared}}/\text{MC}) = 1.1$. The resulting values obtained from a constant fit of the ratios as function of p_T^{ave} are shown in each $|\eta|$ region with statistical uncertainties in Fig. 5.11 (left). It can be seen that the input scaling factor of 1.1 is nicely recovered with this procedure within the statistical uncertainties. Thus, a potential resolution difference in the ratio when comparing data and simulation can be determined with this approach as well.

5.8.2. Validation of the Measured Data-to-Simulation Ratio

In addition to the validation test on simulated events, also the measured data-to-simulation ratio can be directly validated. This is done with a similar approach as described in Section 5.8.1.

For the test of the measured data-to-simulation ratio the individual jet momenta in the simulated PYTHIA QCD sample are scaled with the measured resolution ratio values $c(\text{Data}/\text{MC})$ according to Eq. 5.18. Thus, jets are corrected with different scale factors depending on the respective $|\eta|$ region they belong to. Afterwards, the whole measurement procedure of the data-to-simulation including all corrections for further jet activity and PLI after the smearing procedure of the jet momenta is performed. Since the measured resolution differences between data and simulation have been compensated before the actual re-determination of the data-to-simulation ratio, it is expected to obtain ratios of $c(\text{Data}/\text{MC}_{\text{smeared}}) = 1$. The result of this test is illustrated in Fig. 5.11 (right) with statistical errors. A nice agreement with the expected value of 1 is visible even without the consideration of systematic uncertainties in this consistency test. This again proves that the dijet asymmetry method is well suited to determine resolution differences in the data-to-simulation ratio.

5.9. Systematic Uncertainties

In addition to the uncertainty of the resolution measurement due to statistical effects, there are also systematic components which influence the outcome of the data-to-simulation

²It is necessary to correct all jets which might get relevant for the analysis i.e. become one of the leading three jets. In this analysis the five leading jets in p_T are corrected.

ratio. The different contributions of systematic uncertainties to the measurement are discussed in this section.

All systematic uncertainties are determined by evaluating the shift of the data-to-simulation ratio when varying a certain aspect in the measurement procedure. The uncertainty δc is calculated by the determination of the ratio for a certain shift Δ and comparing it to the nominal ratio

$$\delta c(\text{Data}/\text{MC}) = c(\text{Data}/\text{MC})_{\Delta} - c(\text{Data}/\text{MC}) \quad (5.19)$$

Thus, all systematic uncertainties are determined as absolute shift of the nominal ratio. Typically, an upward and downward shift is evaluated. Resulting uncertainties are afterwards symmetrized by taking the average shift. If however the up- and downward variation both result in either an upward or downward shift, the absolute shift is determined and the maximum of both is taken and quoted as symmetric uncertainty. In case only an up- or downward variation is performed, the systematic uncertainty however is taken as up- and downward shift.

PU reweighting: The trigger dependent pile-up distributions in data which are taken in order to reweight the simulated sample to match the observed pile-up distribution in data are calculated with a nominal minimum bias cross section of 69.4 mb. In order to propagate the uncertainty on the minimum bias cross section to the data-to-simulation ratio, it is increased to 73.5 mb. Hence, the pile-up scenario of the simulated PYTHIA QCD sample is reweighted to the data distributions obtained with this varied minimum bias cross section. Apart from that variation the measurement of the data-to-simulation ratio is performed as for the nominal ratio.

Although the minimum bias cross section is only shifted upwards, the resulting shift in the data-to-simulation ratio is considered as symmetric uncertainty.

Particle-level imbalance: The measured resolution in data and simulation is corrected for an imbalance at particle level due to out-of-cone showering effects based on simulation. In order to account for the uncertainty on σ_{PLI} the PLI-correction factor for each measured resolution value is shifted by $\pm 25\%$. Consequently, the changed ratio is calculated as

$$c(\text{Data}/\text{MC})_{\text{PLI}} = \frac{\sqrt{2} \cdot \sigma_{A,\text{Data}}(\alpha_{max} \rightarrow 0) \ominus f \cdot \sigma_{\text{PLI}}}{\sqrt{2} \cdot \sigma_{A,\text{MC}}(\alpha_{max} \rightarrow 0) \ominus f \cdot \sigma_{\text{PLI}}} \quad (5.20)$$

with $f = 0.75, 1.25$ respectively. The size of this variation is chosen by comparing the size of the PLI correction in simulated events from the nominal PYTHIA sample to the PLI correction estimated from simulated events obtained by the HERWIG++ generator with tune EE3C. This comparison is illustrated in Fig. 5.12. Since the size of the PLI correction agrees within 25% between both generators, the size of this uncertainty is justified.

Jet energy scale: The jet energy scale has been corrected to particle level by the application of dedicated calibration factors which have a certain uncertainty assigned. In order to propagate this uncertainty to the data-to-simulation ratio of the resolution all jet momenta in the simulated sample are shifted up and down by the JEC uncertainty. The jet momenta in data stay unchanged. Afterwards the data-to-simulation is determined again based on the varied jet momenta.

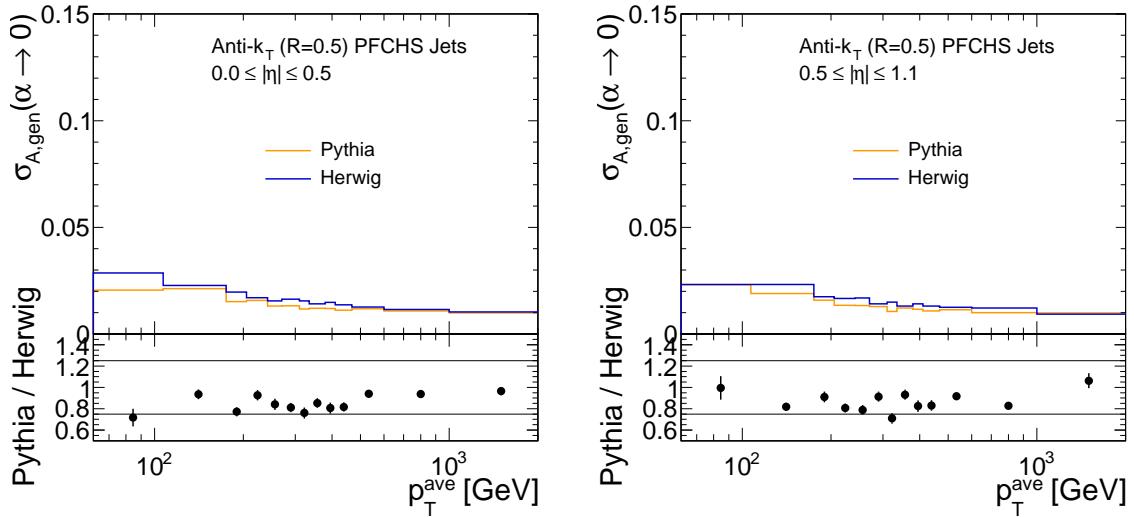


Figure 5.12.: Comparison of the contributions to the asymmetry width due to particle-level imbalance in QCD multijet events generated with PYTHIA (yellow) to the same quantity derived with the HERWIG++ generator (blue) in the two central $|\eta|$ regions as function of p_T^{ave} .

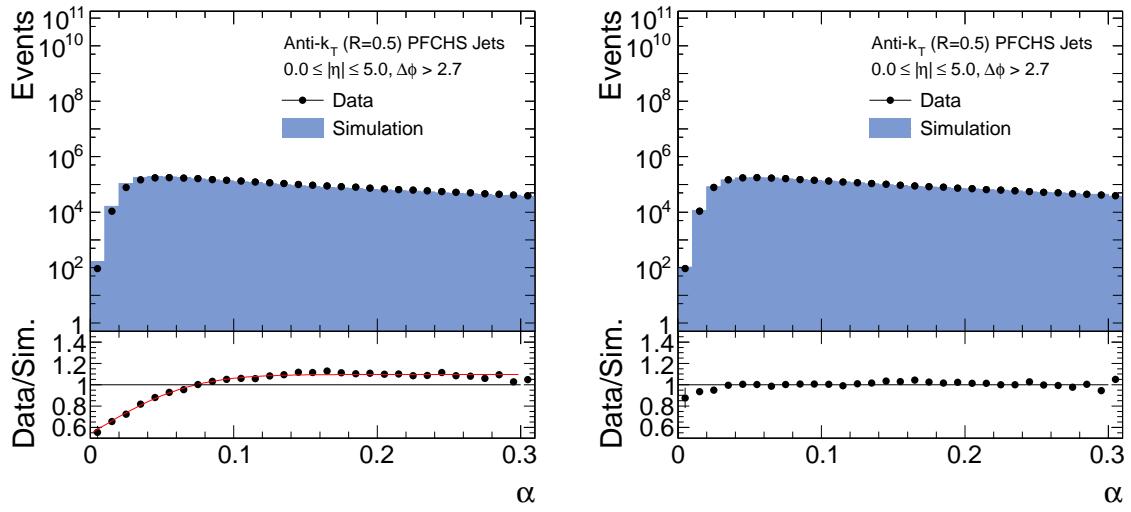


Figure 5.13.: Inclusive α -spectrum before (*left*) and after (*right*) reweighting the α -spectrum of the simulated sample. The red curve in the bottom of the left plot illustrates the function used for the α -spectrum reweighting.

α -spectrum: To account for additional jet activity in the event the measured widths of the asymmetry distributions are extrapolated to zero additional jet activity by a linear function. This linear behaviour is an empirically found relation rather than a theoretically fundamental connection. One influencing factor of the linear behaviour is the functional form of the α -spectrum. It determines how many events are added to the asymmetry distribution in the next higher α interval and consequently how much the asymmetry distribution is broadened. This can result e.g. in a smaller or bigger slope of the linear function or result in a non-linear behaviour. However, such effects might cancel in the ratio as long as the α -spectrum is the same in data and simulation.

The observed inclusive α -spectrum in the simulated sample is compared to the spectrum in data and shown in Fig. 5.13 (left). The bottom part displays the ratio Data/MC. Since the ratio shows that both spectra do not agree the influence of the α -spectrum on the data-to-simulation ratio is evaluated by reweighting the α -spectrum to roughly match the one observed in data. The red curve overlaid in the ratio of the left distribution is used to reweight the events in the simulation. For each event a weight $w(\alpha)$ is calculated according to

$$w(\alpha) = 0.545 \cdot (\text{erf}(13.5 \cdot \alpha - 0.02) + 1) \quad (5.21)$$

where erf indicates the error-function³. This weight is considered as multiplicative factor onto the usual event weight in simulation. For comparison, the reweighted α -spectrum is shown on the right of Fig. 5.13.

α -range: In addition to the functional form of the α -spectrum also the specific range which has been chosen for α can result in a more or less linear behaviour of the measured asymmetry widths. The choice of the linear function implies that the linear behaviour holds also for small α -values. However, especially for small values of α the linear behaviour could not be tested explicitly due to the imposed minimum p_T threshold of 10 GeV for the third jet.

In order to study the linear behaviour of the extrapolation also towards smaller α -values, the minimum p_T cut of 10 GeV for the third jet is dropped and an additional α -interval 0.0 – 0.05 is introduced for the measured asymmetries at detector level as well as the PLI correction. In Fig. 5.14 some exemplary extrapolations of the asymmetry widths in data and simulation with the additional α -interval are shown. The resulting difference to the nominal data-to-simulation ratio is considered as systematic uncertainty.

Non-gaussian tails: As discussed in Sec. 5.5, the width of the asymmetry distribution is calculated as a truncated root-mean-square in order to reject contributions from non-gaussian tails. The truncation was chosen to be 1.5% for both data and simulation. In general it is possible, that the tail contributions in data and simulation differ and consequently do not cancel out in the ratio. In order to evaluate this effect the data-to-simulation ratio is calculated when truncating 5% of the original distribution instead of 1.5%.

³ $\text{erf} = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt$

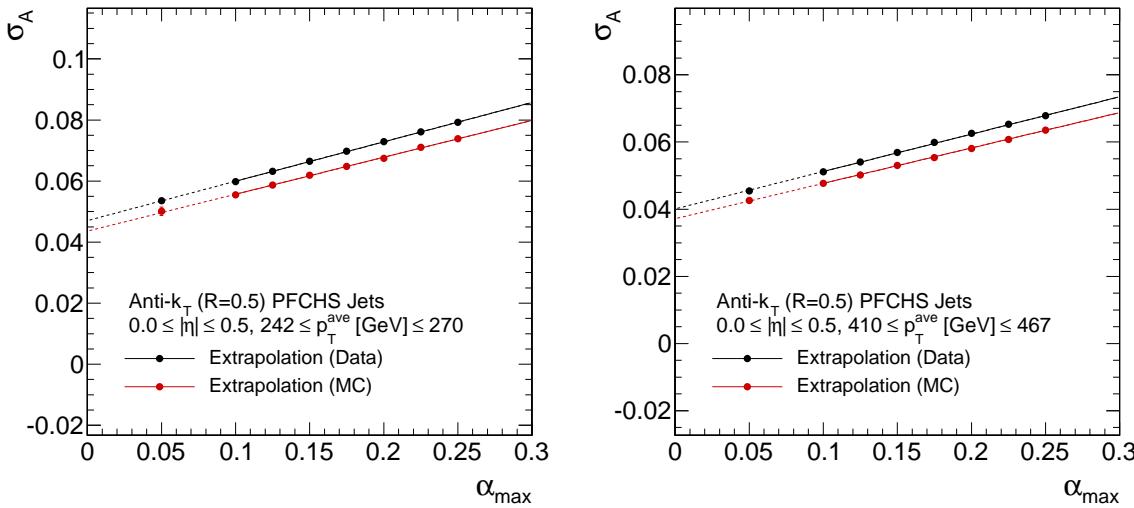


Figure 5.14.: Two exemplary extrapolations of measured values for σ_A in data and simulation to obtain the result at zero additional jet activity when adding one additional α interval 0.0 – 0.05.

Flavour uncertainty: As discussed in Sec. 5.1, the jet response can be quite different for light (u,d,s) and heavy quarks (c,b). If the flavour composition is the same in data and simulation, this flavour difference cancels out in the data-to-simulation ratio. Since it is known that especially the rate of gluon splitting events with $g \rightarrow b\bar{b}$ is off by up to about a factor of two [147], the impact of heavy quarks produced in gluon splitting processes on the data-to-simulation ratio is evaluated by varying the event weights for simulated events with gluon splitting into heavy quarks.

An event is considered as gluon splitting event, if one of the two leading jets undergoes a gluon splitting identified by utilizing generator truth information. All events identified as gluon splitting into heavy quarks get an event weight of 1.5. Afterwards the data-to-simulation ratio is derived as the resolution from data to the resolution in the reweighted simulated sample.

Shape of the data-to-simulation ratio: The data-to-simulation ratio in this measurement is determined by fitting the ratio of $\sigma_{\text{JER, data}}$ to $\sigma_{\text{JER, simulation}}$ with a constant. This assumes that the ratio is flat as function of p_T^{ave} . In order to test this assumption one can fit the $\sigma_{\text{JER, data}}$ and $\sigma_{\text{JER, simulation}}$ distributions with a more model dependent approach.

The following function, denoted *NSC-function*, is used to fit the resolution in simulation

$$f(p_T) = \sqrt{\frac{N^2}{p_T^2} + \frac{S^2}{p_T} + C^2} \quad (5.22)$$

with $p_T = p_T^{\text{ave}}$ and the free parameters N , S and C . This function is chosen in accordance to the parametrization typically used for the description of the relative energy-resolution in calorimeters [1]. At low momenta, the resolution is mainly dominated by electronic noise and pile-up described by the *noise term* N . For increasing

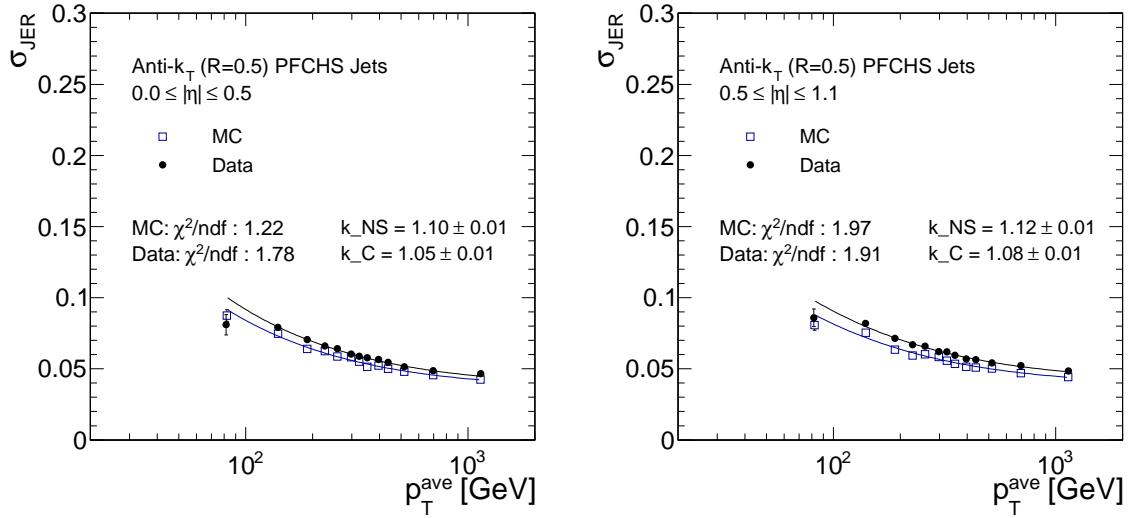


Figure 5.15.: Results of fitting the resolution in data and simulation with the respective NSC-functions in the two central $|\eta|$ interval.

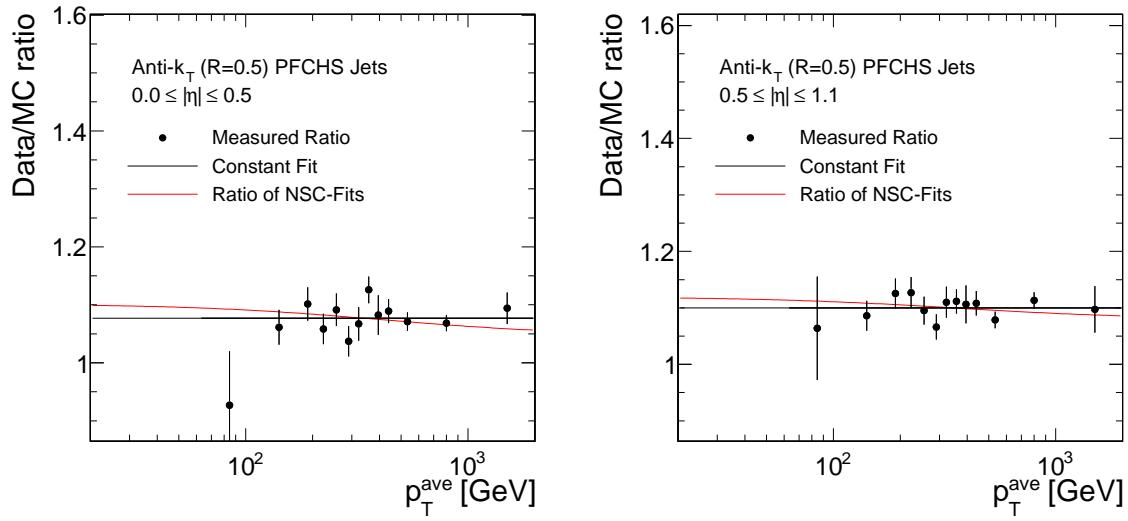


Figure 5.16.: Ratio of the results fitting the resolution in data and simulation with the NSC-functions in the two central $|\eta|$ interval together with the measured ratio and a constant fit.

energies, the resolution is driven by fluctuations of the shower development described by the *stochastic term* S . At high energies however, the resolution is eventually limited by calorimeter miscalibration and non-uniformities described by the *constant term* C . If the data-to-simulation ratio is flat as function of p_T^{ave} one expects to find one common scale factor k_{NSC} for the N , S and C parameters to describe the data. Significant differences between data and the actual modelling of the detector behaviour consequently can result in the necessity to use individual scale factors for the k_N , k_S and k_C parameters in order to describe the data from the fit results of the function to the simulation.

In order to test the assumption of a flat ratio the data points are fitted with the following function

$$f(p_T) = \sqrt{\frac{(k_{NS} \cdot N)^2}{p_T^2} + \frac{(k_{NS} \cdot S)^2}{p_T} + (k_C \cdot C)^2} \quad (5.23)$$

which uses one common scale factor for parameters N and S and another one for C . Since there are no events with low transverse momenta available, no sensitivity to the special characteristics of N is expected and hence no separate scale factor for N is defined. Furthermore this means, that this procedure can not be used to study influences from pile-up. For the determination of k_{NS} and k_C the parameters N , S and C are fixed to the results of the fit to the simulation. Some example fits for the two most central $|\eta|$ intervals are summarized in Fig. 5.15. It turns out that typically the scale factors k_{NS} and k_C are not equal and thus a shape different from a constant could describe the measured data-to-simulation ratios as function of p_T^{ave} as well. In order to illustrate how big this effect is, the ratio of the two fitted NSC-functions for data and simulation is shown together with the measured values and the constant fit in Fig. 5.16. It is visible that the ratio derived from the NSC-fits is also compatible with the measured values for the ratio. Consequently, a trend versus p_T^{ave} could be hidden in the measured points. The difference of the measured scale factors k_{NS} and k_C to the central value of the measurement from the constant fit amounts to about 2% in the two central $|\eta|$ intervals. This is considered as systematic uncertainty for all $|\eta|$ intervals in order to compensate for this potentially hidden shape in the data-to-simulation ratio. Furthermore, this also covers possible differences of the scaling factors for jets with very low or very high transverse momenta which can not be directly measured in this analysis. Overall, this systematic uncertainty constitutes the major contribution to the total systematic uncertainty.

A summary of the data-to-simulation ratios measured with the dijet asymmetry method requiring that the two leading jets belong to the same $|\eta|$ interval together with all systematic uncertainties considered for the ratio is given in Tab 5.3. Due to limited event numbers no results could be determined for the two highest $|\eta|$ regions with this approach. In general, the measurement is limited by systematic effects which amount to uncertainties of about 2 – 2.5% in the central detector part. However, in the $|\eta| = 2.3 – 2.8$ region the measurement is limited by the statistical uncertainty.

Table 5.3.: Summary of the measured data-to-simulation ratios $c(\text{Data}/\text{MC})$ with statistical uncertainty and systematic uncertainty for each uncertainty source in different $|\eta|$ regions up to $|\eta| = 2.8$.

	$ \eta $				
	0.0 – 0.5	0.5 – 1.1	1.1 – 1.7	1.7 – 2.3	2.3 – 2.8
$c(\text{Data}/\text{MC})$	1.077	1.100	1.119	1.205	1.145
Stat. uncertainty	± 0.007	± 0.006	± 0.010	± 0.027	± 0.078
PU reweighting	± 0.004	± 0.003	± 0.004	± 0.004	± 0.013
Particle-level imbalance	± 0.004	± 0.005	± 0.005	± 0.015	± 0.010
Jet energy scale	± 0.005	± 0.007	± 0.008	± 0.015	± 0.020
α -spectrum	± 0.004	± 0.007	± 0.004	± 0.006	± 0.003
α -range	± 0.005	± 0.009	± 0.004	± 0.011	± 0.012
Non-gaussian tails	± 0.004	± 0.003	± 0.003	± 0.013	± 0.007
Jet Flavour	± 0.007	± 0.004	± 0.005	± 0.006	± 0.003
Ratio shape	± 0.022	± 0.022	± 0.022	± 0.024	± 0.023
Total syst. uncertainty	± 0.025	± 0.027	± 0.026	± 0.038	± 0.038

5.10. Extension of the Method to the Forward Detector Region

The presented measurement with results summarized in Tab. 5.3 is utilizing the dijet asymmetry method based on the requirement that the two leading jets in an event belong both to the same $|\eta|$ region. Consequently, the resolution of these two jets is equal which results in the simple relation expressed in Eq. 5.2 between asymmetry width and jet resolution. However, the requirement that both leading jets have to belong to the same $|\eta|$ interval reduces significantly the available number of events. Especially in the forward region of the detector this is a problem as the jets in these events have low transverse momenta and consequently have to be triggered by the lowest p_T^{ave} triggers which are highly prescaled. Hence, the available number of events is small anyway and for the intervals $|\eta| > 2.8$ no data-to-simulation ratio could be determined implying the same- η requirement.

In order to extend the analysis such that the data-to-simulation ratio can also be measured in the forward region of the detector the requirement that $|\eta_{\text{jet},1}| = |\eta_{\text{jet},2}|$ has to be dropped. Instead events are selected where the two leading jets belong to different pseudorapidity regions. However, the resolution $\sigma(p_T^{\text{jet,probe}})$ in a probe interval $|\eta_{\text{probe}}|$ can be determined, if the resolution $\sigma(p_T^{\text{jet,ref}})$ in a reference interval $|\eta_{\text{ref}}|$ is known.

Based on Eq. 5.3, it can be shown that for $\langle p_{T,\text{probe}} \rangle = \langle p_{T,\text{ref}} \rangle = \langle p_T \rangle$ and $\sigma(p_{T,\text{probe}}) \neq$

$\sigma(p_{T,\text{ref}})$ the relation

$$\frac{\sigma(p_{T,\text{probe}})}{\langle p_T \rangle} = \sqrt{4 \cdot \sigma_{A(|\eta_{\text{probe}}| \neq |\eta_{\text{ref}}|)} - \left(\frac{\sigma(p_{T,\text{ref}})}{\langle p_T \rangle} \right)^2} \quad (5.24)$$

is obtained. Here, $\sigma_{A(|\eta_{\text{probe}}| \neq |\eta_{\text{ref}}|)}$ is the width of the asymmetry calculated from the two leading jets in an event of which one jet belongs to a reference $|\eta|$ interval and the probe jet belonging to another $|\eta|$ region. The resolution of the jet in the reference interval has to be determined already with the same-eta requirement. The asymmetry width is meant to be corrected for additional jet activity and the particle-level imbalance determined from $\sigma_{A,\text{gen}(|\eta_{\text{probe}}| \neq |\eta_{\text{ref}}|)}$. The reference interval is preferred to be in the central detector region where the statistical precision from the same-eta measurement has already been sufficient. The statistical uncertainty of $\frac{\sigma(p_{T,\text{ref}})}{\langle p_T \rangle}$ is propagated to the statistical uncertainty of $\frac{\sigma(p_{T,\text{probe}})}{\langle p_T \rangle}$.

As long as both jets belong to the same $|\eta|$ region, residual effects from jet energy scale differences in data and simulation affect both jets the same way and should not have an impact on the resolution measurement. Since residual effects from the jet energy scale, where the mean of the asymmetry is shifted and not exactly located at zero, become more important, if both jets belong to different probe and reference intervals, a slightly modified definition of the asymmetry of

$$A = \frac{p_{T,\text{probe}} - p_{T,\text{ref}}}{p_{T,\text{probe}} + p_{T,\text{ref}}} \quad (5.25)$$

is used for the reference-and-probe interval measurement. The transverse momenta $p_{T,\text{ref}}$ and $p_{T,\text{probe}}$ correspond to the transverse momenta of the reference and probe jet, respectively. Accordingly, in the determination of the asymmetry width σ_A the mean of the distribution A_{mean} is also estimated and the width calculated as

$$\sigma_A = \text{RMS}_{98.5\%} = \sqrt{\frac{1}{y_i} \cdot \sum_i (A_i - A_{\text{mean}})^2} \quad (5.26)$$

with the frequency y_i of the individual asymmetry values A_i and where the sum over i includes all values such that 98.5% of the total asymmetry distribution are covered symmetric around the mean.

The measurement with the reference-and-probe interval selection is performed taking the three innermost $|\eta|$ regions 0.0 – 0.5, 0.5 – 1.1 and 1.1 – 1.7 as reference intervals each. This allows a measurement of the data-to-simulation ratio also in the intervals 2.8 – 3.2 and 3.2 – 5.0. In addition, it provides for the innermost $|\eta|$ intervals an enhancement of the available number of events and thus a further reduction of the statistical uncertainties. The systematic uncertainties of the two new measurements are evaluated according to same-eta measurement by varying a certain aspect in the determination of the data-to-simulation ratio and taking the absolute deviation from the nominal ratio as uncertainty. The variation is done simultaneously for the $\sigma_{A(|\eta_{\text{probe}}| \neq |\eta_{\text{ref}}|)}$ value as well as for $\frac{\sigma(p_{T,\text{ref}})}{\langle p_T \rangle}$ for each uncertainty source.

A summary of the resulting data-to-simulation ratios determined in the measurements with the different reference intervals and the corresponding uncertainties is shown in Table 5.4

for $|\eta_{\text{ref}}| = 0.0 - 0.5$, Table 5.5 for $|\eta_{\text{ref}}| = 0.5 - 1.1$ and Table 5.6 for $|\eta_{\text{ref}}| = 1.1 - 1.7$, respectively.

Table 5.4.: Summary of the measurement with reference bin $|\eta| = 0.0 - 0.5$ showing the nominal data-to-simulation ratio $c(\text{Data}/\text{MC})$ with statistical uncertainty and systematic uncertainty for each uncertainty source in different $|\eta|$ regions.

	$ \eta_{\text{ref}} = 0.0 - 0.5$						
	$ \eta_{\text{probe}} $						
	0.0 – 0.5	0.5 – 1.1	1.1 – 1.7	1.7 – 2.3	2.3 – 2.8	2.8 – 3.2	3.2 – 5.0
$c(\text{Data}/\text{MC})$	-	1.106	1.133	1.227	1.253	1.410	1.171
Stat. uncertainty	-	± 0.008	± 0.009	± 0.025	± 0.047	± 0.068	± 0.116
PU reweighting	-	± 0.002	± 0.001	± 0.001	± 0.021	± 0.025	± 0.007
Particle-level imbalance	-	± 0.005	± 0.005	± 0.008	± 0.009	± 0.006	± 0.004
Jet energy scale	-	± 0.006	± 0.010	± 0.021	± 0.034	± 0.022	± 0.066
α -spectrum	-	± 0.010	± 0.006	± 0.011	± 0.006	± 0.009	± 0.012
α -range	-	± 0.010	± 0.014	± 0.046	± 0.089	± 0.015	± 0.018
Non-gaussian tails	-	± 0.001	± 0.007	± 0.005	± 0.044	± 0.046	± 0.029
Jet Flavour	-	± 0.004	± 0.004	± 0.001	± 0.021	± 0.013	± 0.008
Ratio shape	-	± 0.022	± 0.023	± 0.025	± 0.025	± 0.028	± 0.023
Total syst. uncertainty	-	± 0.028	± 0.030	± 0.058	± 0.112	± 0.067	± 0.079

5.11. Measurement for simulated events obtained with the Herwig++ generator

In order to test if the data-to-simulation ratio is generator independent or features specific differences due to the different fragmentation models used in the generators, the data-to-simulation ratio is not only derived with the PYTHIA QCD multijet sample, but also for simulated events generated with HERWIG++ using tune EE3C.

The applied selection criteria are the same as described in Section 5.4.2 and also the pile-up reweighting procedure is performed as described. Fig. 5.17 shows the inclusive p_T^{ave} spectrum compared in data and simulation for simulated events taken from HERWIG++. The agreement of the p_T^{ave} spectrum between data and simulation looks reasonable and is similar to the spectrum in simulated events from PYTHIA.

The nominal values of the data-to-simulation ratio have been determined with statistical

Table 5.5.: Summary of the measurement with reference bin $|\eta| = 0.5 - 1.1$ showing the nominal data-to-simulation ratio $c(\text{Data}/\text{MC})$ with statistical uncertainty and systematic uncertainty for each uncertainty source in different $|\eta|$ regions.

	$ \eta_{\text{ref}} = 0.5 - 1.1$ $ \eta_{\text{probe}} $						
	0.0 – 0.5	0.5 – 1.1	1.1 – 1.7	1.7 – 2.3	2.3 – 2.8	2.8 – 3.2	3.2 – 5.0
$c(\text{Data}/\text{MC})$	1.081	-	1.111	1.206	1.300	1.356	0.829
Stat. uncertainty	± 0.008	-	± 0.009	± 0.023	± 0.047	± 0.058	± 0.082
PU reweighting	± 0.001	-	± 0.003	± 0.005	± 0.002	± 0.024	± 0.039
Particle-level imbalance	± 0.005	-	± 0.005	± 0.010	± 0.017	± 0.009	± 0.021
Jet energy scale	± 0.008	-	± 0.009	± 0.021	± 0.009	± 0.061	± 0.134
α -spectrum	± 0.007	-	± 0.004	± 0.003	± 0.008	± 0.036	± 0.002
α -range	± 0.004	-	± 0.012	± 0.016	± 0.043	± 0.048	± 0.041
Non-gaussian tails	± 0.004	-	± 0.008	± 0.010	± 0.061	± 0.023	± 0.012
Jet Flavour	± 0.005	-	± 0.007	± 0.009	± 0.013	± 0.042	± 0.007
Ratio shape	± 0.022	-	± 0.022	± 0.024	± 0.026	± 0.027	± 0.017
Total syst. uncertainty	± 0.026	-	± 0.030	± 0.039	± 0.082	± 0.105	± 0.149

uncertainties with the same- η requirement as well as for the three different reference-and-probe jet combinations. Since this is only meant as cross-check, no systematic uncertainties have been studied. The resulting data-to-simulation ratios for simulated events taken from HERWIG++ together with statistical uncertainties for the different measurements are summarized in Tab. 5.7. In general, the obtained data-to-simulation ratios for HERWIG++ agree among each other and are quite similar to the values obtained from PYTHIA. Some further discussion can be found in Sec. 5.12.1.

5.12. Results

5.12.1. Determination of a combined result

In order to determine one data-to-simulation ratio for each $|\eta|$ region the results obtained from the individual measurements are combined into one single scale factor. This combination is for all $|\eta|$ regions – except for the very last $|\eta|$ region ranging from 3.2 – 5.0 which is discussed separately – derived from the measurements done with simulated events from PYTHIA while the HERWIG++ results are only used as cross check.

Table 5.6.: Summary of the measurement with reference bin $|\eta| = 1.1 - 1.7$ showing the nominal data-to-simulation ratio $c(\text{Data}/\text{MC})$ with statistical uncertainty and systematic uncertainty for each uncertainty source in different $|\eta|$ regions.

	$ \eta_{\text{ref}} = 1.1 - 1.7$ $ \eta_{\text{probe}} $						
	0.0 – 0.5	0.5 – 1.1	1.1 – 1.7	1.7 – 2.3	2.3 – 2.8	2.8 – 3.2	3.2 – 5.0
$c(\text{Data}/\text{MC})$	1.084	1.082	-	1.189	1.250	1.432	1.137
Stat. uncertainty	± 0.012	± 0.012	-	± 0.031	± 0.051	± 0.066	± 0.105
PU reweighting	± 0.005	± 0.004	-	± 0.024	± 0.001	± 0.002	± 0.029
Particle-level imbalance	± 0.004	± 0.004	-	± 0.016	± 0.017	± 0.019	± 0.006
Jet energy scale	± 0.012	± 0.010	-	± 0.029	± 0.041	± 0.055	± 0.018
α -spectrum	± 0.006	± 0.008	-	± 0.019	± 0.006	± 0.019	± 0.009
α -range	± 0.016	± 0.022	-	± 0.036	± 0.032	± 0.016	± 0.048
Non-gaussian tails	± 0.010	± 0.010	-	± 0.005	± 0.018	± 0.012	± 0.047
Jet Flavour	± 0.005	± 0.007	-	± 0.012	± 0.015	± 0.031	± 0.054
Ratio shape	± 0.022	± 0.022	-	± 0.024	± 0.025	± 0.029	± 0.023
Total syst. uncertainty	± 0.033	± 0.036	-	± 0.063	± 0.065	± 0.077	± 0.096

The combined data-to-simulation ratios are calculated as weighted mean from the individual measurements with the same- η requirement and the three central-forward combinations. Each measurement is weighted by its statistical uncertainty. For the evaluation of the systematic uncertainties for the combined result the systematic shift of the weighted mean caused by the systematic shifts of the single measurements is determined. This means that for each uncertainty source the individual up and down shifts to the nominal ratio of each measurement are taken and combined as the weighted mean using the statistical uncertainty just as for the nominal mean. Consequentially, the difference of each systematically shifted weighted mean to the nominal weighted mean is the respective systematic uncertainty for the combination.

The $|\eta|$ region from $3.2 - 5.0$ covering the most forward region of the detector – namely the hadronic forward – behaves a bit special compared to the others. In contrast to the other $|\eta|$ regions, the individual measurements from PYTHIA and HERWIG++ have a very large spread of ratios from 0.83 up to 1.32 in this particular interval. If those are combined for this interval, it is in this case reasonable to use all six individual measurements from PYTHIA as well as from HERWIG++ to find a result which covers all of these partly very different ratios. If the combination is done following the same procedure as described for

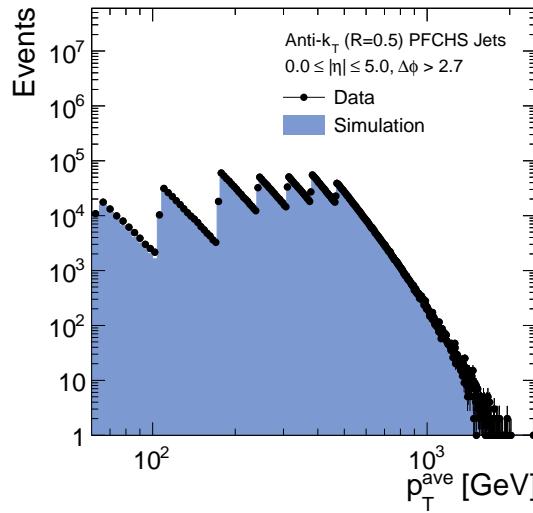


Figure 5.17.: Inclusive p_T^{ave} spectrum of events after the event selection in data and in simulated events generated with HERWIG++.

Table 5.7.: Measured data-to-simulation ratio in various $|\eta|$ regions with statistical uncertainty for simulated events obtained with the HERWIG++ generator.

$ \eta $	$ \eta_1 = \eta_2 $ $c(\text{Data}/\text{MC})$	$ \eta_{\text{ref}} = 0.0 - 0.5$ $c(\text{Data}/\text{MC})$	$ \eta_{\text{ref}} = 0.5 - 1.1$ $c(\text{Data}/\text{MC})$	$ \eta_{\text{ref}} = 1.1 - 1.7$ $c(\text{Data}/\text{MC})$
0.0 – 0.5	1.090 ± 0.008	-	1.091 ± 0.011	1.129 ± 0.017
0.5 – 1.1	1.107 ± 0.008	1.111 ± 0.010	-	1.121 ± 0.016
1.1 – 1.7	1.117 ± 0.013	1.177 ± 0.013	1.149 ± 0.012	-
1.7 – 2.3	1.297 ± 0.038	1.212 ± 0.035	1.229 ± 0.033	1.187 ± 0.041
2.3 – 2.8	1.085 ± 0.080	1.231 ± 0.072	1.178 ± 0.061	1.363 ± 0.087
2.8 – 3.2	-	1.368 ± 0.094	1.259 ± 0.061	1.488 ± 0.112
3.2 – 5.0	-	1.245 ± 0.158	1.124 ± 0.128	1.324 ± 0.215

the other intervals considering all nominal values obtained with PYTHIA and HERWIG++ while taking the systematic uncertainty from the combined PYTHIA results, the total scale factor would be 1.056 ± 0.048 (stat. unc.) ± 0.079 (syst. unc.) ($= 0.092$ total unc.). In order to test the consistency of this combined value with the single measurements the χ^2/ndf is calculated. This results in $\chi^2/\text{ndf} = 4.3$. This shows that the obtained total uncertainty is too small to reasonably cover all individual results. Hence, the systematic

uncertainty is increased to a value of ± 0.185 which gives a total uncertainty of ± 0.191 and results in a $\chi^2/\text{ndf} = 1.0$ such that the single measurements are well represented by this combined result and the individual shifts are covered by the assigned uncertainty. The obtained data-to-simulation ratios for the various $|\eta|$ regions are summarized in Tab. 5.8 together with the statistical and systematic uncertainties. In addition, Fig. 5.18 shows the measured ratios together with the total uncertainty. This is given by the quadratic sum of the statistical and systematic uncertainties. The determined ratios vary from 1.06 to 1.40 and are larger than one in all detector parts. Possible sources of this difference can be e.g. mismodelled noise effects or inhomogeneities of the detector.

Table 5.8.: Measured data-to-simulation ratio in various $|\eta|$ regions with statistical and systematic uncertainty as well as the total uncertainty.

