Probing supersymmetry based on precise jet measurements at the CMS experiment

Dissertation zur Erlangung des Doktorgrades

an der Fakultät für

Mathematik, Informatik und Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von

Kristin Goebel geb. Heine

Hamburg

2014

Tag der Disputation: 5. Februar 2015
Folgende Gutachter empfehlen die Annahme der Dissertation: Prof. Dr. Johannes Haller
Jun.-Prof. Dr. Christian Sander

Abstract

The search for new physics beyond the standard model of particle physics is one of the main goals of the CMS experiment at the CERN Large Hadron Collider. Many theories, for instance supersymmetry, involve the possible production of new coloured particles which feature jets as their experimental signature. Thus, it is important to have a good understanding of jet-related properties in order to allow such searches.
In the first part of this thesis, a measurement of the jet transverse-momentum resolution is presented. This is based on the analysis of proton-proton collision data recorded at a centre-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$ by the CMS experiment. The measurement utilizes the transverse momentum balance of dijet events at particle level. The main focus is on the determination of the data-to-simulation ratio of the jet transverse-momentum resolution which can be used to correct the jet resolution in simulated events to match the one observed in data. This ratio has been determined with a significantly improved precision compared to previous analyses for the pseudorapidity range $0.0 \leq |\eta| \leq 5.0$.
The second part of the thesis focuses on searches for supersymmetry in final states with several jets, missing transverse momentum and no isolated leptons. A search performed with collision data recorded at $\sqrt{s} = 8\,\text{TeV}$ is presented which is mainly sensitive to the production of light-flavour squarks and gluinos as well as the gluino-mediated production of third generation particles. In this analysis, the main challenge arises from a precise determination of background contributions from standard model processes as the analysis is performed in an extreme kinematic phase space. In this thesis, a method to estimate QCD background contributions relying on the jet-p_T response is presented and necessary modifications for a correct prediction of high jet multiplicity events are introduced. In the analysis, results consistent with standard model expectations have been obtained and the production of light-flavour squarks below 780 GeV and that of gluinos up to 1.1–1.2 TeV has been excluded at 95% confidence level for a mass of the lightest supersymmetric particle (LSP) not exceeding 100 GeV in the context of simplified supersymmetric models.
Furthermore, a prospect study investigating different search strategies for the identification of direct pair production of top squarks is shown. This is based on simulated events at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Utilizing algorithms for the identification of boosted hadronically decaying top quarks arising from the decay of heavy top squarks, a search sensitivity for top squark masses up to the 1 TeV range can be obtained for LSP masses less than approximately 300 GeV with the same integrated luminosity as recorded at $\sqrt{s} = 8\,\text{TeV}$. This selection could improve the search sensitivity with respect to existing analyses. Moreover, the identified selection is also suitable to study gluino-mediated production of third-generation squarks and provides a complementary approach to existing multijet analyses.

Kurzfassung

Die Suche nach neuer Physik jenseits des Standardmodells der Teilchenphysik ist eines der Hauptziele des CMS-Experiments am CERN Large Hadron Collider. Zahlreiche Theorien, beispielsweise Supersymmetrie, beinhalten die mögliche Produktion von neuen farbgeladenen Teilchen, welche als experimentelle Signatur Jets aufweisen. Deshalb ist es wichtig, ein gutes Verständnis dieser Objekte zu erlangen, um derartige Suchen zu ermöglichen.
Im ersten Teil dieser Arbeit wird eine Messung der Transversalimpulsauflösung von Jets vorgestellt, welche auf der Analyse von Proton-Proton-Kollisionsdaten basiert, die bei einer Schwerpunktsenergie von $\sqrt{s} = 8\,\text{TeV}$ vom CMS-Experiment aufgezeichnet wurden. Die Messung basiert auf der Transversalimpulsbalance von Zweijetereignissen auf Teilchenebene. Der Fokus liegt dabei auf der Bestimmung des Verhältnisses der Auflösung in Daten zu der Auflösung in simulierten Ereignissen, welches verwendet werden kann, um die Auflösung in simulierten Ereignissen an die in Daten beobachtete anzupassen. Dieses Verhältnis wurde mit einer signifikant verbesserten Präzision im Vergleich zu vorherigen Analysen für einen Pseudorapiditätsbereich von $0.0 \leq |\eta| \leq 5.0$ bestimmt.
Der zweite Teil der Arbeit konzentriert sich auf Suchen nach Supersymmetrie unter Verwendung von Endzuständen mit zahlreichen Jets, fehlendem Transversalimpuls und keinen isolierten Leptonen. Es wird eine Suche vorgestellt, die auf Kollisionsdaten basiert, welche bei einer Schwerpunktsenergie von $\sqrt{s} = 8\,\text{TeV}$ aufgezeichnet wurden. Die Analyse zielt auf Signaturen ab, welche hauptsächlich sensitiv sind auf die Produktion von Squarks der ersten und zweiten Generation und Gluinos, sowie die gluino-induzierte Produktion von Squarks der dritten Generation. Die größte Herausforderung ergibt sich in dieser Analyse durch eine genaue Bestimmung der Untergrundbeiträge aus Standardmodellprozessen, da die Analyse in einem extremen kinematischen Phasenraum durchgeführt wird. In dieser Arbeit wird eine Methode vorgestellt, die Untergrundbeiträge aus QCD Ereignissen basierend auf der Jet-p_T-Response abschätzt. Darüberhinaus wird eingeführt, wie die Methode modifiziert werden kann, um Ereignisse mit hoher Jet-Multiplizität korrekt vorherzusagen. In der Analyse werden Ergebnisse erzielt, die mit der Erwartung aus dem Standardmodell kompatibel sind. Damit wird die Produktion von Squarks der ersten beiden Generationen unter 780 GeV und die von Gluinos unter 1,1–1,2 TeV im Kontext von vereinfachten supersymmetrischen Modellen mit 95% C.L. für eine Masse des leichtesten supersymmetrischen Teilchens (LSP) unter 100 GeV ausgeschlossen.
Weiterhin wird eine Studie vorgestellt, welche unterschiedliche Analysestrategien zur Identifikation von direkt produzierten Top-Squarks untersucht. Diese Studie basiert auf simulierten Ereignissen bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$. Unter Verwendung von Algorithmen zur Identifikation von geboosteten hadronisch zerfallenden Top-Quarks aus den Zerfällen von Top-Squarks, kann mit derselben integrierten Luminosität wie bei einer Schwerpunktsenergie von 8 TeV erzielt wurde, eine Sensitivität der Suche für Top-Squark Massen bis 1 TeV für LSP Massen unter etwa 300 GeV erreicht werden. Diese Selektion kann die Sensitivität der Suche im Vergleich zu bestehenden Analysen verbessern. Darüberhinaus ist die identifizierte Selektion auch geeignet, um gluino-induzierte Produktion von Squarks der dritten Generation zu studieren und bietet einen komplementären Ansatz zu existierenden Multijet-Analysen.

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 . . . 7
2.2 Supersymmetry . . . 8
2.2.1 Natural Supersymmetry . . . 10
2.2.2 The MSSM . . . 11
2.2.3 SUSY-Breaking . . . 12
2.3 Searches for Supersymmetry and Current Constraints . . . 13
2.3.1 Indirect Constraints . . . 13
2.3.2 Searches at pre-LHC Collider Experiments . . . 14
2.3.3 Searches at the LHC . . . 15

3 Experimental Setup **23**
3.1 The Large Hadron Collider . . . 23
3.2 The CMS Experiment . . . 25
3.2.1 Coordinate Conventions and Kinematic Variables . . . 26
3.2.2 Superconducting Magnet . . . 27
3.2.3 Inner Tracking System . . . 27
3.2.4 Calorimeter System . . . 29
3.2.5 Muon System . . . 32
3.2.6 Trigger System . . . 34
3.3 LHC Operation and Data Taking . . . 35

4 Event Simulation **37**
4.1 Event Modeling . . . 37
4.2 Event Generators . . . 39
4.3 Detector Simulation . . . 40

5 Object Reconstruction and Particle Identification **41**
5.1 Global Event Description with the Particle-Flow Algorithm at CMS . . . 41
5.2 Reconstruction of Jets . . . 42
5.2.1 Jet Algorithms . . . 42
5.2.2 Jet Types at CMS . . . 45
5.2.3 Jet Transverse-Momentum Response . . . 46
5.2.4 Jet Energy Calibration . . . 47
5.3 Identification of b-Quark Jets . . . 49
5.4 Identification of Boosted t-Quark Jets . . . 51

6 Measurement of the Jet Transverse-Momentum Resolution 55
6.1 Components of the Jet Response . . . 55
6.2 Basic Concept of the Dijet Asymmetry Method . . . 57
6.3 Application to Realistic Collision Events . . . 57
6.4 Samples and Event Selection . . . 58
6.4.1 Datasets and Triggers . . . 58
6.4.2 Selection Criteria . . . 59
6.5 Definition of the Asymmetry Width . . . 63
6.6 Corrections to the Dijet Asymmetry . . . 64
6.6.1 Correction for Additional Jet Activity . . . 66
6.6.2 Correction for Particle-Level Imbalance . . . 67
6.6.3 Results of the Corrections to the Asymmetry . . . 68
6.7 Determination of the Data-to-Simulation Ratio . . . 69
6.8 Validation of the Method . . . 71
6.8.1 Validation in Simulated Events . . . 71
6.8.2 Validation of the Measured Data-to-Simulation Ratio . . . 72
6.9 Systematic Uncertainties . . . 72
6.10 Extension of the Method to the Forward Detector Region . . . 78
6.11 Measurement for Simulated Events Obtained with the HERWIG++ Generator 80
6.12 Results . . . 82
6.12.1 Determination of a Combined Result . . . 82
6.12.2 Comparison to Other Measurements . . . 84
6.12.3 Impact of the Improved JER Measurement . . . 85
6.13 Adjustment of the MC-Resolution to Data . . . 86

7 Search for New Physics with Jets and Missing Transverse Momentum 89
7.1 Event Selection . . . 90
7.1.1 Data Samples . . . 90
7.1.2 Trigger . . . 90
7.1.3 Event Cleaning . . . 91
7.1.4 Baseline Selection . . . 94
7.1.5 Exclusive Search Regions . . . 95
7.2 Estimation of Non-QCD Backgrounds . . . 95
7.2.1 Invisible Z Background . . . 96
7.2.2 Hadronic-tau Background . . . 99
7.2.3 Lost-Lepton Background . . . 101
7.3 QCD Background Estimation with the Rebalance-And-Smear Method . . . 102
7.3.1 General concept of the Rebalance-and-Smear method . . . 103
7.3.2 Application to Collision Events . . . 107
7.3.3 Validation in Simulated Events . . . 110
7.3.4 Application to Data Events . . . 114
7.3.5 Validation in Data Events . . . 116
7.3.6 Systematic Uncertainties . . . 117
7.3.7 QCD Background Prediction . . . 119
7.4 Results and Interpretation . . . 121
7.4.1 Comparison to Other Measurements . . . 124
7.5 Status of supersymmetry after LHC Run I . . . 126