$ \eta $	$c(\text{Data}/\text{MC})$	stat.	syst.	tot.
0.0 – 0.5	1.079	± 0.005	± 0.026	± 0.026
0.5 – 1.1	1.099	± 0.005	± 0.028	± 0.028
1.1 – 1.7	1.121	± 0.005	± 0.029	± 0.029
1.7 – 2.3	1.208	± 0.013	± 0.045	± 0.046
2.3 – 2.8	1.254	± 0.026	± 0.056	± 0.062
2.8 – 3.2	1.395	± 0.036	± 0.051	± 0.063
3.2 – 5.0	1.056	± 0.048	± 0.185	± 0.191

5.12.2. Comparison to Other Measurements

Earlier analyses using dijet events for data collected at $\sqrt{s} = 7 \text{ TeV}$ have obtained similar results showing data-to-simulation ratios greater than one. A complementary approach is followed using $\gamma + \text{jet}$ events which offer a very precise opportunity to measure the jet resolution due to the excellent resolution of the photon energy. A comparison of the numbers derived in the context of this thesis to the latest results from dijet events obtained at $\sqrt{s} = 7 \text{ TeV}$ [130] and results for $\gamma + \text{jet}$ measurements at $\sqrt{s} = 8 \text{ TeV}$ [133] are shown together in Fig. 5.19.

The results obtained from different methods and for the different center-of-mass energies are well compatible with each other. The main advantage of the dijet measurement presented in this thesis compared to the $\gamma + \text{jet}$ analysis is that one obtains also results for the outermost $|\eta|$ region. In addition, the total uncertainties in the other $|\eta|$ regions are at a same accuracy or even slightly lower.

In comparison to the $\sqrt{s} = 7 \text{ TeV}$ results from dijet events, the total uncertainty could

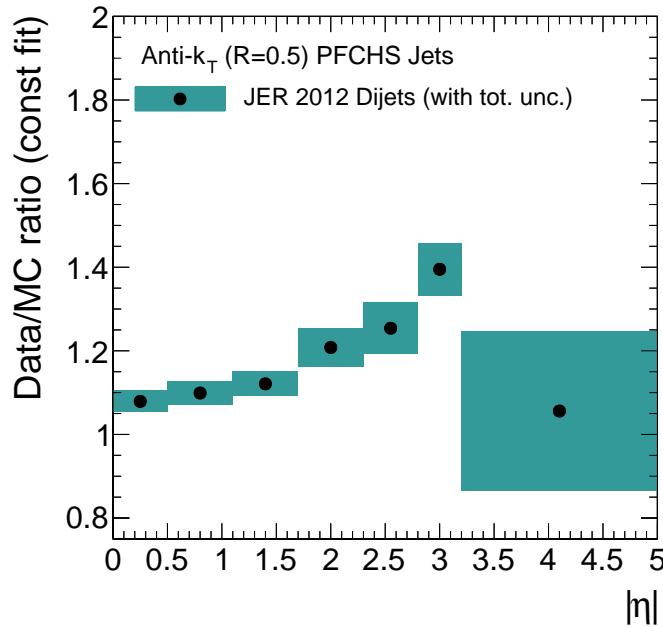


Figure 5.18.: Measured data-to-simulation ratio in various $|\eta|$ regions displayed with total uncertainty.

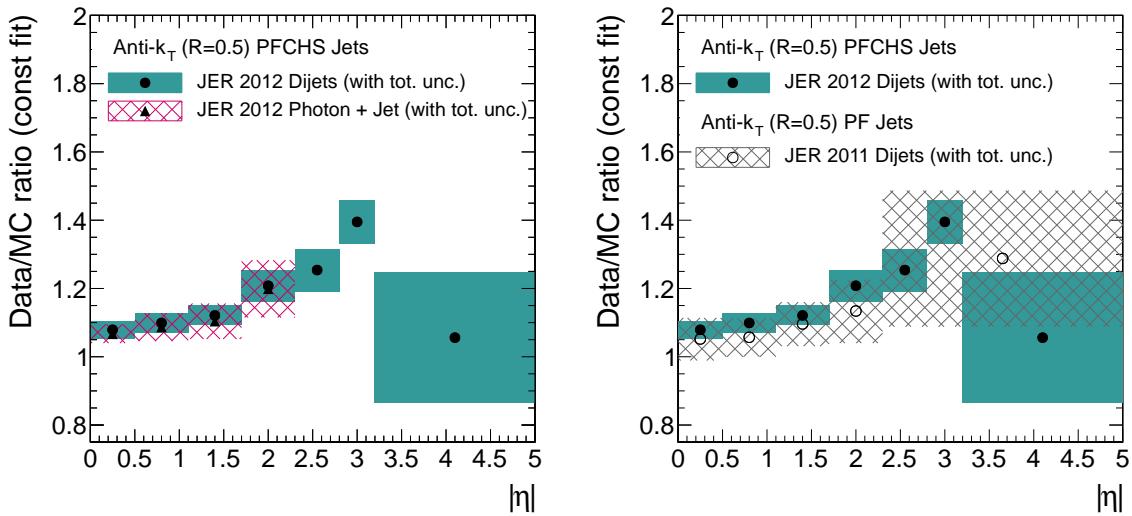


Figure 5.19.: Measured data-to-simulation ratio in various $|\eta|$ regions displayed with total uncertainty. Comparison to results obtained for $\sqrt{s} = 8$ TeV data from γ +jet events (left) and to $\sqrt{s} = 7$ TeV data from dijet events (right).

be significantly reduced by incorporating the correlation among the different inclusive α regions in the extrapolation procedure. In the previous analyses, the statistical uncertainties were underestimated by not considering the above mentioned correlation. This

effect was compensated by introducing a rather conservative systematic uncertainty on the extrapolation to zero additional jet activity. Since the treatment of the statistical uncertainty has been changed in this analysis, the estimation of systematic effects could be adjusted accordingly and lead to the overall reduced uncertainty values summarized in Table 5.8.

5.13. Adjustment of the MC-Resolution to Data

The determined data-to-simulation ratios as given in Tab. 5.18 can be used to adjust the resolution in simulated events to match the measured resolution in data. This can be done with different approaches depending on what the actual purpose of the analysis is.

If a correspondence of the resolution in data and simulation per jet is needed (as it was used in Sec. 5.8.2), each detector-level jet in the simulation can be scaled according to the momentum difference to the respective generator-level jet. Since the resolution is proportional to this momentum difference, an adjustment of the resolution to data corresponds to scaling the momentum difference by the measured scale factor as expressed already in Eq. 5.18.

If however an adjustment of the full response function is needed, the distribution can be adjusted by convolution with a Gaussian. Taking a Gaussian function with appropriate width σ_c , the width of the response function after convolution corresponds to the resolution in data according to

$$\sigma_c = \sqrt{c^2 - 1} \cdot \left(\frac{\sigma_{MC}}{p_T} \right) \quad (5.27)$$

Consequently, this method can only be used, if $c > 1$.

Further discussion about the adjustment of the resolution in simulation to data can be found in Section 6.3 where this is actually applied in the estimation of background contributions arising from QCD multijet events to a search for new physics in an all-jet final state.

6. Search for New Physics in the Multijet and Missing Transverse Momentum Final State at $\sqrt{s} = 8$ TeV

The search for hints of supersymmetric particles is among the important goals of the LHC physics program. As discussed in Sec. 2.3.2, searches performed at $\sqrt{s} = 7$ TeV have constrained the allowed parameter space for light-flavour squarks and gluinos already up to around 800 GeV and 1 TeV for light LSP masses, respectively. However, the increased centre of mass energy from 7 TeV to 8 TeV and the recorded dataset which is around four times larger than at 7 TeV provide the opportunity to extend the reach of such searches into entirely unexplored parameter regions. In Fig. 6.1 the theory cross section for the production of supersymmetric particles is shown as function of the SUSY particle mass. The y-axis on the right indicates how many events are expected in 20 fb^{-1} of pp collision data at the LHC at $\sqrt{s} = 8$ TeV.

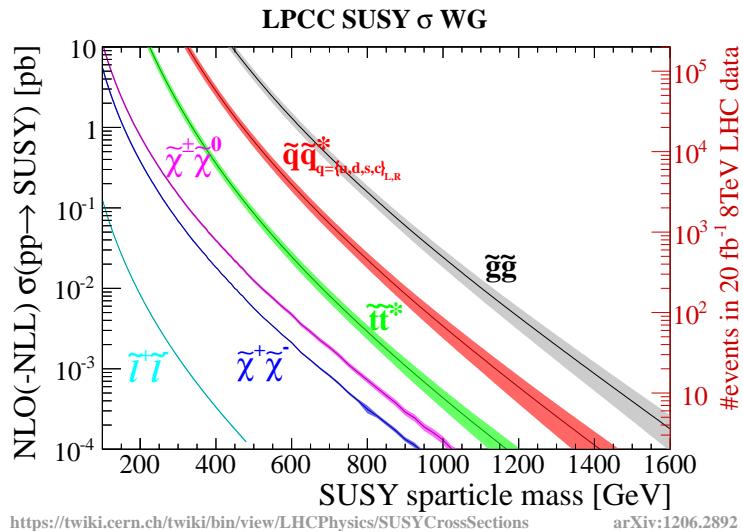


Figure 6.1.: Theory cross sections for selected SUSY processes as function of the SUSY sparticle mass. The y-axis on the right indicates the expected number of events in 20 fb^{-1} of pp collision data at the LHC at $\sqrt{s} = 8$ TeV [60].

It is visible that especially light-squarks and gluinos are expected at a sizable rate even for high masses above 1 TeV. For instance, around 100 pairs of gluinos are expected at a mass of 1200 GeV.

The analysis presented in this Chapter is aiming at a search for supersymmetric cascade decays arising from strongly produced light-flavour squarks or gluinos. Thus, events are

selected based on the scalar sum of the jet transverse momenta (H_T), the missing transverse momentum calculated from the jet momenta (\cancel{H}_T) and the number of jets (N_{Jets}). However, the generic structure makes the analysis in principle sensitive to any new physics model that manifests in final states containing several hard jets accompanied by missing transverse energy.

After the description of the event selection (Sec. 6.1), it is discussed how contributions from standard model processes to the selected final state are estimated. Special emphasis is put on the estimation of the QCD multijet background (Sec. 6.3). Finally, results are presented and interpreted in various simplified supersymmetric models (Sec. 6.4). Parts of this Chapter are taken from [148], having been written by the author. This analysis follows previous inclusive searches [63, 64] and is published in [149].

6.1. Event Selection

6.1.1. Data Samples and Trigger

The analysis is based on pp collision data recorded with the CMS detector at a centre of mass energy of $\sqrt{s} = 8 \text{ TeV}$. This corresponds to an integrated luminosity of 19.5 fb^{-1} for all sub-detectors fully functional.

The data have been collected by triggering on H_T , the scalar sum of the jet transverse momenta, and \cancel{E}_T , the missing transverse energy. An overview of the considered runs and the integrated luminosity is shown together with the respective HLT trigger paths in Tab. 6.1. H_T and \cancel{E}_T are calculated from particle-flow objects at trigger level with nominal thresholds of 350 GeV and 100 GeV, respectively. Jets considered in this calculation are reconstructed with the anti- k_T algorithm and distance parameter $R = 0.5$. The labelling *PFNoPU* indicates that for that particular run also charged-hadron subtraction was applied to jets at trigger level.

In order to determine the offline values for H_T and \cancel{H}_T (calculated according to the definition following in Sec. 6.1.3) for which the triggers reach the plateau efficiency, the trigger efficiencies are measured with respect to a single electron trigger (HLT_Ele27_WP80). In principal, an orthogonal trigger is needed in order to get an unbiased estimate of the trigger efficiency. However, due to the PF-algorithm all subdetectors are used simultaneously

Table 6.1.: Utilized signal trigger paths in individual run ranges listed together with the integrated luminosity.

Trigger path	Run range	Luminosity [fb^{-1}]
HLT_PFHT350_PFMET100	190456 – 196531	0.9
HLT_PFHT350_PFMET100	190782 – 190949	4.4
HLT_PFNoPUHT350_PFMET100	198022 – 198523	6.9
HLT_PFNoPUHT350_PFMET100	198524 – 208686	7.3

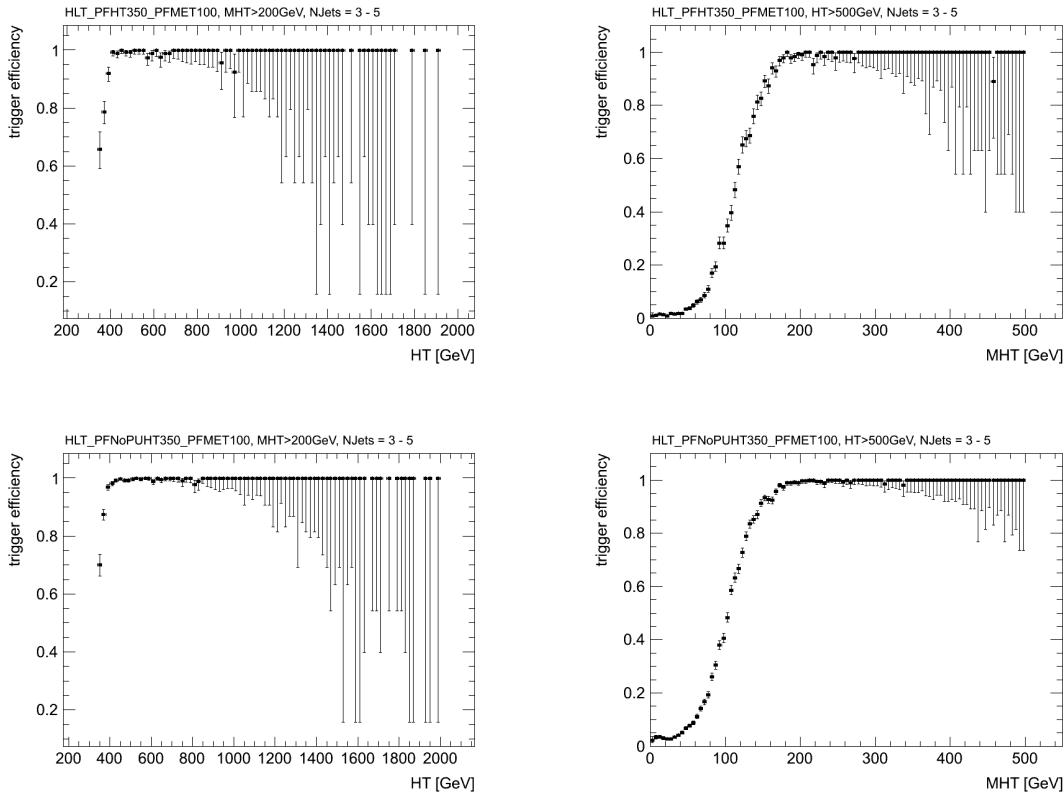


Figure 6.2.: Measured trigger efficiency for paths HLT_PFH350_PFMET100 (*top*) and HLT_PFNNoPUHT350_PFMET100 (*bottom*) as a function of H_T (*left*) and \cancel{H}_T (*right*) shown for $3 \leq N_{\text{jets}} \leq 5$.

to reconstruct the particles in an event. Hence, no independent trigger paths providing enough statistical precision are available. Consequently, only the reach of a plateau efficiency for a certain trigger path can be determined while the absolute efficiency is not accessible. The determination of the trigger plateau efficiency is performed for different jet multiplicity intervals ($N_{\text{jets}} = 3 - 5$, $N_{\text{jets}} = 6 - 7$ and $N_{\text{jets}} \geq 8$). The obtained trigger turn-on curves for the two different HLT paths are shown as function of offline H_T and \cancel{H}_T for jet multiplicity $N_{\text{jets}} = 3 - 5$ in Fig. 6.2. For both trigger paths, the efficiency plateau is reached around offline values of $H_T = 500\text{ GeV}$ and $\cancel{H}_T = 200\text{ GeV}$. The integrated trigger efficiencies for these particular values in different jet multiplicity intervals are summarized with statistical uncertainties in Tab. 6.2. In general, they are close to 100% with small uncertainties below 1%. However, for the highest jet multiplicity selection of $N_{\text{jets}} \geq 8$ only few events where selected such that statistical uncertainties are few 10% large. Though, no hints for a systematic inefficiency have been observed and the signal triggers are considered as fully efficient with an uncertainty of 2% for offline values of $H_T > 500\text{ GeV}$ and $\cancel{H}_T > 200\text{ GeV}$ independent of the jet multiplicity.

In order to validate the basic event selection discussed in Sec. 6.1.2 and 6.1.3 and the

Table 6.2.: Summary of total trigger efficiencies of the signal triggers for selections of offline $H_T > 500 \text{ GeV}$ and $\cancel{H}_T > 200 \text{ GeV}$ in different jet multiplicity intervals.

N_{jets}	HLT_PFHT350_PFMET100	HLT_PFNopUHT350_PFMET100
3 - 5	$99.4^{+0.2}_{-0.3}$	$99.8^{+0.1}_{-0.1}$
6 - 7	$99.1^{+0.7}_{-2.0}$	$100.0^{+0.0}_{-0.6}$
≥ 8	$100.0^{+0.0}_{-36.9}$	$100.0^{+0.0}_{-10.9}$

background estimation methods described in Sec. 6.3 and 6.2, several MC simulation samples are used. The standard model processes for $t\bar{t}$, $W + \text{jets}$, $Z + \text{jets}$, $\gamma + \text{jets}$ and QCD multijet events are generated with the MadGraph5 [96] generator at leading order and are interfaced with the parton-shower model in Pythia 6.4.24 [94]. They are scaled to cross-section predictions at next-to-leading order or even next-to-next-to-leading order when available [150, 151]. The events are processed with the full detector simulation. Furthermore, SUSY signal samples are obtained from simulation. They are generated with MadGraph5 [96] (with up to two additional partons), the CTEQ6L parton distribution functions [136] and processed with the fast detector simulation. Cross sections are determined at NLO with a resummation of soft gluon emission at the accuracy of next-to-leading-log [60, 152–156]. The cross section calculation as well as the generation of signal events for a certain type of sparticle is performed by effectively removing contributions from other sparticles by assuming their mass to be very large.

6.1.2. Event Cleaning

The analysis presented in this Chapter heavily relies on a precise measurement of the momentum imbalance in the event. In order to remove events with large values of fake missing momentum arising from detector noise a dedicated sequence of cleaning filters is applied.

Primary Vertex and Beam Halo: Only events with at least one high-quality primary vertex are considered in the analysis. A primary vertex is classified as good, if it has more than four associated tracks and is located within 24 cm in z and 2 cm in xy direction from the nominal interaction point (*good-vertex filter*). In order to reject events in which protons from the beam interact with residual gas molecules in the beam pipe, the CSC subdetector is used to identify muons moving parallel to the beam (*beam-halo filter*).

Anomalous Calorimeter Signals: Some events are affected by particles hitting the readout electronics or other technical instrumentation and cause anomalous signals in the ECAL or HCAL. For instance, noise in the readout system can fake artificial energy deposits at random times. Such events are identified based on timing and pulse-shape information (*HBHE noise filter*). Furthermore, two 5×5 supercrystal regions in the EE have been observed to give anomalously high energies. They are removed by imposing selections on the deposited energy in the identified supercrystals (*EE bad supercrystal filter*). In order to account for transparency losses in the ECAL

crystals, the system is calibrated with a dedicated laser. However, in the data some crystals are observed which receive unphysically large corrections. Events affected by this unusually large ECAL laser correction factors are rejected (*ECAL laser correction filter*). The HCAL is also monitored by a dedicated laser system. Sometimes the laser fires into the collision bunch-crossing resulting in unwanted signals. These events are removed according to an event list indicating the affected events (*HCAL laser filter*). The jet reconstruction utilizes information from the HO. This is used as identifier of significant leakage beyond the HCAL barrel. However, events with anomalous energy deposits in the HO have to be rejected. Thus, events in which the fraction of the momentum deposited in the HO is $> 40\%$ are removed (*HO filter*).

Dead ECAL Cells: Some single crystals in the ECAL are affected by malfunctioning. These single dead ECAL cells make up around 1% in total and can be responsible for energy losses resulting in large values of fake- H_T . Such events can be identified by using the trigger primitive information to determine how much energy was lost (*TP filter*) or by using the surrounding energy of the masked cells (*BE filter*).

Tracking Failure: In some events the track reconstruction is observed to fail manifesting in large calorimeter energy deposits with lack of associated tracks. This can be caused e.g. by too many seed clusters or by collisions not taking place in the actual center of the detector. Thus, the scalar sum of track momenta associated to the good vertices divided by H_T in the event has to be larger than 10% (*tracking failure filter*) and if at least ten tracks are present in the event, good-quality tracks have to be more than 25% (*beam-scraping filter*). In addition, events with misreconstructed muon momenta in the PF algorithm (*inconsistent muon filter*) or events where soft muons wrongly absorb energy from energetic HCAL towers they traverse (*greedy muon filter*) are rejected. Furthermore, events with coherent noise in the strip tracker can occur. These cause several clusters distributed across the whole detector and lead to the identification of fake tracks. Such events where the track reconstruction aborted can be identified by comparing the number of pixel clusters to the number of strip clusters (*many strip clusters filter*, *too many strip clusters filter*, *log error too many strip clusters filter*). Another failure of track reconstruction occurs sometimes when track seeds from the TOB and TEC are used. Consequently, events are rejected, if a jet with number of charged hadrons above 200 is reconstructed within $0.9 < |\eta| < 1.9$ (*TOB/TEC tracking filter*).

Noise Induced Jets: In order to reject events with fake jets from detector noise, events are discarded if the energy of a jet with $p_T > 30 \text{ GeV}$ is composed of more than 95% from PF photon candidates or more than 90% from PF neutral hadron energy (*PBNR filter*).

6.1.3. Baseline Selection

The physics objects used in the analysis are reconstructed with the PF algorithm. Jets are clustered from the particle-flow objects with the anti- k_T algorithm using $R = 0.5$. Furthermore, they are calibrated as discussed in Sec. 4.2.4 including residual correction factors for data.

The following *baseline selection* criteria are used to pre-define the event sample used for the analysis.

- The number of jets (N_{Jets}) is required to be ≥ 3 . N_{Jets} is defined as the number of jets with $p_T > 50 \text{ GeV}$ and $|\eta| < 2.5$. This requirement is imposed in order to select multijet events.
- The scalar sum of jet momenta (H_T) is required to be $\geq 500 \text{ GeV}$ with

$$H_T = \sum_{\text{jets}} p_T$$

for all jets that have $p_T > 50 \text{ GeV}$ and $|\eta| < 2.5$. This condition selects events with a large visible energy in the event indicating a high energy scale of the hard interaction.

- The absolute value of the negative vectorial sum of the jet momenta (\cancel{H}_T) is required to be $\geq 200 \text{ GeV}$ with

$$\cancel{H}_T = |\vec{\cancel{H}}_T| = \left| - \sum_{\text{jets}} \vec{p}_T \right|$$

for all jets with $p_T > 30 \text{ GeV}$ and $|\eta| < 5.0$. This selection reduces contributions from standard model processes where missing transverse momentum is expected to be small. Especially, QCD multijet background is suppressed.

- In order to suppress events where missing transverse energy is mainly arising from jet mismeasurements, as for QCD multijet events, it is required that \cancel{H}_T is not aligned with the leading three jets. Thus, events with

$$\Delta\phi(\text{jet}_n, \cancel{H}_T) > 0.5 \text{ for } n = 1, 2 \text{ and } \Delta\phi(\text{jet}_3, \cancel{H}_T) > 0.3$$

are selected. The value of 0.5 is chosen according to the jet size parameter. However, this is reduced in case of the third jet in order to retain signal efficiency.

- Background contributions arising from $t\bar{t}$ and $W+\text{jets}$ events are reduced by rejecting events containing isolated electrons or muons with $p_T > 10 \text{ GeV}$. These are required to have a good quality track that can be associated with the primary interaction vertex [157, 158]. The isolation is given as the scalar sum of transverse momenta of PF particles (except for the lepton itself) within a cone of width $\Delta R = 0.3$ for the electron and $\Delta R = 0.4$ for the muon, respectively. This is required to be less than 15% of the transverse momentum of the electron and less than 20% of the transverse momentum of the muon.

The selection described above defines a loose validation region and provides a basis for tighter criteria. In data, 11753 events are selected after the application of the baseline selection including the event cleaning filters. A comparison of selected events in data and simulation after the baseline selection is shown in Fig. 6.3. Although a reasonable data and MC agreement is observed especially in the bulk of the distributions, the background estimation from simulation is not further used in the analysis, but is meant to give an impression of the relative background contributions. Since the analysis is performed in

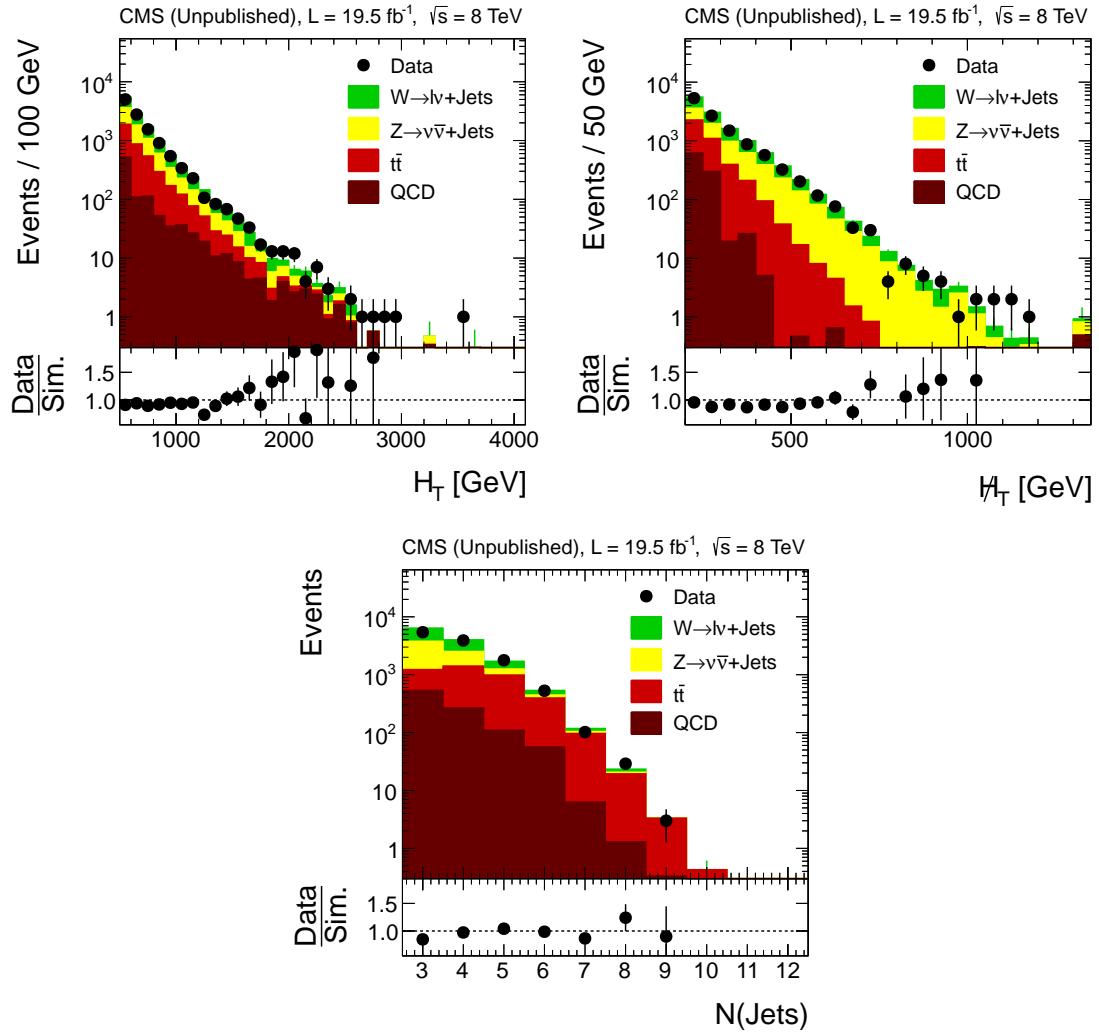


Figure 6.3.: Comparison of selected H_T (top left), \cancel{H}_T (top right) and N_{Jets} (bottom) distributions in data (black dots) and simulated events (shaded curve) found from applying the event cleaning and baseline selection criteria described in Sec. 6.1.2 and 6.1.3. Only statistical uncertainties are shown. Taken from [149].

extreme tails of the H_T , \cancel{H}_T and N_{Jets} phase space with only few events and large theoretical uncertainties in the simulation, SM background contributions are estimated solely with data-based methods.

6.1.4. Exclusive Search Regions

The analysis presented in this Chapter is an extension of previous inclusive searches [63,64]. These have been based on the requirement of at least three jets in the final state. In this analysis, the data is further subdivided into three exclusive jet multiplicity categories: $N_{\text{Jets}} = 3 - 5, 6 - 7, \geq 8$. This enhances the sensitivity of the search towards multijet final

states. These are typically the manifestations of long cascade decays from squarks and gluinos. Furthermore, it enlarges the sensitivity of the analysis to models in which gluinos often decay into top quarks. By requiring a large number of jets this analysis utilizes a complementary approach to other analyses which often use the presence of bottom-quark jets in the final state to discriminate against background [159–162]. This procedure allows to keep the signal efficiency for all-hadronic final states as high as possible. In order to gain sensitivity to a variety of models, the jet categories are further classified according to H_T and \cancel{H}_T . With this approach various exclusive search regions are defined. An overview of the exact definition of all 36 exclusive regions in N_{Jets} , H_T and \cancel{H}_T is given in Tab. 6.3.

Table 6.3.: Exclusive search regions used in the analysis binned in H_T , \cancel{H}_T and N_{Jets} .

N_{Jets}	[3-5]	[6-7]	[≥ 8]
	\cancel{H}_T [GeV]	\cancel{H}_T [GeV]	\cancel{H}_T [GeV]
$500 < H_T$ [GeV] < 800	200-300	200-300	> 200
	300-450	300-450	
	450-600	> 450	
	> 600		
$800 < H_T$ [GeV] < 1000	200-300	200-300	> 200
	300-450	300-450	
	450-600	> 450	
	> 600		
$1000 < H_T$ [GeV] < 1250	200-300	200-300	> 200
	300-450	300-450	
	450-600	> 450	
	> 600		
$1250 < H_T$ [GeV] < 1500	200-300	200-300	> 200
	300-450	300-450	
	> 450	> 450	
H_T [GeV] > 1500	200-300	200-300	> 200
	> 300	> 300	

6.2. Estimation of Non-QCD Backgrounds

As discussed in Sec. 2.3.2, events from the SM processes $Z(\nu\bar{\nu}) + \text{jets}$, $W + \text{jets}$ or semi-leptonic $t\bar{t}$ (where either the lepton is lost or a hadronically decaying τ lepton) and mismeasured QCD multijet events constitute important background contributions to all-hadronic final states. Dedicated data-based methods are employed to estimate their size in the selected data. In this Section those methods are described starting with a short review of the estimation of non-QCD backgrounds (Sec. 6.2.1, Sec. 6.2.2 and Sec. 6.2.3) followed by a detailed discussion of the QCD background estimation (Sec. 6.3).

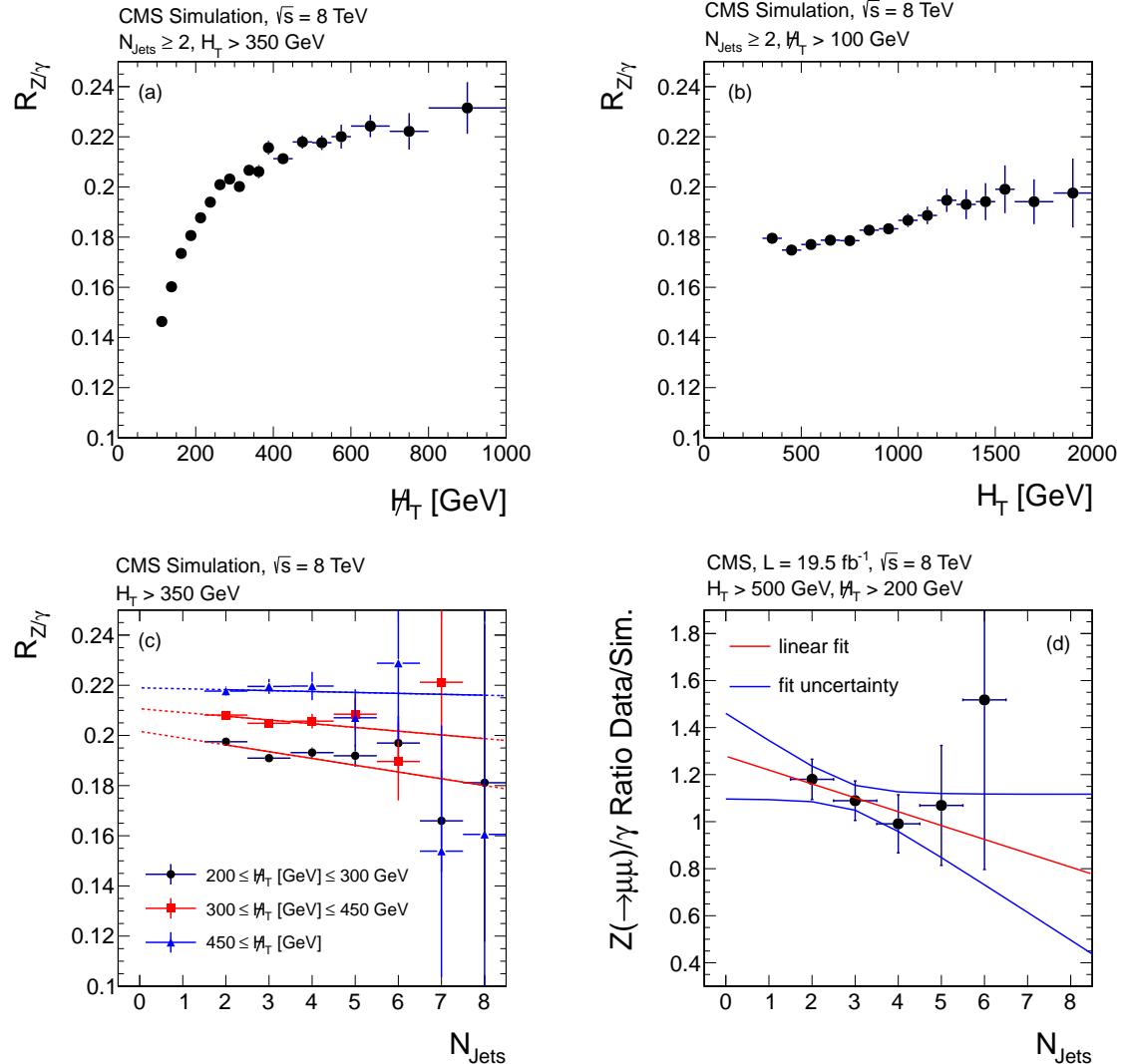


Figure 6.4.: The simulated ratio $R_{Z/\gamma}$ as a function of (a) H_T , (b) H_T , (c) N_{jets} , where the values for three H_T selections are shown with linear fits, and (d) the double ratio of $R_{Z(\mu+\mu-)} / \gamma$, using events from data to those from simulation; the linear fit and its uncertainty band are overlaid. Taken from [149].

6.2.1. Invisible Z Background

The irreducible background contributions arising from $Z(\nu\bar{\nu}) + \text{jets}$ events are estimated using $\gamma + \text{jets}$ events. This is a well suited method as for high transverse momenta of the vector boson the event kinematics are basically the same and the cross sections differ only according to the different couplings [163, 164].

The $\gamma + \text{jets}$ sample is collected by triggering on a γ candidate and large values of H_T . Photon candidates are selected if they satisfy $p_T > 100$ GeV and $|\eta| < 1.44$ or $1.566 < |\eta| < 2.5$. Furthermore, they have to have a shower profile consistent with that of a prompt photon produced directly in the hard interaction. In order to distinguish photons from

misidentified electrons they must not have an associated track in the pixel detector. In addition, photon candidates are required to be isolated meaning that in a cone of $\Delta R = 0.3$ the summed transverse momenta of PF candidates around the momentum direction of the photon candidate are not allowed to exceed a certain value. The number of selected $\gamma + \text{jets}$ events is corrected for photon acceptance, reconstruction and isolation efficiency. Furthermore, the purity of the $\gamma + \text{jets}$ sample which is the fraction of selected photon candidates emerging from direct production has to be taken into account. The number of background photons caused e.g. by misidentified jet fragments, is estimated by exploiting the difference between the shower profile of prompt and background photons. The average purity of the $\gamma + \text{jets}$ sample is measured to be 93%.

Subsequently, the number of $Z(\nu\bar{\nu}) + \text{jets}$ events in data $N_{Z(\nu\bar{\nu})+\text{jets}}^{\text{data}}(H_T, \cancel{H}_T, N_{\text{jets}})$ using the number of selected $\gamma + \text{jets}$ events $N_{\gamma+\text{jets}}^{\text{data}}(H_T, \cancel{H}_T, N_{\text{jets}})$ is obtained according to

$$N_{Z(\nu\bar{\nu})+\text{jets}}^{\text{data}}(H_T, \cancel{H}_T, N_{\text{jets}}) = R_{Z(\nu\bar{\nu})/\gamma}^{\text{MC}}(H_T, \cancel{H}_T, N_{\text{jets}}) \times \frac{R_{Z(\mu\mu)/\gamma}^{\text{data}}}{R_{Z(\mu\mu)/\gamma}^{\text{MC}}} \times N_{\gamma+\text{jets}}^{\text{data}}(H_T, \cancel{H}_T, N_{\text{jets}})$$

with the ratio relating the production cross section of $Z(\nu\bar{\nu}) + \text{jets}$ and $\gamma + \text{jets}$ events $R_{Z(\nu\bar{\nu})/\gamma}^{\text{MC}}(H_T, \cancel{H}_T, N_{\text{jets}})$ determined in simulation and the double ratio $\frac{R_{Z(\mu\mu)/\gamma}^{\text{data}}}{R_{Z(\mu\mu)/\gamma}^{\text{MC}}}$ selected in data and MC to consider the theoretical uncertainty and correct $R_{Z/\gamma}$ for a given jet multiplicity. The missing momentum in the event is emulated by ignoring the momentum of the photon candidate in the calculation of \cancel{H}_T .

The behaviour of $R_{Z/\gamma}$ is examined in simulated events as function of \cancel{H}_T , H_T and N_{jets} . The obtained distributions are shown in Fig. 6.4 (a) – (c). While a strong dependence on \cancel{H}_T for small values ($\lesssim 500$ GeV) is observed, the variation as function of H_T amounts to only $(12 \pm 5)\%$ in the relevant range for this analysis of $H_T = 500 - 1500$ GeV. The ratio as function of the jet multiplicity is determined for different \cancel{H}_T ranges of $\cancel{H}_T = 200 - 300$ GeV, $\cancel{H}_T = 300 - 450$ GeV and $\cancel{H}_T > 450$ GeV. The behaviour in each of these \cancel{H}_T ranges is described by a linear function also displayed in Fig. 6.2.1 (c). It is found that the ratio decreases slightly with increasing jet multiplicity which is consistent with findings from theory [164, 165].

In order to take the theoretical uncertainty on $R_{Z/\gamma}$ into account this phenomenological ratio is also determined for $Z_{\mu^+\mu^-}$ events in data and simulation, respectively. The double ratio of $R_{Z(\mu\mu)/\gamma}^{\text{data}}$ to $R_{Z(\mu\mu)/\gamma}^{\text{MC}}$ is derived as function of the jet multiplicity and shown in Fig. 6.2.1 (d). It is fitted with a linear function and the deviation from unity is considered as correction for the $R_{Z/\gamma}$ ratio in each jet multiplicity selection.

The main sources of uncertainty for the prediction of $Z(\nu\bar{\nu}) + \text{jets}$ events arise from the fit uncertainty to the double ratio which is at the order of 20%, 25% and 45% for the different jet multiplicity intervals, the differences between data and simulation regarding the photon identification and isolation as well as the subtraction of background photons from QCD multijet events.