8 Prospect Studies for a Search for Top Squarks at $\sqrt{s}$ = 13 TeV 129
8.1 Samples of Simulated Events . . . 129
8.2 Sensitivity of a Basic Selection using H_{T} and $\not{E}_{\mathrm{T}}$ . . . 131
8.3 Improvement using B Tagging . . . 135
8.4 Improvement using Top Tagging . . . 137
8.5 Performance Comparison of Various Kinematic Selections . . . 141
8.6 Performance Comparison to Selection Based on Published Top Squark Analysis . . . 148
8.7 Stability Test . . . 151
8.8 Sensitivity to Gluino-Mediated Stop Production . . . 152
8.9 Results and Discussion . . . 154

9 Conclusions 157

A Jet Transverse-Momentum Resolution 159
A.1 Primary Vertex Distributions . . . 159
A.2 Extrapolation Graphs . . . 161
A.3 Correlation of Asymmetry Widths . . . 166
A.4 Detailed Results of the Forward Extension . . . 168

B Search for New Physics with Jets and Missing Transverse Momentum 171
B.1 Signal Trigger Efficiencies . . . 171
B.2 Data-to-Simulation Ratio: Core Resolution . . . 173
B.3 Data-to-Simulation Ratio: Response Tails . . . 174
B.4 Input Truth Resolutions for Rebalancing in Kinematic Fit . . . 175
B.5 Trigger Efficiencies of Control Triggers for QCD Background Prediction . . . 177

1 Introduction

The current knowledge and understanding of the fundamental particles and interactions between them are summarized in the 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. However, there are several fundamental questions still unanswered, like the origin of dark matter or the accommodation of large radiative corrections to the Higgs boson mass. One of such theories, which goes beyond the standard model and could provide solutions to some of these problems, is supersymmetry (SUSY). In general, there are several possibilities to investigate if supersymmetry is realised in nature. However, large parts of the SUSY parameter space can be best explored in collider experiments.
The Large Hadron Collider (LHC) located at CERN[1] is currently the most powerful particle accelerator and provides proton-proton (pp) collisions at a centre-of-mass energy of up to $\sqrt{s} = 8\,\text{TeV}$ to date. In order to further investigate the SM and search for new physics beyond, the CMS experiment has been built. The CMS experiment is a particle detector designed to analyse particle collisions delivered by the LHC and studies that are based on data recorded by this experiment are presented in this thesis.
Many SM and new physics processes, like SUSY, which are subject to the LHC physics program, manifest in final states containing jets – the experimental signature of quarks and gluons. Thus, it is crucial to have a precise knowledge of jet-related quantities, like the jet transverse-momentum resolution. This can be measured utilizing events with a momentum balance in the transverse plane, like γ + jet, Z + jet or dijet events. In this thesis, a measurement of the jet transverse-momentum resolution in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ using dijet events is performed. These events are especially suited as they are produced at a high rate and enable a measurement with a good detector coverage. In contrast to previous analyses, which have been carried out at $\sqrt{s} = 7\,\text{TeV}$, the measurement presented here provides an improved estimate of statistical and systematic uncertainties and has been extended such that the resolution in the forward part of the detector can be determined with higher precision.
In the second part of the thesis, the detailed knowledge about jets and their resolution is exploited in a search for new physics targeting decays of supersymmetric particles. This analysis is also based on proton-proton collision data recorded at $\sqrt{s} = 8\,\text{TeV}$ and makes use of events with missing transverse energy, several hard jets and no isolated leptons. Previous versions of this analysis have been performed at $\sqrt{s} = 7\,\text{TeV}$ and were in particular sensitive to supersymmetric models describing the production of gluinos as well as first and second generation squarks. The analysis presented here is extended to final states with high jet multiplicities in order to be in addition sensitive to gluino-mediated production of third generation squarks. A key feature in this analysis is a precise prediction of standard model background contributions. Due to large theoretical uncertainties, especially background events from QCD multijet processes are difficult to model. These

[1] *European Organization for Nuclear Research* near Geneva, Suisse

arise from mismeasured jets and decays of heavy-flavour quarks. In this thesis, a method relying on the jet-p_{T} response to estimate QCD background contributions is presented and special considerations to correctly predict high jet multiplicity events are discussed.
Analyses of $\sqrt{s} = 8\,\mathrm{TeV}$ data typically exclude gluino and light-flavour squark masses below around 1 TeV. Thus, it is of particular interest to investigate third generation squarks which have weaker mass limits. Especially the next run period of the LHC, starting in 2015 at a centre-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$, provides ideal conditions to further explore direct production of top squarks up to the TeV mass range. In this thesis, various analysis strategies for a search for top squarks at $\sqrt{s} = 13\,\mathrm{TeV}$ are discussed. Special emphasis is put on the study of several kinematic variables and the application of jet substructure tools.

The main focus of the work documented in this thesis is on the measurement of the jet transverse-momentum resolution, the estimation of QCD background contributions to the search for new physics based on final states with jets and missing energy and the prospect studies for top squark searches. The thesis is organized as follows:

Chapter 2: A short introduction to the phenomenology of the standard model as well as to supersymmetry is given. Furthermore, current indirect and direct constraints from collider experiments on supersymmetric models are discussed.

Chapter 3: This chapter provides an overview of the CMS experiment at the LHC including a discussion of data taking at the LHC until to date.

Chapter 4: The simulation of events using Monte Carlo techniques is introduced.

Chapter 5: In this chapter, an introduction to the reconstruction of objects recorded in the particle collisions is given. Furthermore, dedicated algorithms to identify specific particle decays are discussed.

Chapter 6: A measurement of the jet transverse-momentum resolution using dijet event topologies is explained. This measurement is performed for pp collision data as well as simulated events at $\sqrt{s} = 8\,\mathrm{TeV}$.

Chapter 7: A search for supersymmetry in final states containing several hard jets and missing transverse momentum at $\sqrt{s} = 8\,\mathrm{TeV}$ is reviewed. Special emphasis is put on the determination of the QCD multijet background.

Chapter 8: Based on simulated events, prospect studies for a search for top squarks at $\sqrt{s} = 13\,\mathrm{TeV}$ are discussed.

Chapter 9: This chapter provides a short summary of the thesis and main results.

2 Phenomenological Aspects of the Standard Model and Beyond

The standard model of particle physics describes the fundamental particles and interactions between them. It is a theory that successfully predicted the existence of several particles and has been tested extensively, e. g. in electroweak precision measurements at LEP.
Although the SM is a successful theory, there are also 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 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 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 in which the particles are denoted together with their interactions.[1]
Mathematically, the standard model is a quantum field theory in which interactions between particles are described via gauge symmetries. The underlying gauge group of the standard model is

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

in which $SU(3)$ 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.

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

[1] Gravity is not included in the standard model and thus it is not discussed in this thesis.

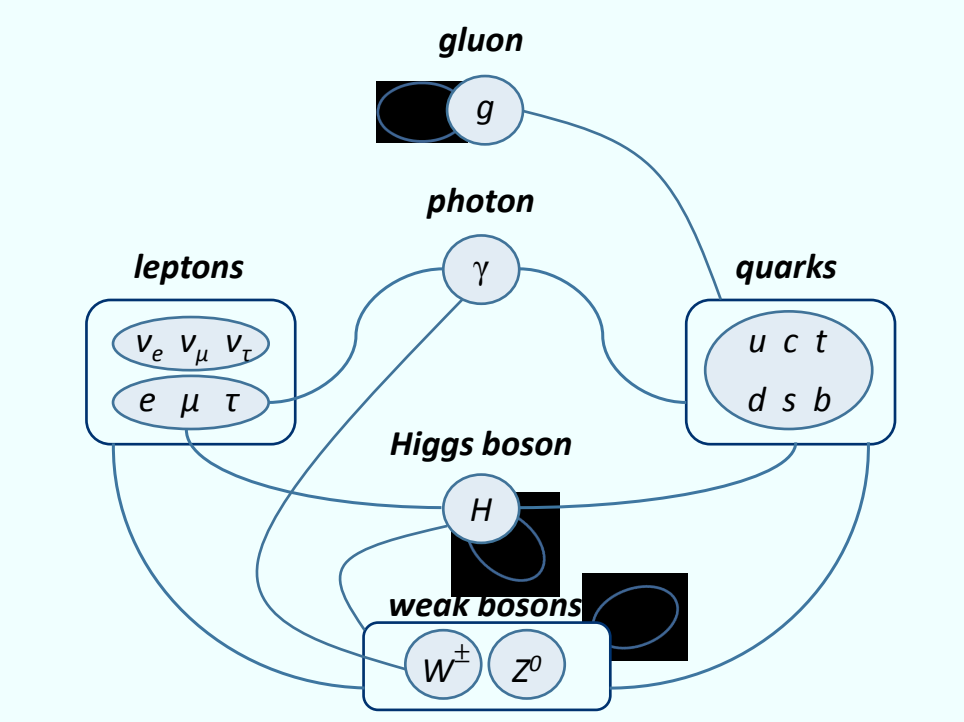

Figure 2.1: Overview of particles contained in the standard model. Blue lines indicate interactions between different particles.

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 arranged in 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 comprises the up- and down-quark (u, d), the second the charm- and strange-quark (c, s) and the third the top- and bottom-quark (t, b). All quarks carry electrical charge, but in contrast to 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 colour charge which comes in three types.

In addition to the attributes described above, fermions are furthermore characterized by the weak isospin. In each generation, left-handed fermions form an isospin-doublet with a weak isospin of $\pm 1/2$ while right-handed components are isospin-singlets 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 exchanged 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 left-handed fermions, i. e. on fermions with non-zero weak isospin, and manifests in charged and neutral currents. Weak interactions preferably take place within one fermion generation. However, since the mass eigenstates in the weak interaction differ from the flavour eigenstates, also transitions between different generations are possible. In the quark-sector, typically a representation is chosen in which 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 a unitary matrix, described by three mixing angles and one CP-violating phase, which indicates the relative strength between individual transitions. Similarly, also in the neutrino sector a mixing between the weak and the mass eigenstates occurs which leads to the phenomenom of *neutrino oscillation* [4–6].

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 couple to themselves. The colour charge exists in three different states commonly denoted as *red*, *green* and *blue*.
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 referred to as *confinement* [7] and describes the fact that coloured objects can not exist freely. Actually, 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, i. e. the coupling strength is small, when the distance decreases. This feature is known as *asymptotic freedom* [8, 9].
A typical example for a hadron is the proton. In a simplified picture, it is composed of three quarks: two up quarks and one down quark (*valence quarks*). However, the valence quarks continuously exchange gluons which can exchange further gluons or split into quark-antiquark pairs (*sea quarks*). The constituents of the proton are commonly denoted *partons* and the internal proton structure is described by *parton-distribution functions* (PDFs) specifying the momentum fraction of the proton carried by individual partons.

First proposed by Salam, Glashow and Weinberg [10, 11], 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 boson are formed by mixing of these massless states. While the charged $W^\pm$ 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 according to

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{2.1}$$

with the weak mixing angle θ_W. This mixing angle relates also the electromagnetic coupling strength e and the weak coupling strength g according to

$$e = g \sin\theta_W \; . \tag{2.2}$$

The fields W^μ couple only to left-handed fermions such that the same holds also for $W^\pm$. Since, however, the B^μ couples to left- and right-handed states, a coupling to left- and right-handed fermions takes place for γ and Z^0. Unlike the photon, the $W^\pm$ and Z vector bosons are massive with masses[2] 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.

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 do 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* [12–14] which is based on the concept of spontaneous symmetry breaking. The main idea behind this meachnism is that while in general the principle of local gauge invariance is obeyed, 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 Φ described by a potential

$$\mathcal{V}(\Phi) = \mu^2 \Phi^+ \Phi^- \; + \; \lambda (\Phi^+ \Phi^-)^2$$

with parameters μ and λ. Choosing μ^2 to be negative and λ positive, the potential has a non-zero minimum value with the vacuum expectation value

$$v = \sqrt{\frac{-\mu^2}{2\lambda}} \; .$$

Expansion of the Higgs field around this vacuum expectation value eventually leads

[2] In this thesis, natural units are used, i. e. $\hbar = c = 1$. Thus, also particle masses and momenta have the dimension of energies.

to a new spin-0 particle, the scalar *Higgs boson*, which is a quantum excitation of one of the components of the Higgs field. Furthermore, the masses of gauge bosons are generated by the couplings to the Higgs field according to

$$m_\gamma = 0, \quad m_W = \frac{v}{2}g, \quad m_Z = \frac{m_W}{\cos\theta_W}, \quad m_H = \sqrt{-2\mu^2}\ .$$

Similarly, the Higgs mechanism introduces mass terms for fermions

$$m_f = G\frac{v}{\sqrt{2}}$$

resulting from Yukawa couplings to the Higgs field with coupling constants G_i.
The discovery of a new boson at a mass of around 125 GeV has been announced by the ATLAS and CMS collaborations in 2012 [15, 16]. As all properties of this new boson are consistent with SM predictions for the Higgs boson so far (cf. for instance [17–20]), 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 very successful so far leading to several discoveries while withstanding numerous precision tests, it is known to be an incomplete theory. Some of the shortcomings of the SM are:

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 anitmatter exist to equal amounts in the universe which is in fact not the case. A theory, which would be able to explain the observed 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 [21].

Unification of couplings: The unification of the electromagnetic and the weak force leads to the question if it is 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 require 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.

Nature of dark matter: There exist several cosmological observations that indicate that the matter described by the SM makes up only 4.9% of the universe [22]. 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 [23].

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. For a fermion with mass m_f and Yukawa coupling λ_f to the Higgs field, the corrections to m_H^2 are

$$\Delta m_H^2 \propto -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{\mathrm{UV}}^2 \propto m_f^2 \tag{2.3}$$

with an ultraviolet cut-off scale Λ_{UV}. Typically, this cut-off scale is interpreted as the energy at which new physics enter. If it is chosen to be the Planck scale, the Higgs mass is several orders of magnitudes larger than the electroweak scale and thus would require an enormous amount of fine-tuning at each order of perturbation theory to yield the expected Higgs mass around $\mathcal{O}(100)\,\mathrm{GeV}$.

2.2 Supersymmetry

In order to overcome the weaknesses of the SM and to provide explanations for so far unsolved problems, 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. The first supersymmetric four-dimensional quantum field theory has been introduced by Wess and Zumino in 1974 [24].
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. [25, 26].

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

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

The supersymmetric fermionic and bosonic partner particles are called *superpartners* 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 colour degrees of freedom, except for the spin. Commonly, supersymmetric particles are denoted *sparticles*.

In a general supersymmetric theory fulfilling the criteria of gauge invariance and renormalisability, processes are allowed which violate either lepton or baryon number conservation. However, a baryon and lepton number violation would imply for instance a rapid decay of protons. The lower limit on the proton lifetime is found to be 5.9×10^{33} years at 90% confidence level [27] 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 occur.[3] 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 at 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 provides some elegant solutions to open questions as raised in Sec. 2.1.1:

- 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. For instance for the fermion contributions described in Eq. 2.3, the quadratically divergent terms are canceled by contributions with opposite sign that arise from a scalar with same mass and thus the same coupling strength to the Higgs field. Since the same cancelation occurs for bosons vice versa, SUSY is able to provide a solution to the hierarchy problem. However, no observation of such kind of supersymmetric particles with exact same masses as their SM correpondents has been made such 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 $\mathcal{O}(1\,\text{TeV})$, which is typically referred to as *natural supersymmetry*. This is the main argument why one would expect masses of supersymmetric particles to be in the TeV range, well within the reach of the LHC. Some more considerations about natural SUSY follow in Sec. 2.2.1.
- Considering the existence of low scale supersymmetric particles, the coupling constants of the forces meet in one point when extrapolating the couplings from the electroweak to the GUT scale. 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 enter. In general, this hints to the possibility of a grand unification.
- R-parity conserving SUSY models provide a suitable dark matter candidate. As discussed previously, each decay of supersymmetric particles finally leads to the existence of an LSP which is stable. Thus, it is an adequate DM candidate when it is only weakly interacting.
- Supersymmetry is in principal also suited to explain the observed matter-antimatter asymmetry in the universe as especially the description of supersymmetry breaking can involve CP-violating phases. However, these phases are strongly constraint by experimental results (cf. for instance [28] for a comprehensive review).

[3] There exist also several R-parity violating SUSY models which are not in contradiction to the observed proton lifetime (cf. for instance [26]). However, these models are not subject of this thesis and thus not discussed.

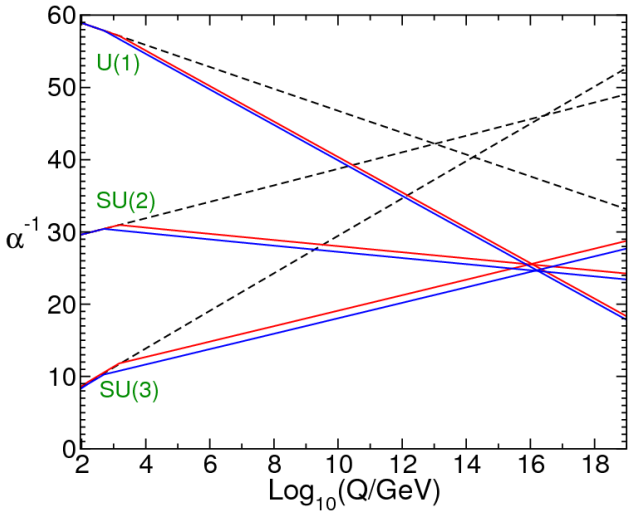

Figure 2.2: Comparison of the renormalization group evolution of the couplings α_a^{-1} in the SM (dashed lines) and the MSSM (solid lines) including two-loop effects. The masses of the supersymmetric particles in the MSSM are considered as a common threshold changing between 500 GeV and 1.5 TeV while the strong coupling constant $\alpha_3(m_Z)$ is varied between 0.117 and 0.121. Taken from [26].

- SUSY might also provide an indication of the nature of gravity. As denoted in Sec. 2.1.1, it is currently not possible to unify general relativity and quantum theory. However, supersymmetry is a crucial requirement for *string theories* which are the only suited candidates for a quantum field theory of gravity to date (see, e. g. [29]).

2.2.1 Natural Supersymmetry

As introduced above, SUSY models are considered natural if they provide a solution to the hierarchy problem. Commonly, certain measures are introduced to estimate the naturalness of a supersymmetric model.
One example is the Veltman definition of naturalness [30]. This states that radiative contributions should not exceed tree-level effects in size, regarding the mass of a scalar particle. Another definition is given by Babieri and Giudice [31] which do not restrict the magnitude of the radiative corrections, but the sensitivity of the physical mass of a scalar m to small changes in the bare couplings λ_0. This constraint is often imposed by quantifying the amount of necessary fine-tuning Δ according to

$$\left| \frac{\lambda_0}{m^2} \frac{\partial m^2}{\partial \lambda_0} \right| < \Delta \, . \tag{2.4}$$

While the value of Δ has been set to ~ 10 in the past, it was increased over the years so that now also values of ~ 100 or even ~ 1000 are often considered as acceptable values [32]. Such conditions can be used in order to derive constraints on the spectrum of superpartners which regulate the hierarchy problem. Here, it turns out that not all sparticles are equally important and thus not all have to be situated at the same mass scale. Some typical mass ranges for sparticles considered as natural are given in Sec. 2.2.2.

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	$\tilde{W}^\pm\ \tilde{H}_u^+\ \tilde{H}_d^-$	$\tilde{\chi}_1^\pm\ \tilde{\chi}_2^\pm$
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. Adapted from [26].

2.2.2 The MSSM

In general, it is possible to have theories with more than one supersymmetry transformation N. However, the smallest possible supersymmetric extension of the SM including the full particle spectrum and interactions, i. e. a $N = 1$ supersymmetry, is realised in the MSSM. An overview of the respective particle content is given in Tab. 2.1. This extension is minimal in the sense that it introduces the least feasible number of additional particles to the existing SM particles meaning that each SM fermion gets one superpartner. 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 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*. Furthermore, 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 needed with one doublet H_u giving 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. Within the MSSM, the mass of the lightest Higgs boson is restricted to be smaller than the Z-boson mass at tree level. However, due to radiative corrections, which mainly arise from the top sector, this limit is enhanced and results in an upper bound of [26]

$$m_{h^0} \lesssim 135\,\text{GeV} .$$

Consequently, significant contributions from the top squark mass are required to push the mass of the lightest Higgs boson up to a value of around 125 GeV. This is somewhat in tension to the requirement of having a top squark mass close to the top mass in order to solve the hierarchy problem. However, it is still possible to accommodate a Higgs mass of 125 GeV without the necessity to decouple the top squark or add new dynamics to the MSSM. These scenarios are referred to as *maximal mixing* [33].
Similar to the SM, the gauge eigenstates of the SUSY theory are not necessarily equal to 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 called *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.
In order to obtain a natural realization of the MSSM, it is important that certain particles have a particular mass scale [32]. 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 to the Higgs mass. These give typically the largest contributions since the top quark is the heaviest particle in the SM. Assuming a maximal accepted fine-tuning of $\Delta \lesssim 10$, the top squark mass is supposed to be around 400 GeV. Given these conditions, also light higgsinos are expected with a typical mass scale around 200 GeV. Since gluinos yield loop corrections to the top squark mass, they are expected to not significantly exceed the 1 TeV range as a second order effect.