6.2.2. Hadronic τ Background

Background contributions arising from $W + \text{jets}$ and $t\bar{t}$ events with a hadronically decaying τ lepton are estimated using a $\mu + \text{jets}$ control sample. Since $\mu + \text{jets}$ and $\tau_h + \text{jets}$ events arise from the same physics process, they feature the same kinematics except for

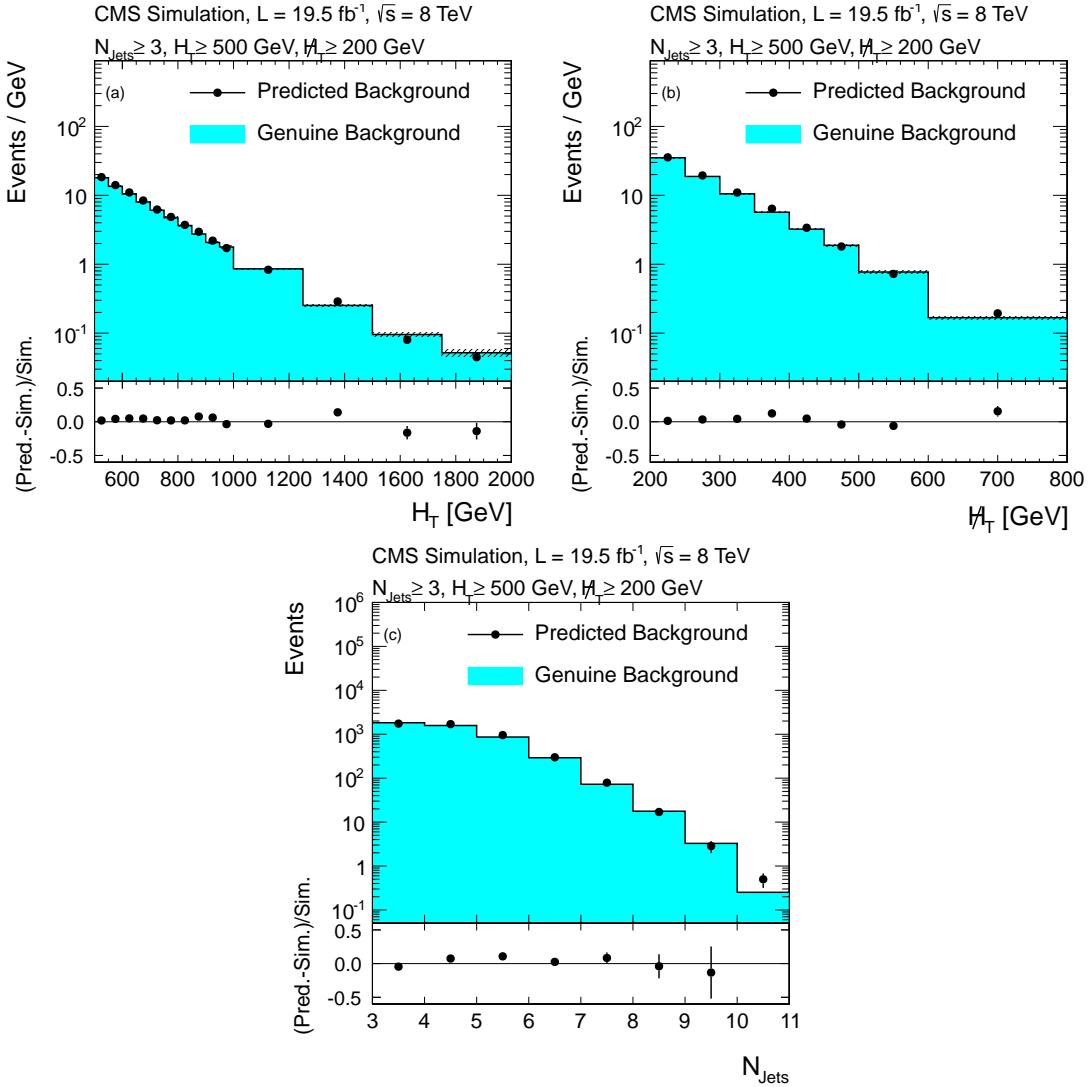


Figure 6.5.: Predicted (a) H_T , (b) \cancel{H}_T , and (c) N_{Jets} distributions found from applying the hadronic-tau background evaluation method to simulated $t\bar{t}$ and $W + \text{jets}$ events (solid points) in comparison to the genuine $t\bar{t}$ and $W + \text{jets}$ background from simulation (shaded curve). Only statistical uncertainties are shown. Taken from [149].

the different response of the detector to a μ and a τ_h .

The $\mu + \text{jets}$ sample is selected by triggering on a single muon or a muon accompanied by at least two jets. Furthermore, instead of applying a lepton veto, exactly one μ with $p_T > 20 \text{ GeV}$ and $|\eta| < 2.1$ is required. In order to prevent the control sample from signal contamination a selection on the transverse mass $m_T = \sqrt{2p_T^\mu E_T [1 - \cos(\Delta\phi)]}$ of $m_T \leq 100 \text{ GeV}$ is imposed with the azimuthal angle $\Delta\phi$ between the direction of the muon four-momentum and the E_T vector.

The difference between the μ and the τ_h is taken into account by replacing the muon by

a simulated τ_h jet. This is done by randomly sampling the transverse momentum of the τ_h jet from the response $p_T^{\text{jet}}/p_T^\tau$ obtained from simulation a reconstructed jet with p_T^{jet} to a generated hadronically decaying τ lepton with p_T^τ in the (η, ϕ) -space. The generated τ lepton has to fulfill $p_T > 20 \text{ GeV}$ and $|\eta| < 2.1$ and the distance for the matching is chosen to be $\Delta R < 0.2$ for $\text{tau-}p_T < 50 \text{ GeV}$ and $\Delta R < 0.1$ otherwise. The response is obtained from simulated $t\bar{t}$ and $W + \text{jets}$ events and subsequently mixed according to the cross sections of these processes in order to emulate what happens in data where both $t\bar{t}$ and $W + \text{jets}$ events occur. The random sampling of the response is repeated one hundred times for each event to sample the complete response function. If an event in the control sample is obtained from a prescaled trigger, the repetition of the sampling is increased according to the prescale factor. Technically, sampling means that the four-momentum of the muon is scaled with the proper value from the τ_h response.

In the following, H_T , \cancel{H}_T and N_{Jets} are recalculated for each event including the transverse momentum of the τ_h jet and all selection criteria as described in Sec. 6.1.3 are applied to the sample. The background contribution due to hadronic- τ events is obtained for all search regions by further correcting the event yields for the trigger efficiency, muon reconstruction and isolation efficiency, kinematic and detector acceptance as well as the ratio of branching fractions of $W \rightarrow \tau_h\nu$ to $W \rightarrow \mu\nu$ events. The statistical uncertainty of the background prediction is estimated with pseudo-experiments using a bootstrap method [166].

The validity of this background estimation procedure is tested by comparing the event yields obtained from applying the prediction method to simulated events from $t\bar{t}$ and $W + \text{jets}$ events to the respective genuine background. This comparison is shown as function of H_T , \cancel{H}_T and N_{Jets} in Fig. 6.5 after the baseline selection. Although the agreement is quite reasonable and hence the method is supposed to work reliable, uncertainties of 10% are considered for $N_{\text{Jets}} = 3 - 5$ and 20% for jet multiplicities $N_{\text{Jets}} = 6 - 7$ and $N_{\text{Jets}} \geq 8$. These uncertainties mainly reflect the statistical precision of the validation test.

Further systematic uncertainties taken into account for the hadronic τ background prediction are covering differences in data and MC for the muon isolation and reconstruction efficiencies as well as uncertainties on the kinematic and geometric acceptance, the τ_h jet response and the acceptance of the transverse mass cut.

6.2.3. Lost-Lepton Background

Similar to background events from hadronic tau decays, the background contribution due to a failed veto of a light lepton is estimated from a $\mu + \text{jets}$ control sample. This is selected with the same trigger as used for the search. The sample is selected by requiring exactly one well-reconstructed and isolated muon with $p_T > 10 \text{ GeV}$. Furthermore, the same transverse mass requirement of $m_T < 100 \text{ GeV}$ as for the hadronic tau background is applied.

The number of events in the zero-lepton search regions can be estimated from the single-muon sample by weighting the events according to the lepton reconstruction and isolation efficiencies as well as the detector and kinematic acceptance of the muons. The respective efficiencies and acceptances are obtained from simulated $t\bar{t}$ and $W + \text{jets}$ events and determined in intervals of H_T , \cancel{H}_T and N_{Jets} .

In order to determine the number of events due to unidentified leptons the events in the

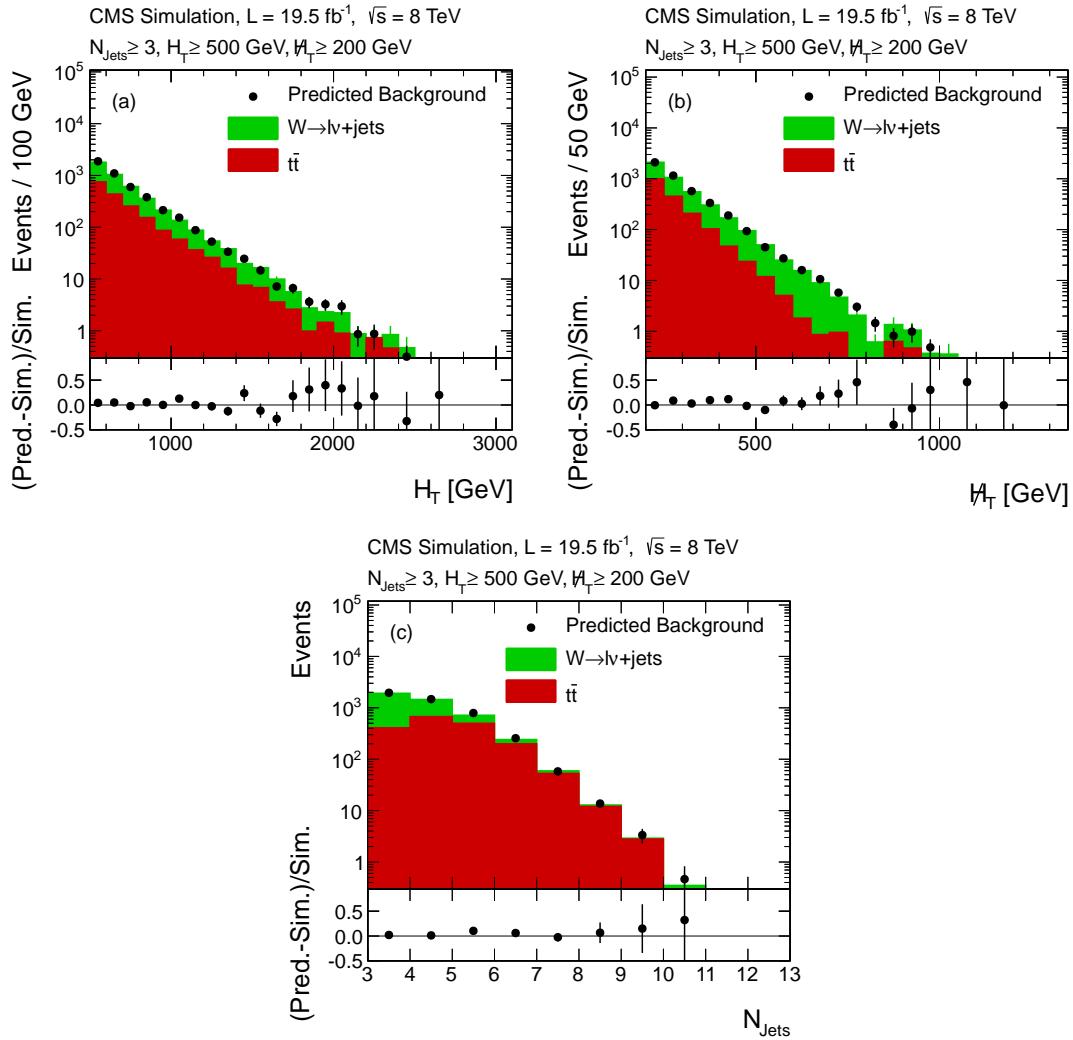


Figure 6.6.: Predicted (a) H_T , (b) \cancel{H}_T , and (c) N_{jets} distributions found from applying the lost lepton background evaluation method to simulated $t\bar{t}$ and $W + \text{jets}$ events (solid points) in comparison to the genuine $t\bar{t}$ and $W + \text{jets}$ background from simulation (shaded curves). Only statistical uncertainties are shown. Taken from [149].

control sample are weighted according to

$$\frac{1}{\epsilon_{\text{iso}}^\mu} \times \frac{1 - \epsilon_{\text{reco}}^{e,\mu}}{\epsilon_{\text{reco}}^\mu}$$

and with

$$\frac{\epsilon_{\text{reco}}^{e,\mu}}{\epsilon_{\text{reco}}^\mu} \times \frac{1 - \epsilon_{\text{iso}}^{e,\mu}}{\epsilon_{\text{iso}}^\mu}$$

to account for non-isolated leptons. Both equations are based on the reconstruction $\epsilon_{\text{reco}}^{e,\mu}$ and isolation $\epsilon_{\text{iso}}^{e,\mu}$ efficiencies of the electron and muon, respectively. The method is val-

idated in simulation by comparing the predicted event yields for lost-lepton events in $t\bar{t}$ and $W + \text{jets}$ from a single-muon control sample after the baseline selection to the genuine background. This comparison is illustrated in Fig. 6.6 as function of H_T , \cancel{H}_T and N_{Jets} and shows a good overall agreement. An uncertainty of 15% is assigned to jet multiplicities 3–5 and 40% to the others in order to account for the statistical precision of this validation test.

Main other uncertainties of the lost-lepton background prediction arise from the lack of sufficient events in the control sample in each defined search region, differences in lepton reconstruction and isolation efficiency between data and simulation, impact on the acceptance when varying the used PDFs and the acceptance of the transverse mass cut.

6.3. QCD Background Estimation with the Rebalance-And-Smear Method

Background contributions arising from events with no intrinsic missing energy, except for neutrinos in jets arising for instance from electroweak decays of heavy-flavour quarks, are estimated with the *Rebalance-and-Smear* (R+S) method. The main contribution is caused by QCD multijet events which also caused the naming ‘QCD background’. However, minor contributions originate also from fully-hadronic decaying $t\bar{t}$, $W + \text{jets}$ and $Z + \text{jets}$ events. The high values of missing transverse energy which cause these events to contribute to the analysed signal regions is consequently arising from severe jet mismeasurements.

Typically, QCD background is the most difficult background to model for all-hadronic SUSY searches, as a precise description of the underlying particle-level jet spectrum is needed and how this manifests in the detector. Especially the former suffers from large theoretical uncertainties in particular in the extreme kinematic phase space the analysis is performed in. To overcome this, the data-based R+S-method was developed and already successfully used in previous analyses [63, 64]. In this thesis, particularly improvements of the R+S method are discussed that became inevitable to handle the main changes of this analysis at $\sqrt{s} = 8 \text{ TeV}$ manifesting in the extension of the search regions in several jet multiplicities and the increased amount of pileup compared to the 7 TeV analyses. Consequently, the R+S method is adjusted such that it leads to precise results regarding also this new conditions. In general, the method is based on the assumption that if the momenta of particle-level jets in an event are known, the reconstructed jet momenta can be modeled by a per-jet resolution function.

After the discussion of the general concept of the R+S method in Sec. 6.3.1, the adjustment of the procedure to the actual conditions for 8 TeV data are introduced in Sec. 6.3.2. Furthermore, systematic uncertainties are discussed in Sec. 6.3.3 and the results of the actual QCD background prediction are presented in Sec. 6.3.4.

6.3.1. General concept of the Rebalance-and-Smear method

As stated above, QCD background contributions arise from jet mismeasurements in the detector. Thus, this background contribution can be estimated by retracing the measurement of multijet events. In order to do this, the prediction of the QCD background is performed in two subsequent steps. First, the events are *rebalanced*, as described below, such that the missing energy in the event is removed and almost ideal multijet events

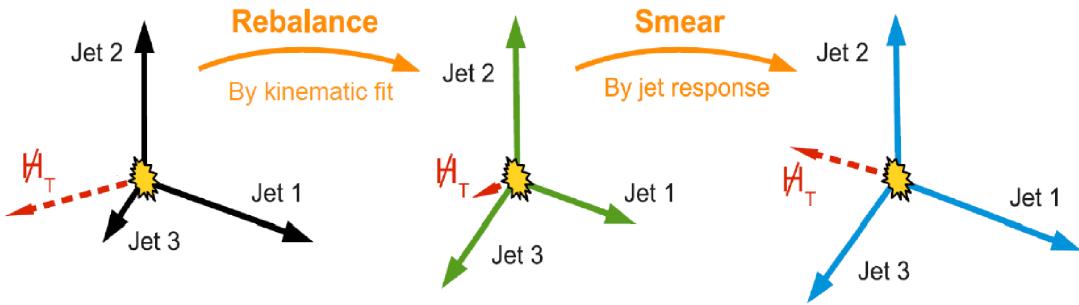


Figure 6.7.: Outline of the two steps performed in the R+S method for estimation of QCD background events [167].

denoted *seed events* are obtained. The resulting seed events reflect the event kinematics before the actual detector measurement is performed and are thus estimators of the true particle-level jet momenta. In a second step, all jets in the event are *smeared* with the full jet response function, i. e. jet momenta are scaled with a factor randomly drawn according to the jet response distribution, to model the interaction of the multijet state with the detector. Smeared events contain the whole event kinematics of the studied background contribution, such that these can be used to derive contributions to various kinematic distributions, like H_T or \tilde{H}_T , so that the R+S method can be used to really predict event kinematics rather than only event counts for the selected search regions. The general outline of the R+S method is illustrated in Fig. 6.7.

Response Templates: As indicated above, the R+S method crucially relies on a precise parametrization of the jet response to both perform the rebalancing and the jet smearing. The MC-truth response is derived for simulated QCD multijet events with the full detector simulation for p_T^{gen} and $|\eta^{\text{gen}}|$ intervals summarized in Tab. 6.4. Although a very fine binning is chosen, the response in each interval is averaged over a certain part of the p_T^{ave} spectrum. Thus, the response distribution tends to overestimate the width of the resolution for high- p_T jets while it behaves the other way round for low- p_T jets. Reconstructed jets at detector-level are AK5 jets with charged-hadron subtraction applied and calibrated according to the description in Sec. 4.2.4. Furthermore, the pileup scenario of the simulated sample is reweighted to match the one observed in data, as explained in Sec. 5.4.2.

As explained in Sec. 5.1, the truth jet response is derived by performing an unambiguous matching of reconstructed jet i to generated jet i using here $\Delta R < 0.1$. In order to avoid tails from splitting and merging effects of the jet reconstruction, any further reconstructed or generated jet $j \neq i$ around a matched pair is vetoed in a cone of size $R < 0.7$ by requiring

$$\frac{p_T^{\text{GenJet}_j}}{p_T^{\text{GenJet}_i}} < 0.05 \quad (6.1)$$

and

$$p_T^{\text{Jet}_j} < 30 \text{ GeV} \text{ and } \frac{p_T^{\text{Jet}_j}}{p_T^{\text{Jet}_i}} < 0.05. \quad (6.2)$$

The obtained jet response distributions are averaged over all jets in an event not

Table 6.4.: Overview of the $|\eta^{\text{gen}}|$ and p_T^{gen} interval boundaries used for the MC-truth response determination used as input for the R+S method.

$ \eta^{\text{gen}} $
0, 0.3, 0.5, 0.8, 1.1, 1.4, 1.7, 2.0, 2.3, 2.8, 3.2, 4.1, 5.0
p_T^{gen} [GeV]
0, 20, 30, 50, 80, 120, 170, 230, 300, 380, 470, 570, 680, 800, 1000, 1300, 1700, 2200, 2800, 3500

separating them according to their rank, i. e. their position in a descending p_T order, or according to the jet flavour. Thus, the flavour composition reflects that of an average QCD multijet sample.

The truth response templates to be used for applying the R+S method in simulated events, e. g. for validation tests, are determined as described above. However, when using the truth response templates for the actual QCD background predictions in data, they have to be corrected for potential data to simulation jet resolution differences. As seen in Chap. 5, the resolution in data is typically worse than in simulation. Thus, the determined truth response templates are adjusted accordingly. This correction is done for the Gaussian core and the non-Gaussian tails separately. First, the response function is splitted into the respective core and tail parts. This is done by fitting the response distribution with a Gaussian in the range of ± 1 RMS around the mean which is then subtracted from the total response distribution in order to obtain the tail parts and vice versa. The correction factors for the core resolution are applied by convoluting the MC-truth response with a Gaussian of width σ_c according to Eq. 5.27. The considered correction factors are listed in Tab. B.1 in the appendix and correspond to the data-to-simulation ratios obtained from dijet data at $\sqrt{s} = 7$ TeV, as illustrated in Fig. 5.19 (right).¹ The residual tail contributions are scaled according to the correction factors ρ_{tail} listed in Tab. B.2 in the appendix derived from dijet asymmetry parts fulfilling ($\mathcal{A} > 2\sigma_c$) [130].

Rebalance Procedure: As stated above, the first step in the R+S method is to create a sample of seed events that serve as estimator of the true particle-level jets by performing a rebalancing of the multijet events. This rebalancing is done based on a *kinematic fit* [168].

This is based on the hypothesis that for a given event all measured and unmeasured quantities fulfill certain kinematic constraints, like energy and momentum conservation. However, due to the uncertainties of these quantities the constraints are not exactly fulfilled. Thus, the constraints can be used to adjust the measured values within the uncertainties to meet the event hypothesis. This is done on an event-by-event basis by performing a least-square fit considering the kinematic constraints by Lagrange multipliers which provide a general method to determine local extrema of

¹Although the correction factors for the data-to-simulation ratio derived in the context of this thesis are more precise, these have not been available at that time when the analysis presented in this chapter has been performed.

non-linear functions of many variables. Mathematically, the likelihood function

$$L(\vec{y}, \vec{a}, \vec{\lambda}) = d\vec{y}^T C^{-1} d\vec{y} + 2 \sum_{k=1}^m \lambda_k f_k(\vec{y}, \vec{a}) \quad (6.3)$$

with $d\vec{y}^i = y_{\text{measured}}^i - y_{\text{predicted}}^i$, covariance matrix C , the Lagrange multipliers λ_k and the kinematic constraints f_k is minimised.

In this particular case, the measured jet four-momenta are fitted using the constraint of transverse momentum balance while the uncertainties are quantified by the jet resolution. However, the resolution for the angular components is not explicitly determined and for simplification set to a fixed tiny value, such that in fact no proper angular fit is performed. The uncertainties for the jet transverse momenta are approximated by the Gaussian MC-truth resolution, since the procedure is insensitive to the minor contributions of the non-Gaussian tails. For each response template the Gaussian core is extracted, as described above, by performing a fit with a Gaussian function within ± 1 RMS around the mean. The obtained resolutions are illustrated as function of p_T^{gen} for two example $|\eta^{\text{gen}}|$ intervals in Fig. These are fitted with

$$\frac{\sigma_{\text{MC}}(p_T)}{p_T} = \sqrt{\text{sgn}(N) \cdot \left(\frac{N}{p_T}\right)^2 + S^2 \cdot p_T^{m-1} + C^2} \quad (6.4)$$

where N, S, C and m are free parameters. This function is a modified version of Eq. 5.22 introduced for the characterizarion of the resolution in Chap. 5. It is adjusted to describe more precisely the resolution development for particle-flow jets. The term $\text{sgn}(N)$ considers the improved momentum resolution at low p_T due to the employed tracking information. Since also at medium p_T the tracking information still compensates for non-linearities of the calorimeters, the parameter m is introduced. The fitted functions are also illustrated in Fig. ... and used as input for the kinematic fit.

All events for which the fit converged within ($|\mathcal{H}_T^x| + |\mathcal{H}_T^y| < 0.02$) are then kept as seed events. Thus, the imbalance in each multijet event is removed by actually scaling the jet transverse momenta within the range of the respective resolution. Contributions to the seed sample from non-QCD multijet SM processes or even signal events in data do not have to be treated special, since this rebalancing procedure turns each of those events into QCD-like balanced topologies. Furthermore, the seed sample can be generated from an inclusive sample and their is no necessity to apply certain preselection cuts. Typically, selections suppressing high tails of missing energy tend to bias the QCD kinematics. This is due to the fact that for QCD events usually H_T and \mathcal{H}_T are correlated quantities, as high values of \mathcal{H}_T caused by severe jet mismeasurements can only occur, if there is a certain amount of energy in the event. Consequently, selection cuts removing high- \mathcal{H}_T tails also remove parts of the H_T spectrum resulting in an overall underestimation of QCD contributions to the high H_T and \mathcal{H}_T tails. Thus, the rebalancing with a kinematic fit allows to generate an unbiased seed sample.

Response Smearing: The second step of the R+S method is made up of the smearing procedure. Here, all jets of a seed event are smeared with the full jet response

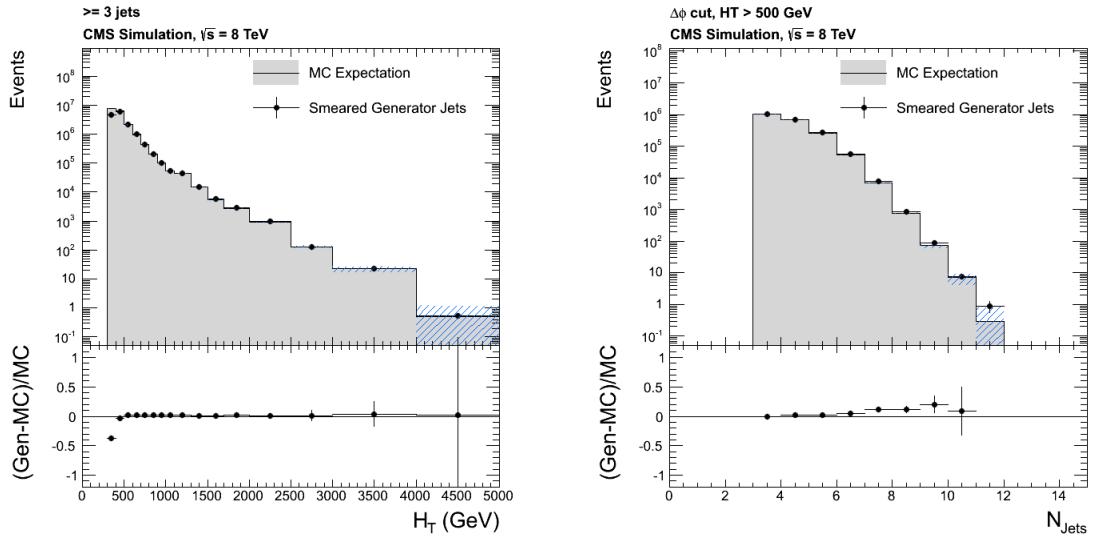


Figure 6.8.: Generated jets obtained from a QCD multijet sample generated with MADGRAPH smeared with the truth response templates are compared to the expectation from full simulation. This comparison is shown for H_T (left) with a loose pre-selection of at least three jets and N_{jets} (right) after the $\Delta\phi$ and the baseline H_T cut.

distributions including non-Gaussian tails. This means that the magnitudes of the transverse momenta of the jets are scaled with a factor that is randomly obtained from the jet response function of the respective p_T and $|\eta|$ interval to model the reconstructed transverse momentum. These smeared events hence resemble the full QCD event kinematics and thus contributions from QCD events to the search regions can be estimated by imposing the respective selection cuts to the smeared events. Similarly to the hadronic-tau background prediction, also the R+S method makes use of a bootstrap method in the smearing step. Instead of using all events only once, each seed event is smeared $N = 100$ times. The mean of these N predictions is taken as final result while the statistical uncertainty is obtained as the standard deviation of this set of predictions. This definition of the statistical uncertainty on the prediction ignores the statistical fluctuations of the seed sample. However, since the seed sample is very large, this uncertainty indeed is negligible to good approximation.

In order to validate the smearing procedure, generated QCD multijet events obtained from the MADGRAPH generator are smeared as described above and compared to fully simulated events at reconstruction level. This is performed on a sample which has a loose pre-selection at detetcor-level of $N_{\text{jets}} \geq 2$ and $H_T > 350$ GeV applied. The result of this generator jet smearing is shown in Fig. 6.8.

Overall the distributions derived from the smeared generator jets reproduce nicely the expectation from the simulation and tends to a slight overprediction of events for high jet multiplicities at the level of 10%. Such potential disagreements between the predicted and expected quantities are here denoted *non-closure* of the method. In general, non-closure can occur due to the limited granularity of the response tem-

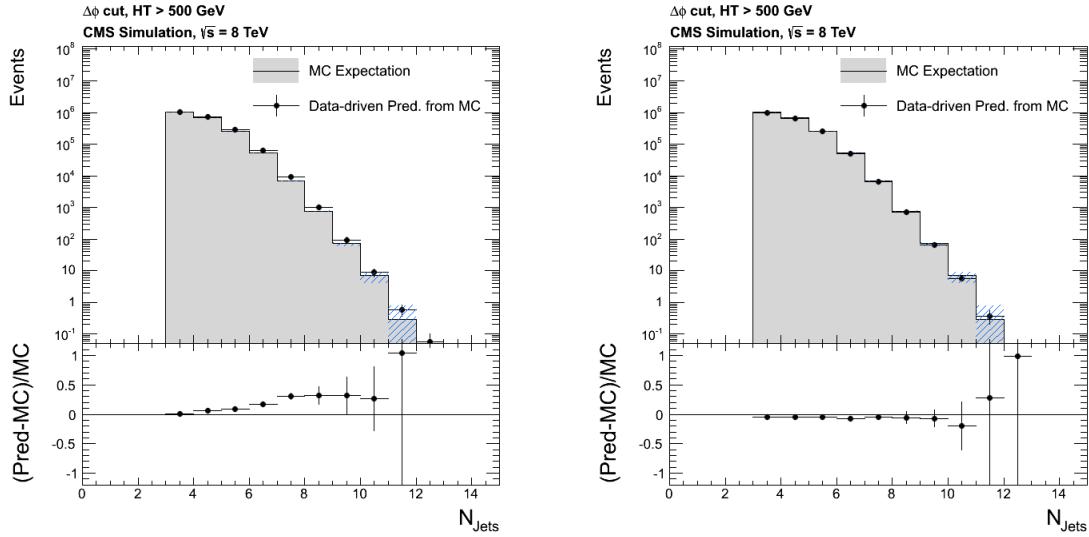


Figure 6.9.: Prediction of QCD background on a Monte Carlo sample compared to the expectation from full simulation showing the N_{jets} distribution for a p_{T} cut of 10 GeV on jets considered in the rebalancing *without* the application of a correction factor for the rebalancing (left) and *with* a correction of the rebalancing procedure (right) as described in the text.

plate binning and the averaged flavour composition of the response. In the end, the non-closure of the method is quantified for the whole R+S method including also the rebalancing step (see Sec. 6.3.2) and is assigned as systematic uncertainty to the QCD background prediction (see Sec. 6.3.3). This covers also the deviations observed in Fig. 6.8 for the smearing of generator jets.

6.3.2. Application to Realistic Collision Events

The presence of pile-up affects the rebalancing and needs to be taken particularly care of. Jets assigned to an event, especially soft jets, do not necessarily have to belong to the hard interaction, but might arise from pile-up events. Therefore, it is necessary to discard jets below a certain p_{T} threshold in the rebalancing step in order to not balance pile-up jets against the hard interaction. This p_{T} threshold can be chosen such that a good inclusive closure on Monte Carlo is observed. Nevertheless, by not taking all jets into account in the rebalancing, also soft jets belonging to the hard interaction are not considered. By doing this, additional imbalance in the event is introduced. It can be observed that this affects the kinematic fit such that especially jets with small p_{T} (< 100 GeV) are on average rebalanced to too high momenta. This results then in the fact that more jets pass the threshold of 50 GeV than expected. As a consequence, the number of predicted QCD events in the higher jet multiplicity bins is too large. The resulting trend in the N_{jets} distribution for a p_{T} cut of 10 GeV is illustrated in the left plot of Fig. 6.9. In order to account for this unavoidable threshold effect, an empirical correction factor (f) is introduced. This is done by comparing the p_{T} -spectrum of the rebalanced jets to the p_{T} -spectrum of matched generated jets when rejecting jets below 10 GeV in the rebalanc-

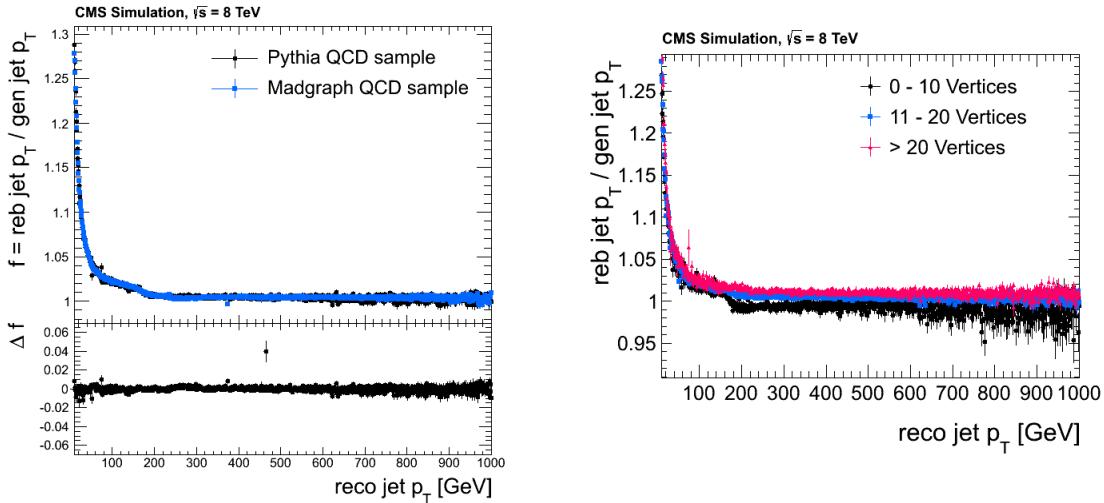


Figure 6.10.: Comparison of the rebalance correction factor determined on the PYTHIA and the MADGRAPH QCD Monte Carlo (left) and rebalance correction factor for different primary vertex bins (right).

ing procedure. The mean of this ratio is taken as correction factor and evaluated versus the momentum of the matched reconstructed jets. The correction factor is used such that in the rebalancing procedure each jet momentum is scaled by $1/f$ before the kinematic fit is performed. This adjusted method then leads to correct predictions of the QCD event yields also in the high jet multiplicity bins as demonstrated in the right plot of Fig. 6.9. The correction factor is derived by using generator-truth information and hence can only be determined on Monte Carlo samples. This is performed on the flat PYTHIA QCD Monte Carlo sample and on the MADGRAPH Monte Carlo sample. The difference of the correction factors derived from these two samples is found to be of the order of $1 - 2\%$ cf. Fig. 6.10, left. In addition, the correction factor was also determined for different bins of primary vertices and thus for different pile-up conditions. It is observed that the functional form of the correction factor is independent of the pile-up conditions cf. Fig. 6.10, right. This justifies the assumption that the observed harder jet p_T -spectrum after the rebalancing without the correction factor is just due to the acceptance of the jet p_T cut on the jets considered for the rebalancing. Therefore, the correction factor derived from Monte Carlo is applied to data where the same p_T cut value of 10 GeV for jets considered in the rebalancing is chosen. Since a remaining non-closure of the method due to the rebalancing step is covered by the total non-closure uncertainty (discussed in Sec. 6.3.3), no additional uncertainty is assigned.

The quality of the R+S method to predict background contributions from QCD multijet events is tested on Monte Carlo samples by several closure tests in different kinematic regions. Therefore, the data-based prediction is applied to Monte Carlo events and compared to the results from full simulation. The different closure tests as function of H_T for various jet multiplicity bins for a low H_T ($= 500 - 1000\text{ GeV}$) and a high H_T ($\geq 1000\text{ GeV}$) selection are shown in Fig. 6.11. These closure tests are used to determine the remaining bias of the R+S method.

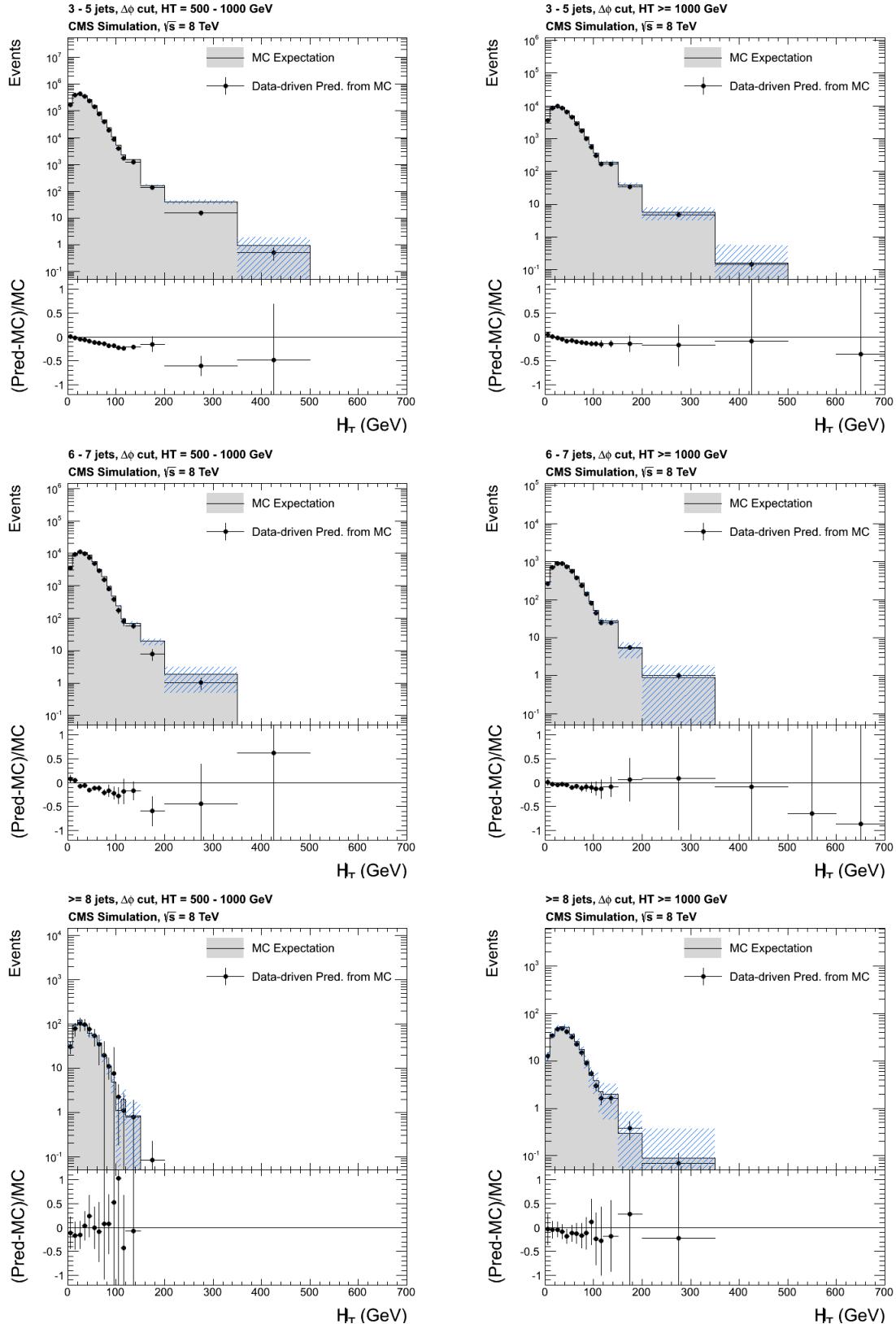


Figure 6.11.: Prediction of QCD background on a Monte Carlo sample compared to the expectation from full simulation. The closure test is shown for various jet multiplicity bins and low (left) or high (right) H_T selections.

Table 6.5.: Summary of non-closure uncertainties with their statistical error derived for the signal region (left) and two control regions with $\cancel{H}_T = 100 - 200$ GeV (second column) and inverted $\Delta\phi$ criterium (third column). The fourth column is used as additional cross check region as described in the text. The numbers marked in bold letters are taken as the final non-closure uncertainties of the method.

		Signal region	Control region 1	Control region 2	Cross check region
N_{jets}	H_T (GeV)	$\cancel{H}_T > 200$ GeV $\Delta\phi$ cut	$\cancel{H}_T = 100 - 200$ GeV $\Delta\phi$ cut	$\cancel{H}_T > 200$ GeV $\Delta\phi$ cut inverted	$\cancel{H}_T = 100 - 200$ GeV $\Delta\phi$ cut inverted
3 – 5	500 – 1000	(60.4 \pm 9.8)%	(22.6 \pm 1.6)%	(20.1 \pm 6.0)%	(2.8 \pm 1.3)%
6 – 7	500 – 1000	(43.1 \pm 46.5)%	(25.4 \pm 11.1)%	(59.3 \pm 96.0)%	(4.6 \pm 20.0)%
≥ 8	500 – 1000	–	(8.9 \pm 90.1)%	(86.0 \pm 38.2)%	(12.2 \pm 66.4)%
3 – 5	≥ 1000	(17.1 \pm 35.0)%	(14.4 \pm 3.1)%	(14.5 \pm 8.9)%	(5.1 \pm 1.7)%
6 – 7	≥ 1000	(5.5 \pm 108.0)%	(10.9 \pm 8.8)%	(14.5 \pm 42.9)%	(3.0 \pm 7.0)%
≥ 8	≥ 1000	(19.4 \pm 276.0)%	(21.8 \pm 28.6)%	(40.4 \pm 293.5)%	(21.1 \pm 42.6)%

The first choice for the determination of remaining biases is to calculate the difference between prediction and expectation for the signal region (defined by $\cancel{H}_T > 200$ GeV and the application of the $\Delta\phi$ cut) and take the observed difference as systematic uncertainty or even correct the prediction for a statistically significant non-closure of the method. The calculated differences with their statistical errors for the signal region are summarized in Tab. 6.5 (left column).

There is one bin (for 3 – 5 jets and $H_T = 500 - 1000$ GeV) where the signal region shows a statistically significant non-closure. In Fig. 6.12, this particular distribution is shown for the MADGRAPH QCD sample (left) and the PYTHIA sample (right). In the region between $\cancel{H}_T = 200 - 350$ GeV, the MADGRAPH sample shows an underprediction of $\approx 60\%$, while the PYTHIA sample tends to a statistically not significant overprediction. Therefore, it is hard to judge, if it is really a systematic effect or just a statistical fluctuation. In order to treat this conservatively, the result is not corrected for this effect, but the whole 60% deviation in the madgraph sample is considered as systematic uncertainty. As QCD is not the dominant background contribution in this search bin, it will hardly affect the final result.