2.2.3 SUSY-Breaking

As discussed above, SUSY particles are in general expected to have the same mass as their corresponding SM partner particle. Since, however, none of such particles has been observed so far, this implies that supersymmetry in fact must be broken and that sparticles are actually heavier than the SM counterparts.
Typically, SUSY breaking is introduced such that so-called *soft breaking* terms, i. e. of positive mass dimension, are added to the theory in addition to the terms determining the gauge and Yukawa couplings. The SUSY breaking is assumed to take place in a hidden sector which does not couple directly to the visible sector represented by the supermultiplets. In this case, no specific breaking mechanism must be assumed, but only a mechanism

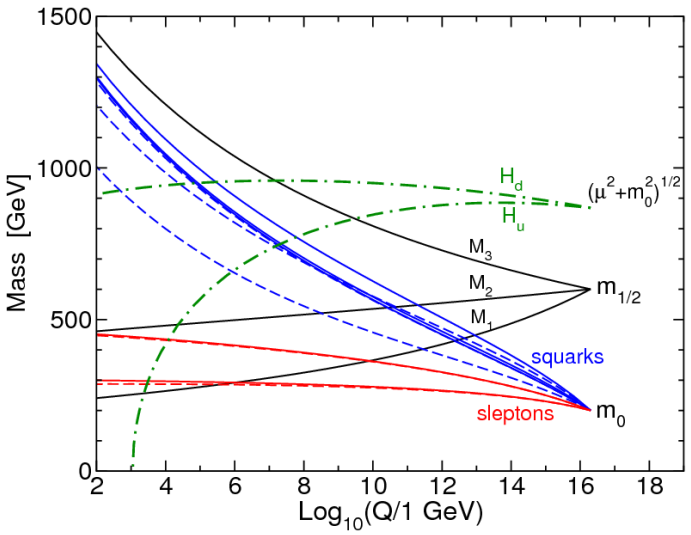

Figure 2.3: Evolution of scalar and gaugino mass parameters in the MSSM with mSUGRA boundary conditions imposed at $Q_0 = 2 \times 10^{16}$ GeV. The parameter $\mu^2 + m_{H_u}^2$ runs negative, provoking electroweak symmetry breaking [26].

to describe the mediation of the supersymmetry breaking from the hidden to the visible sector. In total, the MSSM, including soft SUSY breaking, features several new phases, mixing angles and masses which add another 105 free parameters to the already existing parameters of the SM [34].
Two SUSY breaking scenarios, that are often studied, are either based on gravity-mediated or gauge-mediated interactions and known as *minimal supergravity* (mSUGRA) [35, 36] or *constrained MSSM* (CMSSM) [37, 38] and *gauge-mediated supersymmetry breaking* (GMSB) [39,40]. 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, $\tan\beta = v_u/v_d$ and the sign of the higgsino mass parameter μ. In Fig. 2.3, the evolution of the corresponding mass parameters to the electroweak scale is 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.

2.3.1 Indirect Constraints

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 contribu-

tions 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 [41–44]. Furthermore, precise measurements of rare processes in B-meson decays, like $B_s^0 \to \mu^+\mu^-$, are in good agreement with expectations from the SM [45–47]. The most compelling difference between experimental results and SM prediction is currently observed for the anomalous magnetic moment of the muon [1, 48, 49]. Here, deviations at the level of 3.6σ occur. However, discussions about the accuracy of the SM calculation are ongoing [50].
Further constraints arise from astrophysical and cosmological studies. Several observations suggest that a considerable amount of cold dark matter contributes to the composition of the universe [22, 51]. Good candidates are weakly interacting massive particles (WIMPs) which could be the neutralino in SUSY models where it is the LSP. Consequently, also the observed cold dark matter relic density can put constraints on the MSSM parameter space assuming that it is caused by a neutralino LSP. Thus, constraints on supersymmetry can be derived in particular from direct and indirect dark matter searches (cf. for instance [52–59]).

2.3.2 Searches at pre-LHC 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 assuming R-parity conservation. As discussed in Sec. 2.2, R-parity conservation implies 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 aim at a specific kind of supersymmetric particle. If searches are designed 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

The main focus of SUSY searches at HERA has been on R-parity violating models [60–62]. However, also searches for R-parity conserving supersymmetric models have been performed [63,64] and in these models the excluded region extends to around 60–70 GeV for squark masses and around 40 GeV for the LSP mass, depending on the specific assumptions made to contrain the MSSM parameter space.

LEP

At LEP, various searches for supersymmetry were performed targeting different species of supersymmetric particles in R-parity conserving models [65]. Scalar leptons and quarks are mainly pair-produced in the s-channel via Z bosons and photons. In case of selectrons, also the t-channel exchange of neutralinos yields important contributions. The energy scale at LEP opens a parameter space where sparticles with quite high masses are produced such that they predominantly decay into the respective SM partner particles (except for the scalar top, since the top quark is too heavy) and the lightest neutralino. Furthermore, also cascade decays are possible. Typical final states contain missing energy and a pair of acoplanar leptons (jets) where the direction of the first lepton (jet) is not in the plane defined by the direction of the second lepton (jet) and the beam direction.
Similarly, neutralinos and charginos are expected to be pair-produced via Z/γ s-channel exchange or t-channel selectron or sneutrino exchange, respectively. Typically, charginos decay into $\tilde{\chi}_1^0 l\nu$ or $\tilde{\chi}_1^0 qq'$ while in neutralino pairs ($\tilde{\chi}_1^0\tilde{\chi}_2^0$) the $\tilde{\chi}_2^0$ decays into $\tilde{\chi}_1^0\nu\bar{\nu}$, $\tilde{\chi}_1^0 l^+ l^-$ or $\tilde{\chi}_1^0 qq'$. 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 signatures 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 about 100–200 GeV over a 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 [65].

Tevatron

The Tevatron accelerator made another SUSY parameter space accessible for searches as the centre-of-mass energy exceeded that of HERA and LEP by at least one order of magnitude. 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 coloured 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 in Sec. 2.3.3.
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 GeV and $m_{1/2} = 125$–165 GeV. This allows to exclude gluinos below around 280–300 GeV for all squark masses and squarks below 380 GeV independent of the gluino mass [66–68].

2.3.3 Searches at the LHC

The absence of SUSY-like signals at any collider experiment performed previously to the start of the LHC and exclusion limits at the order of a few hundred GeV on sparticle masses activated a variety of searches for supersymmetry at the LHC. These are targeting the various different production and decay modes of supersymmetric partices and can

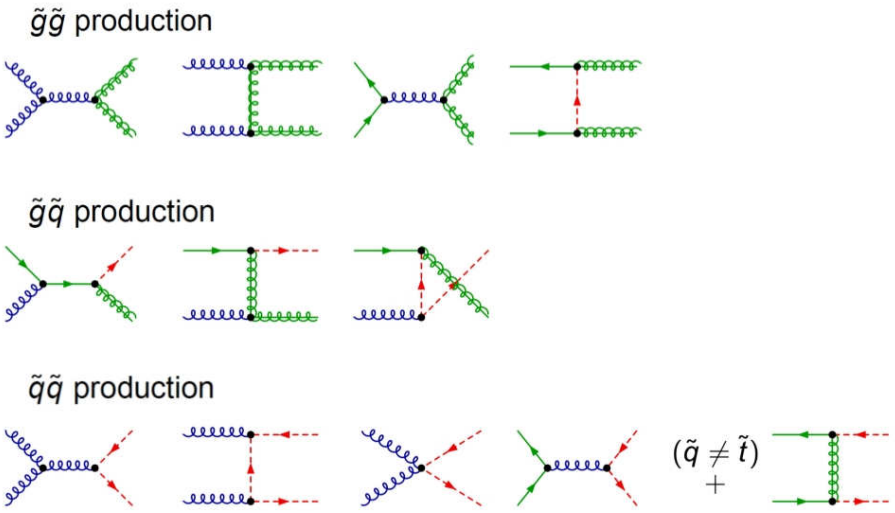

Figure 2.4: Example diagrams for the production of supersymmetric particles in hadron collisions at parton level.

be classified into searches for squarks and gluinos, third generation sfermions and electroweakinos. Of particular interest are searches for coloured particles, as the LHC is, as well as the Tevatron, a hadron collider. In this thesis, searches for supersymmetry are presented that are based on jet final states accompanied by missing transverse energy and no isolated leptons. Thus, these so-called *hadronic signatures* are discussed in the following.

Production Modes

At leading order, sparticles in R-parity conserving models are predominantly produced in processes like [69–73]

$$pp \to \tilde{g}\tilde{g}, \tilde{g}\tilde{q}, \tilde{q}\tilde{q}\,. \tag{2.5}$$

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 partners of the light-flavour (u, d, s, c) quarks with suppressed chiralities of the squarks $\tilde{q} = (\tilde{q}_L, \tilde{q}_R)$. 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 for instance by quark radiation or gluon loops, typically up to next-to-leading order (NLO). Production cross sections for the processes $pp \to \tilde{g}\tilde{g}$, $pp \to \tilde{q}\tilde{q}$ and $pp \to \tilde{t}\tilde{t}$ are illustrated in Fig. 2.5 for various sparticle masses as a function of the centre-of-mass energy. For instance, for a gluino with a mass of 1.5 TeV the gluino pair-production cross section is expected to be at the order of 10^{-4} pb at $\sqrt{s} = 7$ TeV. 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, squark-pair production (and also associated squark-gluino production) are favoured in case of large SUSY masses and low collider energies.

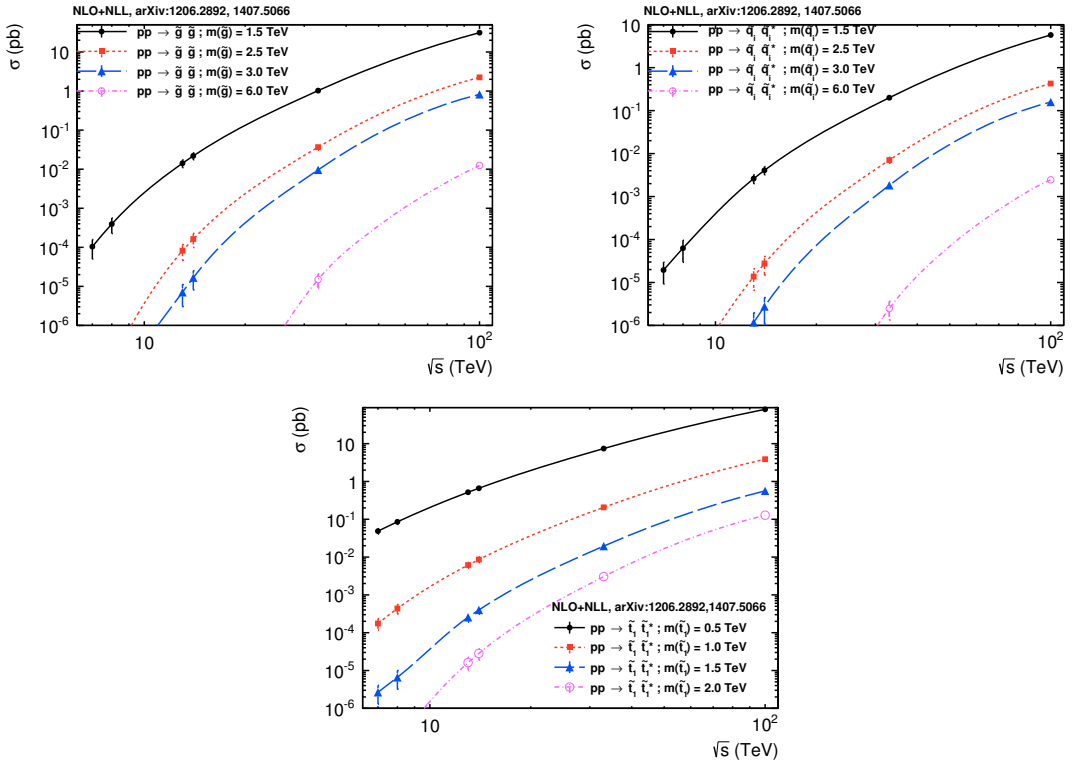

Figure 2.5: SUSY production cross sections of processes $pp \to \tilde{g}\tilde{g}$ (*top left*), $pp \to \tilde{q}\tilde{q}$ (*top right*) and $pp \to \tilde{t}\tilde{t}$ (*bottom*) displayed for different sparticle masses shown as a function of the centre-of-mass energy [74, 75].

Decay Channels

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. In Fig. 2.6, some example diagrams for possible decay modes of squarks and gluinos are illustrated. Here, the three-body decays of the gluinos in the upper two diagrams denote effective couplings. These occur in case squark masses are decoupled from the rest of the particle spectrum, i. e. that their masses are significantly larger than that of the gluinos. Gluinos decay preferrably into the LSP and a quark pair. This quark pair can for instance either be a pair of light-flavour quarks or a pair of top quarks. The latter represents the case of gluino-mediated top squark production. In case of squarks, a preferred decay to the LSP and one quark is expected. While a light-flavour quark is expected from the decay of a light-flavour squark, the final state of direct production of top squarks is made up of the LSP and a top quark.
In all such cases, a multijet final state accompanied by missing transverse momentum with no isolated leptons is expected as experimental signature.[4]

[4] In general, final state topologies including leptons can occur as well, for instance in cascade decays or semi-leptonic decays of final state top quarks. However, since the analyses presented in this thesis concentrate on all-jet final states, final states including leptons are not discussed.

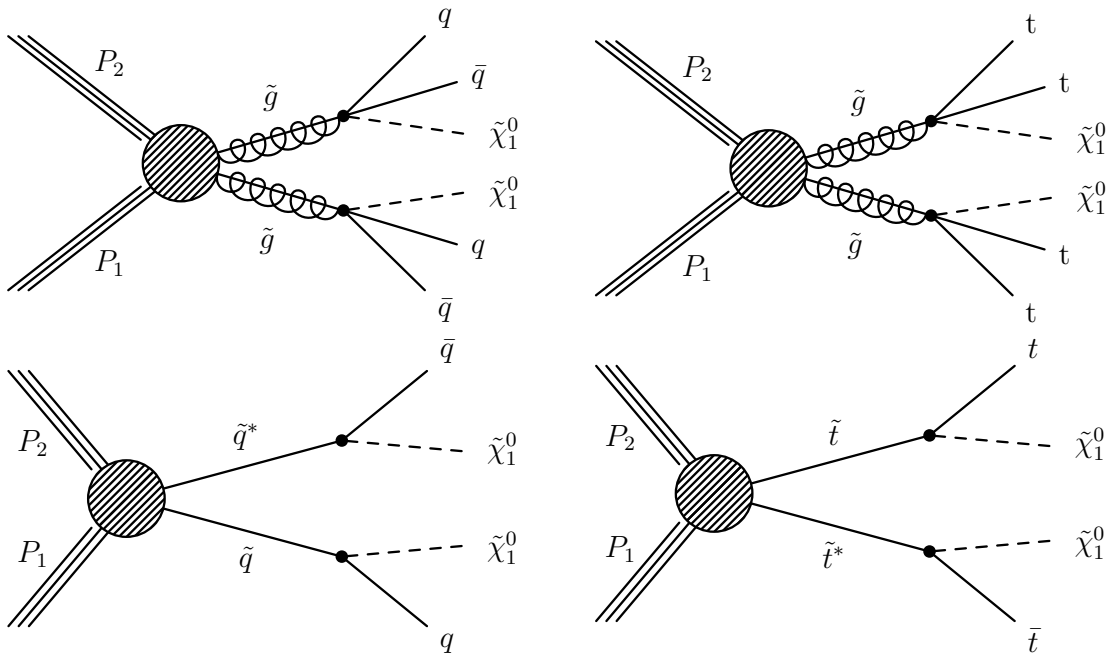

Figure 2.6: Example diagrams of different SUSY decay channels showing $\tilde{g} \to q\bar{q}\tilde{\chi}_1^0$ (*top left*), $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$ (*top right*), $\tilde{q} \to q\tilde{\chi}_1^0$ (*bottom left*) and $\tilde{t} \to t\tilde{\chi}_1^0$ (*bottom right*) [76].

Background Processes in Hadronic SUSY Searches

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 have to be considered as background which is a crucial task in each SUSY analysis. In case of multijet $+\not{E}_{\mathrm{T}}$ searches these are typically $Z(\nu\bar{\nu})+$jets events in which large genuine $\not{E}_{\mathrm{T}}$ is caused by the neutrinos. This background is denoted *invisible Z background* in this thesis. Furthermore, events with intrinsic missing energy stem from $W+$jets and $t\bar{t}$ events. The top quark decays almost exclusively into a W boson and a b quark. Experimentally, the decay of the top-quark is characterized by the decay of the W boson which 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 gives rise to nine different W decay modes and actually two thirds of top quark decays result exclusively in hadrons. However, since the targeted final state is assumed to contain a significant amount of missing energy only the semi-leptonic top quark decays relevantly contribute as background events. Thus, $W+$ 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 considered as possible background. This source of background events is referred to as *lost-lepton background*. Furthermore, also $W+$ 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 contain no intrinsic missing energy, severly mismeasured jets can give rise

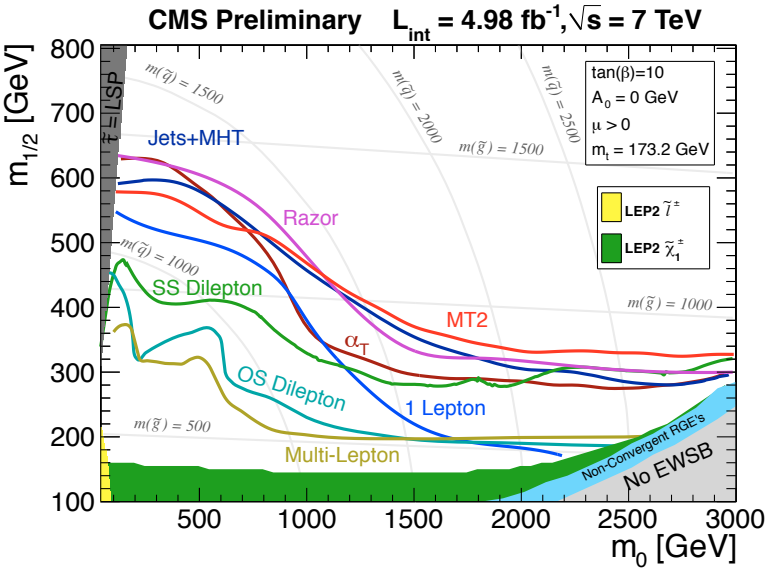

Figure 2.7: Interpretation of searches for supersymmetry at the CMS experiment within the CMSSM. Shown are the 95% C.L. exclusion limits in the m_0 and $m_{1/2}$ plane for various searches performed using different final state topologies [76].

to large amounts of $\not{E}_{\mathrm{T}}$ 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 [77, 78].

Results at $\sqrt{s} = 7$ TeV

As soon as first collision data have been obtained at the LHC at a centre-of-mass energy of $\sqrt{s} = 7$ TeV, searches for supersymmetry were carried out based on various final states. Like for previous collider experiments, however, no hints of new physics have been found and the results were interpreted in various SUSY models by setting exclusion limits [76,79]. In Fig. 2.7, interpretations of CMS searches for supersymmetry, based on various different hadronic and leptonic final states, are summarized in the context of the CMSSM. For comparison, also the exclusion curves from the LEP experiments [65] are illustrated in the $m_0/m_{1/2}$-plane which 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 hadronic states as described above. However, interpreting search results only in the context of the CMSSM carries some risks. The simplified assumption of universal gaugino masses at the GUT scale does not allow all mass patterns and signatures that are in general possible within the MSSM. Consequently, the CMSSM imposes for some SUSY topologies too strong constraints.
Thus, results of SUSY searches are, in addition to interpretations in the CMSSM, also interpreted in the context of *simplified models* [80–83]. 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 dominant features of SUSY events that are common for several SUSY and SUSY-like models. The characterization

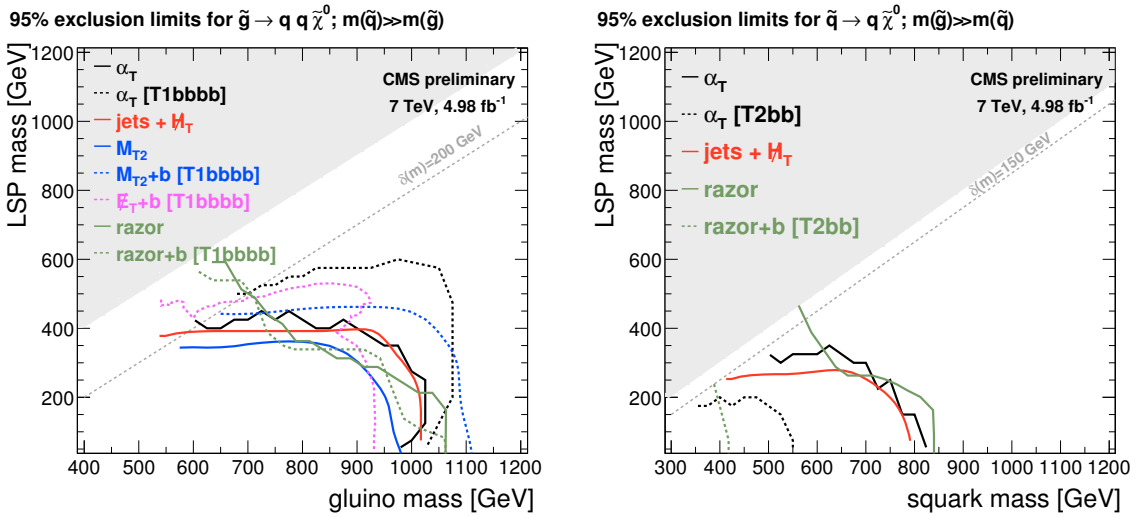

Figure 2.8: Interpretation of searches for supersymmetry at the CMS experiment with simplified models in $\tilde{g} \to qq\tilde{\chi}^0$ (*left*) and $\tilde{q} \to q\tilde{\chi}^0$ (*right*) topologies. Shown are the 95% C.L. upper limits on the produced particle and LSP masses. The grey area represents the region where the respective decay mode is forbidden [83].

of basic properties allows a comparison of search results to any (more complex) model and provides a suitable framework for reinterpretations of results from SUSY searches. A simplified model is described by a set of particles, their masses and a certain sequence of the particle 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%. Interpretations of SUSY searches at the CMS experiment within the context of these simplified models are shown in Fig. 2.8 for $\tilde{g} \to qq\tilde{\chi}^0$ and $\tilde{q} \to q\tilde{\chi}^0$. These illustrate the 95% confidence level upper limit on the product of the cross section and branching fraction as a function of the sparticles masses. Hence, the values of cross section 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.