Since the number of events in the signal region is low for all bins except the one discussed above and thus the statistical uncertainties are large and do not lead to reasonable conclusions about the closure of the method, two sidebands of the signal region are studied. The prediction is compared to the full simulation either for control region 1 defined by $\cancel{H}_T = 100 - 200$ GeV (second column in Tab. 6.5) or for control region 2 defined by an inverted $\Delta\phi$ criterium (third column in Tab. 6.5).

- If the differences in both control regions are statistically significant, the larger one is considered as systematic uncertainty.
- If only one of the numbers is statistically significant, it has to be made sure that the assigned uncertainty by taking this number, e.g. coming from control region 1, is not too small, as a remaining bias might come from the application of the $\Delta\phi$ cut. Therefore, the chosen value is compared to the numbers in the corresponding cross

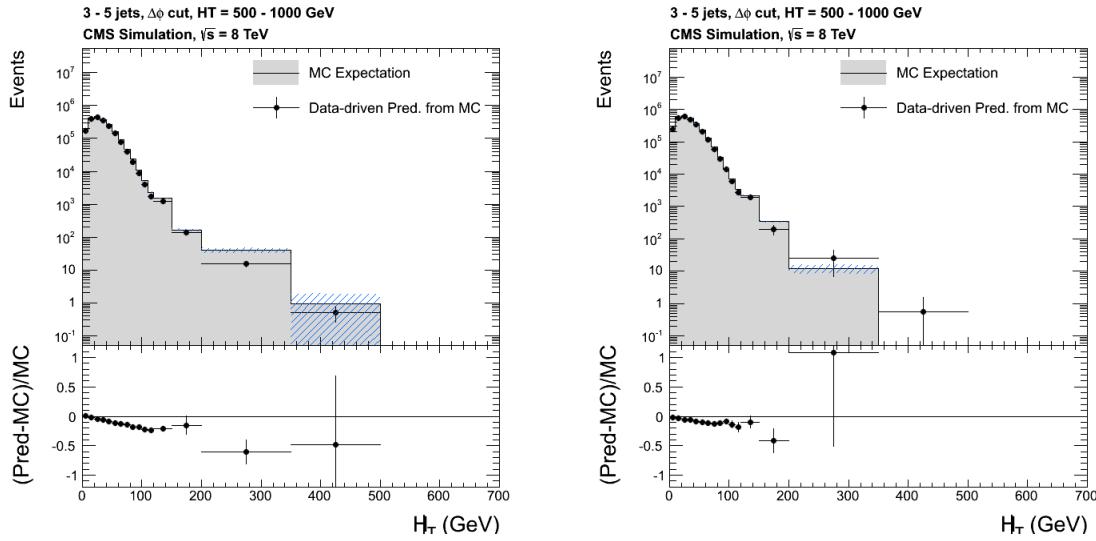


Figure 6.12.: Prediction of QCD background on a Monte Carlo sample compared to the expectation from full simulation. The closure test is shown for 3 – 5 jets and $H_T = 500 – 1000$ GeV on the madgraph QCD sample (left) and the pythia QCD sample (right).

check region bin (right column of Tab. 6.5). If the values in the cross check region are smaller than the chosen uncertainty value from the control region, the number from the control region is considered as systematic error. Otherwise take largest number from cross check region.

- If none of the numbers is statistically significant, take the number with highest precision and proceed as above by comparing this value to the numbers in the cross check region. If the cross check region does not show larger values, take the number with highest precision, otherwise take largest number from cross check region.

The uncertainty which is then finally chosen by the procedure described above as the non-closure uncertainty of the method, is printed in bold letters in Tab. 6.5.

Furthermore, the method can also be validated directly in data. The QCD background prediction is performed on a QCD enriched data control sample collected by a set of prescaled and unprescaled PFHT triggers. Thus, it has to be made sure that the events are collected such that the background prediction is not biased and the highest possible statistical precision is obtained. This is done as follows: For each event, the trigger with the lowest prescale is determined. The event then gets a weight according to this prescale factor. Since only one prescaled PFHT trigger exists, this assignment is unambiguous. This leads to a smooth H_T spectrum going down to the lowest trigger threshold. The prescale weighted seed H_T distribution is shown in Fig. 6.13. The usage of prescaled triggers is important in order to collect also events with H_T lower than 500 GeV which enter the signal region through a fluctuation to large response values. Nevertheless the events collected by the prescaled trigger come along with high event weights and spoil artificially the prediction when they enter the signal region since they lead to a substantially higher

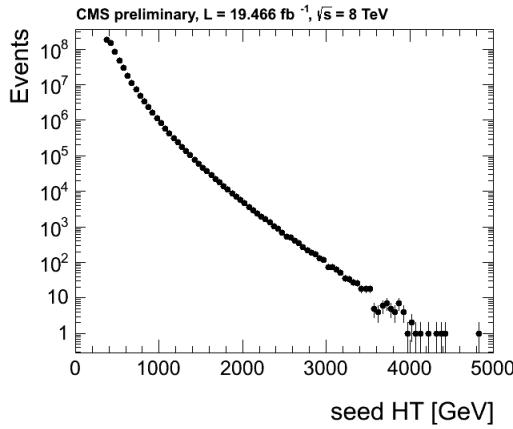


Figure 6.13.: Seed H_T spectrum of full 2012 dataset used as input for the R+S method after correcting for trigger prescales.

standard deviation. This problem is solved by smearing the events not only $N = 100$ times, but $N = \text{prescale factor} \times 100$ times and weighting them accordingly with one.

The performance of the R+S method can also be tested directly on data by comparing the prediction with selected data events in QCD enriched phase-space regions. This is done for $H_T > 1000$ GeV and an inverted $\Delta\phi$ criterium in order to select a QCD enriched data sample as depicted in Fig. 6.14. Nevertheless no perfect agreement is expected since contaminations from other backgrounds do still contribute. Especially the deviations in the \cancel{H}_T distribution increase with increasing values of \cancel{H}_T as QCD is not dominant here, so that a comparison is only reasonable for the low \cancel{H}_T region.

6.3.3. Systematic Uncertainties

Different uncertainties of the R+S method have been evaluated:

- **Core of response functions:** The uncertainties on the factors of the Gaussian core resolution in Tab. B.1 accounting for the differences in data and Monte Carlo are propagated to the prediction. This is done by shifting the scaling factors by $\pm 1\sigma$ up and down. The variations are between 10 to 30 % in the search bins where QCD leads to significant contributions.
- **Tail of response functions:** The uncertainties on the scaling factors of the non-Gaussian tails are also propagated to the prediction. The typical size of this uncertainty in QCD dominated regions is between 20 to 35 % which results in the main systematic uncertainty of the method.
- **Non-closure:** The systematic uncertainty due to possible non-closure and remaining biases of the R+S method is evaluated as described in section 6.3.2. This uncertainty also covers uncertainties on the rebalance correction factor and the jet- p_T cut

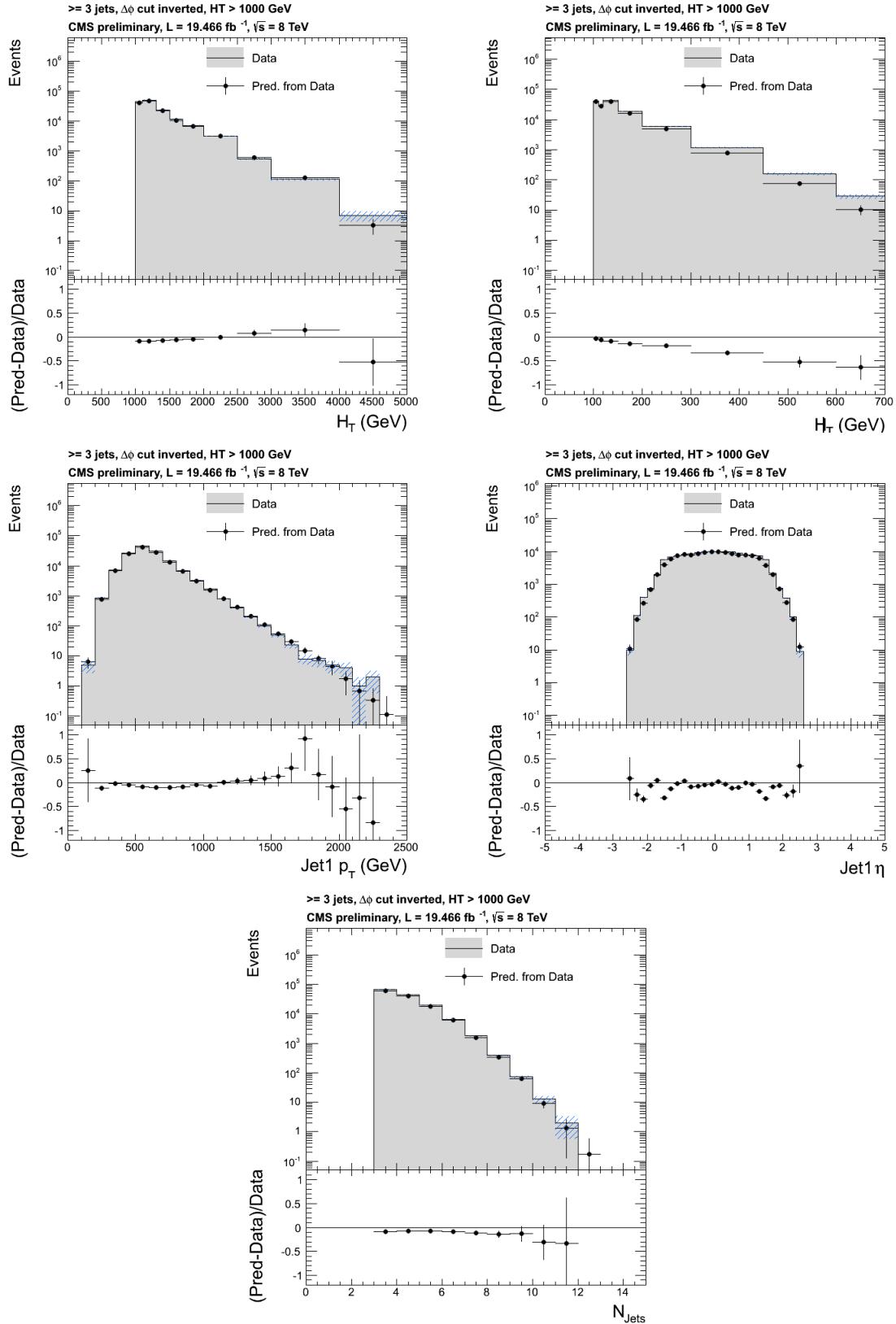


Figure 6.14.: Prediction of QCD background on data compared to the expectation from data for a QCD enriched control region with $H_T > 1000 \text{ GeV}$, inverted $\Delta\phi$ cut and at least three jets.

value chosen for the jets considered in the rebalancing. It is evaluated for each jet multiplicity bin which is separated into a low and a high H_T (> 1000 GeV) region.

- **Pile-up:** In general, the R+S method is relatively safe against pile-up since very soft jets are neglected in the rebalancing procedure. Additionally, it is expected that pile-up is an issue especially for *soft* jets which do not contribute significantly to the high H_T search bins because they are mainly populated due to heavily mis-measured *hard* jets. Therefore, a valid approach is to evaluate possible remaining pile-up effects in the low H_T region for $H_T > 100$ GeV. This approach also ensures that statistical limitations are reduced.

The general approach is to study the difference in the behaviour of the prediction in data and Monte Carlo for different pile-up conditions. Hence, the sample is divided into three different bins of primary vertices (0 – 10, 11 – 20, > 20). The prediction for each vertex bin (for $H_T > 500$ GeV) is calculated and normalized to the number of seed events contributing to that particular bin. It is assumed that pile-up effects are negligible in the lowest primary vertex bin. Therefore the distributions in data and Monte Carlo are normalized to each other such that they have the same yield in the first primary vertex bin. Then, the absolute difference between data and Monte Carlo prediction is calculated for the second and the third vertex bin, multiplied each with the seed events in data for that bin and summed up. The input distributions for this calculation are illustrated in Fig. 6.15. The ratio of this pile-up dependent fraction of the prediction calculated as described above and the total prediction is taken as pile-up uncertainty. It is evaluated for each jet multiplicity bin separately.

6.3.4. QCD Background Prediction

The final prediction for QCD background contributions derived with the R+S method is summarized in Tab. 6.6 for jet multiplicity bins 3–5, 6–7 and ≥ 8 . The quoted total systematic uncertainty is obtained by adding the single contributions in quadrature. Some search bins show very large uncertainties of 100 % or sometimes even larger for the core and tail scaling uncertainties. This occurs only in bins where QCD is almost negligible and is caused by the fact that the statistical variations are so large that they also dominate the evaluation of systematic effects. As a conservative estimate the largest observed variation is taken as systematic uncertainty for core and tail scaling.

6.4. Results and Interpretation

The selected number of events in 19.5 fb^{-1} of data together with the predicted event yields for the various SM background contributions estimated as discussed in Sec. 6.2 and Sec. 6.3 are listed in Tab. 6.7 for all 36 exclusive search regions. The displayed uncertainties for the background predictions are the total uncertainties. These have been obtained by adding statistical and systematic uncertainties in quadrature. Furthermore, the obtained yields in data and the predicted background are visualized in Fig. 6.16. The ratio in the bottom displayes the difference between observed data events and predicted background normalized to the background prediction. In general, the data are consistent with the SM expectation. The largest deviation occurs in the search region for $N_{\text{Jets}} = 6 - 7$, $H_T = 500 - 800$ GeV and $H_T \geq 450$ GeV. However, this is insignificant when including the

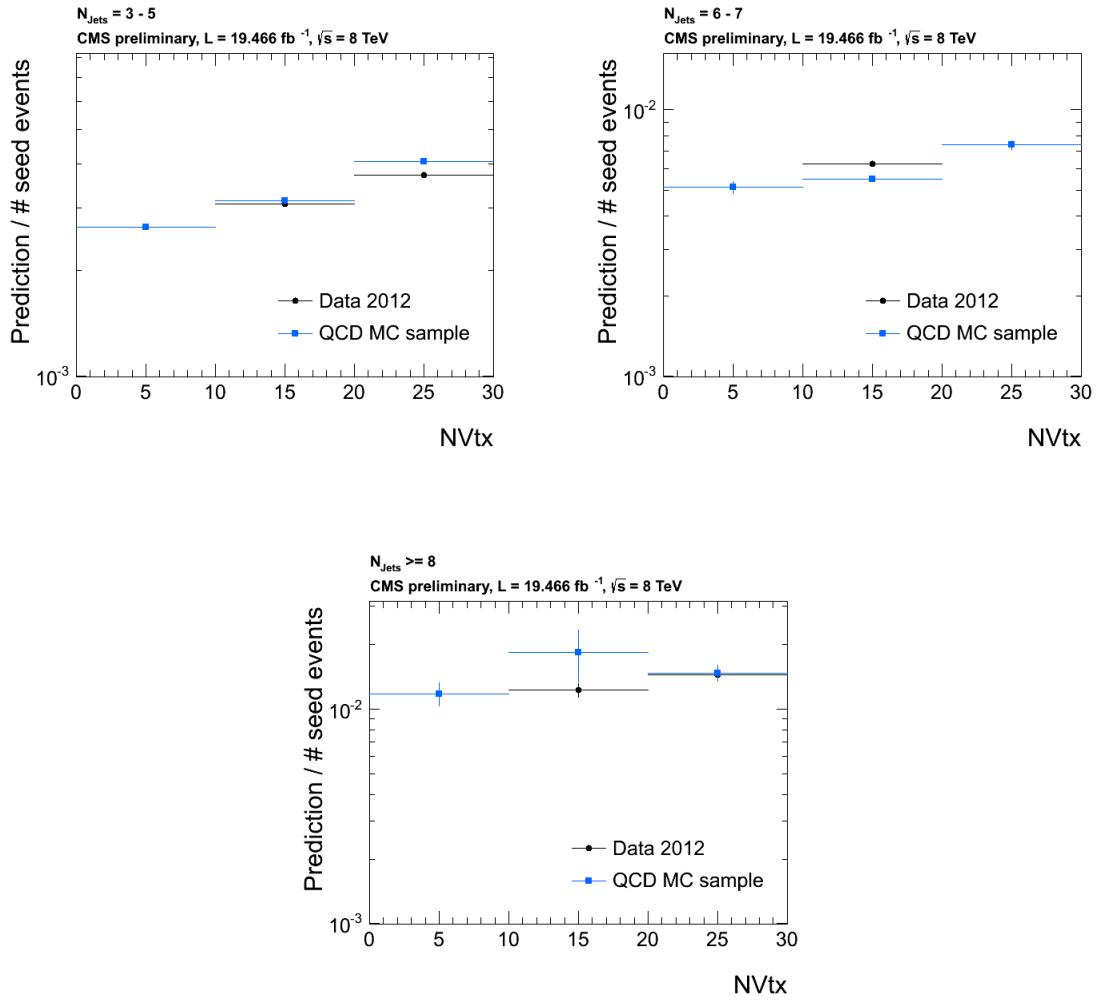


Figure 6.15.: Prediction of QCD background in data and Monte Carlo as a function of primary vertices (NVtx) normalized to the number of contributing seed events used for the determination of the pile-up uncertainty as described in Sec. 6.3.3.

Table 6.6.: Predicted event yields for the QCD background in the search regions defined by H_T , \cancel{H}_T and N_{jets} shown together with statistical and systematic uncertainties. The uncertainties of the different systematic uncertainty sources are added in quadrature to obtain the total systematic uncertainties.

N_{jets}	H_T [GeV]	\cancel{H}_T [GeV]	Pred.	stat. unc.	Core [%]	Tail [%]	Bias [%]	PU [%]	syst. unc.
3-5	500-800	200-300	307.4	± 18.5	+13.0 -12.2	36.0 -34.4	± 60.4	± 2.9	+220 -217
3-5	500-800	300-450	34.5	± 5.8	+7.3 -10.5	22.9 -31.6	± 60.4	± 2.9	+22.4 -23.8
3-5	500-800	450-600	1.3	± 1.2	+24.2 -16.7	+37.9 -26.5	± 60.4	± 2.9	+1.0 -0.9
3-5	500-800	> 600	0.1	± 0.3	+55.6 -55.6	+55.6 -55.6	± 60.4	± 2.9	+0.09 -0.09
3-5	800-1000	200-300	91.7	± 10.2	+14.7 -13.8	+33.2 -33.5	± 60.4	± 2.9	+64.7 -64.7
3-5	800-1000	300-450	9.9	± 3.2	+5.2 -5.1	+29.8 -27.1	± 60.4	± 2.9	+6.7 -6.6
3-5	800-1000	450-600	0.8	± 0.9	+65.5 -65.5	+65.5 -65.5	± 60.4	± 2.9	+0.9 -0.8
3-5	800-1000	> 600	0.1	± 0.4	+75.0 -41.7	+8.3 -41.7	± 60.4	± 2.9	+0.1 -0.1
3-5	1000-1250	200-300	59.0	± 7.2	+19.0 -14.6	+34.7 -31.7	± 14.5	± 2.9	+24.9 -22.4
3-5	1000-1250	300-450	5.1	± 2.2	+12.2 -8.8	+32.0 -16.9	± 14.5	± 2.9	+1.9 -1.2
3-5	1000-1250	450-600	0.5	± 0.7	+35.3 -3.9	+23.5 -5.9	± 14.5	± 2.9	+0.2 -0.1
3-5	1000-1250	> 600	0.1	± 0.3	+41.7 -41.7	+41.7 -41.7	± 14.5	± 2.9	+0.1 -0.1
3-5	1250-1500	200-300	31.2	± 5.3	+18.3 -19.1	+30.3 -29.7	± 14.5	± 2.9	+12.0 -11.9
3-5	1250-1500	300-450	2.3	± 1.3	+16.3 -5.3	+38.3 -30.4	± 14.5	± 2.9	+1.0 -0.8
3-5	1250-1500	> 450	0.2	± 0.5	+0.0 -8.3	+54.2 -8.3	± 14.5	± 2.9	+0.1 -0.1
3-5	>1500	200-300	35.1	± 6.1	+19.6 -20.0	+23.3 -29.4	± 14.5	± 2.9	+11.9 -13.5
3-5	>1500	> 300	2.4	± 1.4	+39.9 -39.9	+39.9 -39.9	± 14.5	± 2.9	+1.4 -1.4
6-7	500-800	200-300	18.2	± 3.9	+8.9 -12.5	+37.4 -33.5	± 25.4	± 8.0	+8.5 -8.1
6-7	500-800	300-450	1.9	± 1.4	+31.9 -31.9	+31.9 -31.9	± 25.4	± 8.0	+1.0 -1.0
6-7	500-800	> 450	0.01	± 0.1	+400.0 -100.0	+400.0 -100.0	± 25.4	± 8.0	+0.1 -0.01
6-7	800-1000	200-300	13.13	± 3.4	+15.0 -8.2	+33.7 -30.0	± 25.4	± 8.0	+6.0 -5.3
6-7	800-1000	300-450	2.0	± 1.1	+5.1 -20.0	+30.8 -28.2	± 25.4	± 8.0	+0.8 -0.9
6-7	800-1000	> 450	0.2	± 0.4	+46.7 -46.7	+46.7 -46.7	± 25.4	± 8.0	+0.1 -0.1
6-7	1000-1250	200-300	11.9	± 3.8	+5.9 -12.7	+33.5 -35.8	± 10.9	± 8.0	+4.4 -4.4
6-7	1000-1250	300-450	1.5	± 1.3	+31.8 -31.8	+31.8 -31.8	± 10.9	± 8.0	+0.7 -0.7
6-7	1000-1250	> 450	0.1	± 0.3	+100.0 -100.0	+100.0 -100.0	± 10.9	± 8.0	+0.2 -0.1
6-7	1250-1500	200-300	6.8	± 3.0	+12.0 -11.9	+32.6 -32.4	± 10.9	± 8.0	+2.5 -2.5
6-7	1250-1500	300-450	0.9	± 1.0	+54.4 -54.4	+54.4 -54.4	± 10.9	± 8.0	+0.7 -0.7
6-7	1250-1500	> 450	0.09	± 0.3	+44.4 -44.4	+44.4 -44.4	± 10.9	± 8.0	+0.06 -0.06
6-7	>1500	200-300	8.0	± 2.8	+20.9 -15.4	+31.5 -25.8	± 10.9	± 8.0	+3.1 -2.6
6-7	>1500	> 300	0.8	± 0.9	+47.0 -47.0	+47.0 -47.0	± 10.9	± 8.0	+0.6 -0.6
≥ 8	500-800	> 200	0.14	± 0.38	+71.4 -71.4	+71.4 -71.4	± 86.0	± 33.4	+0.19 -0.14
≥ 8	800-1000	> 200	0.54	± 0.69	+33.3 -33.3	+33.3 -33.3	± 86.0	± 33.4	+0.56 -0.54
≥ 8	1000-1250	> 200	0.73	± 0.78	+19.2 -1.4	+56.2 -27.4	± 86.0	± 33.4	+0.59 -0.44
≥ 8	1250-1500	> 200	0.54	± 0.75	+55.6 -55.6	+55.6 -55.6	± 86.0	± 33.4	+0.52 -0.52
≥ 8	>1500	> 200	0.89	± 0.94	+65.2 -65.2	+65.2 -65.2	± 86.0	± 33.4	+0.95 -0.89

Table 6.7.: Predicted event yields for the different background components in the search regions defined by H_T , \cancel{H}_T and N_{Jets} . The uncertainties of the different background sources are added in quadrature to obtain the total uncertainties. Taken from [149].

N_{Jets}	Selection	H_T [GeV]	\cancel{H}_T [GeV]	$Z \rightarrow \nu\bar{\nu}$	$t\bar{t}/W$ $\rightarrow e, \mu + X$	$t\bar{t}/W$ $\rightarrow \tau_h + X$	QCD	Total background	Data
3-5	500-800	200-300		1821 ± 387	2211 ± 448	1749 ± 210	307 ± 219	6088 ± 665	6159
3-5	500-800	300-450		994 ± 218	660 ± 133	590 ± 69	35 ± 24	2278 ± 266	2305
3-5	500-800	450-600		273 ± 63	77 ± 17	66.3 ± 9.5	$1.3^{+1.5}_{-1.3}$	418 ± 66	454
3-5	500-800	> 600		42 ± 10	9.5 ± 4.0	5.7 ± 1.3	$0.1^{+0.3}_{-0.1}$	57.4 ± 11.2	62
3-5	800-1000	200-300		216 ± 46	278 ± 62	192 ± 33	92 ± 66	777 ± 107	808
3-5	800-1000	300-450		124 ± 26	113 ± 27	84 ± 12	9.9 ± 7.4	330 ± 40	305
3-5	800-1000	450-600		47 ± 11	36.1 ± 9.9	24.1 ± 3.6	$0.8^{+1.3}_{-0.8}$	108 ± 15	124
3-5	800-1000	> 600		35.3 ± 8.8	9.0 ± 3.7	10.3 ± 2.0	$0.1^{+0.4}_{-0.1}$	54.8 ± 9.7	52
3-5	1000-1250	200-300		76 ± 17	104 ± 26	66.5 ± 9.9	59 ± 25	305 ± 41	335
3-5	1000-1250	300-450		39.3 ± 8.9	52 ± 14	41 ± 11	5.1 ± 2.7	137 ± 20	129
3-5	1000-1250	450-600		18.1 ± 4.7	6.9 ± 3.2	6.8 ± 2.0	$0.5^{+0.7}_{-0.5}$	32.3 ± 6.1	34
3-5	1000-1250	> 600		17.8 ± 4.8	2.4 ± 1.8	2.5 ± 0.8	$0.1^{+0.3}_{-0.1}$	22.8 ± 5.2	32
3-5	1250-1500	200-300		25.3 ± 6.0	31.0 ± 9.5	21.3 ± 4.1	31 ± 13	109 ± 18	98
3-5	1250-1500	300-450		16.7 ± 4.3	10.1 ± 4.4	13.7 ± 7.1	2.3 ± 1.6	42.8 ± 9.5	38
3-5	1250-1500	> 450		12.3 ± 3.5	2.3 ± 1.7	2.7 ± 1.2	$0.2^{+0.5}_{-0.2}$	17.6 ± 4.1	23
3-5	>1500	200-300		10.5 ± 2.9	16.7 ± 6.2	23.5 ± 5.6	35 ± 14	86 ± 17	94
3-5	>1500	> 300		10.9 ± 3.1	9.7 ± 4.3	6.6 ± 1.4	2.4 ± 2.0	29.7 ± 5.8	39
6-7	500-800	200-300		22.7 ± 6.4	133 ± 59	117 ± 25	18.2 ± 9.2	290 ± 65	266
6-7	500-800	300-450		9.9 ± 3.2	22 ± 11	18.0 ± 5.1	1.9 ± 1.7	52 ± 12	62
6-7	500-800	> 450		0.7 ± 0.6	$0.0^{+3.2}_{-0.0}$	$0.1^{+0.5}_{-0.1}$	$0.0^{+0.1}_{-0.0}$	$0.8^{+3.3}_{-0.6}$	9
6-7	800-1000	200-300		9.1 ± 3.0	56 ± 25	46 ± 11	13.1 ± 6.6	124 ± 29	111
6-7	800-1000	300-450		4.2 ± 1.7	10.4 ± 5.5	12.0 ± 3.6	1.9 ± 1.4	28.6 ± 6.9	35
6-7	800-1000	> 450		1.8 ± 1.0	2.9 ± 2.5	1.2 ± 0.8	$0.1^{+0.4}_{-0.1}$	6.0 ± 2.8	4
6-7	1000-1250	200-300		4.4 ± 1.7	24 ± 12	29.5 ± 7.8	11.9 ± 6.0	70 ± 16	67
6-7	1000-1250	300-450		3.5 ± 1.5	8.0 ± 4.7	8.6 ± 2.7	1.5 ± 1.5	21.6 ± 5.8	20
6-7	1000-1250	> 450		1.4 ± 0.8	$0.0^{+3.6}_{-0.0}$	$0.6^{+0.8}_{-0.6}$	$0.1^{+0.4}_{-0.1}$	$2.2^{+3.8}_{-1.1}$	4
6-7	1250-1500	200-300		3.3 ± 1.4	11.5 ± 6.5	6.4 ± 2.7	6.8 ± 3.9	28.0 ± 8.2	24
6-7	1250-1500	300-450		1.4 ± 0.8	3.5 ± 2.6	3.5 ± 1.9	$0.9^{+1.3}_{-0.9}$	9.4 ± 3.6	5
6-7	1250-1500	> 450		0.4 ± 0.4	$0.0^{+2.5}_{-0.0}$	$0.1^{+0.5}_{-0.1}$	$0.1^{+0.3}_{-0.1}$	$0.5^{+2.6}_{-0.4}$	2
6-7	>1500	200-300		1.3 ± 0.8	10.0 ± 6.9	2.0 ± 1.2	7.8 ± 4.0	21.1 ± 8.1	18
6-7	>1500	> 300		1.1 ± 0.7	3.2 ± 2.8	2.8 ± 1.9	$0.8^{+1.1}_{-0.8}$	7.9 ± 3.6	3
≥ 8	500-800	> 200		$0.0^{+0.8}_{-0.0}$	1.9 ± 1.5	2.8 ± 1.4	$0.1^{+0.4}_{-0.1}$	$4.8^{+2.3}_{-2.1}$	8
≥ 8	800-1000	> 200		0.6 ± 0.6	4.8 ± 2.9	2.3 ± 1.2	$0.5^{+0.9}_{-0.5}$	$8.3^{+3.4}_{-3.3}$	9
≥ 8	1000-1250	> 200		0.6 ± 0.5	$1.4^{+1.5}_{-1.4}$	2.9 ± 1.3	$0.7^{+1.0}_{-0.7}$	$5.6^{+2.3}_{-2.1}$	8
≥ 8	1250-1500	> 200		$0.0^{+0.9}_{-0.0}$	5.1 ± 3.5	1.4 ± 0.9	$0.5^{+0.9}_{-0.5}$	$7.1^{+3.8}_{-3.6}$	5
≥ 8	>1500	> 200		$0.0^{+0.7}_{-0.0}$	$0.0^{+4.2}_{-0.0}$	2.4 ± 1.4	$0.9^{+1.3}_{-0.9}$	$3.3^{+4.7}_{-1.7}$	2

probability to observe a statistical fluctuation as large or larger in any of the search regions.

Since no significant excess over the SM prediction is observed, the results are interpreted in several simplified supersymmetric models of pair production of light-flavour squarks or gluinos. The LSP is denoted as $\tilde{\chi}_1^0$. Several different decay modes are studied in the parameter space of the LSP and the squark or gluino which are

$$(a) \tilde{q} \rightarrow q + \tilde{\chi}_1^0$$

in case of light-flavour squarks and

$$(b) \tilde{g} \rightarrow q\bar{q} + \tilde{\chi}_1^0$$

$$(c) \tilde{g} \rightarrow t\bar{t} + \tilde{\chi}_1^0$$

$$(d) \tilde{g} \rightarrow q\bar{q} + \tilde{\chi}_1^\pm/\tilde{\chi}_2^0 \text{ where } \tilde{\chi}_1^\pm \rightarrow W + \tilde{\chi}_1^0 \text{ and } \tilde{\chi}_2^0 \rightarrow Z + \tilde{\chi}_1^0$$

for decays of the gluino. The branching ratios are assumed to be 100% for the different decay modes, except for case (d) where the decay via $\tilde{\chi}_1^+$, $\tilde{\chi}_1^-$ and $\tilde{\chi}_2^0$ is considered with equal probabilities.

Exclusion limits are derived with the modified CLs [169–171] approach and denote the 95% confidence level (CL) upper limit on the production cross section of the signal. The profile likelihood ratio is used as test statistics which is derived from the combined likelihood of the acceptance, efficiencies and uncertainties for the signal as well as the background predictions for all search regions. The uncertainties considered for the signal acceptance and efficiency in the limit setting procedure are

- The uncertainty on the integrated luminosity amounts to 2.6% [141].
- An uncertainty of 2% is considered for a possible trigger inefficiency (cf. Sec. 6.1.1).
- Due to the event cleaning criteria an uncertainty of 3% is considered.
- The propagation of the uncertainties on the jet energy calibration and resolution amount to uncertainties of 2 – 8% and 1 – 2% in the signal acceptance, respectively.
- The systematic variation of PDFs [172] leads to uncertainties of 1 – 8% in the signal acceptance.
- In order to match the measured rate of initial-state radiation in data the simulation of the signal events is corrected accordingly [173]. For model parameter points with small differences between the LSP and the gluino or squark mass this results in an uncertainty of 22% whereas it is usually less than a few percent for others.

Finally, the resulting exclusion limits for the above described processes (a)-(d) are shown in Fig. 6.17(a)-(d), respectively. The expected (*dashed*) and observed (*solid*) 95% CL upper limits are shown accordingly in the gluino-LSP and squark-LSP mass plane for the signal production cross sections. The one-standard-deviation uncertainty for the theory prediction is obtained by varying the renormalization and factorization scale by a factor of two and incorporating CTEQ6.6 [174] and MSTW2008 [175] as alternative PDF sets. By considering conservatively the observed limit minus the one sigma theory uncertainty, pair production of squarks of the first two generations is excluded below 780 GeV for a LSP

mass less than 200 GeV. However, if only one light squark is accessible the limit decreases to 400 GeV for LSP masses below 80 GeV. Similarly, the pair production of gluinos could be excluded for the three different decay modes (b)-(d) in case of a LSP mass less than 100 GeV for gluino masses up to 1.16 TeV, 1.13 TeV and 1.21 TeV, respectively.

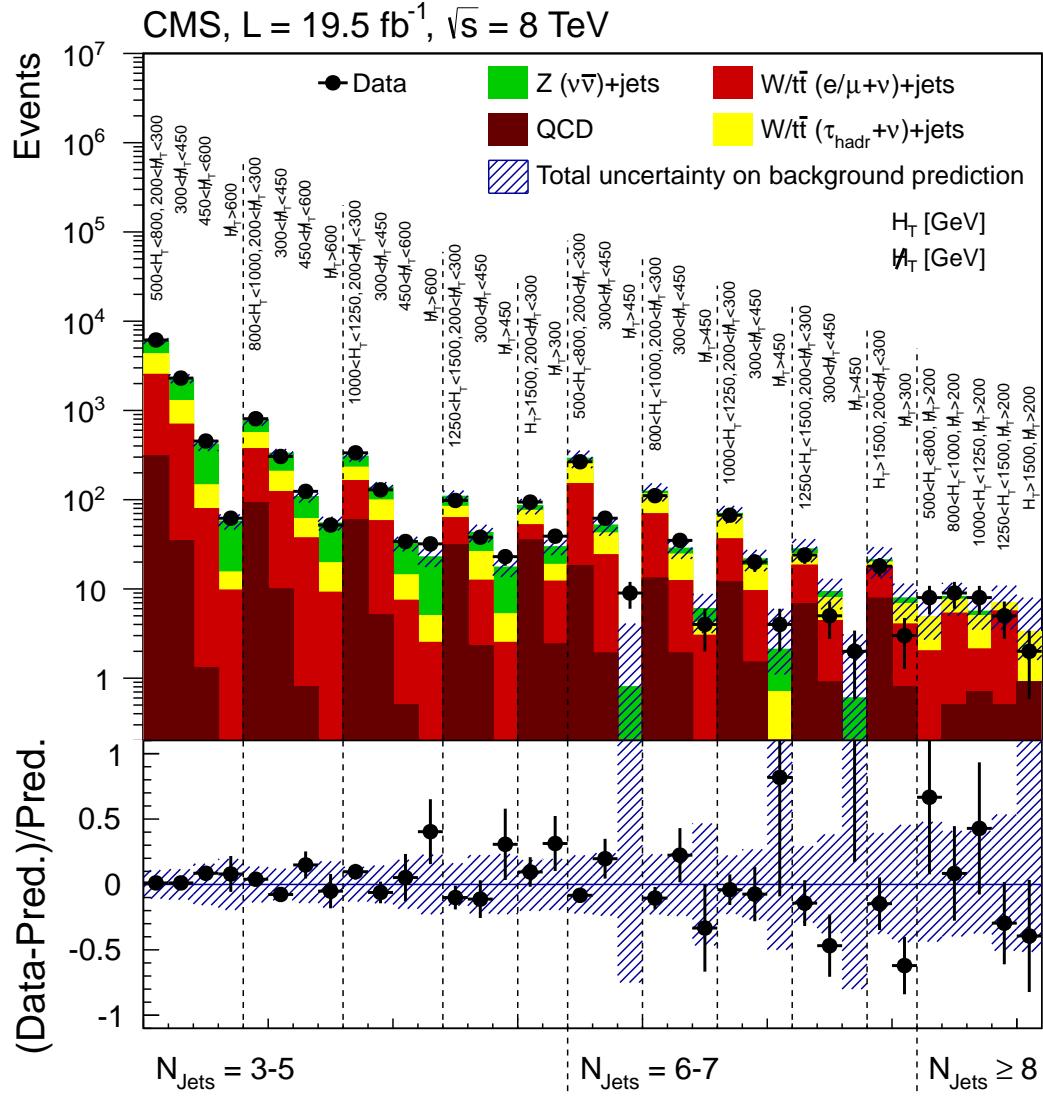


Figure 6.16.: Summary of the observed number of events in each of the 36 search regions in comparison to the corresponding background prediction. The hatched region shows the total uncertainty of the background prediction. Taken from [149].

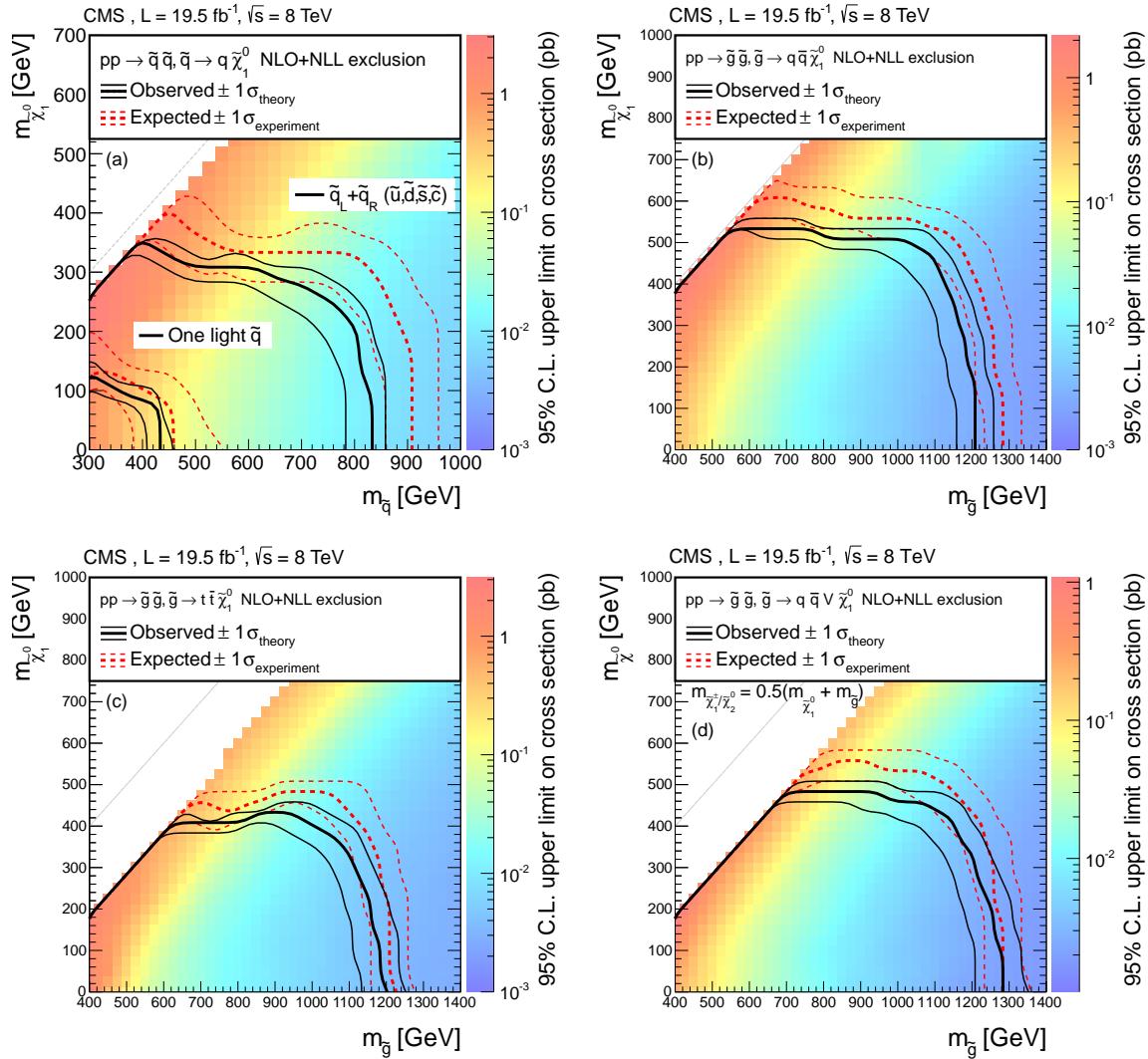


Figure 6.17.: The observed and expected 95% CL upper limits on the (a) squark-squark and (b-d) gluino-gluino production cross sections in either the $m(\text{squark})$ - $m(\text{LSP})$ or the $m(\text{gluino})$ - $m(\text{LSP})$ plane obtained with the simplified models. For the squark-squark production the upper set of curves corresponds to the scenario when the first two generations of squarks are degenerate and light, while the lower set corresponds to only one light accessible squark. Taken from [149].

6.4.1. Comparison to Other Measurements

The exclusion limits obtained with the analysis presented here exceed the exclusion limits derived from the 7 TeV analysis (cf. Fig. 2.8). Especially, the limit on the gluino mass is improved by around 200 GeV for light LSPs. Furthermore, the extension of the analysis into the N_{Jets} plane provided a good sensitivity towards the gluino-mediated production of third generation squarks and to decays involving W and Z bosons.

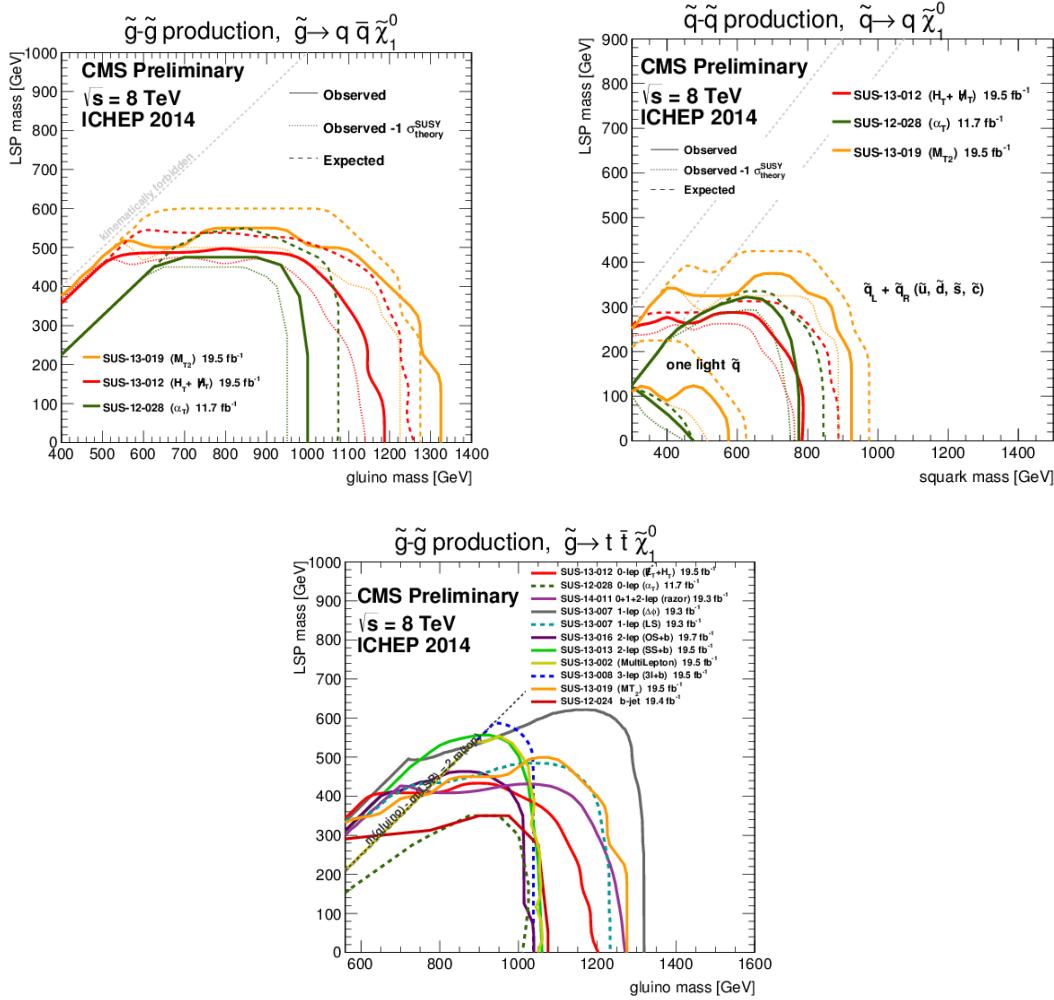


Figure 6.18.: Comparison of various exclusion limits derived by different CMS analyses for the process $\tilde{g} \rightarrow q\bar{q} + \tilde{\chi}_1^0$ (top left), $\tilde{q} \rightarrow q + \tilde{\chi}_1^0$ (top right) and $\tilde{g} \rightarrow t\bar{t} + \tilde{\chi}_1^0$ (bottom). Taken from [65]

6.5. Status of natural supersymmetry after LHC Run-I

7. Prospect Studies for a Search for Top Squarks in Events with Jets and Missing Transverse Momentum at $\sqrt{s} = 13$ TeV

The second run period of the LHC starting in 2015 at a centre of mass energy of $\sqrt{s} = 13$ TeV provides the excellent opportunity to further investigate the question, if supersymmetry is realised in nature at the TeV scale. As discussed in Section 6.5, the current limits exclude stop quarks with masses up to around 750 GeV for LSP masses below 100 GeV in case of direct stop production. Since in case of natural supersymmetry the stop quark mass is expected to not exceed the 1 TeV significantly, it is thus of particular interest to probe the stop mass range ≥ 750 GeV for light LSPs.

The targeted process of this analysis is illustrated in a diagram in Fig. 7.1 below.

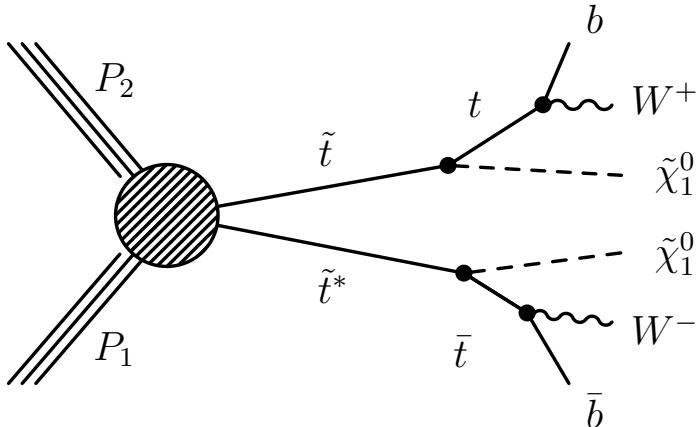


Figure 7.1.: Schematic diagram of the direct pair production of stop squarks in pp collisions with subsequent decay into a top quark and the LSP. Taken from [65].

Here, the pair production of top squarks is shown with subsequent decay into a top quark and the LSP. Since the top quark decays into a b quark and a W boson, the expected final state signature further depends on the decay modes of the W boson. In this analysis, only final states with fully hadronic top decays are considered. Consequently, this analysis targets a jet final state accompanied by missing transverse energy caused by the LSPs. As discussed in Section 2.3.2, background contributions from the SM to such signatures arise mainly from QCD multijet events, $W + \text{jets}$, $Z + \text{jets}$ and $t\bar{t}$ events.

In this Chapter, various analysis strategies for a search for top squarks at $\sqrt{s} = 13$ TeV are discussed and compared in order to address mass scenarios with large mass differences between the stop quark mass and the LSP. Furthermore, the performance of studied

Table 7.1.: Overview of simulated background and signal samples used in the analysis and corresponding production cross sections.

Process		σ	Precision
$t\bar{t}$		0.805 [nb]	NNLO
$W + \text{jets}$	$H_T = [0, 50]$ GeV	99.92 [nb]	LO
	$H_T = [50, 150]$ GeV	15.98 [nb]	LO
	$H_T = [150, 300]$ GeV	1.328 [nb]	LO
	$H_T = [300, \infty]$ GeV	0.169 [nb]	LO
$Z + \text{jets}$	$H_T = [0, 100]$ GeV	22.0 [nb]	LO
	$H_T = [100, 300]$ GeV	0.951 [nb]	LO
	$H_T = [300, \infty]$ GeV	0.0396 [nb]	LO
QCD multijet	$H_T = [100, 250]$ GeV	22 930 [nb]	LO
	$H_T = [250, 500]$ GeV	465 [nb]	LO
	$H_T = [500, 1000]$ GeV	18.66 [nb]	LO
	$H_T = [1000, \infty]$ GeV	0.536 [nb]	LO
Stop-pair production	$m_{\tilde{t}} = 600$ GeV	0.17460 [pb]	NLO
	$m_{\tilde{t}} = 700$ GeV	0.06705 [pb]	NLO
	$m_{\tilde{t}} = 800$ GeV	0.02833 [pb]	NLO
	$m_{\tilde{t}} = 900$ GeV	0.01289 [pb]	NLO
	$m_{\tilde{t}} = 1000$ GeV	0.00615 [pb]	NLO
	$m_{\tilde{t}} = 1100$ GeV	0.00307 [pb]	NLO
Gluino-pair production	$m_{\tilde{g}} = 1300$ GeV	0.0211 [pb]	NLO
	$m_{\tilde{g}} = 1500$ GeV	0.0064 [pb]	NLO
	$m_{\tilde{g}} = 1700$ GeV	0.0021 [pb]	NLO

selections is compared to the existing all-hadronic stop search performed by the CMS experiment at $\sqrt{s} = 8$ TeV which is published in [176]. In addition, it is interesting to study, if selections that are found to be suited for direct stop production are also sensitive to the gluino-mediated production of third generation quarks.

7.1. Data samples

The studies presented in this chapter are based on simulated samples at $\sqrt{s} = 13$ TeV. However, the pileup scenario considered for the samples is the same as for conditions present at $\sqrt{s} = 8$ TeV including out-of-time pileup according to a bunch spacing of 50 ns. Although these are not the pileup conditions expected for $\sqrt{s} = 13$ TeV, the selections studied here typically are based on objects with high transverse momenta such that influences from pileup are supposed to be negligible in good approximation. Nevertheless, possible disturbing effects and performance degradation due to pileup necessarily have to

be studied as soon as first collision data at $\sqrt{s} = 13$ TeV are available.

The processes considered in the analysis for background and signal events are summarized in Tab. 7.1. $W + \text{jets}$, $Z + \text{jets}$ and QCD multijet events are generated with MadGraph5 using the PDF CTEQ6L1 while $t\bar{t}$ events are generated with POWHEG and the MSTW2008 PDF. For all samples the showering process is performed with PYTHIA6 Tune Z2-star. The process $W + \text{jets}$ includes decays with $W \rightarrow l\nu$ with up to two jets. Similarly, also the process $Z + \text{jets}$ includes up to two jets. Here, decays $Z \rightarrow \nu\nu$ are modelled. For these processes, cross sections are considered at leading order. The scale factors to NLO cross sections are 1.04 and 1.07 for $W + \text{jets}$ and $Z + \text{jets}$, respectively.

For background processes $W + \text{jets}$, $Z + \text{jets}$ and QCD, samples are generated for different H_T selections at generator level. This procedure ensures that especially for the kinematic regions under study which typically concentrate on high- p_T objects a sufficient number of events is generated.

In order to study signal events, two different processes are generated. On the one hand direct stop-pair production is considered in which the \tilde{t} always decays to a stable neutralino $\tilde{\chi}_0^1$ and a top quark with mass $m_t = 172.5$ GeV including all decay channels of the top quark. These samples are generated with different neutralino masses of 50, 100, 150, 200, 250, 300 and 350 GeV. On the other hand gluino-pair production is generated with a decay $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_0^1$. For the top quark only fully hadronic decays are simulated. The cross section given in Tab. 7.1 for gluino-pair production is corrected for the hadronic branching fraction of the top quark. The neutralino mass is fixed to 50 GeV.

All samples are normalized to an integrated luminosity of 19.5 fb^{-1} which corresponds to the same amount of data recorded at $\sqrt{s} = 8$ TeV.

7.2. Sensitivity of a Basic Selection using H_T and \cancel{E}_T

The targeted signature of the direct pair production of stops in the all-hadronic channel is based on jets and missing transverse momentum similar to the search presented in Chapter 6. Thus, a very similar baseline selection is employed as a basis for further studies. An advantage of such a choice is that a synchronized baseline selection would allow to utilize the same triggers to collect the data. The applied selections are

- Background contributions arising from $t\bar{t}$ and $W + \text{jets}$ events are reduced by rejecting events containing isolated electrons or muons with $p_T > 10$ GeV and $|\eta| < 2.5$. These are required to have a good quality track that can be associated with the primary interaction vertex [157, 158]. The isolation is given as the scalar sum of transverse momenta of PF particles (except for the lepton itself) within a cone of width $\Delta R = 0.3$ for the electron and $\Delta R = 0.4$ for the muon, respectively. This is required to be less than 20% of the transverse momentum of the electron and less than 15% of the transverse momentum of the muon.
- The number of jets (N_{Jets}) is required to be ≥ 3 . N_{Jets} is defined as the number of jets with $p_T > 50$ GeV and $|\eta| < 2.5$. This requirement is imposed in order to select multijet events as expected from the two top quarks.

- The scalar sum of jet momenta (H_T) is required to be ≥ 500 GeV with

$$H_T = \sum_{\text{jets}} p_T$$

for all jets that have $p_T > 50$ GeV and $|\eta| < 2.5$. This condition selects events with a large visible energy in the event indicating a high energy scale of the hard interaction.

- The missing transverse energy \cancel{E}_T calculated from the PF candidates is required to be ≥ 200 GeV. This selection reduces contributions from standard model processes where missing transverse momentum is expected to be small. Especially, QCD multijet background is suppressed.
- In order to suppress events where missing transverse energy is mainly arising from jet mismeasurements, as for QCD multijet events, it is required that \cancel{E}_T is not aligned with the leading three jets. Thus, events with

$$\Delta\phi(\text{jet}_n, \cancel{E}_T) > 0.5 \text{ for } n = 1, 2 \text{ and } \Delta\phi(\text{jet}_3, \cancel{E}_T) > 0.3$$

are selected. The value of 0.5 is chosen according to the jet size parameter. However, this is reduced in case of the third jet in order to retain signal efficiency.

Since only simulated events are used, no dedicated event cleaning filters are applied as it is necessary for data (cf. Sec 6.1.2). The selection described here is denoted *baseline* selection in the following. In Tab. 7.2 the event yields of background and two signal processes obtained from simulation normalized to an integrated luminosity of 19.5 fb^{-1} are shown after various requirements together with the statistical uncertainty. It is visible that the \cancel{E}_T and $\Delta\phi$ requirement reject most of the background events. After the baseline selection the background is composed quite equally of all four SM processes. Furthermore, in Fig. 7.2, the obtained spectra of H_T , \cancel{E}_T and N_{Jets} after applying the baseline selection are shown for the SM backgrounds and two selected signal points. The signal points represent mass values of 600 GeV and 1100 GeV for the stop mass while the LSP mass is in both cases 50 GeV. These two signal points illustrate the difference in the kinematic properties of events for low and high stop masses. Typically, the H_T and \cancel{E}_T spectrum get harder for higher stop masses while the shape of the N_{Jets} spectrum stays nearly unchanged.

In order to investigate how this baseline selection can be further improved to gain sensitivity to the model of interest the evolution of the signal and background efficiencies is studied when changing specific selections in the analysis. In general, the signal and background efficiencies $\epsilon_{\text{sig/bg}}$ are defined according to

$$\epsilon = \frac{\text{no. of selected events}}{\text{no. of all events}} \quad (7.1)$$

for the number of signal and background events, respectively. In Fig. 7.3 the development of the signal versus background efficiencies are shown for the stop mass 600 GeV and 1100 GeV signal points and $t\bar{t}$ as well as the total background events. The curves are

Table 7.2.: Event yields and cut flow from MC simulated samples after various requirements described in the text. All numbers are scaled to 19.5 fb^{-1} . Only statistical uncertainties are shown in the table. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass.

Process	e/ μ veto	jet counting	H_T	\cancel{E}_T	$\Delta\phi$
$t\bar{t}$	10712200 ± 9170	6431190 ± 7105	1274570 ± 3163	34616.8 ± 521.3	16444.9 ± 359.3
$W + \text{jets}$	713523000 ± 159183000	1522960 ± 20432.5	196599 ± 1259.5	25104.5 ± 295.7	12520 ± 211.6
$Z + \text{jets}$	203002000 ± 408457	364126 ± 4090.1	60043.1 ± 212.4	16245.7 ± 104.9	11858.0 ± 91.1
QCD multijet	$4.96 \cdot 10^{11} \pm 6.89 \cdot 10^8$	$1.01 \cdot 10^{10} \pm 8.58 \cdot 10^7$	$4.43 \cdot 1^8 \pm 3.54 \cdot 10^6$	109456 ± 19090	20013 ± 18085
Signal (600, 50)	2252.7 ± 8.8	1994.7 ± 8.3	1478.3 ± 7.1	1157.9 ± 6.3	1013.1 ± 5.9
Signal (1100, 50)	42.9 ± 0.2	37.3 ± 0.2	36.0 ± 0.1	33.2 ± 0.1	29.2 ± 0.1

obtained from increasing the cut value for the denoted variable by keeping the selection for all other variables fixed. The curve of a variable with good separation power runs close to the lower right corner. Here, the performance of H_T , \cancel{E}_T and N_{Jets} is compared. In addition to \cancel{E}_T , also $\cancel{\cancel{E}}_T$, as defined in Sec. 6.1.3, is shown in this comparison. Both variables show a very similar performance such that in principle both variables could be employed with similar results. However, the performance of \cancel{E}_T seems to be slightly better and thus this variable is used in these studies rather than $\cancel{\cancel{E}}_T$. Furthermore, as stated above the same trigger as for the analysis presented in Chapter 6 could be used to collect the data. Since this trigger is based on \cancel{E}_T , the choice of the offline selections is closer to the online definition when using \cancel{E}_T instead of $\cancel{\cancel{E}}_T$. As can be seen in Fig. 7.3, \cancel{E}_T shows in general the best performance when comparing these variables. Furthermore, also H_T provides a good separation power and is only in case of the stop mass of 600 GeV inferior to N_{Jets} . However, when only $t\bar{t}$ background is considered, the jet multiplicity is not suitable as discriminating variable since the N_{Jets} spectra of signal and $t\bar{t}$ background are almost identical. As in general the kinematic features of $t\bar{t}$ background are closest to the signal, it is of special importance to identify selections that can reduce this background. Consequently, \cancel{E}_T and H_T are the preferred variables to distinguish signal and background processes. In the following, an analysis strategy close to the one described in Chapter 6 is pursued. Events selected with the baseline selection are further categorized according to their H_T and \cancel{E}_T values in exclusive search regions shown in Tab. 7.3.

In order to determine the expected exclusion reach, the uncertainties for the individual background processes have to be considered. These are chosen in correspondence to the uncertainties obtained for similar kinematic regimes in the published all-hadronic stop analysis [176]. The respective uncertainties considered for the different processes are

- QCD multijet events: 100%
- $Z + \text{jets}$: 50%
- $W + \text{jets}$: 20%
- $t\bar{t}$: 20% + additional 20% in the high \cancel{E}_T search regions

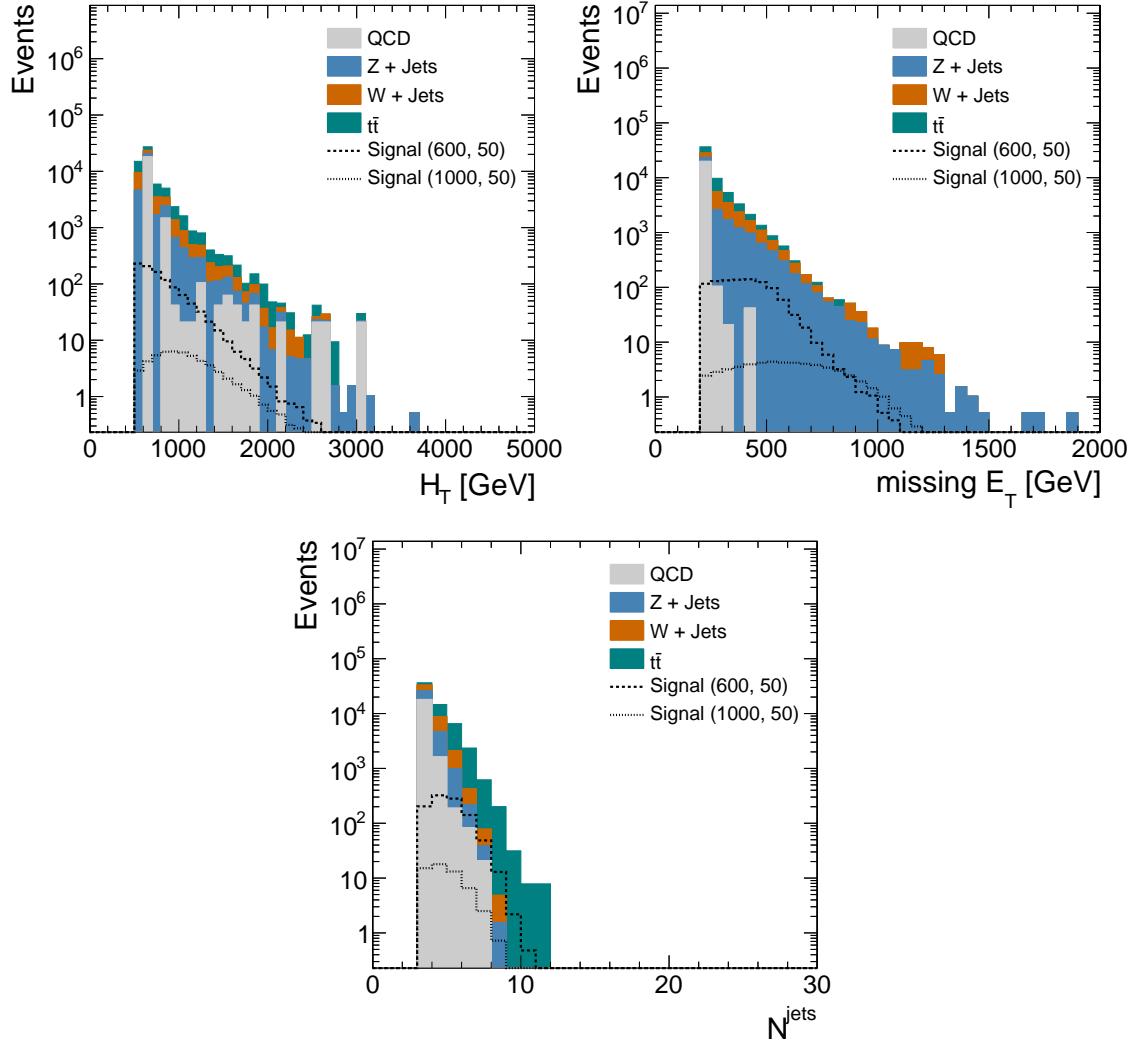


Figure 7.2.: Comparison of selected H_T (top left), \cancel{E}_T (top right) and N_{jets} (bottom) distributions in simulated events found from applying the baseline selection criteria. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass.

which are meant to be total uncertainties such that the actual statistical uncertainty of the number of simulated MC events is not considered explicitly. Based on the selected event yields and the estimated uncertainties the 95% confidence-level expected upper limit is calculated as an asymptotic CL_s limit [177]. The obtained exclusion curve is shown in Fig. 7.4 for the signal strength μ which is the excluded production cross section divided by the theoretical cross section for direct stop production as function of the stop quark mass. The LSP mass is fixed to 50 GeV in order to study the sensitivity of the selection for signal scenarios with large mass splittings between stop and neutralino mass. A particular mass point can be excluded, if the expected limit drops below one. However, it turns out that this baseline selection cuts and the subsequent binning in exclusive search

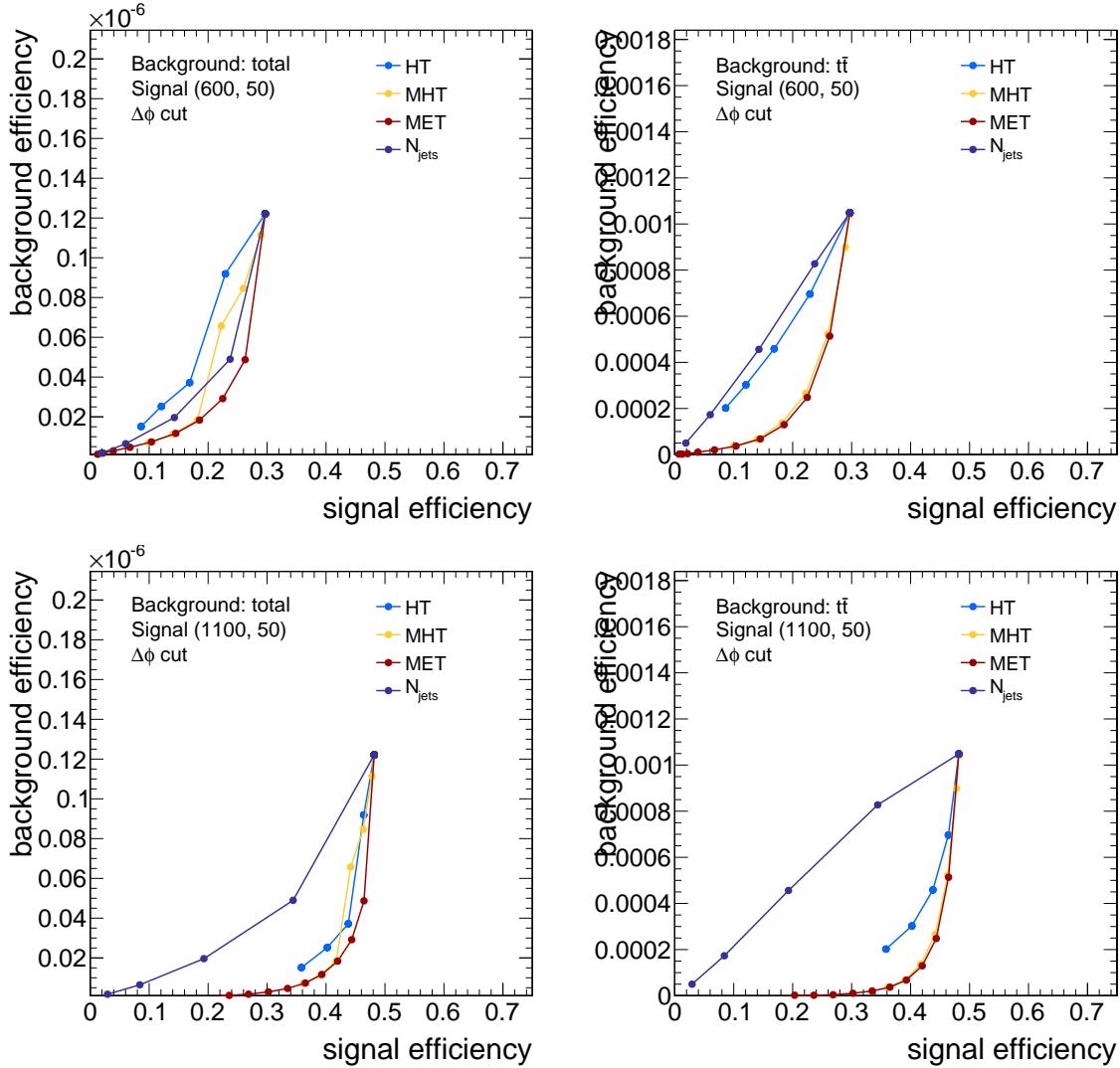


Figure 7.3.: Evolution of the signal versus background efficiency for a stop mass of 600 GeV (top) and 1100 GeV (bottom) where the background is the sum of all backgrounds (left) or the $t\bar{t}$ background (right).

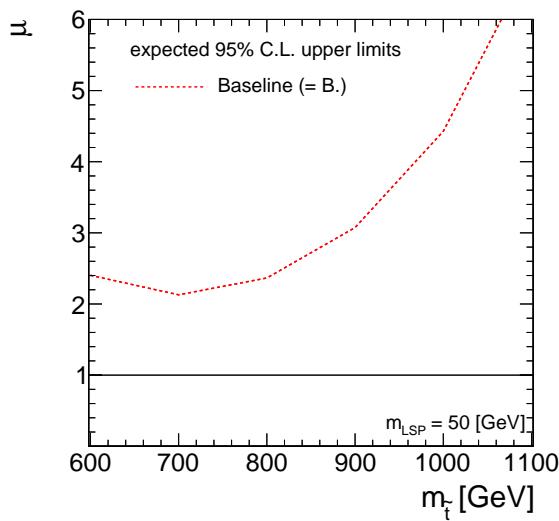
regions is not yet sensitive enough to probe any of the selected mass points. Thus, possible improvements of the analysis are discussed in the following sections.

7.3. Sensitivity Improvement using b Tagging

As illustrated in Fig. 7.1, the targeted signal final state involves the presence of bottom quarks emerging from the decay of the top quarks. Thus, a very common thing to possibly enhance the sensitivity of such an analysis is to employ b-tagging techniques to identify the b quarks in the final state. Typical b-tagging algorithms for the identification of b-quark jets used within the CMS experiment have been discussed in Section 4.3. In this analysis

Table 7.3.: Exclusive search regions used in the analysis binned in H_T and \cancel{E}_T .

	\cancel{E}_T [GeV]
$H_T = 500 - 1000$ GeV	200 - 400
	> 400
$H_T > 1000$ GeV	200 - 400
	> 400

Figure 7.4.: Expected 95% upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV.

b-quark jets are identified based on the CSV-algorithm using the medium working point. In Fig. 7.5 the b-tag multiplicity, i. e. the number of b-tagged jets in the event is illustrated. As expected the maximum of this distribution for signal and $t\bar{t}$ events is around two, since two b-quark jets are expected from the two decaying top quarks. Values above two or b-tagged jets in backgrounds not containing real b-quarks can be explained by the misidentification rate of the tagging algorithm.

The resulting distributions for H_T , \cancel{E}_T and N_{Jets} after requiring the presence of at least one b-tagged jet in the event are illustrated in Fig. C.1 in App. C. It turns out that all backgrounds except for $t\bar{t}$ can be significantly reduced when requiring the presence of at least one b-quark jet. This is expected, since the $t\bar{t}$ background is the only background containing real b quarks.

Consequently, the expected exclusion reach of the analysis is supposed to improve when applying the b-tag requirement in addition to the baseline selection. Thus, the same exclusive search regions as defined in Tab. 7.3 are used considering the same total uncertainties as described in Sec. 7.2 for a performance comparison of this improved selection with respect to the baseline requirements. The expected limits for the baseline selection as well as the baseline selection and the b-tag requirement are shown in Fig 7.6. It turns out

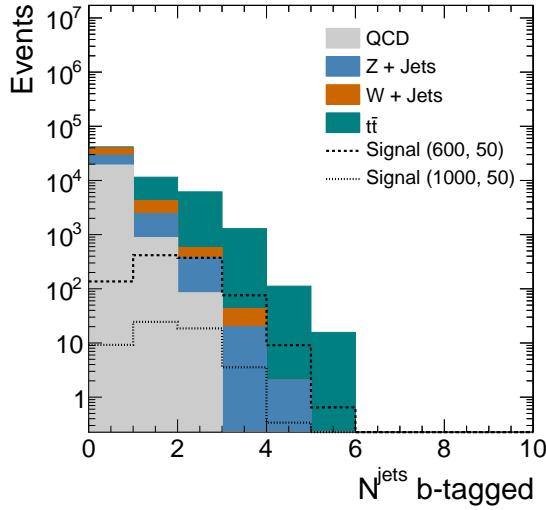


Figure 7.5.: B-tag multiplicity after applying the baseline selection in simulated events. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass.

that the b-tag requirement significantly improves the sensitivity of the analysis towards the specified top squark mass range. Especially lower masses could be tested with such an analysis strategy. However, since the focus of this analysis is put on higher stop quark masses, it is discussed in the next section how this mass range can be addressed better.

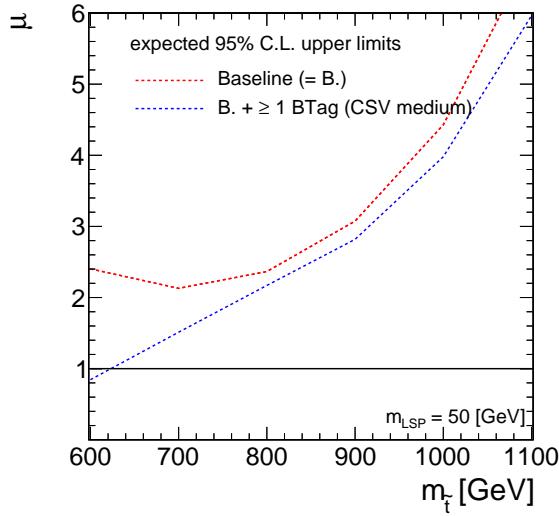


Figure 7.6.: Expected 95% upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV.

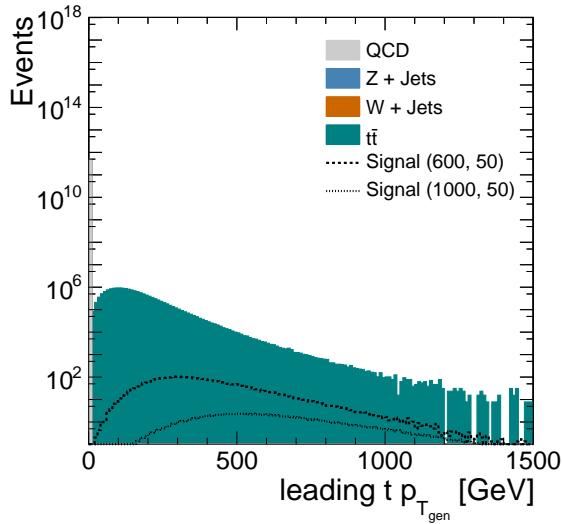


Figure 7.7.: Transverse momentum spectrum of the leading generated top quarks without applying any selection criteria. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass.

7.4. Sensitivity Improvement using top Tagging

In order to gain a better understanding of the kinematics of the investigated final state, the p_T spectrum of the leading generated top quark is illustrated in Fig. 7.7 for $t\bar{t}$ background and two selected signal points without applying any selection criteria. It turns out that the p_T spectrum of the signal is significantly harder than that of the $t\bar{t}$ background. For instance for a stop mass of 1000 GeV the maximum lies at a p_T range around 400-500 GeV. As discussed in Sec. 4.4, decay products emerging from the decay of a top quark with large transverse momentum do not show up as separated objects in the detector, but merge into jets with large radius parameter. Following Equation 4.4 the opening angle of decay products from a top quark with transverse momentum between 400-500 GeV is supposed to be $R = 0.8$. Thus, such topologies are well suited to utilize the top tagging techniques described in Sec. 4.4.

The performance of the CMS- and the HEP-top-tagging algorithms are investigated based on the 13 TeV simulation samples and reviewed in the following.

Top Tagging Efficiency Studies: In order to evaluate the performance of the top-tagging algorithms, the top tagging efficiencies and misidentification rates are derived. While the top tagging efficiency is determined for $t\bar{t}$ events, the QCD multijet sample is used to measure the misidentification rate.

The top tag efficiency is defined as the number of hadronically decaying generated top quarks matched to a top-tagged CA-jet divided by the number of all generated hadronically decaying generated top quarks. A successful match is identified by requiring the ΔR of a generated top quark and a top-tagged CA-jet to be less than the jet radius parameter which is $R = 0.8$ for the CMS top tagger and $R = 1.5$ for the HEP top tagger, respectively. The obtained efficiency as function of the transverse momentum of the generated hadronically decaying top quark is shown in the left

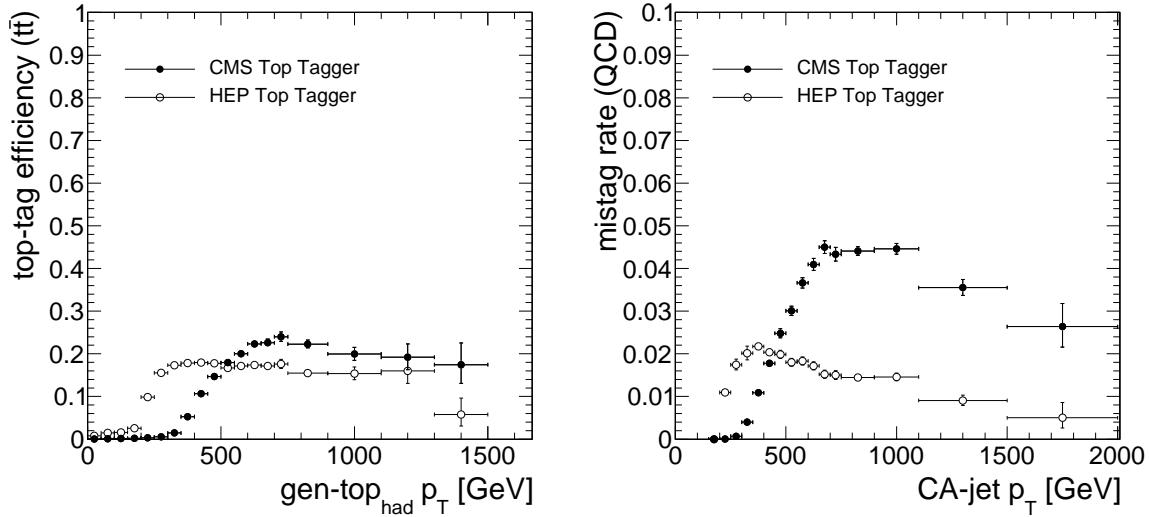


Figure 7.8.: Top tag efficiency for $t\bar{t}$ events as function of the transverse momentum of the generated top quark (*left*) and misidentification rate for QCD multijet events as function of the CA-jet p_T (*right*). In case of the CMS top tagger CA-jets with a distance parameter of $R = 0.8$ are used while the HEP top tagger is based on CA-jets with $R = 1.5$.

part of Fig. 7.8 for both the CMS and the HEP top tagger. It is visible that the turn-on of the HEP top tagger starts already around 200 GeV while the CMS tagger begins to become efficient not before around 400 GeV. However, the efficiency of the HEP top tagger lies around 20% in the plateau region while the plateau efficiency of the CMS top tagger is in general higher with a value of around 25%. This behaviour results mainly from the different jet sizes and selection criteria which are in case of the HEP top tagger optimized to be sensitive already in the p_T range around 200-300 GeV.

The misidentification rate can be evaluated by dividing the number of top-ttagged CA-jets by the number of all CA-jets. Again, in case of the CMS top tagger the jet distance parameter is $R = 0.8$ and for the HEP top tagger it is $R = 1.5$. The misidentification rate as function of the respective CA-jet transverse momentum is shown in the right part of Fig. 7.8. Here, a similar feature as for the efficiency curve is observed. The HEP top tagger shows a certain misidentification rate already for lower transverse momenta than the CMS top tagger. However, in general the misidentification rate of the HEP top tagger is significantly smaller with a plateau around 1.5% compared to the CMS top tagger which shows a misidentification rate of up to 4-5% in the plateau.

Consequently, in analyses which rely on top quarks with moderate transverse momenta and suffer mainly from QCD background the HEP top tagger is a good choice since the efficiency has an early turn-on and the misidentification rate is small. However, when the main background is top-like, as it is the case for top squarks, the more important property is to have an efficiency as high as possible to retain signal efficiency while the actual misidentification rate only plays a minor role.

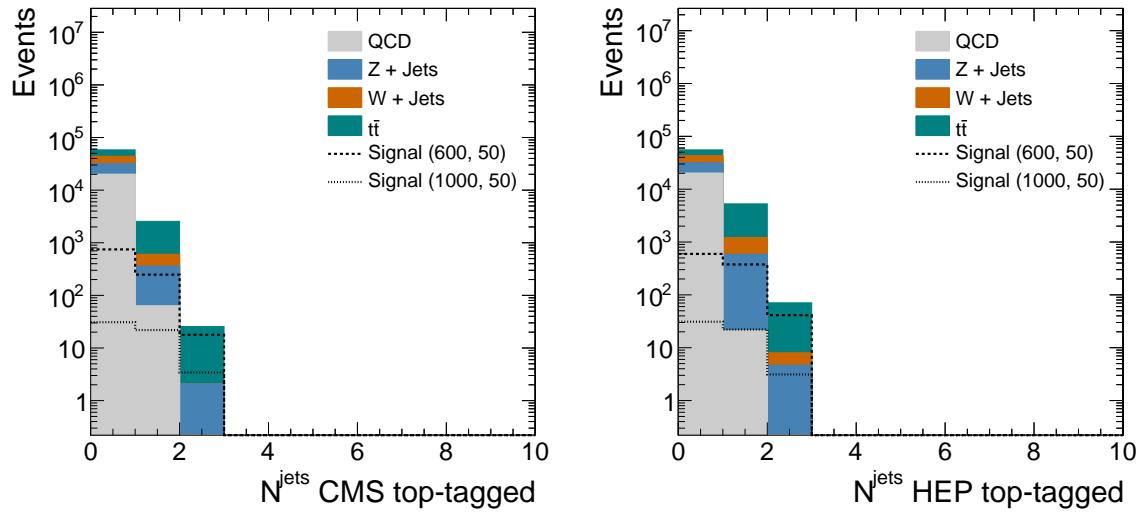


Figure 7.9.: Top tag multiplicity for the CMS top tagger (*left*) and the HEP top tagger (*right*) after application of the baseline selection. Only CA-jets of the corresponding jet size with a transverse momentum above 150 GeV are considered.