However, interpretations in simplified models typically target only well-defined isolated SUSY topologies and thus, like the CMSSM, do also not account for all possible decay patterns in the MSSM. Consequently, also interpretations in more general models are desirable. One such example of a more generic SUSY model is the pMSSM [84]. The pMSSM is a 19-parameter realization of the MSSM and captures most of the features of general R-parity conserving SUSY models and covers a wide diversity of possible SUSY topologies. The MSSM is constrained by assuming that there is no new source of CP-violation, that no flavour changing neutral currents occur and that the first two sfermion generations are degenerate. Interpretations of CMS SUSY searches performed at $\sqrt{s} = 7$ TeV within the pMSSM are published in [85]. A discussion of updated results follows in Sec. 7.5.

Although the SUSY parameter space has been investigated already extensively with the LHC data obtained at $\sqrt{s} = 7\,\text{TeV}$ and exclusion limits on sparticle masses enter the TeV range, searches for supersymmetry stay a very important field within the CMS experiment also for $\sqrt{s} = 8\,\text{TeV}$. 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 production cross sections profit a lot from the collider energy increase such that a new parameter space is accessible. 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 with data obtained from 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.

3.1 The Large Hadron Collider

The LHC [86, 87] is a ring accelerator designed to provide particle collisions of hadrons. It is built in the tunnel of the former LEP [88] 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 pre-accelerated in various smaller accelerators, which are: 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 collisions to the experiments with centre-of-mass energies up to 14 TeV in order to explore physics processes at novel energy regimes. For a certain process, the expected number of events N is given by the product of the cross section σ and the integral $L = \int \mathcal{L}\, dt$ of the instantaneous luminosity $\mathcal{L}$ over time, such that

$$N = \sigma \cdot L \,. \tag{3.1}$$

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

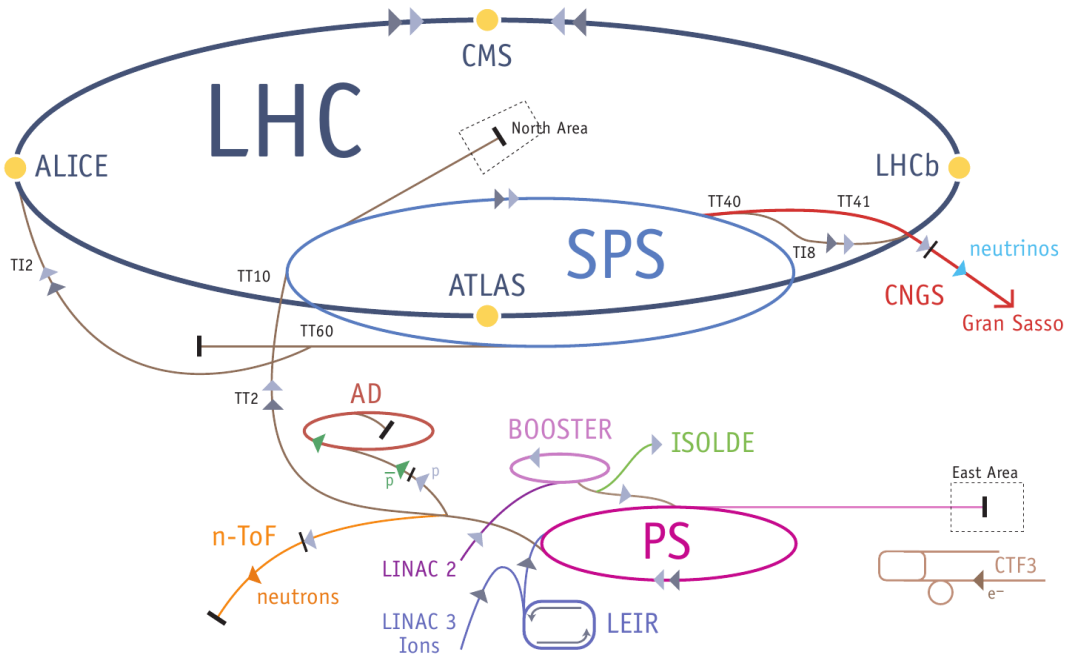

Figure 3.1: Illustration of the CERN accelerator complex including the injector chain of the LHC ring [89].

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

$$\mathcal{L} = \frac{f n_1 n_2}{4\pi\sigma_x\sigma_y} \cdot F \tag{3.2}$$

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 $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. 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 [91] and CMS [92, 93] 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 [94], 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 [95] detector with the main emphasis on the physics of strongly interacting matter at extreme energy densities, like quark-gluon plasma.

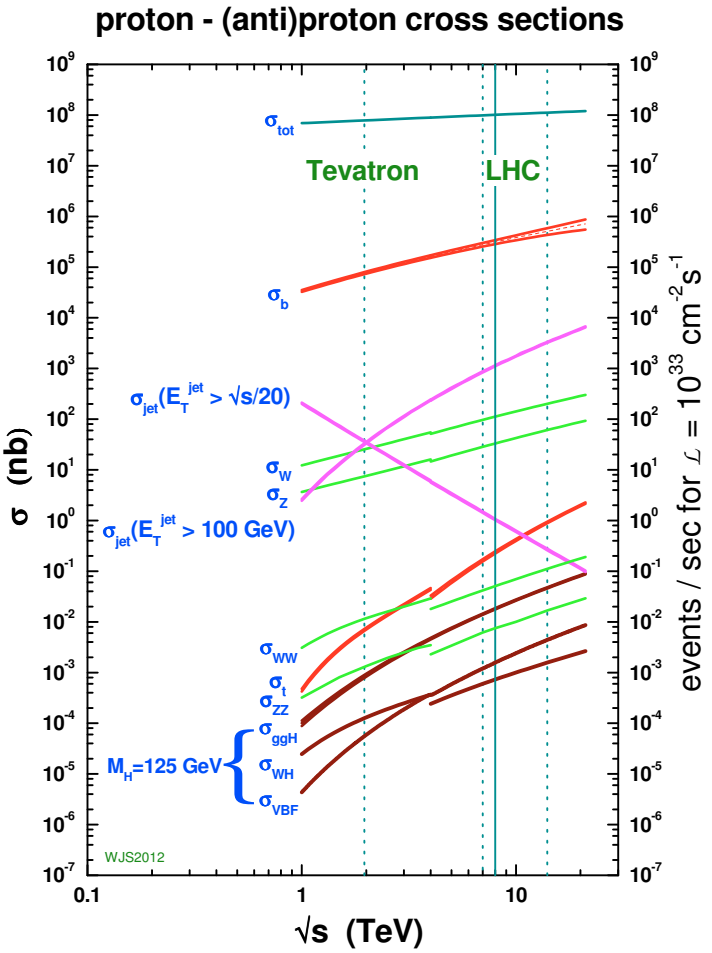

Figure 3.2: Summary of cross sections for various standard model processes in proton-antiproton ($\sqrt{s} < 4\,\text{TeV}$) and proton-proton ($\sqrt{s} > 4\,\text{TeV}$) collisions as a function of the centre-of-mass energy. The right axis displays the correponding event rate at a luminosity of $10^{33}\,\text{cm}^{-2}\text{s}^{-1}$ [90].

3.2 The CMS Experiment

The CMS detector is one of the two general-purpose experiments at the LHC. In addition to tests of the SM at the TeV scale, studies of the nature of elektroweak symmetry breaking and searches for so far unknown effects are the primary targets of this experiment. 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. Like other high-energy particle experiments, the CMS detector makes use of tracking detectors and calorimeters

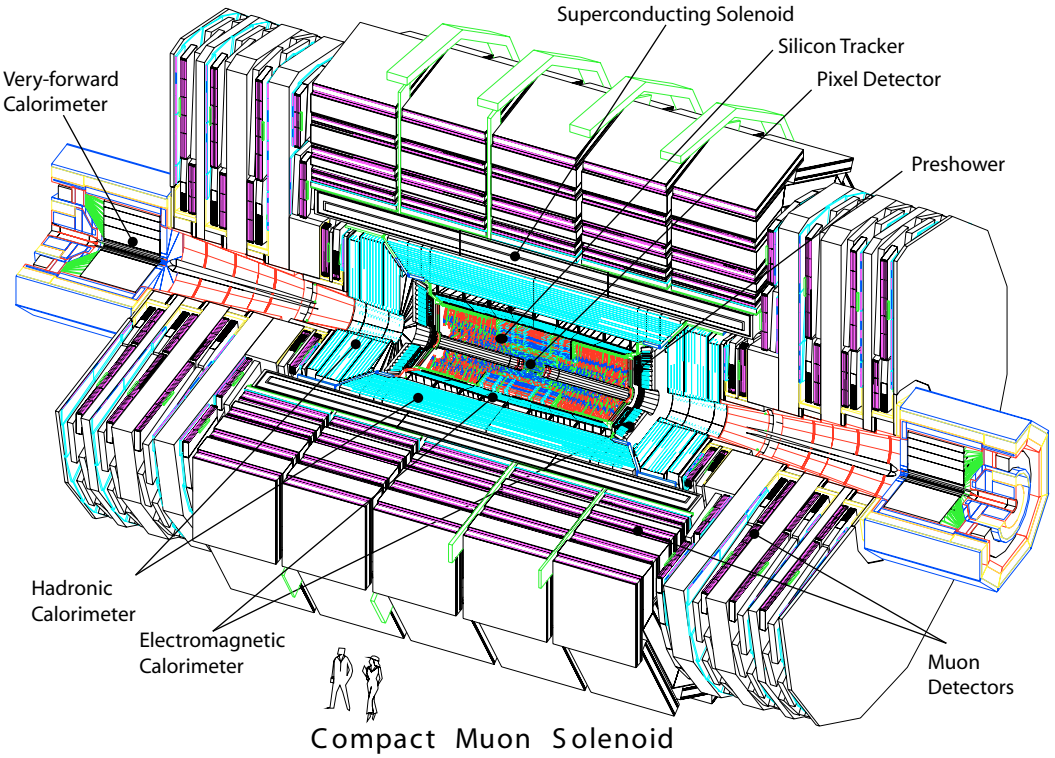

Figure 3.3: A perspective view of the CMS detector [92].

to measure momenta of particles, energy depositions and flight directions in order to identify the objects emerging from the particle collisions.
The following sections comprise a description of the CMS detector which exhibits a total weight of 12 500 t and has a length of 21.6 m and a diameter of 14.6 m. A detailed discussion of the detector design can be found in [92, 93].