The impact on the analysis when applying top tag requirements can be studied when looking for instance at the top tag multiplicity. In Fig. 7.9 the top tag multiplicity is shown for the CMS top tagger (left) and the HEP top tagger (right) after the application of the baseline selection. Here, only CA-jets of the corresponding jet size with a transverse momentum above 150 GeV are considered since the taggers are not supposed to be efficient for smaller momenta. The distributions exhibit that although also a significant amount of signal events does not have a CA-jet identified as top jet by the respective algorithm the background can be significantly reduced when requiring at least one top-tagged jet. In general the number of signal events with one or two top tags is larger when using the HEP top tagger than for the CMS tagger, but also the number of background events from $t\bar{t}$ shows a larger value for the HEP top tagger case since more $t\bar{t}$ events with lower top transverse momenta are selected.

In order to test the impact of the top tagging requirement on the search sensitivity, the expected limits are derived again based on the search regions defined in Tab. 7.3 with the same uncertainties as assumed in Sec. 7.2. The results are illustrated in Fig 7.15.

It turns out that for the HEP top tagger as well as for the CMS top tagger the usage of a top tag requirement in addition to the baseline selections significantly improves the search sensitivity by a factor of two to three. Hence, mass points up to $(m_{\tilde{t}}, m_{\text{LSP}}) = (700, 50)$ or even $(800, 50)$ can be probed. This selection is also better than the result obtained when adding a b-tag requirement as discussed in Sec. 7.4 which allows to probe stop masses around 600 GeV for an LSP mass of 50 GeV. In general, the selection based on the CMS top tagger performs better than the selection involving the HEP top tagger, since the signal to background ratio is higher as discussed above.

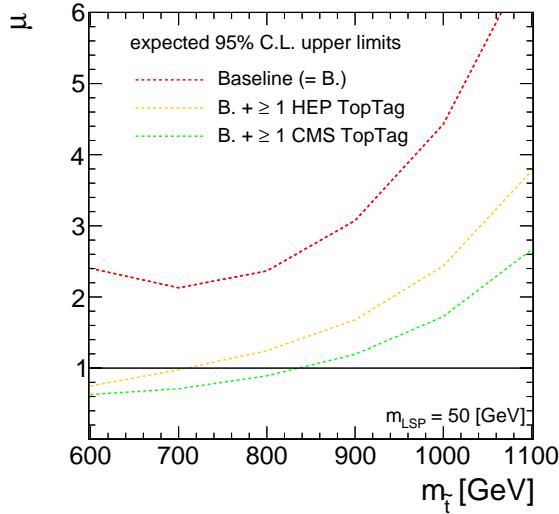


Figure 7.10.: Expected 95% upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV.

7.5. Performance Comparison of Various Kinematic Selections

The sensitivity of the analysis for top squark searches based on the baseline requirements discussed in Sec. 7.2 can be improved already significantly when employing in addition b-tag and top-tag requirements, as shown in Sec. 7.4 and 7.4. Nevertheless, the sensitivity might still be improved when imposing further kinematic selections.

In order to address the specific kinematics of the various background contributions, first the background composition is studied. In Fig. 7.11 the background and signal processes are shown according to the decay mode of the process after the application of the baseline selection (left) and the baseline selection with an additional requirement of ≥ 1 CMS Top Tag (right). These channels are defined according to the top quark properties based on generator information. The first channel includes all non-top backgrounds and the all-hadronic top decays. Channels two to four exhibit the semi-leptonic top decays which are involving electrons (channel two), muons (channel three) and tau leptons (channel four). The distributions exhibit that after the application of the top-tagging requirement the main background contribution arises from $t\bar{t}$ events. Here, contributions from lost-leptons (channel two and three) and hadronically decaying tau leptons (channel four) occur in about equal amounts with slightly more hadronic tau events. Since in general the number of events is small after application of the top tag requirement, kinematic selections and their performance are investigated when applying baseline selection criteria. Nonetheless, since $t\bar{t}$ is known to be the largest background after top tagging, special emphasis is put on reducing the $t\bar{t}$ background contributions.

Several kinematic variables exist that have been successfully used already in various susy searches at $\sqrt{s} = 7$ and 8 TeV:

α_T : The α_T variable is mainly used to reject QCD multijet events which have no intrinsic

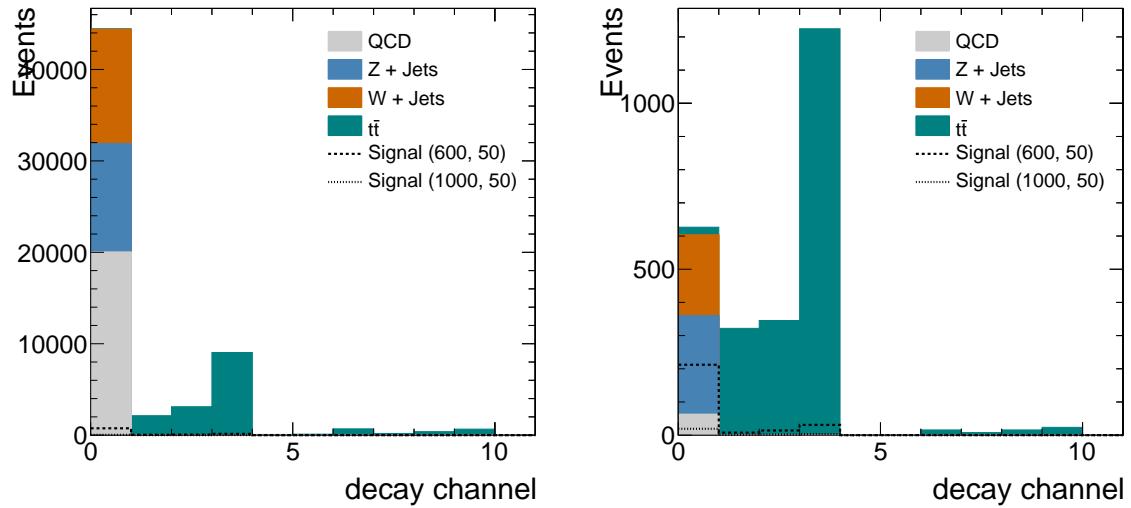


Figure 7.11.: Comparison of decay modes of signal and background processes defined according to the top quark properties based on generator information after the baseline selection (*left*) and after the baseline selection and the requirement of ≥ 1 CMS Top Tag (*right*). For a definition of the decay channels see text.

\cancel{E}_T [159, 178]. In case of dijet events, it is defined as:

$$\alpha_T = \frac{E_T^{j_2}}{M_T} \text{ with } M_T = \sqrt{\left(\sum_{i=1}^2 E_T^{j_i} \right)^2 - \left(\sum_{i=1}^2 p_x^{j_i} \right)^2 - \left(\sum_{i=1}^2 p_y^{j_i} \right)^2} \quad (7.2)$$

with the transverse energy $E_T^{j_2}$ of the less energetic jet and the transverse mass M_T of the dijet system. Since in case of boosted top quark decays the two top quarks are supposed to be represented each by one fatjet, the α_T variable is in this studies calculated from the two leading CA8 jets. The distribution of the α_T variable after application of the baseline selection requirements is shown in Fig. 7.12 (top left). Typically, α_T is 0.5 for an ideal dijet event with $E_T^{j_1} = E_T^{j_2}$ in which each jet momentum is large compared to its mass. If an imbalance occurs due to a jet mismeasurement, α_T drops below 0.5 while it is greater than 0.5 when the two jets recoil against real \cancel{E}_T , as for instance resulting from LSPs.

M_{T2} : The M_{T2} variable represents a generalized version of the transverse mass [161, 179–181]. In events with pair-produced particles which decay further and eventually contain undetectable particles in the decay products, as e. g. LSPs, the event kinematics are underconstrained and thus a classical transverse mass can not be determined. The M_{T2} variable is defined for two identical decay chains as

$$M_{T2}(m_{\tilde{\chi}}) = \min_{\vec{p}_T^{\tilde{\chi}(1)} + \vec{p}_T^{\tilde{\chi}(2)} = \vec{p}_T^{\text{miss}}} \left[\max \left(M_T^{(1)}, M_T^{(2)} \right) \right] \quad (7.3)$$

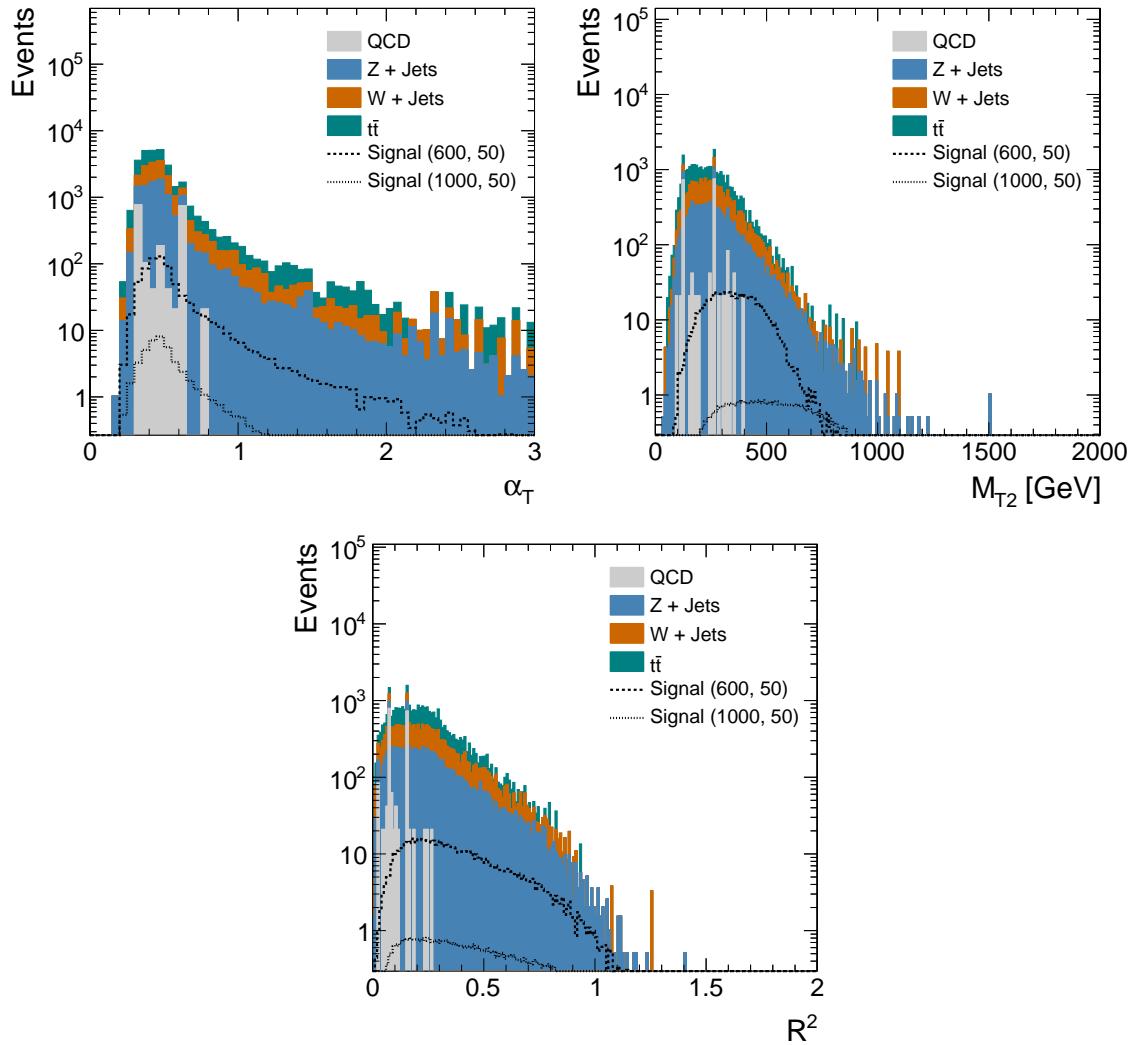


Figure 7.12.: Distribution of the α_T variable (*top left*), M_{T2} (*top right*) and the razor variable R^2 (*bottom*) after application of the baseline selection requirements for background and two selected signal samples. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass.

with the two transverse masses ($i = 1, 2$)

$$(M_T^{(i)})^2 = (m^{\text{vis}(i)})^2 + m_{\tilde{\chi}}^2 + 2 \left(E_T^{\text{vis}(i)} E_T^{\tilde{\chi}(i)} - \vec{p}_T^{\text{vis}(i)} \cdot \vec{p}_T^{\tilde{\chi}(i)} \right) \quad (7.4)$$

described by the transverse momenta $\vec{p}_T^{\text{vis}(i)}$, transverse energies $E_T^{\text{vis}(i)}$ and masses $m^{\text{vis}(i)}$ for the visible systems and the unknown transverse momenta of the LSPs $\vec{p}_T^{\tilde{\chi}(i)}$. Experimentally, the momenta $\vec{p}_T^{\tilde{\chi}(i)}$ are not accessible separately. Thus, a minimization on trial LSP masses fulfilling the constraint given by \vec{p}_T^{miss} , the missing

transverse momentum¹ is performed. This minimization is carried out to make sure that M_{T2} does not exceed the mass of the parent particle. For the correct value of $m_{\tilde{\chi}}$, the distribution of M_{T2} is expected to have an endpoint at the mass of the parent particle.

In this studies, the two visible systems are assumed to be described by the two leading CA8 jets in the event. The mass of the neutralinos is considered to be zero in the calculation of M_{T2} .

The obtained distributions for M_{T2} in background and signal samples are shown in Fig 7.12 (top right). They exhibit that the maximum of M_{T2} is higher for signal events than for background such that this variable is supposed to be well suited to distinguish signal and background events.

Razor variables: The kinematic razor variables are used to quantify the generic process of pair production of two heavy particles which subsequently decay into visible products usually represented by (large) jets and undetected particles [162, 182, 183]. These variables are used to test, if the two jets represent the visible part of the decay of two heavy objects.

The razor variables are defined as

$$M_R = \sqrt{[(|\vec{p}^{j1}| + |\vec{p}^{j2}|)^2 - (p_z^{j1} + p_z^{j2})^2]} \quad (7.5)$$

$$M_T^R = \sqrt{\frac{1}{2} (\cancel{E}_T(p_T^{j1} + p_T^{j2}) - \cancel{E}_T \cdot (\vec{p}^{j1} + \vec{p}^{j2}))} \quad (7.6)$$

$$R = \frac{M_T^R}{M_R} \quad (7.7)$$

with the transverse momenta of the two jets p_T^{j1} and p_T^{j2} . For signal events M_T^R has an endpoint and R a maximum of approximately one.

For these studies, the two leading CA8 jets and \cancel{E}_T are used to compute the razor variables. In Fig. 7.12 the variable R^2 is illustrated (bottom) after the application of baseline selection criteria.

In addition to these variables, which are already well established in searches for supersymmetry, also other kinematic quantities can be considered

Transverse mass m_T : As shown in Fig. 7.11, background contributions from $t\bar{t}$ events arise predominantly from semi-leptonic top quark decays. In such decays the missing transverse energy is mainly arising from the leptonically decaying top quark. This is caused by neutrinos from the decay of the W boson and becomes even more prominent for lost-lepton events in which also the undetected lepton contributes to the missing energy. In both cases however, the missing energy is expected to point into the direction of the leptonic top decay which is supposed to be accompanied by a b-quark jet also stemming from the top decay. Thus, the missing transverse energy and the closest b-tagged jet in $\Delta\phi$ can be utilized to calculate a transverse

¹More commonly denoted by \cancel{E}_T in this thesis.

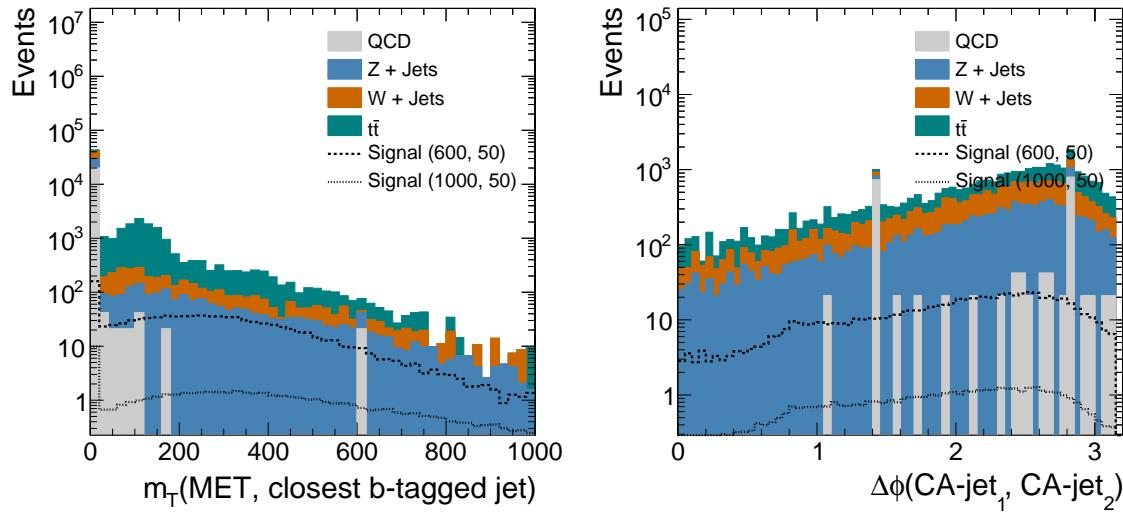


Figure 7.13.: Distribution of the m_T variable (left) and $\Delta\phi(\text{CA-jet}_1, \text{CA-jet}_2)$ (right) after application of the baseline selection requirements for background and two selected signal samples. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass.

mass according to

$$m_T = \sqrt{2p_T^{\text{jet}} E_T \cdot (1 - \cos(\Delta\phi(\text{jet}, E_T)))} \quad (7.8)$$

The distribution obtained after applying the baseline selection is shown for background and two selected signal samples in Fig. 7.13 (left). Here, the transverse mass is calculated from the missing transverse momentum and the closest b-tagged anti- k_T jet with $R = 0.5$ as identified by the CSV algorithm with medium working point and transverse momentum greater than 30 GeV. It is treated as zero in case no b-tagged jet could be identified. The distributions exhibit that this transverse mass variable has a peak in case of $t\bar{t}$ events close to the top quark mass while it is relatively flat for signal events.

$\Delta\phi(\text{CA-jet}_1, \text{CA-jet}_2)$: The selection of events with a back-to-back topology has been used in chapter 5 in order to identify dijet events by requiring $\Delta\phi > 2.7$. A similar situation can be expected to occur when having event topologies with boosted top quark decays in $t\bar{t}$ events. For high transverse momenta a $t\bar{t}$ event features a dijet-like structure of two fatjets balanced against each other. For signal events however, such a topology is not expected, since the event is balanced against genuine E_T arising from the LSPs. The selected $\Delta\phi$ distributions after applying baseline selection criteria are shown in Fig. 7.13 (right).

In order to quantify the quality of these various kinematic selections, the separation power of the different variables is tested again by studying the background versus signal efficiency curves as introduced in Sec. 7.2. The resulting distributions after applying baseline selection criteria are shown in Fig. 7.14. For comparison also the signal versus background

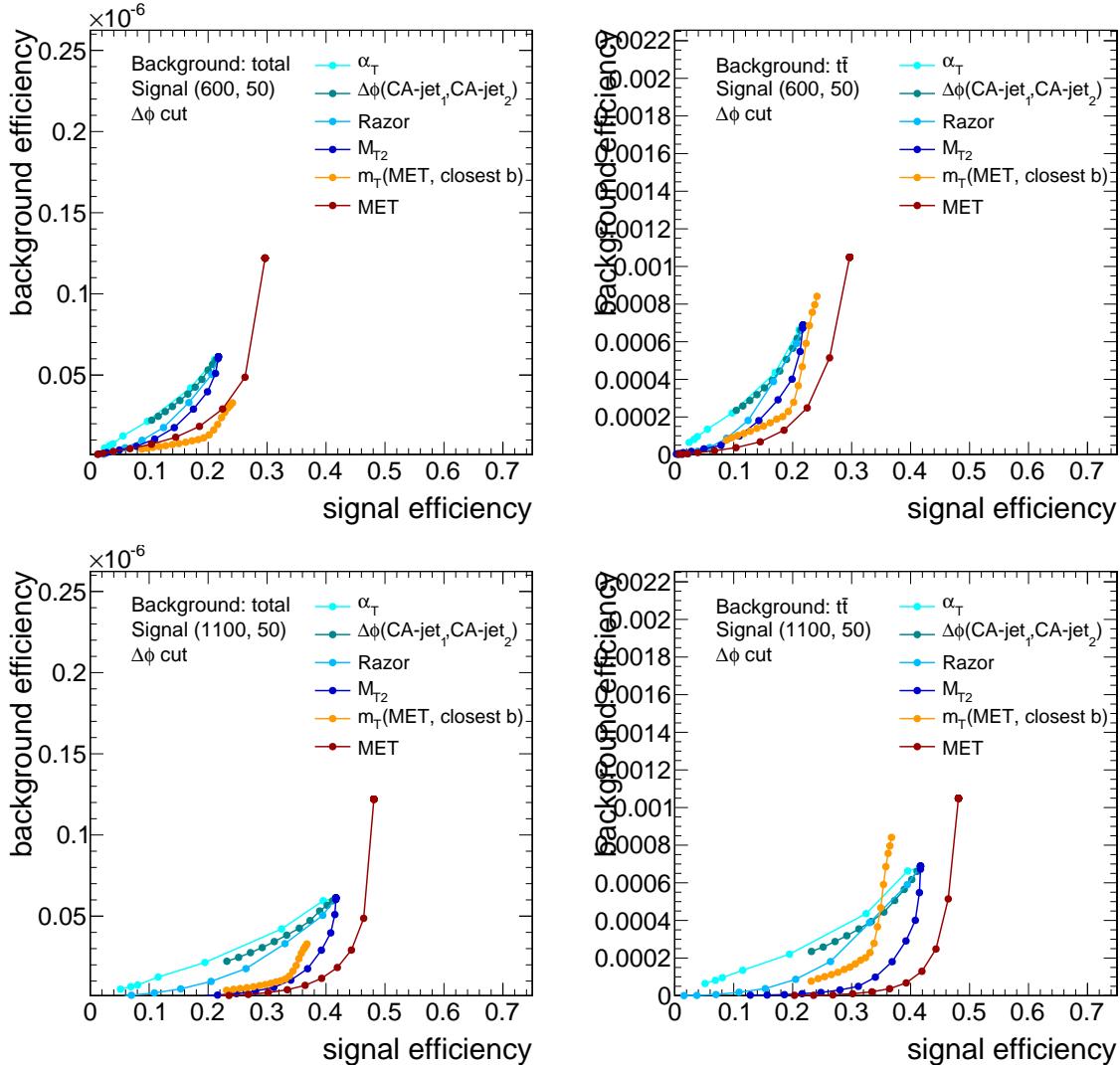


Figure 7.14.: Evolution of the signal versus background efficiency for a stop mass of 600 GeV (*top*) and 1100 GeV (*bottom*) where the background is the sum of all backgrounds (*left*) or the $t\bar{t}$ background (*right*).

efficiency curve for \cancel{E}_T is illustrated.

The calculation of α_T , $\Delta\phi(\text{CA-jet}_1, \text{CA-jet}_2)$, M_{T2} and the razor variable R^2 imply that there exist two CA8 jets with $p_T > 150$ GeV in the event. Consequently, the starting point of that scan curves is different than that of the \cancel{E}_T curve which represents the signal and background efficiencies after the baseline selection. Similarly the curve for m_T implies that each event has at least one b-tagged jet. For each event which does not possess a b-tagged jet, the transverse mass is considered as zero. The first point of the m_T curve corresponds to a requirement of $m_T > 20$ GeV.

The distributions exhibit that among the variables based on the presence of two CA8 jets M_{T2} performs best especially for high top squark masses. This statement applies to the total background as well as to the $t\bar{t}$ background only. Furthermore, the selection based

on m_T shows a nice separation power as well but preferentially for smaller stop masses. It is visible that the curve for m_T shows a distinct kink. This corresponds to a selection keeping only events with m_T greater than the top quark mass. When imposing even tighter selection requirements on m_T , the separation power decreases rapidly.

Consequently, the sensitivity of the analysis is tested adding further selection criteria to the baseline requirements after the requirement of ≥ 1 CMS top tag. In Fig. 7.15 the expected exclusion limits are compared for various selections as function of the top squark mass for a LSP mass of 50 GeV. Exclusion curves are derived based on the same exclusive search regions as defined in Tab. 7.3 considering the uncertainties for each background source as described in Sec. 7.2.

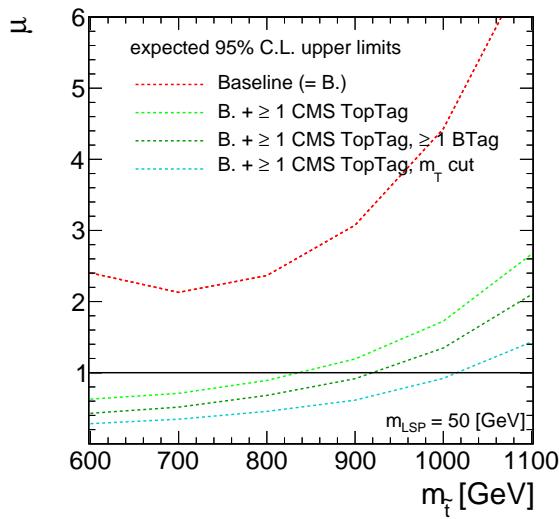


Figure 7.15.: Expected 95% upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV.

Here, different selections are compared. The exclusion curves derived from the baseline requirements is shown and indicates the sensitivity based on a very basic selection as discussed in Sec. 7.2. Furthermore, the selection discussed in Sec. 7.4 in which at least one CMS top tag is required in addition to the baseline selection is illustrated in light green. This selection can be easily combined with a b-tag requirement as discussed in Sec. 7.4 and is shown as dark green curve in Fig. 7.15. Hence, the combination of b-tagging and top tagging can further improve the search sensitivity. Finally, the baseline and top tagging selection is combined with a requirement of $m_T > 180$ GeV which further reduces especially $t\bar{t}$ background events. It turns out that the latter performs best among those described selections.

7.6. Performance Comparison to Selection Based on Published Stop Analysis at $\sqrt{s} = 8$ TeV

In order to get a better understanding of the quality of the studied selections, a comparison to the analysis criteria used in an analysis based on the same final state that was successfully performed at $\sqrt{s} = 8$ TeV is carried out. The respective analysis is published in [176].

The comparison is done by performing the same selections as done in the offical analysis based on the simulated samples discussed in Sec. 7.1. Jets in this analysis are clustered with the anti- k_T algorithm with a distance parameter of 0.5 and corrected for pile-up effects by applying charged-hadron subtraction. The pre-selection criteria in the analysis are:

- Events with isolated electrons and muons with $p_T > 10$ GeV, as described in Sec. 7.2, are vetoed.
- Events have to have at least five jets with $p_T > 30$ GeV and $|\eta| < 2.4$. The two highest p_T jets further have to have $p_T > 70$ GeV while the next two highest jets in p_T must fulfill $p_T > 50$ GeV.
- There has to be at least one b-tagged jet in the event based on the CSV algorithm with medium working point.
- The minimum azimuthal angle between the three highest jets and the missing transverse momentum has to be $\Delta\phi(\text{jet}_n, \cancel{E}_T) > 0.5$, $n = 1, 2$ and $\Delta\phi(\text{jet}_3, \cancel{E}_T) > 0.3$.²

The only difference between these selection criteria and the published analysis is related to the lepton veto. The published analysis requires to have no events with identified and isolated electrons and muons with $p_T > 5$ GeV while here only events with electrons and muons with $p_T > 10$ GeV are vetoed. However, since the 10 GeV lepton veto is the same lepton veto requirement as for the other selections studied in this chapter, it is easier to compare the quality of the different selections applied to the sample after selecting the all-jet final state.

In addition to those pre-selection requirements, further selection criteria are imposed in order to reconstruct the hadronically-decaying top quarks. The set of five or more jets in the event is separated into all possible combinations of three jets and a remnant which must contain at least one b-tagged jet. These sets are used to reconstruct the two expected top quarks in the event: one is based on one of the trijet combinations and denoted *fully-reconstructed* top while the other is based on the remnant system and referred to as *partially-reconstructed* top. Details on the reconstruction process of the two top quark systems can be found in [176]. In particular this process involves, that the fully-reconstructed top quark has to satisfy the criteria described in Eq. 4.6, 4.7 and 4.8 with $R_{\min} = 0.85 \cdot (m_W/m_{\text{top}})$, $R_{\max} = 1.25 \cdot (m_W/m_{\text{top}})$, $m_W = 80.4$ GeV and $m_{\text{top}} = 173.1$ GeV. If there is more than one trijet system satisfying these criteria the combination with $m^{3\text{-jet}}$ closest to m_{top} is selected. The selected remnant system is denoted Rsys.

After the successful identification of the two top quark systems according to the above mentioned criteria, further topological requirements are used. These are:

²The missing transverse energy in the published analysis is denoted with p_T^{miss} .

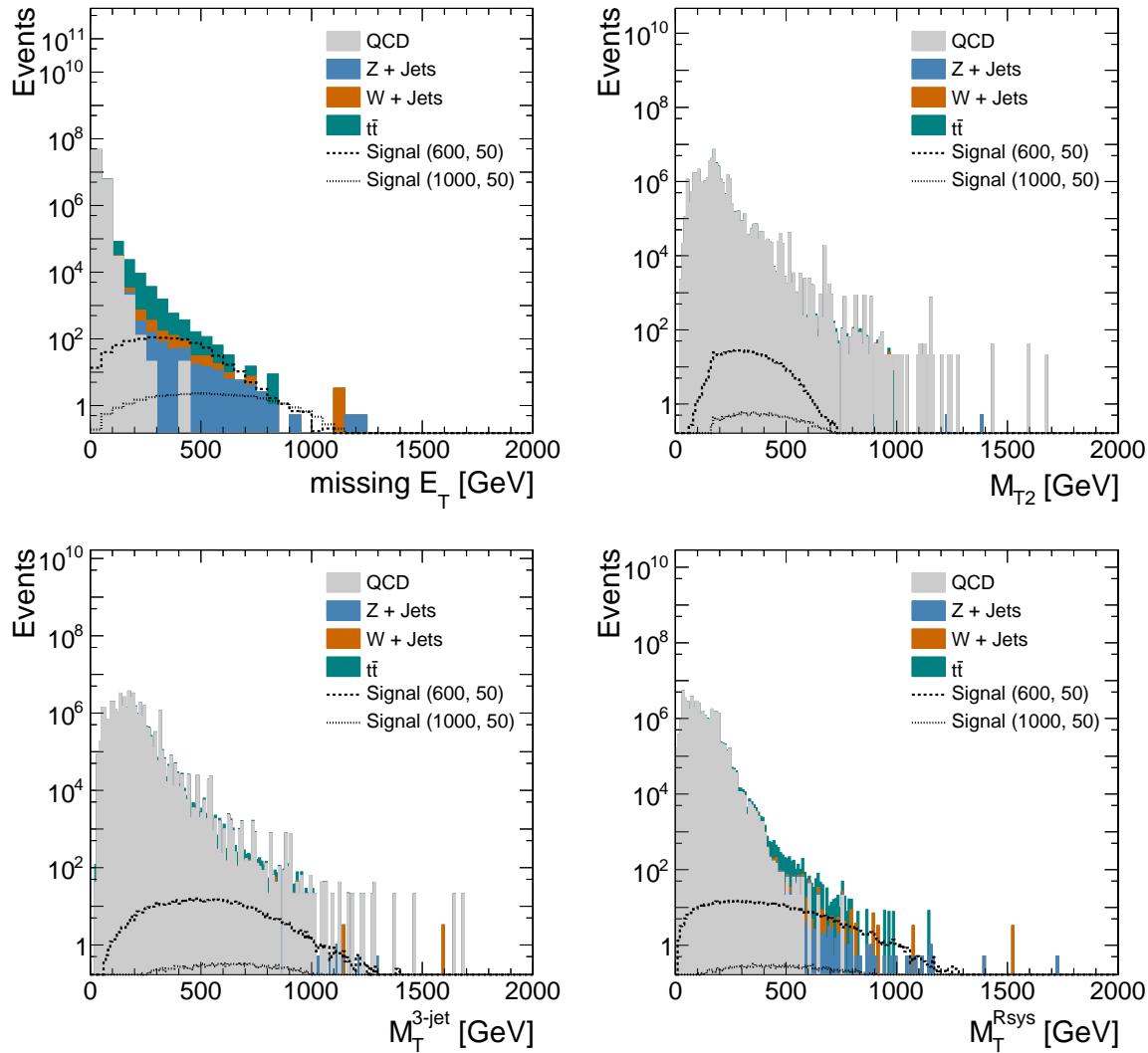


Figure 7.16.: Comparison of \cancel{E}_T , M_{T2} , $M_T^{3\text{-jet}}$ and M_T^{Rsys} distributions for signal and background events after applying pre-selection criteria. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass.

- $\cancel{E}_T > 200$ GeV
- The variable M_{T2} , as discussed in Sec. 7.5, is required to be ≥ 300 GeV. It is calculated from the four momenta of the fully- and the partially-reconstructed top quark as well as \cancel{E}_T assuming the invisible particles to be massless.
- $(0.5 \cdot M_T^{3\text{-jet}} + M_T^{\text{Rsys}}) \geq 500$ GeV. $M_T^{3\text{-jet}}$ and M_T^{Rsys} denote the transverse mass of the fully-reconstructed and the remnant system, respectively, which are calculated according to

$$(M_T^{3\text{-jet}})^2 = (m^{3\text{-jet}})^2 + 2(E_T^{3\text{-jet}} \cancel{E}_T - p_T^{3\text{-jet}} \cancel{E}_T \cos\Delta\phi)$$

and

$$(M_T^{\text{Rsys}})^2 = (m^{\text{Rsys}})^2 + 2(E_T^{\text{Rsys}} \not{E}_T - p_T^{\text{Rsys}} \not{E}_T \cos\Delta\phi).$$

A comparison of the \not{E}_T , M_{T2} , $M_T^{\text{3-jet}}$ and M_T^{Rsys} distributions for signal and background events after the pre-selection is shown in Fig. 7.16. These exhibit that those topological variables are able to reject several background events while keeping good acceptance for signal events.

In order to probe different points of the parameter space, events are further categorized into four inclusive search regions:

- $\not{E}_T > 200$ GeV, $N_{\text{b-jets}} \geq 1$
- $\not{E}_T > 350$ GeV, $N_{\text{b-jets}} \geq 1$
- $\not{E}_T > 200$ GeV, $N_{\text{b-jets}} \geq 2$
- $\not{E}_T > 350$ GeV, $N_{\text{b-jets}} \geq 2$

Uncertainties in the expected limit calculation performed for each of these categories to compare the reach to the selections discussed above are considered to be the same as those assumed in Sec. 7.2. This again makes it easier to compare the actual performance of various selections. From each of the four search regions the expected limit giving the best sensitivity to a specific mass point is considered in the comparison. This is for the scenarios tested here with stop masses ranging from 600 – 1100 GeV and a LSP mass of 50 GeV, the selection requiring $\not{E}_T > 350$ GeV, $N_{\text{b-jets}} \geq 2$. The actual comparison of the expected limits is illustrated in Fig 7.17.

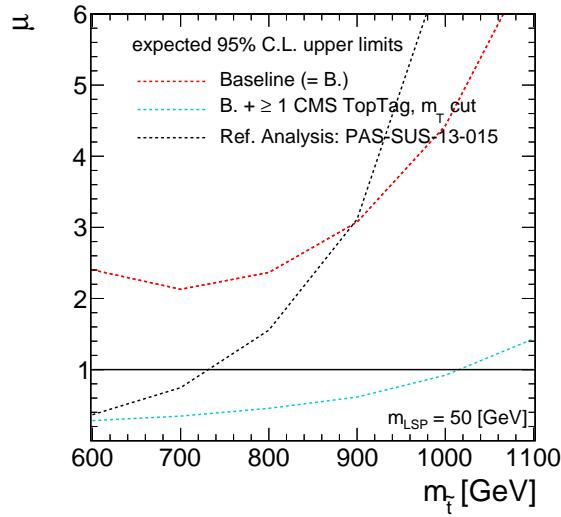


Figure 7.17.: Expected 95% upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV.

The distributions exhibit that the selection based on the requirements used in [176] shows a quite good sensitivity for low stop masses while the sensitivity drops rapidly towards

higher stop masses. However, the selection proposed in Sec. 7.5 performs better for all mass scenarios.

7.7. Sensitivity to Gluino-Mediated Stop Production

The selections studied in this chapter are mainly aiming at identifying events from direct pair production of top squarks which subsequently decay into a top quark and a LSP each. Special emphasis is put on identifying mass scenarios with large mass splittings between top squark and LSP which is suitable for employing top tagging techniques. However, the studied selections possibly offer an interesting alternative to classical selections targeting gluino-mediated production of third generation squarks. In those scenarios pair produced gluinos are considered that subsequently decay into a pair of top quarks and a LSP. Thus, the final state exhibits four top quarks. For gluino masses exceeding 1 TeV this potentially gives rise to also boosted top quark decays in the final state. Consequently, the sensitivity of the best performing selection criteria based on the baseline selection, at least one CMS top-tagged jet and a transverse mass selection of $m_T > 180$ GeV is also evaluated with respect to the gluino-mediated stop production. This is done for three different mass scenarios, as specified in Sec. 7.1, with gluino masses of 1.3 TeV, 1.5 TeV and 1.7 TeV for a LSP mass of 50 GeV, respectively. Based on the same search strategy as before using exclusive H_T and \cancel{E}_T search regions, the derived expected exclusion curve for gluino-mediated stop production is illustrated in Fig. 7.18.

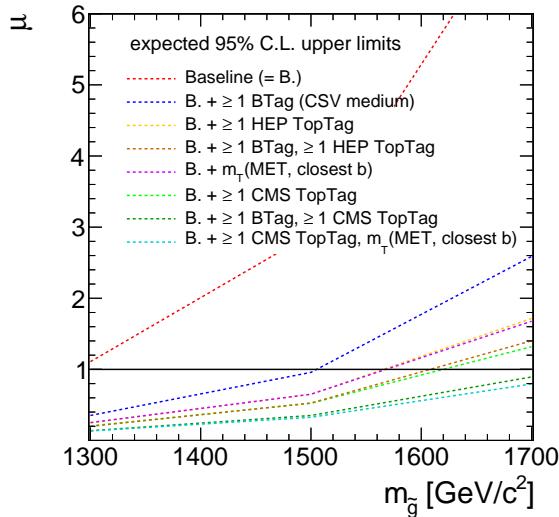


Figure 7.18.: Expected 95% upper limit for signal strength versus $m_{\tilde{g}}$. The LSP mass is chosen to be 50 GeV.

The distributions exhibit that the derived selection described in Sec. 7.5 is also sensitive to the tested mass scenarios of gluino-mediated stop production. Thus, selections involving top tagging are also offering the opportunity to probe susy scenarios than direct stop production as long as top quarks with high transverse momenta are present in the final

state. Such selections could serve as complementary approach to search strategies like those employed in chapter 6 for further supersymmetry searches at $\sqrt{s} = 13$ TeV.

7.8. Results and Discussion

Selections based on different kinematic properties of direct stop pair production have been studied and analysis strategies were identified that allow to probe direct stop production up to 1 TeV in case of an LSP mass of 50 GeV (cf. Sec. 7.5). However, the sensitivity of this selection targeting large mass splittings between stop and LSP can be evaluated also for other mass scenarios. Thus, the same search strategy and assigned uncertainties have been used to probe the expected exclusion reach for the same stop masses as before but when considering the LSP mass to be larger up to 350 GeV. The derived exclusion curves are summarized in Fig. 7.19 in the $m_{\tilde{t}}$ versus m_{LSP} plane.

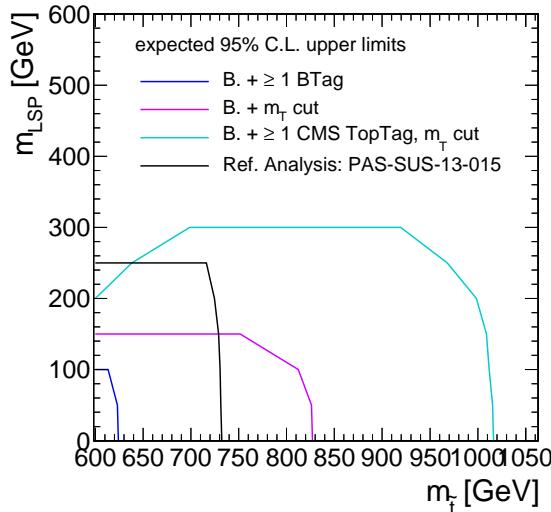


Figure 7.19.: Expected 95% upper limit for $m_{\tilde{t}}$ versus m_{LSP} . Note that the x-axis starts at 600 GeV.

The selections illustrated there are baseline requirements plus at least one medium b-tagged jet (dark blue), baseline requirements and $m_T > 180$ GeV (magenta), the selection following [176] discussed in Sec. 7.6 (black) and the selection based on the baseline selection employing top tagging and $m_T > 180$ GeV (light blue). The areas below the lines can be excluded. Consequently, for the blue selection stop masses up to 1 TeV can be tested for LSP masses not exceeding 300 GeV.

In general, the conclusions found in previous sections hold true also for the two dimensional exclusion curves. The best sensitivity for a wide range of mass point configurations is achieved by the selection introduced in Sec. 7.5. Although the selection following the already published analysis at $\sqrt{s} = 8$ TeV is able to probe stop masses up to 700-800 GeV also for other LSP masses than 50 GeV, the sensitivity towards larger stop masses is not achieved. Consequently, these studies suggest that either the existing analysis has to be improved in order to test the whole stop mass range up to 1 TeV or a search strategy as

proposed in this chapter should be employed.

However, when deciding on a particular analysis strategy for direct stop production at $\sqrt{s} = 13$ TeV it has to be kept in mind that the studies presented here, are based on some simplifications. The pileup scenario considered in the simulation is not the actual expected situation during the next running period, but uses the pileup conditions from $\sqrt{s} = 8$ TeV. As discussed already in Sec. 7.1 this should in principle not affect the analysis shown here too much. Furthermore, the samples have been processed based on the fast detector simulation only. Although this has shown a comparable performance to the full simulation for several quantities in the past, as discussed in Sec. 3.4, no dedicated studies have been performed yet, if those holds true also for all jet substructure variables used for the top tagging algorithms. Nevertheless, the derived efficiency and misidentification rates shown in Fig. 7.8 show a similar performance to those quantities at $\sqrt{s} = 8$ TeV (cf. Sec. 4.4). Consequently, the performance improvement of the analysis due to an employment top tagging is expected to be estimated at the correct order of magnitude. In addition to those simplified assumptions also the considered uncertainties for the different background processes are based on ad hoc assumptions. Although they have been chosen in correspondence to the considered uncertainties in [176], it is not yet studied how precisely the individual background contributions can be estimated for the extreme kinematic regions targeted by the top tagging selections when employing a still to be defined data-based background estimation procedure.

Since though all those simplifications affect the various selections studied in this chapter in a similar manner, the sensitivity of various selections relative to each other should not be affected by this. Nevertheless, the absolute expected analysis reach might be better or worse by a few 10 GeV than found in these studies. However, as discussed above most of the simplifications should have only a minor impact on the presented selections. Consequently, the expected data at $\sqrt{s} = 13$ TeV supposedly further shed light on the question of a realization of natural supersymmetry in data manifesting in top squarks below the TeV mass range.

8. Conclusions

Supersymmetry is among the favoured extensions of the standard model of particle physics and one of the main target of searches for new physics within the CMS experiment. Since especially coloured SUSY particles are expected at a high rate at the LHC which predominantly manifest in final states containing jets and missing transverse momentum, it is promising to perform searches based on such signatures which require in addition a good understanding of jet properties.

In this thesis, a measurement of the jet transverse momentum resolution in dijet events corresponding to data with an integrated luminosity of 19.8 fb^{-1} recorded at $\sqrt{s} = 8 \text{ TeV}$ in 2012 by the CMS experiment has been presented. A similar approach to previous analyses was used, but the precision of the measurement could be significantly improved. This was achieved by a correct treatment of statistical uncertainties by considering correlations among inclusive distributions which subsequently also allowed a revised treatment of systematic uncertainties which have been conservatively overestimated in the past. Furthermore, the method has been extended to be able to measure the resolution in the forward part of the detector with higher precision. For $|\eta| = 0.0 - 5.0$ the data-to-simulation ratio of the resolution has been measured. Since no significant trend as function of p_T^{ave} was apparent, the ratio is determined as function of $|\eta|$ only. The ratios obtained for the various $|\eta|$ regions increase from 1.079 ± 0.026 in the central region up to 1.395 ± 0.063 at $|\eta| = 2.8 - 3.2$ and drop again for the outermost region $|\eta| = 3.2 - 5.0$ down to 1.056 ± 0.191 . These values can be utilized to adjust the resolution in simulation to match the one observed in data.

In the second part of this thesis, a search for supersymmetry in events with $N_{\text{Jets}} = 3 - 5, 6 - 7, \geq 8$ and large missing momentum in proton-proton collisions at $\sqrt{s} = 8 \text{ TeV}$ in data corresponding to an integrated luminosity of 19.5 fb^{-1} has been presented. Special emphasis is put on the determination of the QCD multijet background which relies strongly on the precise knowledge of the jet response. These QCD background contributions are estimated directly from multijet events in data by modelling momentum mismeasurements based on the jet response. A similar approach has been used already in earlier versions of this analysis where the search has been performed inclusive in the jet multiplicity requiring at least three jets. With the extension of the analysis to multijet events, an adjustment of the method to predict QCD multijet background contributions became inevitable. A dedicated correction to the existing method has been introduced in order to predict the jet multiplicity correctly. Moreover the assignment of systematic uncertainties has been revised in order to consider e.g. the challenging conditions due to pile-up appropriately. In total, the QCD multijet background could be estimated with a precision of approximately 70 – 80% in search regions with non-negligible QCD background contributions. The observed number of events in data are consistent with the expected number of events from standard model processes such that exclusion limits are derived for various simplified supersymmetric models. In the context of this simplified models the production of squarks below 780 GeV and that of gluinos up to 1.1 – 1.2 TeV can be excluded at 95% CL for

LSP masses not exceeding 100 GeV.

In addition to inclusive searches targeting gluinos and squarks, CMS has also performed searches for direct production of top squarks with data obtained at $\sqrt{s} = 8$ TeV. With this data direct production of top squarks decaying into top and LSP could be excluded up to top squarks masses of approximately 750 GeV for LSP masses below 100 GeV. In order to extend the mass reach of such searches to even higher top squark masses in the next run period of the LHC which will start in 2015 with a center of mass energy of 13 TeV suitable selection criteria for such analyses have been investigated in the third part of this thesis. Studies presented here are based on events with several jets, large momentum imbalance and no leptons. One key aspect is the application of dedicated algorithms for the identification of decay products emerging from boosted top decays in order to separate possible SUSY signal events from standard model background. Various different discriminating variables have been investigated. Moreover the sensitivity of several selections – also that one of the all-hadronic stop search performed at $\sqrt{s} = 8$ TeV – has been compared by deriving the expected exclusion reach for data corresponding to 19.5 fb^{-1} at $\sqrt{s} = 13$ TeV. Selection criteria could be identified extending the mass reach of direct stop searches up to roughly 1 TeV for LSP masses below 100 GeV.

Summary
in den
Kontext
"Natural
SUSY"
einbetten

A. Jet Transverse-Momentum Resolution

A.1. Primary Vertex Distributions

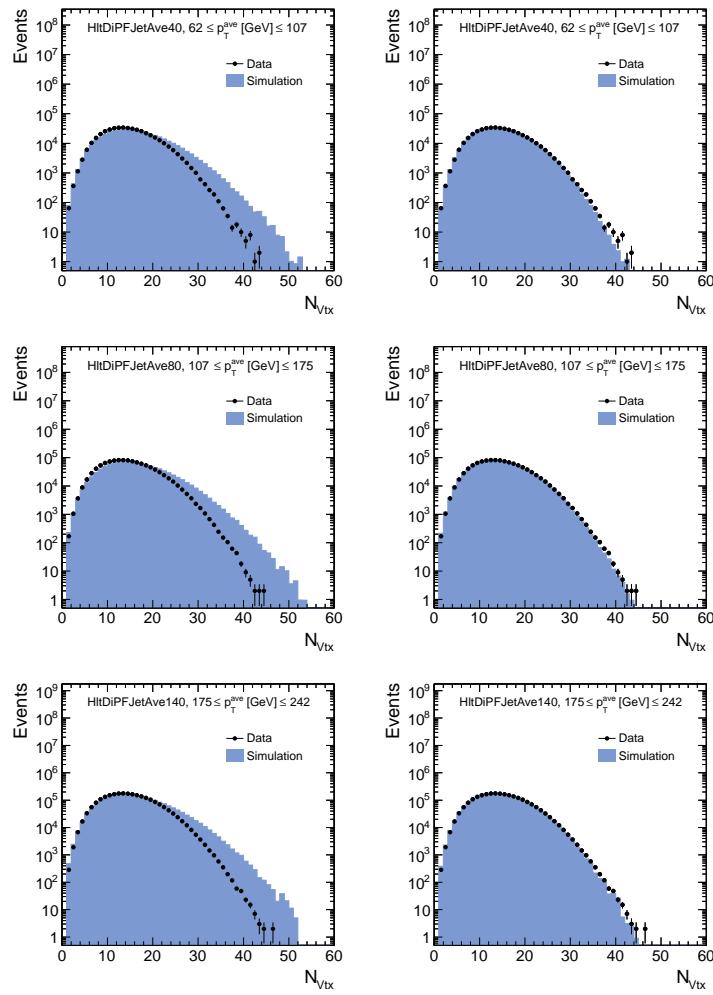


Figure A.1.: Distribution of number of primary vertices in data (black dots) and simulation (blue histogram) before (*left*) and after (*right*) reweighting of the pile-up scenario in simulation for trigger paths HltDiPFJetAve40 (*top*), HltDiPFJetAve80 (*middle*) and HltDiPFJetAve140 (*bottom*).

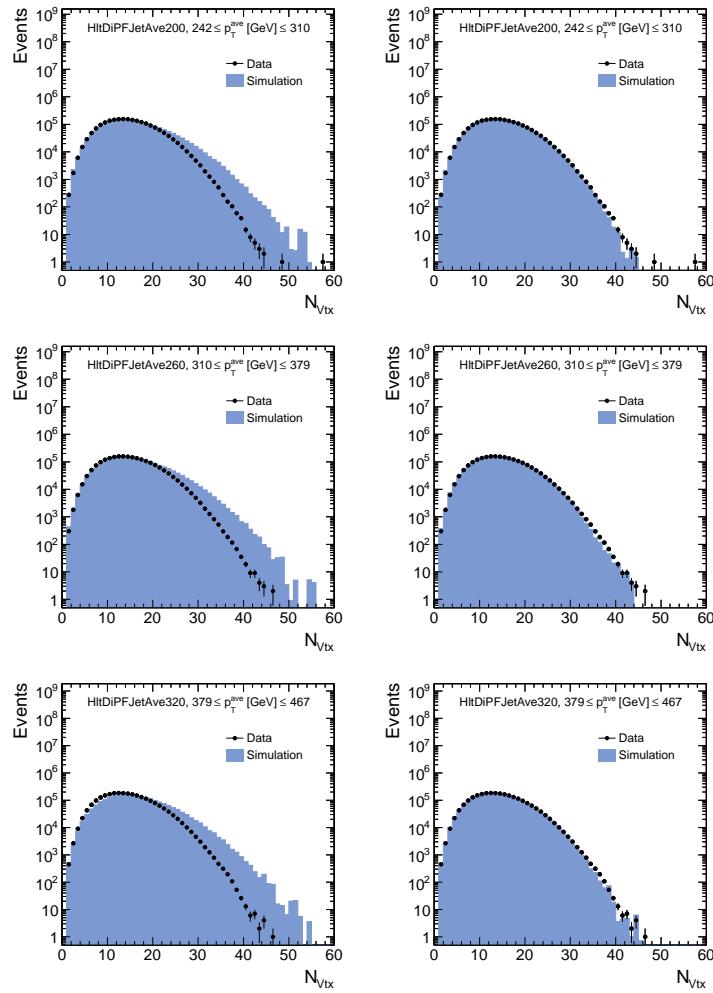


Figure A.2.: Distribution of number of primary vertices in data (black dots) and simulation (blue histogram) before (*left*) and after (*right*) reweighting of the pile-up scenario in simulation for trigger paths HltDiPFJetAve200 (*top*), HltDiPFJetAve260 (*middle*) and HltDiPFJetAve320 (*bottom*).

A.2. Extrapolation Graphs

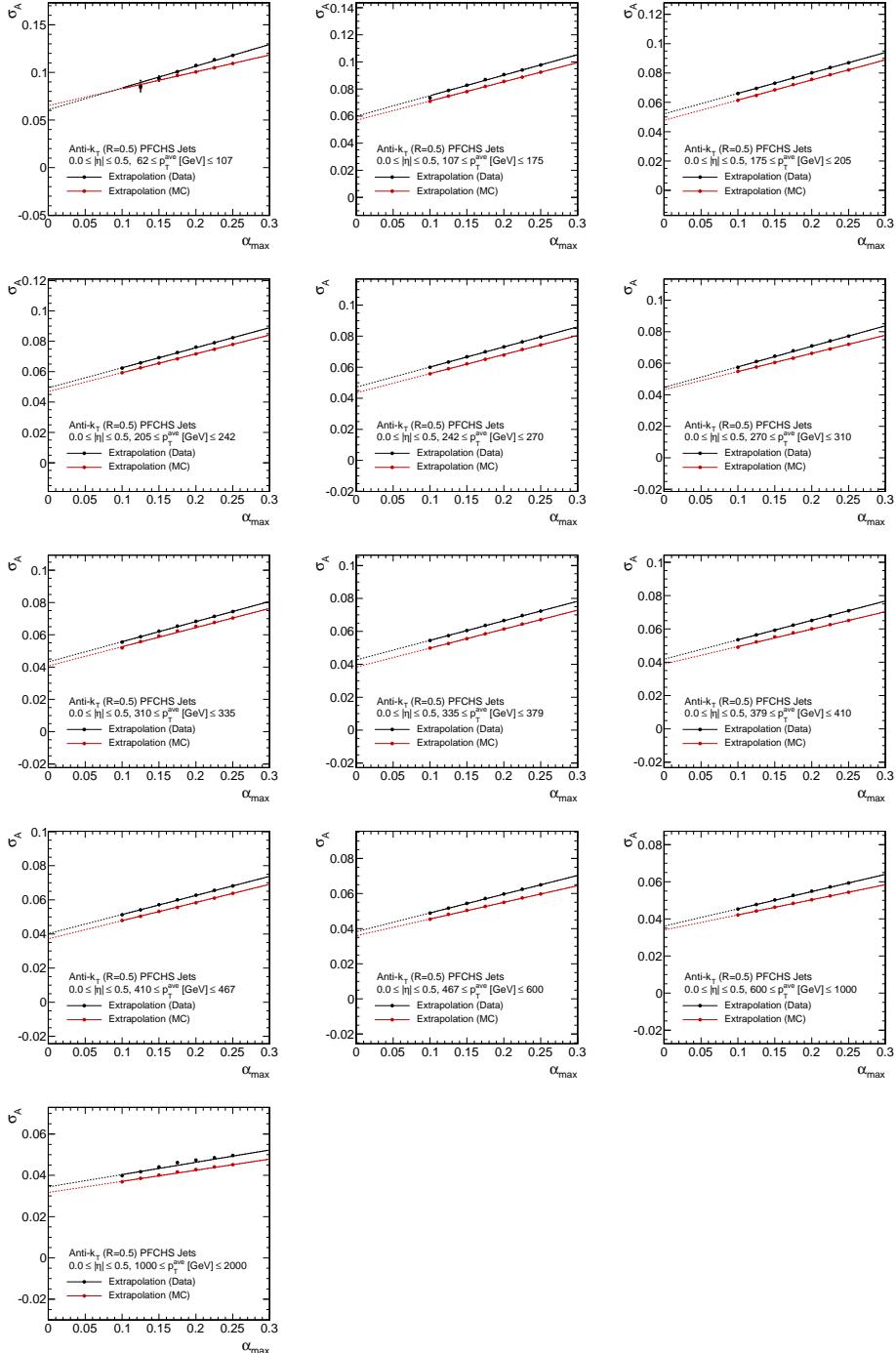


Figure A.3.: Extrapolations of measured values for σ_A in data and simulation to obtain the result at zero additional jet activity for $0.0 \leq |\eta| \leq 0.5$.

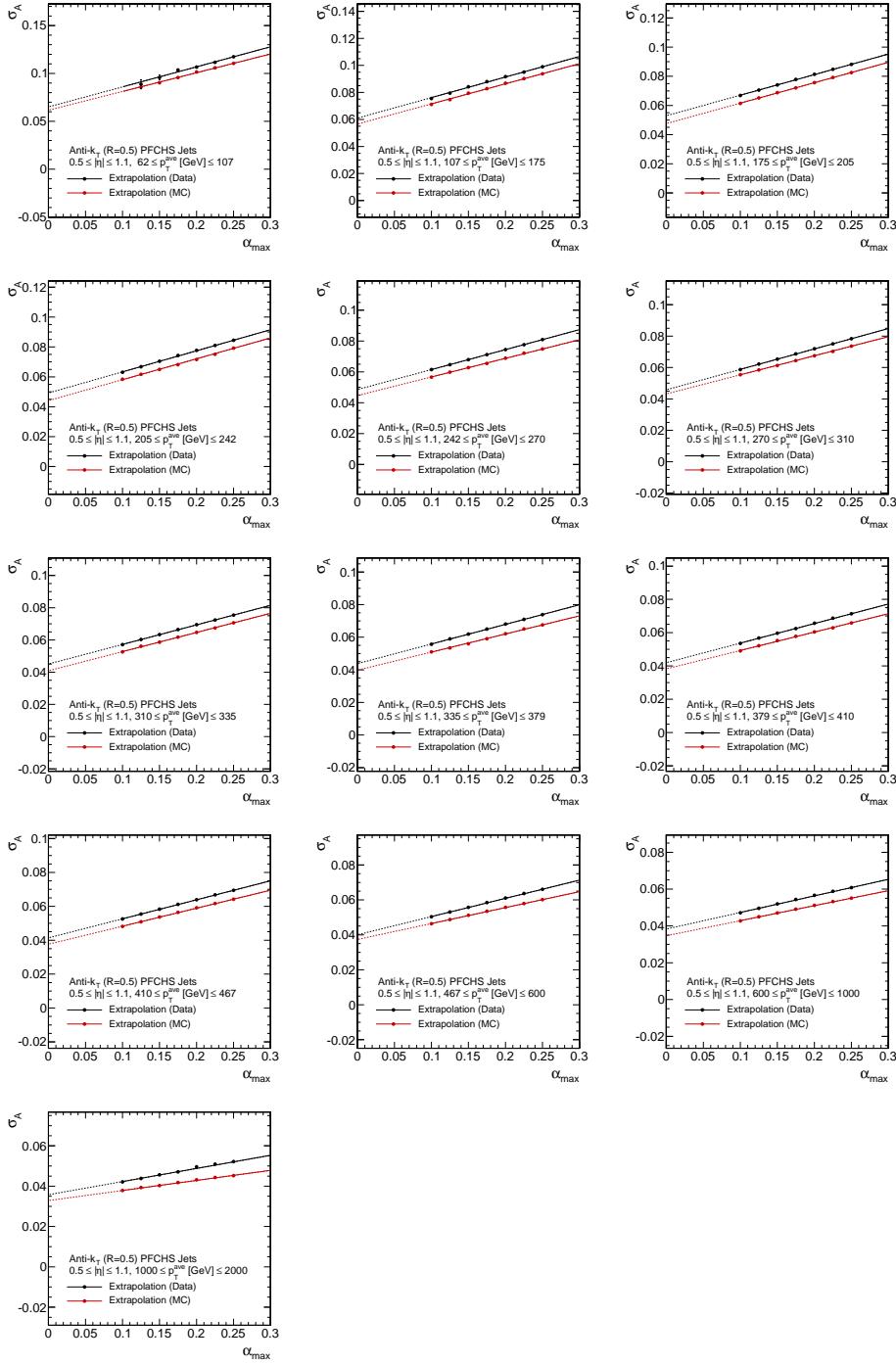


Figure A.4: Extrapolations of measured values for σ_A in data and simulation to obtain the result at zero additional jet activity for $0.5 \leq |\eta| \leq 1.1$.

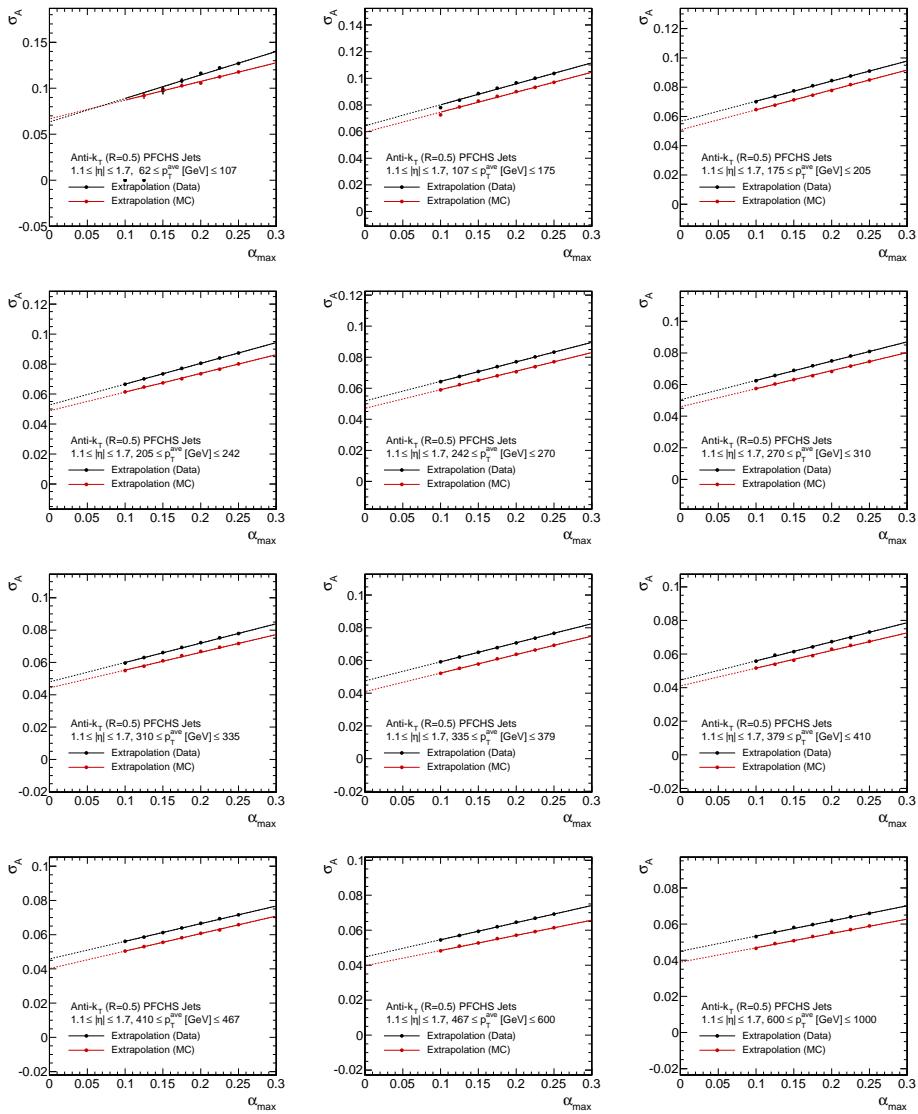


Figure A.5.: Extrapolations of measured values for σ_A in data and simulation to obtain the result at zero additional jet activity for $1.1 \leq |\eta| \leq 1.7$.

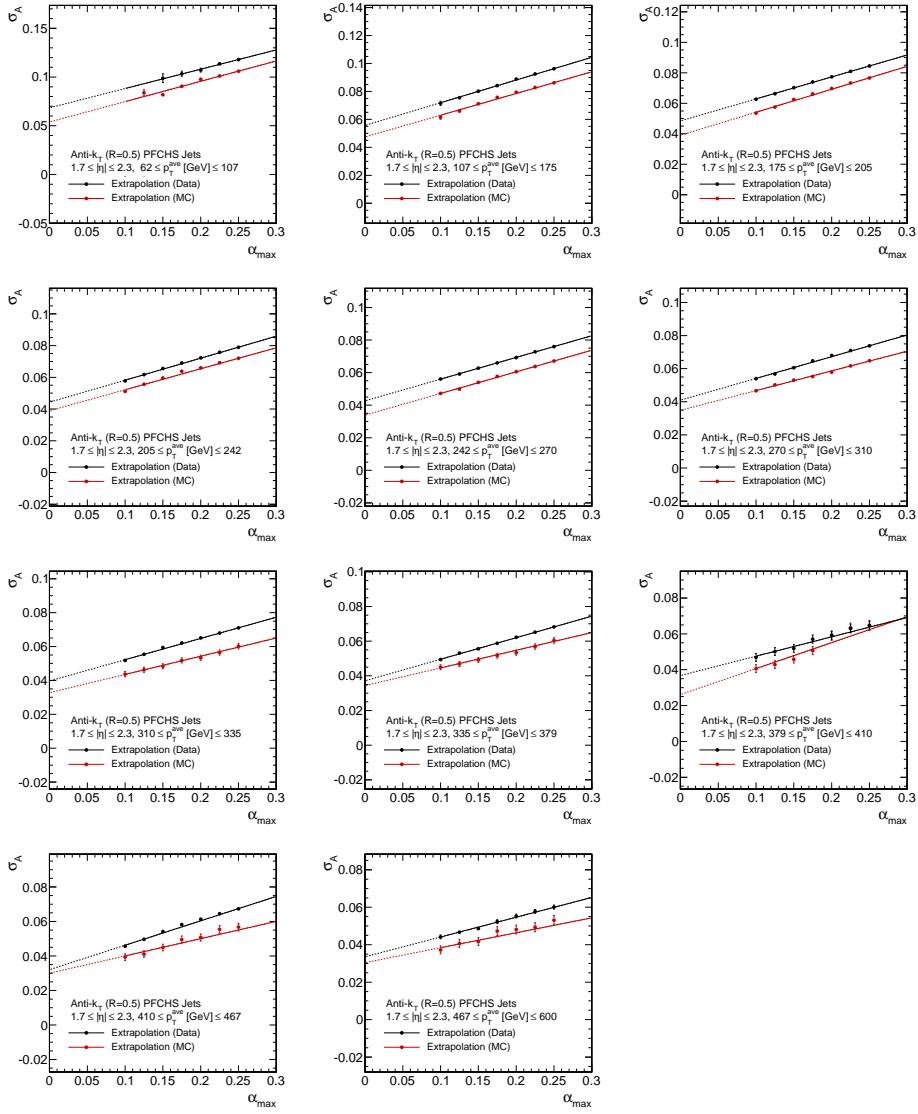


Figure A.6.: Extrapolations of measured values for σ_A in data and simulation to obtain the result at zero additional jet activity for $1.7 \leq |\eta| \leq 2.3$.

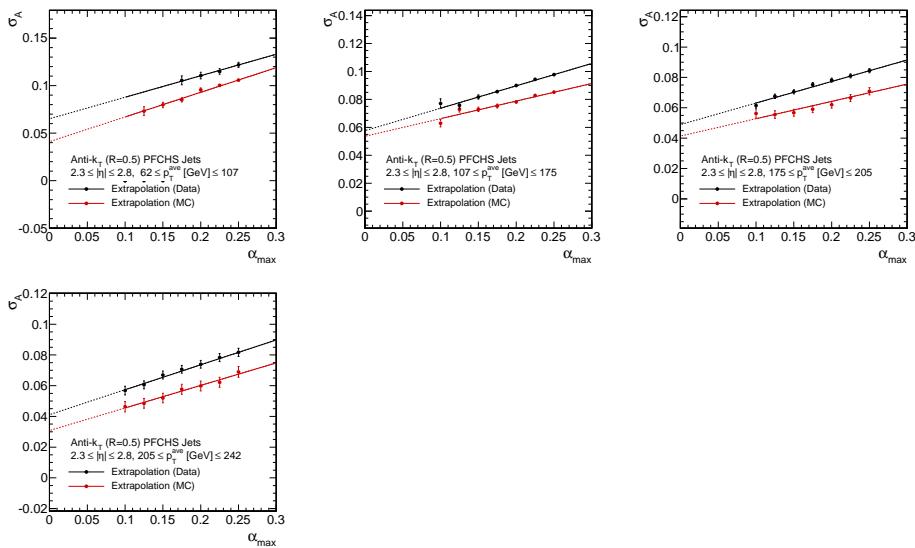


Figure A.7.: Extrapolations of measured values for σ_A in data and simulation to obtain the result at zero additional jet activity for $2.3 \leq |\eta| \leq 2.8$.

A.3. Correlation of Asymmetry Widths

Due to the selection of inclusive α -intervals the measured standard deviations of asymmetry distributions σ_A in the same (p_T^{ave} , $|\eta|$)-interval are correlated for different values of α_{max} . Thus, an expression for this correlation is needed in order to consider such dependencies in the extrapolation fits.

Given a distribution of random variables x_i , with $i = 1, \dots, N$ and mean $\mu = 0$, the estimator of the variance of the distribution is given by

$$\hat{\sigma}_x^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 .$$

Here, the values of x_i correspond to the asymmetry values for a certain $\alpha_{max,i}$ selection. Given another distribution of random variables y_j , with $j = 1, \dots, M$ and mean $\mu = 0$ for values of y_j independent of x_i , the variance of the overall distribution of the x_i and y_j values is given by

$$\hat{\sigma}_{xy}^2 = \frac{1}{N+M} \left(\sum_i x_i^2 + \sum_j y_j^2 \right) = \frac{N}{N+M} \hat{\sigma}_x^2 + \frac{M}{N+M} \hat{\sigma}_y^2 .$$

The values of y_j correspond to the additional asymmetry values compared to x_i when selecting events with $\alpha_{max,j}$ ($\alpha_{max,j} > \alpha_{max,i}$).

In general, the covariance for two random variables X and Y is defined as

$$\text{cov}(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle$$

where $\langle \rangle$ denotes the expectation value, i.e. the average. In this particular case, the covariance of $\hat{\sigma}_x^2$ and $\hat{\sigma}_{xy}^2$ shall be estimated which is

$$\text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_{xy}^2) = \frac{N}{N+M} \text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_x^2) + \frac{M}{N+M} \text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_y^2) = \frac{N}{N+M} \text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_x^2) ,$$

since the values x_i and y_j are independent resulting in $\text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_y^2) = 0$. The expression $\text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_x^2)$ is the variance of the estimator $\hat{\sigma}_x^2$.

In order to convert $\text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_{xy}^2)$ into $\text{cov}(\hat{\sigma}_x, \hat{\sigma}_{xy})$, i.e. finding the covariance of the estimators for the standard deviation $\hat{\sigma}_x$ rather than for the variance $\hat{\sigma}_x^2$, it is used that the standard deviation is the square root of the variance. The standard rules for uncertainty propagation in case of correlated uncertainties state that

$$\Sigma^f = A \Sigma^x A^T$$

with the variance-covariance matrix Σ^f for function f and the variance-covariance matrix Σ^x for variables x .

Applying this to the function

$$f(X, Y) = (\sqrt{X}, \sqrt{Y})$$

with

$$A = \begin{pmatrix} \frac{1}{2\sqrt{X}} & 0 \\ 0 & \frac{1}{2\sqrt{Y}} \end{pmatrix}$$

gives the relation

$$\text{cov}(\sqrt{X}, \sqrt{Y}) = \frac{\text{cov}(X, Y)}{4\sqrt{XY}}.$$

Consequently, it is

$$\text{cov}(\hat{\sigma}_x, \hat{\sigma}_{xy}) = \frac{\text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_{xy}^2)}{4\hat{\sigma}_x \hat{\sigma}_{xy}} = \frac{N}{N+M} \text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_x^2) \frac{1}{4\hat{\sigma}_x \hat{\sigma}_{xy}}.$$

Using a standard result for the *variance of the variance* in case of a normal distribution which is

$$\text{cov}(\hat{\sigma}_x^2, \hat{\sigma}_x^2) = \frac{2\hat{\sigma}_x^4}{N}$$

and the variance of the estimator $\hat{\sigma}_x$ of a normal distribution given by

$$\text{cov}(\hat{\sigma}_x, \hat{\sigma}_x) = \frac{\hat{\sigma}_x^2}{2N} = (\Delta \hat{\sigma}_x)^2$$

the result is determined to be

$$\text{cov}(\hat{\sigma}_x, \hat{\sigma}_{xy}) = \frac{N}{N+M} \frac{\hat{\sigma}_x}{\hat{\sigma}_{xy}} (\Delta \hat{\sigma}_x)^2$$

which is the same as expressed in Eq. 5.13.

B. Search for New Physics with Jets and Missing Transverse Momentum

B.1. Data-to-Simulation Ratios: Core

Table B.1.: Data-to-simulation ratio in various $|\eta|$ regions with total uncertainty as used in the QCD background prediction with the R+S method for the correction of the core response.

$ \eta $	$c(\text{Data}/\text{MC})$
0.0 – 0.5	$1.052^{+0.063}_{-0.062}$
0.5 – 1.1	$1.057^{+0.057}_{-0.056}$
1.1 – 1.7	$1.096^{+0.065}_{-0.064}$
1.7 – 2.3	$1.134^{+0.094}_{-0.092}$
2.3 – 5.0	$1.288^{+0.020}_{-0.020}$

B.2. Data-to-Simulation Ratios: Tails

Table B.2.: Data-to-simulation ratio ρ_{tail} of the fractional tail-size together with the total uncertainty in different $|\eta| \times p_T^{\text{ave}}$ intervals for the tail regions derived from the dijet asymmetry parts with ($\mathcal{A} > 2\sigma_c$) used in the QCD background prediction with the R+S method for the correction of the tail parts of the response [130].

$ \eta $	p_T^{ave} [GeV]	$\rho_{\text{tail}} (\mathcal{A} = 2\sigma_c)$
0.0 – 0.5	45 – 220	0.953 ± 0.283
0.0 – 0.5	220 – 270	1.418 ± 0.502
0.0 – 0.5	270 – 312	1.156 ± 0.349
0.0 – 0.5	312 – 360	1.305 ± 0.350
0.0 – 0.5	360 – 498	1.342 ± 0.393
0.0 – 0.5	498 – 1500	1.353 ± 0.350
0.5 – 1.1	45 – 220	1.096 ± 0.374
0.5 – 1.1	220 – 294	1.083 ± 0.372
0.5 – 1.1	294 – 360	1.195 ± 0.325
0.5 – 1.1	360 – 1500	1.248 ± 0.424
1.1 – 1.7	45 – 220	0.965 ± 0.333
1.1 – 1.7	220 – 335	1.035 ± 0.295
1.1 – 1.7	335 – 1500	1.358 ± 0.327
1.7 – 2.3	45 – 220	0.938 ± 0.286
1.7 – 2.3	220 – 1500	1.196 ± 0.425
2.3 – 5.0	45 – 220	1.069 ± 0.77

C. Project Studies for Stop Quarks at $\sqrt{s} = 13 \text{ TeV}$

C.1. Kinematic Distributions

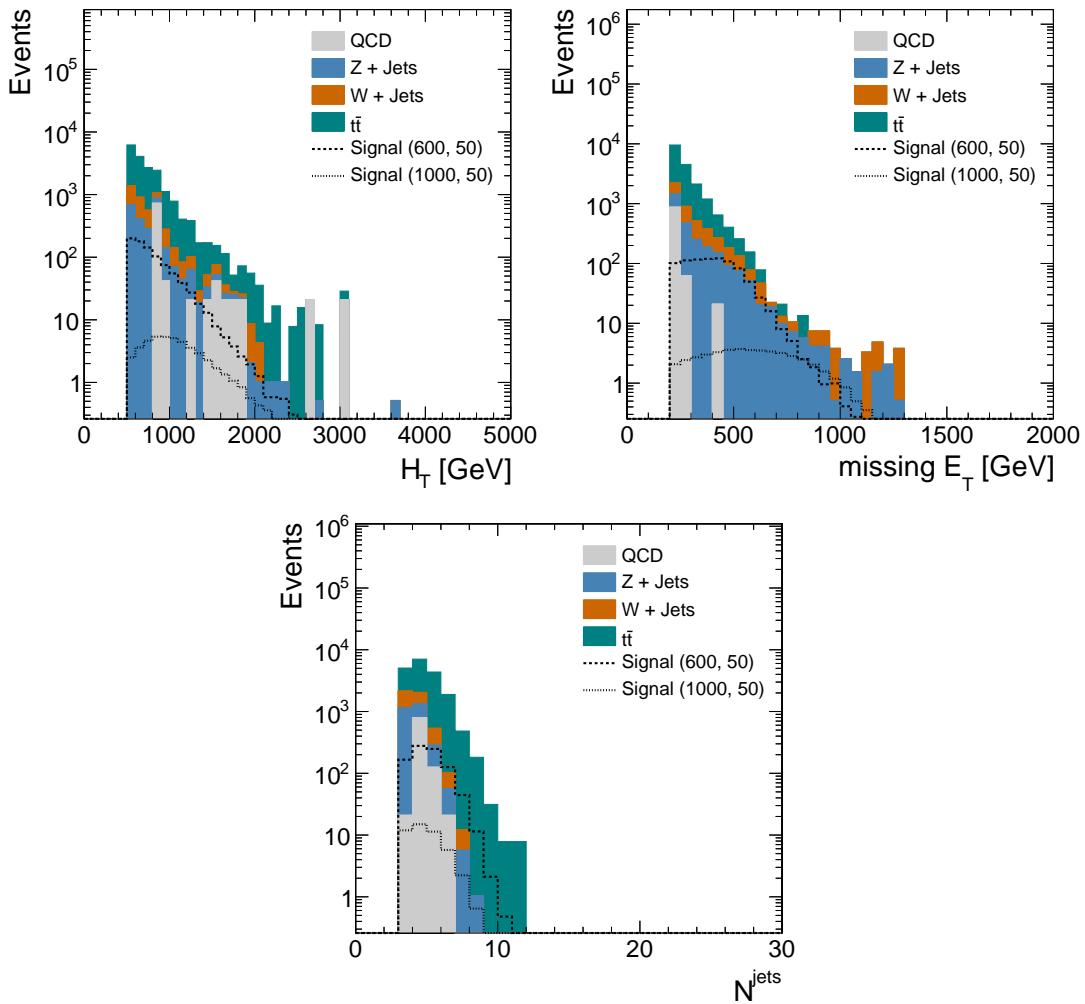


Figure C.1.: Comparison of selected H_T (top left), \cancel{E}_T (top right) and N_{Jets} (bottom) distributions in simulated events found from applying the baseline selection criteria and requiring at least one b-tagged jet. The signal points are labeled as (X, Y) where X is the top squark mass and Y is the LSP mass.

Bibliography

- [1] J. B. et al. (Particle Data Group)), “Review of Particle Physics”, *Phys. Rev. D* **86** (Jul, 2012) 010001. doi:10.1103/PhysRevD.86.010001.
- [2] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10** (Jun, 1963) 531 – 533. doi:10.1103/PhysRevLett.10.531.
- [3] M. Kobayashi and T. Maskawa, “*CP*-Violation in the Renormalizable Theory of Weak Interaction”, *Progress of Theoretical Physics* **49** (1973), no. 2, 652–657. doi:10.1143/PTP.49.652.
- [4] R. Alkofer and J. Greensite, “Quark Confinement: The Hard Problem of Hadron Physics”, *J.Phys.* **G34** (2007) S3, arXiv:hep-ph/0610365. doi:10.1088/0954-3899/34/7/S02.
- [5] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Phys. Rev. Lett.* **30** (Jun, 1973) 1346 – 1349. doi:10.1103/PhysRevLett.30.1346.
- [6] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Non-Abelian Gauge Theories”, *Phys. Rev. Lett.* **30** (Jun, 1973) 1343 – 1346. doi:10.1103/PhysRevLett.30.1343.
- [7] S. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl.Phys.* **22** (1961) 579–588. doi:10.1016/0029-5582(61)90469-2.
- [8] S. Weinberg, “A Model of Leptons”, *Phys.Rev.Lett.* **19** (1967) 1264–1266. doi:10.1103/PhysRevLett.19.1264.
- [9] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508 – 509. doi:10.1103/PhysRevLett.13.508.
- [10] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321 – 323. doi:10.1103/PhysRevLett.13.321.
- [11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* **13** (1964) 585 – 587. doi:10.1103/PhysRevLett.13.585.
- [12] ATLAS Collaboration Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys.Lett.* **B716** (2012) 1–29, arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.