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

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

which is widely used in experimental particle physics. A pseudorapidity $\eta = 0$ corresponds to the direction perpendicular to the beam while $|\eta| \to \infty$ 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

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

At the LHC, the hard interaction, i. e. the actual momentum transfer, is not taking place between the protons as a whole, but rather between the partons. Since the partons carry an unknown fraction of the proton momenta, the initial conditions of the primary collisions are also unknown. 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*. The transverse momentum of a particle is defined as

$$p_\mathrm{T} = \sqrt{p_x^2 + p_y^2} \tag{3.5}$$

with the components p_x and p_y of the momentum vector in the x and y direction. Momentum conservation in the transverse plane is then used to constrain the final state.
The momentum imbalance in the transverse plane ($\vec{\not{E}}_\mathrm{T}$) is determined as the negative vector sum of the momenta of all N particles in the event

$$\vec{\not{E}}_\mathrm{T} = -\sum_{i=1}^{N} \vec{p}_{\mathrm{T,i}}\,. \tag{3.6}$$

The absolute value of the vector momentum imbalance $\not{E}_\mathrm{T}$ is typically termed *missing transverse momentum* or *missing transverse energy* and serves as estimate for the sum of the transverse momenta of all 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 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 the sign of the 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 used for the return of the magnetic flux and closes the magnetic field lines. By instrumenting it, this offers the opportunity to 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 primary as well as secondary vertices. Due to the location close to the interaction point, the tracking system has to

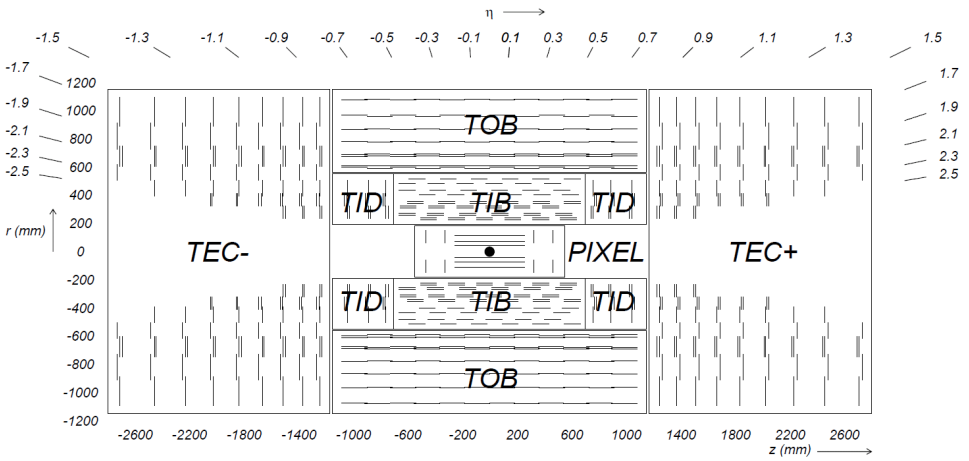

Figure 3.4: Sketch of the CMS tracking system in a rz-view. Each tracker module is represented by one line [92].

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 x 150 $\mu\mathrm{m}^2$ providing similar track resolution quality in r–ϕ 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–ϕ 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 high tracking efficiency [96]. The impact parameter resolution is of the order of $\lesssim 35\,\mu\mathrm{m}$ in the plane perpendicular to the beam (for particles with $p_\mathrm{T} > 10\,\mathrm{GeV}$) and reaches 75 µm 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 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 Calorimeter System

The calorimetry of the CMS experiment is composed of the electromagnetic and the hadron calorimeter and measures the deposited energy as well as the position of incoming particles. A typical characteristic of calorimeters is given by the ratio e/h. This quantifies the relation between the detection efficiencies of electromagnetic (e) and hadronic (h) energy deposits in a particle shower. Often it is determined from the ratio of the calorimeter response to pions and electrons π/e of the same energy

$$\frac{\pi}{e} = \frac{f_{\text{em}}\, e + (1 - f_{\text{em}})h}{e} = \frac{1 + (e/h - 1)f_{\text{em}}}{e/h} \tag{3.7}$$

with the *electromagnetic fraction* f_{em}, i. e. the fraction of the hadronic shower transferred into an electromagnetic component via the decay of neutral pions into two photons. Since f_{em} depends on the energy, the same holds for π/e and the calorimeter response is said to be *non-linear*.
Furthermore, the calorimeter performance can be characterized by the relative calorimeter energy-resolution

$$\frac{\sigma(E)}{E} = \frac{N}{E} \oplus \frac{S}{\sqrt{E}} \oplus C \tag{3.8}$$

which improves with increasing energy E of the measured particle. At low momenta, the resolution is mainly dominated by electronic noise, described by the *noise term* N. For increasing energies, the resolution is driven by fluctuations of the shower development described by the *stochastic term* S and at high energies, the resolution is eventually limited by calorimeter miscalibration and non-uniformities described by the *constant term* C.

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[2] X_0 and a small Molière radius[3] R_m 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.
An overview of the ECAL layout is shown in Fig. 3.5. The individual sub-components are as follows:

[2] The radiation length corresponds to the mean distance after which an electron traversing a material has lost all but $1/e$ of its energy.
[3] The Molière radius corresponds to the radius of a cylinder containing on average 90% of the energy deposition of an electromagnetic shower.

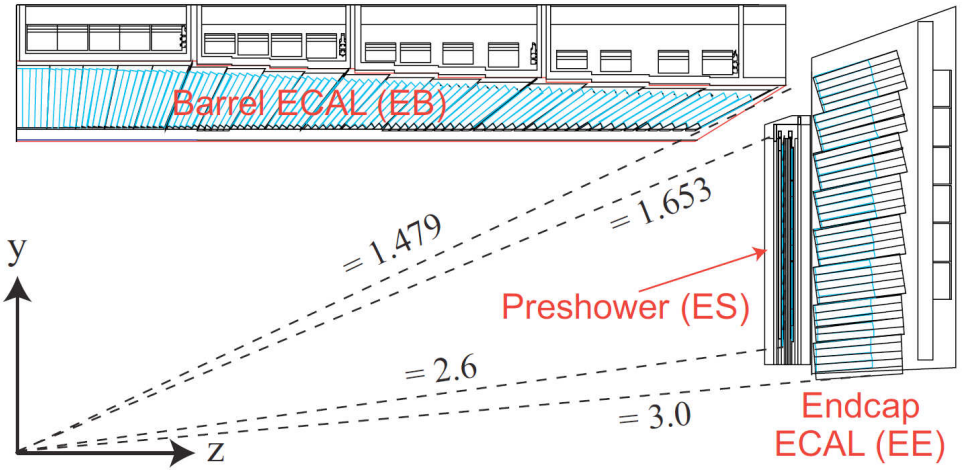

Figure 3.5: View of one quarter section of the CMS electromagnetic calorimeter in a yz-view [93].

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 halfs. 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 $1.653 < |\eta| < 2.6$ and 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). With its high granularity, it offers the possibility to identify neutral pions decaying into two collimated photons. These constitute an important background contribution in the search for the Higgs boson in the $H \to \gamma\gamma$ decay channel.

The performance of the ECAL has already been evaluated based on test-beam results [97, 98]. The ratio e/h has been found to be 1.6 while the relative resolution of electrons with energy E is determined to be

$$\frac{\sigma(E)}{E} = \frac{0.124}{E/\text{GeV}} \oplus \frac{0.036}{\sqrt{E/\text{GeV}}} \oplus 0.0026 \,. \tag{3.9}$$

Thus, the typical relative energy resolution for electrons with a transverse momentum of 120 GeV with this calorimeter configuration is in the order of 0.5%.
In addition to the calibration of the absolute energy scale, especially channel-to-channel

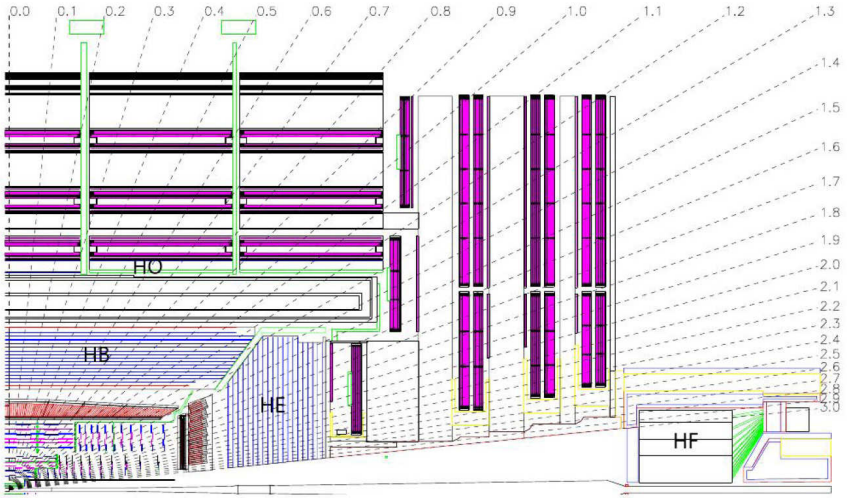

Figure 3.6: Longitudinal view of one quarter of the CMS detector showing the location of the individual HCAL sub-detector parts [92].

effects referred to as *intercalibration* have to be accounted for. This intercalibration is performed based on $\pi^0 \to \gamma\gamma, W \to e\nu$ and $Z \to ee$ events and results in a crystal-intercalibration accuracy of 0.6% [99]. 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.

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 build 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. Thus, an additional calorimeter component is installed outside the solenoid in the barrel part to reduce effects from shower leakage, i. e. compensate for hadronic showers that are not fully 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 wegdes. 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 $10.6\,\lambda_I$ 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).
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.

The HCAL performance has been measured based on test-beam data as well [97, 98]. From the measurement of the HCAL response to pions and electrons, the ratio $e/h = 1.4$ is extracted and the relative energy resolution of the combined ECAL and HCAL system is determined as

$$\frac{\sigma(E)}{E} = \frac{1.2}{\sqrt{E/\mathrm{GeV}}} \oplus 0.069 \tag{3.10}$$

which corresponds to a relative energy resolution of roughly 14% for pions with an incident energy of $E = 100\,\mathrm{GeV}$.
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 continous update of this calibrations is performed using isolated energetic particles, e. g. from decays of W or Z bosons.