- [13] CMS Collaboration Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys.Lett.* **B716** (2012) 30–61, [arXiv:1207.7235](https://arxiv.org/abs/1207.7235). doi:10.1016/j.physletb.2012.08.021.
- [14] ATLAS Collaboration Collaboration, “Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC”, *Phys.Lett.* **B726** (2013) 88–119, [arXiv:1307.1427](https://arxiv.org/abs/1307.1427). doi:10.1016/j.physletb.2013.08.010.
- [15] ATLAS Collaboration Collaboration, “Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector”, [arXiv:1408.5191](https://arxiv.org/abs/1408.5191).
- [16] ATLAS Collaboration Collaboration, “Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector”, [arXiv:1408.7084](https://arxiv.org/abs/1408.7084).
- [17] CMS Collaboration Collaboration, “Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model”, Technical Report CMS-PAS-HIG-14-009, CERN, Geneva, 2014.
- [18] M. Beyer, ed., “CP Violation in Particle, Nuclear, and Astrophysics”. Lecture Notes in Physics. Springer, 2010. ISBN 3-642-07830-3.
- [19] Planck Collaboration Collaboration, “Planck 2013 results. XVI. Cosmological parameters”, *Astron.Astrophys.* (2014) [arXiv:1303.5076](https://arxiv.org/abs/1303.5076). doi:10.1051/0004-6361/201321591.
- [20] G. Bertone, D. Hooper, and J. Silk, “Particle dark matter: Evidence, candidates and constraints”, *Phys.Rept.* **405** (2005) 279–390, [arXiv:hep-ph/0404175](https://arxiv.org/abs/hep-ph/0404175). doi:10.1016/j.physrep.2004.08.031.
- [21] I. J. Aitchison, “Supersymmetry and the MSSM: An Elementary introduction”, [arXiv:hep-ph/0505105](https://arxiv.org/abs/hep-ph/0505105).
- [22] S. P. Martin, “A Supersymmetry primer”, *Adv.Ser.Direct.High Energy Phys.* **21** (2010) 1–153, [arXiv:hep-ph/9709356](https://arxiv.org/abs/hep-ph/9709356).
- [23] (Super-Kamiokande Collaboration) Collaboration, “Search for proton decay via $p \rightarrow \nu K^+$ using 260kiloton · year data of Super-Kamiokande”, *Phys. Rev. D* **90** (Oct, 2014) 072005. doi:10.1103/PhysRevD.90.072005.
- [24] S. Dimopoulos and D. W. Sutter, “The Supersymmetric flavor problem”, *Nucl.Phys.* **B452** (1995) 496–512, [arXiv:hep-ph/9504415](https://arxiv.org/abs/hep-ph/9504415). doi:10.1016/0550-3213(95)00421-N.
- [25] A. H. Chamseddine, R. L. Arnowitt, and P. Nath, “Locally Supersymmetric Grand Unification”, *Phys.Rev.Lett.* **49** (1982) 970. doi:10.1103/PhysRevLett.49.970.
- [26] L. Alvarez-Gaume, J. Polchinski, and M. B. Wise, “Minimal Low-Energy Supergravity”, *Nucl.Phys.* **B221** (1983) 495. doi:10.1016/0550-3213(83)90591-6.

- [27] G. L. Kane, C. F. Kolda, L. Roszkowski et al., “Study of constrained minimal supersymmetry”, *Phys.Rev.* **D49** (1994) 6173–6210, [arXiv:hep-ph/9312272](https://arxiv.org/abs/hep-ph/9312272).
[doi:10.1103/PhysRevD.49.6173](https://doi.org/10.1103/PhysRevD.49.6173).
- [28] H. Baer, C. Balazs, A. Belyaev et al., “Updated constraints on the minimal supergravity model”, *JHEP* **0207** (2002) 050, [arXiv:hep-ph/0205325](https://arxiv.org/abs/hep-ph/0205325).
- [29] M. Dine and W. Fischler, “A Phenomenological Model of Particle Physics Based on Supersymmetry”, *Phys.Lett.* **B110** (1982) 227.
[doi:10.1016/0370-2693\(82\)91241-2](https://doi.org/10.1016/0370-2693(82)91241-2).
- [30] L. Alvarez-Gaume, M. Claudson, and M. B. Wise, “Low-Energy Supersymmetry”, *Nucl.Phys.* **B207** (1982) 96. [doi:10.1016/0550-3213\(82\)90138-9](https://doi.org/10.1016/0550-3213(82)90138-9).
- [31] The ALEPH, DELPHI, L3, OPAL Collaborations, the LEP Electroweak Working Group, “Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP”, *Phys. Rept.* **532** (2013) 119, [arXiv:1302.3415](https://arxiv.org/abs/1302.3415).
- [32] J. Erler, “Tests of the Electroweak Standard Model”, *J.Phys.Conf.Ser.* **485** (2014) 012010, [arXiv:1209.3324](https://arxiv.org/abs/1209.3324). [doi:10.1088/1742-6596/485/1/012010](https://doi.org/10.1088/1742-6596/485/1/012010).
- [33] M. Ciuchini, E. Franco, S. Mishima et al., “Electroweak Precision Observables, New Physics and the Nature of a 126 GeV Higgs Boson”, *JHEP* **1308** (2013) 106, [arXiv:1306.4644](https://arxiv.org/abs/1306.4644). [doi:10.1007/JHEP08\(2013\)106](https://doi.org/10.1007/JHEP08(2013)106).
- [34] Gfitter Group Collaboration, “The global electroweak fit at NNLO and prospects for the LHC and ILC”, *Eur.Phys.J.* **C74** (2014) 3046, [arXiv:1407.3792](https://arxiv.org/abs/1407.3792).
[doi:10.1140/epjc/s10052-014-3046-5](https://doi.org/10.1140/epjc/s10052-014-3046-5).
- [35] CMS Collaboration Collaboration, “Measurement of the $B(s)$ to $\mu^+ \mu^-$ branching fraction and search for B_0 to $\mu^+ \mu^-$ with the CMS Experiment”, *Phys.Rev.Lett.* **111** (2013) 101804, [arXiv:1307.5025](https://arxiv.org/abs/1307.5025).
[doi:10.1103/PhysRevLett.111.101804](https://doi.org/10.1103/PhysRevLett.111.101804).
- [36] LHCb collaboration Collaboration, “Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and search for $B^0 \rightarrow \mu^+ \mu^-$ decays at the LHCb experiment”, *Phys.Rev.Lett.* **111** (2013) 101805, [arXiv:1307.5024](https://arxiv.org/abs/1307.5024).
[doi:10.1103/PhysRevLett.111.101805](https://doi.org/10.1103/PhysRevLett.111.101805).
- [37] CMS and LHCb Collaborations Collaboration, “Combination of results on the rare decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ from the CMS and LHCb experiments”, Technical Report CMS-PAS-BPH-13-007. CERN-LHCb-CONF-2013-012, CERN, Geneva, Jun, 2014.
- [38] Muon G-2 Collaboration, “Final report of the muon E821 anomalous magnetic moment measurement at BNL”, *Phys. Rev.* **D73** (2006) 072003,
[arXiv:hep-ex/0602035](https://arxiv.org/abs/hep-ex/0602035). [doi:10.1103/PhysRevD.73.072003](https://doi.org/10.1103/PhysRevD.73.072003).
- [39] K. Hagiwara, R. Liao, A. D. Martin et al., “ $(g - 2)_\mu$ and $\alpha(M_Z^2)$ re-evaluated using new precise data”, *J.Phys.* **G38** (2011) 085003, [arXiv:1105.3149](https://arxiv.org/abs/1105.3149).
[doi:10.1088/0954-3899/38/8/085003](https://doi.org/10.1088/0954-3899/38/8/085003).

- [40] K. A. Olive et al. (Particle Data Group), “Review of Particle Physics”, *Chin. Phys. C* **38** (2014) 090001. doi:10.1088/1674-1137/38/9/090001.
- [41] M. Davier, A. Hoecker, B. Malaescu et al., “Reevaluation of the Hadronic Contributions to the Muon g-2 and to alpha(MZ)”, *Eur.Phys.J. C* **71** (2011) 1515, arXiv:1010.4180. doi:10.1140/epjc/s10052-012-1874-8, 10.1140/epjc/s10052-010-1515-z.
- [42] D. G. Cerdeno and A. M. Green, “Direct detection of WIMPs”, *arXiv preprint arXiv:1002.1912* (2010).
- [43] M. Cirelli, G. Corcella, A. Hektor et al., “PPPC 4 DM ID: A Poor Particle Physicist Cookbook for Dark Matter Indirect Detection”, *JCAP* **1103** (2011) 051, arXiv:1012.4515. doi:10.1088/1475-7516/2012/10/E01, 10.1088/1475-7516/2011/03/051.
- [44] e. Bennett, C. L., “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results”, *Astrophysical Journal Supplement Series* (2012) arXiv:1212.5226.
- [45] H1 Collaboration, “A Search for Selectrons and Squarks at HERA”, *Phys. Lett. B* **380** (1996) 461, arXiv:hep-ex/9605002. doi:10.1016/0370-2693(96)00640-5.
- [46] ZEUS Collaboration Collaboration, “Search for selectron and squark production in e^+p collisions at HERA”, *Phys.Lett. B* **434** (1998) 214–230, arXiv:hep-ex/9806019. doi:10.1016/S0370-2693(98)00817-X.
- [47] ALEPH Collaboration, “Absolute Mass Lower Limit for the Lightest Neutralino of the MSSM from e^+e^- data at \sqrt{s} up to 209 GeV”, *Phys. Lett. B* **583** (2004) 247. See also references therein. doi:10.1016/j.physletb.2003.12.066.
- [48] DELPHI Collaboration, “Searches for Supersymmetric Particles in e^+e^- Collisions up to 208 GeV and Interpretation of the Results within the MSSM”, *Eur. Phys. J. C* **31** (2003) 421, arXiv:hep-ex/0311019. See also references therein. doi:10.1140/epjc/s2003-01355-5.
- [49] L3 Collaboration, “Search for Scalar Leptons and Scalar Quarks at LEP”, *Phys. Lett. B* **580** (2004) 37, arXiv:hep-ex/0310007. See also references therein. doi:10.1016/j.physletb.2003.10.010.
- [50] OPAL Collaboration, “Search for Chargino and Neutralino Production at $\sqrt{s} = 192$ GeV to 209 GeV at LEP”, *Eur. Phys. J. C* **35** (2004) 1, arXiv:hep-ex/0401026. See also references therein. doi:10.1140/epjc/s2004-01780-x.
- [51] ALEPH, DELPHI, L3 and OPAL Collaborations, “Joint SUSY Working Group”, *LEPSUSYWG/02-06-2* (2004). <http://lepsusy.web.cern.ch/lepsusy>.
- [52] CDF Collaboration Collaboration, “Inclusive Search for Squark and Gluino Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ -TeV”, *Phys.Rev.Lett.* **102** (2009) 121801, arXiv:0811.2512. doi:10.1103/PhysRevLett.102.121801.

- [53] DØ Collaboration, Abazov et al., “Search for squarks and gluinos in events with jets and missing transverse energy using 2.1 fb^{-1} of $p\bar{p}$ collision data at $\sqrt{s} = 1.96 \text{ TeV}$ ”, *Phys. Lett.* **B 660** (2008) 449, [arXiv:0712.3805](https://arxiv.org/abs/0712.3805).
- [54] DØ Collaboration, “Search for associated production of charginos and neutralinos in the trilepton final state using 2.3 fb^{-1} of data”, *Phys. Lett.* **B 680** (2009), no. 1, 34 – 43. doi:[DOI: 10.1016/j.physletb.2009.08.011](https://doi.org/10.1016/j.physletb.2009.08.011).
- [55] G. L. Kane and J. Leveille, “Experimental Constraints on Gluino Masses and Supersymmetric Theories”, *Phys.Lett.* **B112** (1982) 227.
doi:[10.1016/0370-2693\(82\)90968-6](https://doi.org/10.1016/0370-2693(82)90968-6).
- [56] P. Harrison and C. Llewellyn Smith, “Hadroproduction of Supersymmetric Particles”, *Nucl.Phys.* **B213** (1983) 223. doi:[10.1016/0550-3213\(83\)90510-2](https://doi.org/10.1016/0550-3213(83)90510-2).
- [57] E. Reya and D. Roy, “SUPERSYMMETRIC PARTICLE PRODUCTION AT p anti-p COLLIDER ENERGIES”, *Phys.Rev.* **D32** (1985) 645.
doi:[10.1103/PhysRevD.32.645](https://doi.org/10.1103/PhysRevD.32.645).
- [58] S. Dawson, E. Eichten, and C. Quigg, “Search for Supersymmetric Particles in Hadron - Hadron Collisions”, *Phys.Rev.* **D31** (1985) 1581.
doi:[10.1103/PhysRevD.31.1581](https://doi.org/10.1103/PhysRevD.31.1581).
- [59] H. Baer and X. Tata, “Component Formulae for Hadroproduction of Left-handed and Right-handed Squarks”, *Phys.Lett.* **B160** (1985) 159.
doi:[10.1016/0370-2693\(85\)91484-4](https://doi.org/10.1016/0370-2693(85)91484-4).
- [60] M. Kramer, A. Kulesza, R. van der Leeuw et al., “Supersymmetry production cross sections in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ ”, [arXiv:1206.2892](https://arxiv.org/abs/1206.2892).
- [61] C. Borschensky, M. Krmer, A. Kulesza et al., “Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33$ and 100 TeV ”, [arXiv:1407.5066](https://arxiv.org/abs/1407.5066).
- [62] ATLAS Collaboration, CDF Collaboration, CMS Collaboration, D0 Collaboration, “First combination of Tevatron and LHC measurements of the top-quark mass”, [arXiv:1403.4427](https://arxiv.org/abs/1403.4427).
- [63] CMS Collaboration, “Search for new physics with jets and missing transverse momentum in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ ”, *Journal of High Energy Physics* **2011** (2011) 1–46, [arXiv:1106.4503](https://arxiv.org/abs/1106.4503). doi:[10.1007/JHEP08\(2011\)155](https://doi.org/10.1007/JHEP08(2011)155).
- [64] CMS Collaboration, “Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 7 \text{ TeV}$ ”, *Phys.Rev.Lett.* **109** (2012) 171803, [arXiv:1207.1898](https://arxiv.org/abs/1207.1898).
doi:[10.1103/PhysRevLett.109.171803](https://doi.org/10.1103/PhysRevLett.109.171803).
- [65] CMS Collaboration, “CMS Supersymmetry Physics Results”, *Public CMS Wiki* (2014). <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>, Topic revision: r279.

- [66] N. Arkani-Hamed, P. Schuster, N. Toro et al., “MAMBOSET: The Path from LHC Data to the New Standard Model via On-Shell Effective Theories”, [arXiv:hep-ph/0703088](https://arxiv.org/abs/hep-ph/0703088).
- [67] J. Alwall, P. Schuster, and N. Toro, “Simplified Models for a First Characterization of New Physics at the LHC”, *Phys.Rev.* **D79** (2009) 075020, [arXiv:0810.3921](https://arxiv.org/abs/0810.3921). doi:[10.1103/PhysRevD.79.075020](https://doi.org/10.1103/PhysRevD.79.075020).
- [68] J. Alwall, M.-P. Le, M. Lisanti et al., “Model-Independent Jets plus Missing Energy Searches”, *Phys.Rev.* **D79** (2009) 015005, [arXiv:0809.3264](https://arxiv.org/abs/0809.3264). doi:[10.1103/PhysRevD.79.015005](https://doi.org/10.1103/PhysRevD.79.015005).
- [69] CMS Collaboration Collaboration, “Interpretation of Searches for Supersymmetry with simplified Models”, *Phys.Rev.* **D88** (2013), no. 5, 052017, [arXiv:1301.2175](https://arxiv.org/abs/1301.2175). doi:[10.1103/PhysRevD.88.052017](https://doi.org/10.1103/PhysRevD.88.052017).
- [70] O. S. Brüning, P. Collier, P. Lebrun et al., “LHC Design Report”, volume 1: The LHC Main Ring. CERN, Geneva, 2004. CERN-2004-003-V-1.
- [71] L. Evans and P. Bryant, “LHC Machine”, *Journal of Instrumentation* **3** (2008), no. 08, S08001. doi:[10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [72] CERN, “LEP Design Report”, volume 2: The LEP Main Ring. CERN, Geneva, 1984. CERN-LEP-84-01.
- [73] CERN, “LHC: the guide (English version)”. CERN-Brochure-2009-003-Eng, 2009.
- [74] 2012. W.J. Stirling, private communication.
- [75] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003. doi:[10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [76] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004. doi:[10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [77] CMS Collaboration, “CMS Detector Performance and Software”, volume 1 of *CMS Physics Technical Design Report*. CERN, Geneva, 2006. CMS-TDR-008-1. CERN-LHCC-2006-001.
- [78] LHCb Collaboration, “The LHCb Detector at the LHC”, *JINST* **3** (2008) S08005. doi:[10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [79] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002. doi:[10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [80] CMS Collaboration, “The CMS tracker system project”, volume 5 of *Technical Design Report CMS*. CERN, Geneva, 1997. CMS-TDR-005. CERN-LHCC-98-006.
- [81] CMS Collaboration, “Luminosity Information for the 2012 Proton Run”, 2012. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [82] “LHC machine outreach”, 2013. <http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/>.

- [83] M. Lamont, “Status of the LHC”, *J.Phys.Conf.Ser.* **455** (2013) 012001.
[doi:10.1088/1742-6596/455/1/012001](https://doi.org/10.1088/1742-6596/455/1/012001).
- [84] M. H. Seymour and M. Marx, “Monte Carlo Event Generators”, [arXiv:1304.6677](https://arxiv.org/abs/1304.6677).
- [85] A. Buckley, J. Butterworth, S. Gieseke et al., “General-purpose event generators for LHC physics”, *Phys.Rept.* **504** (2011) 145–233, [arXiv:1101.2599](https://arxiv.org/abs/1101.2599).
[doi:10.1016/j.physrep.2011.03.005](https://doi.org/10.1016/j.physrep.2011.03.005).
- [86] N. Metropolis and S. Ulam, “The Monte Carlo Method”, *Journal of the American Statistical Association* **44** (1949), no. 247, 335–341. [doi:10.2307/2280232](https://doi.org/10.2307/2280232).
- [87] M. Dobbs and J. Hansen, “HepMC 2 — a C++ Event Record for Monte Carlo Generators: User Manual Version 2.06”, *Comput. Phys. Commun.* **134** (2001) 41.
<http://lcgapp.cern.ch/project/simu/HepMC>.
- [88] H1 and ZEUS Collaboration, “Combined Measurement and QCD Analysis of the Inclusive $e^\pm p$ Scattering Cross Sections at HERA”, *JHEP* **01** (2010) 109,
[arXiv:0911.0884](https://arxiv.org/abs/0911.0884). [doi:10.1007/JHEP01\(2010\)109](https://doi.org/10.1007/JHEP01(2010)109).
- [89] J. C. Collins and D. E. Soper, “The Theorems of Perturbative QCD”, *Ann.Rev.Nucl.Part.Sci.* **37** (1987) 383–409.
[doi:10.1146/annurev.ns.37.120187.002123](https://doi.org/10.1146/annurev.ns.37.120187.002123).
- [90] J. C. Collins, D. E. Soper, and G. F. Sterman, “Factorization of Hard Processes in QCD”, *Adv.Ser.Direct.High Energy Phys.* **5** (1988) 1–91, [arXiv:hep-ph/0409313](https://arxiv.org/abs/hep-ph/0409313).
- [91] B. Andersson, G. Gustafson, G. Ingelman et al., “Parton fragmentation and string dynamics”, *Physics Reports* **97** (1983), no. 2 - 3, 31 – 145.
[doi:\[http://dx.doi.org/10.1016/0370-1573\\(83\\)90080-7\]\(http://dx.doi.org/10.1016/0370-1573\(83\)90080-7\)](http://dx.doi.org/10.1016/0370-1573(83)90080-7).
- [92] D. Amati and G. Veneziano, “Preconfinement as a property of perturbative QCD”, *Physics Letters B* **83** (1979), no. 1, 87 – 92.
[doi:\[http://dx.doi.org/10.1016/0370-2693\\(79\\)90896-7\]\(http://dx.doi.org/10.1016/0370-2693\(79\)90896-7\)](http://dx.doi.org/10.1016/0370-2693(79)90896-7).
- [93] A. Bassetto, M. Ciafaloni, and G. Marchesini, “Color singlet distributions and mass damping in perturbative QCD”, *Physics Letters B* **83** (1979), no. 2, 207 – 212. [doi:\[http://dx.doi.org/10.1016/0370-2693\\(79\\)90687-7\]\(http://dx.doi.org/10.1016/0370-2693\(79\)90687-7\)](http://dx.doi.org/10.1016/0370-2693(79)90687-7).
- [94] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **0605** (2006) 026, [arXiv:hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175).
[doi:10.1088/1126-6708/2006/05/026](https://doi.org/10.1088/1126-6708/2006/05/026).
- [95] M. Bahr, S. Gieseke, M. Gigg et al., “Herwig++ Physics and Manual”, *Eur.Phys.J.* **C58** (2008) 639–707, [arXiv:0803.0883](https://arxiv.org/abs/0803.0883). [doi:10.1140/epjc/s10052-008-0798-9](https://doi.org/10.1140/epjc/s10052-008-0798-9).
- [96] J. Alwall, P. Demin, S. de Visscher et al., “MadGraph/MadEvent v4: The New Web Generation”, *JHEP* **0709** (2007) 028, [arXiv:0706.2334](https://arxiv.org/abs/0706.2334).
[doi:10.1088/1126-6708/2007/09/028](https://doi.org/10.1088/1126-6708/2007/09/028).

- [97] J. Alwall, R. Frederix, S. Frixione et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **1407** (2014) 079, [arXiv:1405.0301](#).
[doi:10.1007/JHEP07\(2014\)079](#).
- [98] C. Oleari, “The POWHEG-BOX”, *Nucl.Phys.Proc.Suppl.* **205-206** (2010) 36–41, [arXiv:1007.3893](#). [doi:10.1016/j.nuclphysbps.2010.08.016](#).
- [99] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **0411** (2004) 040, [arXiv:hep-ph/0409146](#).
[doi:10.1088/1126-6708/2004/11/040](#).
- [100] S. Agostinelli, J. Allison, K. Amako et al., “G4 – a simulation toolkit”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250 – 303.
[doi:10.1016/S0168-9002\(03\)01368-8](#).
- [101] CMS Collaboration Collaboration, “The fast simulation of the CMS detector at LHC”, *J.Phys.Conf.Ser.* **331** (2011) 032049.
[doi:10.1088/1742-6596/331/3/032049](#).
- [102] CMS Collaboration Collaboration, “Comparison of the Fast Simulation of CMS with the first LHC data”, CMS-DP-2010-039.
- [103] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss} ”, *CMS Physics Analysis Summary CMS-PAS-PFT-09-001* (2009).
- [104] W. Adam, B. Mangano, T. Speer et al., “Track Reconstruction in the CMS tracker”, Technical Report CMS-NOTE-2006-041, CERN, Geneva, Dec, 2006.
- [105] G. P. Salam, “Towards Jetography”, *Eur.Phys.J.* **C67** (2010) 637–686, [arXiv:0906.1833](#). [doi:10.1140/epjc/s10052-010-1314-6](#).
- [106] G. P. Salam and G. Soyez, “A Practical Seedless Infrared-Safe Cone jet algorithm”, *JHEP* **0705** (2007) 086, [arXiv:0704.0292](#).
[doi:10.1088/1126-6708/2007/05/086](#).
- [107] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev. D* **48** (Oct, 1993) 3160 – 3166.
[doi:10.1103/PhysRevD.48.3160](#).
- [108] Y. L. Dokshitzer, G. Leder, S. Moretti et al., “Better jet clustering algorithms”, *JHEP* **9708** (1997) 001, [arXiv:hep-ph/9707323](#).
[doi:10.1088/1126-6708/1997/08/001](#).
- [109] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross-sections in deep inelastic scattering”, [arXiv:hep-ph/9907280](#). DESY-THESIS-2000-049.
- [110] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_T jet clustering algorithm”, *Journal of High Energy Physics* **2008** (2008), no. 04, 063, [arXiv:0802.1189](#).
[doi:10.1088/1126-6708/2008/04/063](#).

- [111] CMS Collaboration Collaboration, “Jet Energy Corrections and Uncertainties. Detector Performance Plots for 2012.”,,
- [112] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, [arXiv:1111.6097](https://arxiv.org/abs/1111.6097).
- [113] M. Cacciari and G. P. Salam, “Dispelling the N^3 myth for the k_t jet-finder”, *Phys.Lett.* **B641** (2006) 57–61, [arXiv:hep-ph/0512210](https://arxiv.org/abs/hep-ph/0512210).
[doi:10.1016/j.physletb.2006.08.037](https://doi.org/10.1016/j.physletb.2006.08.037).
- [114] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *Journal of Instrumentation* **6** (2011) P11002, [arXiv:1107.4277](https://arxiv.org/abs/1107.4277). [doi:10.1088/1748-0221/6/11/P11002](https://doi.org/10.1088/1748-0221/6/11/P11002).
- [115] CMS Collaboration, “The Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS”, *CMS Physics Analysis Summary CMS-PAS-JME-09-002* (2009).
- [116] CMS Collaboration Collaboration, “8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb^{-1} of data in CMS”,,
- [117] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Physics Letters B* **659** (2008), no. 1-2, 119 – 126. % tt
[doi:<http://dx.doi.org/10.1016/j.physletb.2007.09.077>](http://dx.doi.org/10.1016/j.physletb.2007.09.077).
- [118] CMS Collaboration Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, [arXiv:1211.4462](https://arxiv.org/abs/1211.4462).
[doi:10.1088/1748-0221/8/04/P04013](https://doi.org/10.1088/1748-0221/8/04/P04013).
- [119] CMS Collaboration Collaboration, “Performance of b tagging at $\sqrt{s}=8 \text{ TeV}$ in multijet, ttbar and boosted topology events”, Technical Report CMS-PAS-BTV-13-001, CERN, Geneva, 2013.
- [120] J. M. Butterworth, A. R. Davison, M. Rubin et al., “Jet substructure as a new Higgs search channel at the LHC”, [arXiv:0810.0409](https://arxiv.org/abs/0810.0409).
- [121] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Techniques for improved heavy particle searches with jet substructure”, *Phys.Rev.* **D80** (2009) 051501, [arXiv:0903.5081](https://arxiv.org/abs/0903.5081). [doi:10.1103/PhysRevD.80.051501](https://doi.org/10.1103/PhysRevD.80.051501).
- [122] D. Krohn, J. Thaler, and L.-T. Wang, “Jet Trimming”, *JHEP* **1002** (2010) 084, [arXiv:0912.1342](https://arxiv.org/abs/0912.1342). [doi:10.1007/JHEP02\(2010\)084](https://doi.org/10.1007/JHEP02(2010)084).
- [123] CMS Collaboration Collaboration, “Boosted Top Jet Tagging at CMS”.,
- [124] CMS Collaboration Collaboration, “Boosted Top Jet Tagging at CMS”.,
- [125] CMS Collaboration Collaboration, “A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging”, Technical Report CMS-PAS-JME-09-001, CERN, 2009. Geneva, Jul, 2009.
- [126] D. E. Kaplan, K. Rehermann, M. D. Schwartz et al., “Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks”, *Phys.Rev.Lett.* **101** (2008) 142001, [arXiv:0806.0848](https://arxiv.org/abs/0806.0848). [doi:10.1103/PhysRevLett.101.142001](https://doi.org/10.1103/PhysRevLett.101.142001).

- [127] T. Plehn, M. Spannowsky, M. Takeuchi et al., “Stop Reconstruction with Tagged Tops”, *JHEP* **1010** (2010) 078, [arXiv:1006.2833](#).
[doi:10.1007/JHEP10\(2010\)078](#).
- [128] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP* **1103** (2011) 015, [arXiv:1011.2268](#). [doi:10.1007/JHEP03\(2011\)015](#).
- [129] J. Thaler and K. Van Tilburg, “Maximizing Boosted Top Identification by Minimizing N-subjettiness”, *JHEP* **1202** (2012) 093, [arXiv:1108.2701](#).
[doi:10.1007/JHEP02\(2012\)093](#).
- [130] M. Schröder, “Quality of Jet Measurements and Impact on a Search for New Physics at CMS”. PhD thesis, Universität Hamburg, Germany, 2012.
DESY-THESIS-2012-042.
- [131] F. Pandolfi, D. del Re, M. Voutilainen, “Jet Response and Resolution Measurement with Photon+Jet Events at $\sqrt{s} = 7$ TeV”, *CMS AN-2010/141* (2010).
- [132] F. Pandolfi, D. del Re, M. Voutilainen, “Update on Jet Response and Resolution Measurements with Photon+Jet Events at $\sqrt{s} = 7$ TeV”, *CMS AN-2011/004* (2011).
- [133] T. Lenz, C. Sander, P. Schleper et al., “Jet Energy Resolution Measurement using Photon+Jet Events at $\sqrt{s} = 8$ TeV”, *CMS AN-2013/179* (2013).
- [134] K. Goebel, J. Haller, J. Ott et al., “Jet Transverse Momentum Resolution Measurement using Dijet Events at $\sqrt{s} = 8$ TeV”, *CMS Analysis Note CMS-AN-13-416* (2013). Internal documentation.
- [135] CMS Collaboration, “Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV”, *JHEP* **1109** (2011) 109, [arXiv:1107.0330](#). [doi:10.1007/JHEP09\(2011\)109](#).
- [136] J. Pumplin, D. Stump, J. Huston et al., “New generation of parton distributions with uncertainties from global QCD analysis”, *JHEP* **0207** (2002) 012, [arXiv:hep-ph/0201195](#). [doi:10.1088/1126-6708/2002/07/012](#).
- [137] D0 Collaboration, “High- p_T jets in $\bar{p}p$ collisions at $\sqrt{s} = 630$ GeV and 1800 GeV”, *Phys.Rev.* **D64** (2001) 032003, [arXiv:hep-ex/0012046](#).
[doi:10.1103/PhysRevD.64.032003](#).
- [138] D0 Collaboration, “Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV”, *Phys.Rev.* **D85** (2012) 052006, [arXiv:1110.3771](#).
[doi:10.1103/PhysRevD.85.052006](#).
- [139] ATLAS Collaboration, “Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector”, *Eur.Phys.J.* **C73** (2013) 2306, [arXiv:1210.6210](#). [doi:10.1140/epjc/s10052-013-2306-0](#).
- [140] CMS Collaboration, “Measurement of Differential Jet Cross Sections at $\sqrt{s} = 7$ TeV with the CMS Detector”, *CMS Physics Analysis Summary CMS-PAS-QCD-11-004* (2012).

- [141] CMS Collaboration Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update”, Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013.
- [142] “Hamburg Wiki Calibration”. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/HamburgWikiAnalysisCalibration>, Jan., 2014.
- [143] CMS Collaboration, “Calorimeter Jet Quality Criteria for the First CMS Collision Data”, *CMS Physics Analysis Summary CMS-PAS-JME-09-008* (2008).
- [144] CMS Collaboration, “Jet Performance in pp Collisions at 7 TeV”, *CMS Physics Analysis Summary CMS-PAS-JME-10-003* (2010).
- [145] 2013. J. Ott, private communication.
- [146] F. James and M. Roos, “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”, *Comput.Phys.Commun.* **10** (1975) 343–367. doi:[10.1016/0010-4655\(75\)90039-9](https://doi.org/10.1016/0010-4655(75)90039-9).
- [147] CMS Collaboration Collaboration, “Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV”, *JHEP* **1103** (2011) 136, arXiv:1102.3194. doi:[10.1007/JHEP03\(2011\)136](https://doi.org/10.1007/JHEP03(2011)136).
- [148] C. Autermann and other, “Search for new physics in multijets and missing momentum final state at 8 TeV”, *CMS Analysis Note CMS-AN-12-350* (2012). Internal documentation.
- [149] CMS Collaboration Collaboration, “Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s}=8$ TeV”, *JHEP* **1406** (2014) 055, arXiv:1402.4770. doi:[10.1007/JHEP06\(2014\)055](https://doi.org/10.1007/JHEP06(2014)055).
- [150] N. Kidonakis, “Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution”, *Phys.Rev.* **D82** (2010) 114030, arXiv:1009.4935. doi:[10.1103/PhysRevD.82.114030](https://doi.org/10.1103/PhysRevD.82.114030).
- [151] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through $O(\alpha_s)^{**2}$ ”, *Phys.Rev.* **D74** (2006) 114017, arXiv:hep-ph/0609070. doi:[10.1103/PhysRevD.74.114017](https://doi.org/10.1103/PhysRevD.74.114017).
- [152] W. Beenakker, R. Hopker, M. Spira et al., “Squark and gluino production at hadron colliders”, *Nucl.Phys.* **B492** (1997) 51–103, arXiv:hep-ph/9610490. doi:[10.1016/S0550-3213\(97\)80027-2](https://doi.org/10.1016/S0550-3213(97)80027-2).
- [153] A. Kulesza and L. Motyka, “Threshold Resummation for Squark-Antisquark and Gluino-Pair Production at the LHC”, *Phys. Rev. Lett.* **102** (Mar, 2009) 111802. doi:[10.1103/PhysRevLett.102.111802](https://doi.org/10.1103/PhysRevLett.102.111802).
- [154] A. Kulesza and L. Motyka, “Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC”, *Phys. Rev. D* **80** (Nov, 2009) 095004. doi:[10.1103/PhysRevD.80.095004](https://doi.org/10.1103/PhysRevD.80.095004).

- [155] W. Beenakker, S. Brensing, M. Kramer et al., “Soft-gluon resummation for squark and gluino hadroproduction”, *JHEP* **0912** (2009) 041, [arXiv:0909.4418](https://arxiv.org/abs/0909.4418).
[doi:10.1088/1126-6708/2009/12/041](https://doi.org/10.1088/1126-6708/2009/12/041).
- [156] W. Beenakker, S. Brensing, M. Kramer et al., “Squark and Gluino Hadroproduction”, *Int.J.Mod.Phys.* **A26** (2011) 2637–2664, [arXiv:1105.1110](https://arxiv.org/abs/1105.1110).
[doi:10.1142/S0217751X11053560](https://doi.org/10.1142/S0217751X11053560).
- [157] CMS Collaboration, “Electron reconstruction and identification at $\sqrt{s} = 7 \text{ TeV}$ ”, *CMS Physics Analysis Summary CMS-PAS-EGM-10-004* (2010).
- [158] CMS Collaboration, “Performance of muon identification in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ ”, *CMS Physics Analysis Summary CMS-PAS-MUO-10-002* (2010).
- [159] CMS Collaboration Collaboration, “Search for supersymmetry in hadronic final states with missing transverse energy using the variables α_T and b-quark multiplicity in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ ”, *Eur.Phys.J.* **C73** (2013) 2568, [arXiv:1303.2985](https://arxiv.org/abs/1303.2985). [doi:10.1140/epjc/s10052-013-2568-6](https://doi.org/10.1140/epjc/s10052-013-2568-6).
- [160] CMS Collaboration Collaboration, “Search for gluino mediated bottom- and top-squark production in multijet final states in pp collisions at 8 TeV”, *Phys.Lett.* **B725** (2013) 243–270, [arXiv:1305.2390](https://arxiv.org/abs/1305.2390).
[doi:10.1016/j.physletb.2013.06.058](https://doi.org/10.1016/j.physletb.2013.06.058).
- [161] CMS Collaboration Collaboration, “Search for supersymmetry in hadronic final states using MT2 with the CMS detector at $\sqrt{s} = 8 \text{ TeV}$ ”, Technical Report CMS-PAS-SUS-13-019, CERN, Geneva, 2014.
- [162] CMS Collaboration Collaboration, “Exclusion limits on gluino and top-squark pair production in natural SUSY scenarios with inclusive razor and exclusive single-lepton searches at 8 TeV.”, Technical Report CMS-PAS-SUS-14-011, CERN, Geneva, 2014.
- [163] S. Ask, M. Parker, T. Sandoval et al., “Using gamma+jets Production to Calibrate the Standard Model $Z(\nu\bar{\nu})$ +jets Background to New Physics Processes at the LHC”, *JHEP* **1110** (2011) 058, [arXiv:1107.2803](https://arxiv.org/abs/1107.2803).
[doi:10.1007/JHEP10\(2011\)058](https://doi.org/10.1007/JHEP10(2011)058).
- [164] Z. Bern, G. Diana, L. Dixon et al., “Missing Energy and Jets for Supersymmetry Searches”, *Phys.Rev.* **D87** (2013), no. 3, 034026, [arXiv:1206.6064](https://arxiv.org/abs/1206.6064).
[doi:10.1103/PhysRevD.87.034026](https://doi.org/10.1103/PhysRevD.87.034026).
- [165] Z. Bern, G. Diana, L. Dixon et al., “Driving Missing Data at Next-to-Leading Order”, *Phys.Rev.* **D84** (2011) 114002, [arXiv:1106.1423](https://arxiv.org/abs/1106.1423).
[doi:10.1103/PhysRevD.84.114002](https://doi.org/10.1103/PhysRevD.84.114002).
- [166] B. Efron, “The Jackknife, the bootstrap and other resampling plans”. Number 38 in Regional Conference Series in applied mathematics. Society for Industrial and applied mathematics, Philadelphia, Pa., 1982.
- [167] 2014. M. Schroeder, private communication.

- [168] J. D'Hondt, S. Lowette, O. L. Buchmller et al., “Fitting of Event Topologies with External Kinematic Constraints in CMS”, Technical Report CMS-NOTE-2006-023, CERN, Geneva, Jan, 2006.
- [169] A. L. Read, “Presentation of search results: the CL_s technique”, *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10, 2693 – 2704.
[doi:10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [170] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **434** (1999), no. 2-3, 435 – 443. [doi:10.1016/S0168-9002\(99\)00498-2](https://doi.org/10.1016/S0168-9002(99)00498-2).
- [171] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report ATL-PHYS-PUB-2011-011. CMS-NOTE-2011-005, CERN, Geneva, 2011.
- [172] M. Botje, J. Butterworth, A. Cooper-Sarkar et al., “The PDF4LHC Working Group Interim Recommendations”. 2011.
- [173] CMS Collaboration, “Search for top-squark pair production in the single-lepton final state in pp collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **73** (2013) 2677, [arXiv:1308.1586](https://arxiv.org/abs/1308.1586). [doi:10.1140/epjc/s10052-013-2677-2](https://doi.org/10.1140/epjc/s10052-013-2677-2).
- [174] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao et al., “Implications of CTEQ global analysis for collider observables”, *Phys.Rev. D* **78** (2008) 013004, [arXiv:0802.0007](https://arxiv.org/abs/0802.0007).
[doi:10.1103/PhysRevD.78.013004](https://doi.org/10.1103/PhysRevD.78.013004).
- [175] A. Martin, W. Stirling, R. Thorne et al., “Parton distributions for the LHC”, *Eur.Phys.J. C* **63** (2009) 189–285, [arXiv:0901.0002](https://arxiv.org/abs/0901.0002).
[doi:10.1140/epjc/s10052-009-1072-5](https://doi.org/10.1140/epjc/s10052-009-1072-5).
- [176] CMS Collaboration Collaboration, “Search for top squarks in multijet events with large missing momentum in proton-proton collisions at 8 TeV”, Technical Report CMS-PAS-SUS-13-015, CERN, Geneva, 2013.
- [177] J. Ott. www.theta-framework.de.
- [178] CMS Collaboration Collaboration, “Search for supersymmetry in final states with missing transverse energy and 0, 1, 2, or at least 3 b-quark jets in 7 TeV pp collisions using the variable alphaT”, *JHEP* **1301** (2013) 077, [arXiv:1210.8115](https://arxiv.org/abs/1210.8115).
[doi:10.1007/JHEP01\(2013\)077](https://doi.org/10.1007/JHEP01(2013)077).
- [179] C. G. Lester and D. J. Summers, “Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders”, *Phys. Lett. B* **463** (1999) 99–103, [arXiv:hep-ph/9906349](https://arxiv.org/abs/hep-ph/9906349). [doi:10.1016/S0370-2693\(99\)00945-4](https://doi.org/10.1016/S0370-2693(99)00945-4).
- [180] A. Barr, C. Lester, and P. Stephens, “ m_{T2} : the Truth behind the glamour”, *J. Phys. G* **29** (2003) 2343–2363, [arXiv:hep-ph/0304226](https://arxiv.org/abs/hep-ph/0304226).
[doi:10.1088/0954-3899/29/10/304](https://doi.org/10.1088/0954-3899/29/10/304).

- [181] CMS Collaboration Collaboration, “Search for supersymmetry in hadronic final states using MT2 in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **1210** (2012) 018, [arXiv:1207.1798](https://arxiv.org/abs/1207.1798). doi:[10.1007/JHEP10\(2012\)018](https://doi.org/10.1007/JHEP10(2012)018).
- [182] CMS Collaboration Collaboration, “Inclusive search for supersymmetry using the razor variables in pp collisions at $\sqrt{s} = 7$ TeV”, *Phys.Rev.Lett.* **111** (2013), no. 8, 081802, [arXiv:1212.6961](https://arxiv.org/abs/1212.6961). doi:[10.1103/PhysRevLett.111.081802](https://doi.org/10.1103/PhysRevLett.111.081802).
- [183] CMS Collaboration Collaboration, “Search for supersymmetry with razor variables in pp collisions at $\sqrt{s}=7$ TeV”, [arXiv:1405.3961](https://arxiv.org/abs/1405.3961).