3.2.5 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

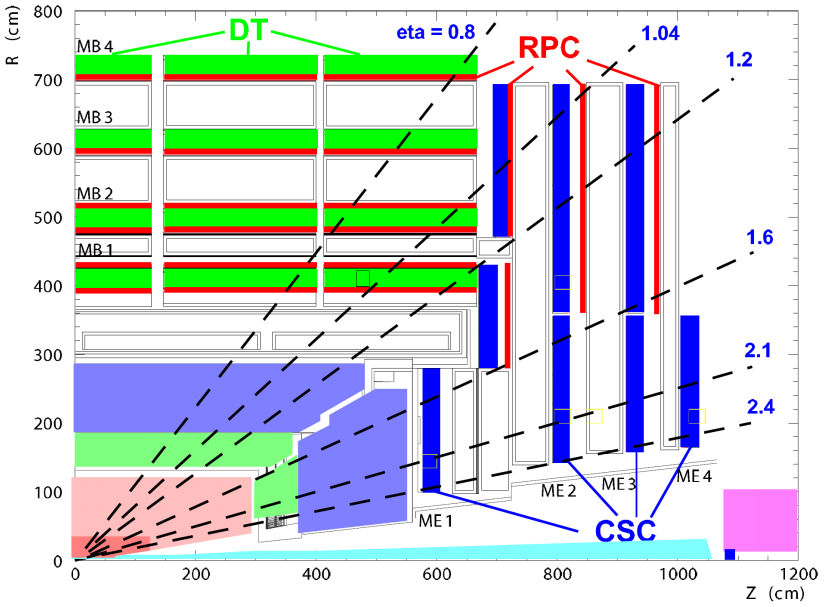

Figure 3.7: View of one quarter section of the muon system in the rz-plane [93].

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 m^2 detection planes are employed. The layout of the muon system is illustrated in Fig. 3.7.
Three different types of gaseous detectors are used to achieve a good muon momentum resolution given the different radiation conditions and variations in the homogenity of the magnetic field 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 likewise residual effects from the magnetic field. 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 chambers 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 and radiation losses in the material, the resolution for muons with transverse momenta below $\approx$ 200 GeV is in general better based on the inner tracking system than for the muon system. However, at higher transverse momenta the track-curvature measurement in the inner tracking system is limited and thus a combination with the measurement in the muon system is beneficial because of the longer lever arm. 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%.

3.2.6 Trigger System

The LHC operating at design conditions provides 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 events. However, as illustrated in Fig. 3.2, the event rate of interesting events is orders of magnitudes smaller than the total inelastic proton-proton cross section. Thus, already in the trigger system a fast event preselection is performed that allows 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 $\not{E}_\mathrm{T}$. 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 a positive L1 trigger decision to the front-end electronics, is 3.2 µs. 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.

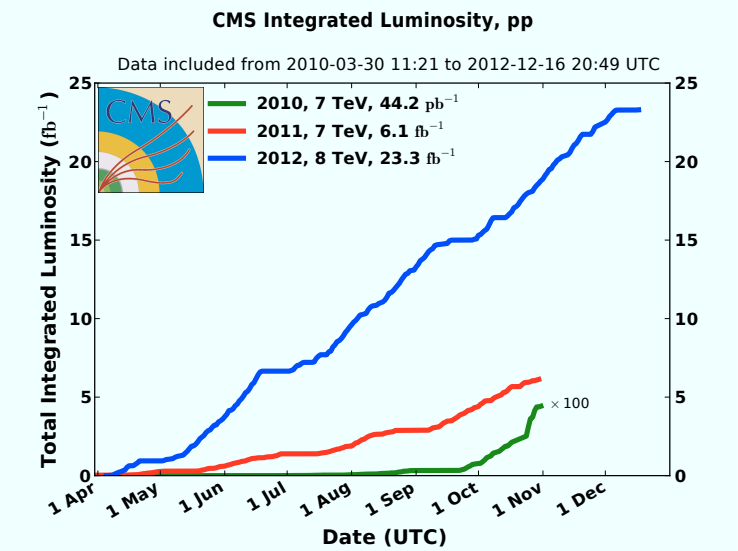

Figure 3.8: Cumulative integrated luminosity versus day delivered to CMS during stable beams for *pp* collisions. Different data-taking periods are indicated as follows: 2010 (green), 2011 (red) and 2012 (blue) [100].

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 continously 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 of a few hundred Hz 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 centre-of-mass energy of 7 TeV finally happened end of March 2010. In the following running period in 2010, data corresponding

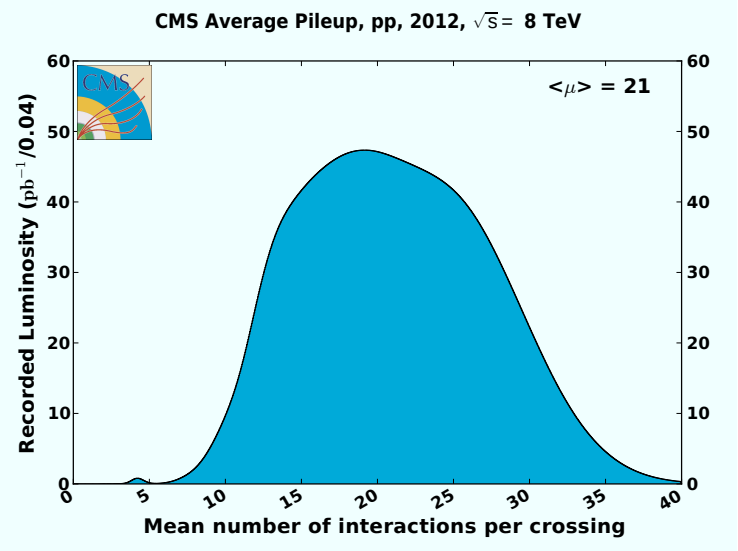

Figure 3.9: Mean number of interactions per bunch crossing in *pp* collisions at a centre-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$ [100].

to an integrated luminosity of $44.2\,\text{pb}^{-1}$ 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 delivered a total amount of $6.1\,\text{fb}^{-1}$ *pp* collision data 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, $23.3\,\text{fb}^{-1}$ of 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. Typically, these periods are referred to as *LHC Run I*. During most of the operation in 2012, the LHC was circulating 1380 bunches per beam with a spacing 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 [100, 101]. In general, such conditions result in multiple interactions per bunch crossing known as *pileup* (PU). Typically, 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 occuring 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. Dedicated techniques in order to mitigate effects from pileup are employed in the CMS experiment. Some of these are discussed in Chap. 5.

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 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 kinematic event properties or the detector performance. In particular, they allow to estimate event kinematics of new physics events that can subsequently be used to derive for instance 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 CMS detector.

4.1 Event Modeling

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 [102]. For convenience, events obtained from simulation are denoted by the label 'MC' in this thesis. A broader overview of event simulation and respective generators for LHC physics can be found e. g. in [103, 104].
Typically, the generation of an event follows several subsequent steps which are illustrated in Fig. 4.1:

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. Thus, two protons interact when there is a momentum transfer Q taking place between two individual partons. The probability for individual partons to take part in the hard interaction is parametrized by the parton-distribution functions which have been determined experimentally for instance in deep-inelasting scattering at the HERA collider [106]. 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* [107, 108]. 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

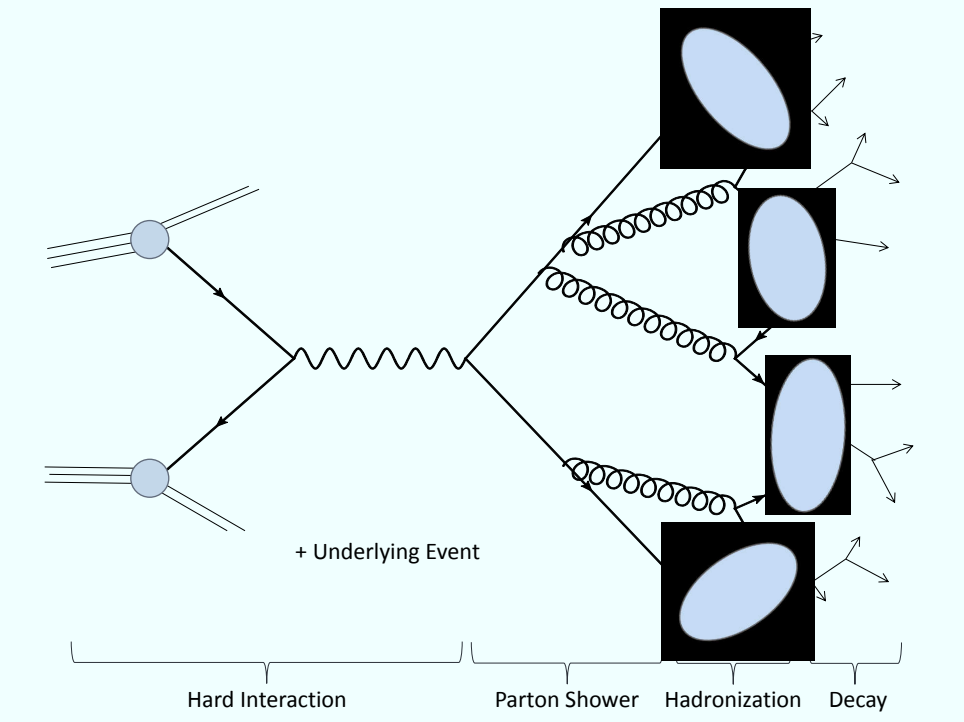

Figure 4.1: Sketch of individual steps in the generation process of simulated events. Adapted from [105].

section for a process $ab \to n$ is given according to

$$\sigma = \sum_{a,b} \int_0^1 \int_0^1 dx_a \, dx_b \int f_a(x_a, \mu_F) f_b(x_b, \mu_F) \, d\hat{\sigma}_{ab\to n}(\mu_F, \mu_R) \, . \tag{4.1}$$

Here, $f(x, \mu_F)$ are the PDFs of the interacting protons which depend on the factorization scale μ_F while $\sigma_{ab\to n}$ 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 [109]) and *cluster models* (cf. for instance [110, 111]) 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.

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 ($c\tau \gtrsim 10\,\text{mm}$). 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, theses 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. Particles arising from the hadronization and decay are denoted with *generator-level* particles and labelled 'gen' within this thesis.

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) [112–114]. The UE activity can be for instance caused by additional interactions occuring 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.

4.2 Event Generators

The various sub-processes described in Sec. 4.1 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 [115]: 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 $2 \to 3$ processes. Nonetheless, more than three final state particles can arise for instance from the parton shower. This parton shower 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++ [116]: 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 [117]: 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 $2 \to n$ 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 [118]: The POWHEG programme provides accurate QCD computations based on matrix elements up to next-to-leading order following the concept proposed in [119]. It can be interfaced to shower programs like PYTHIA or HERWIG++ in order to perform the subsequent showering process.

4.3 Detector Simulation

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 [120] 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* [121]. Especially in analyses in which many signal samples are needed in order 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 [122]. 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 in the next chapter.

5 Object Reconstruction and Particle Identification

Particles produced in pp collisions traverse the detector and interact with the detector components in a characteristic manner, e. g. by producing hits in the inner tracker or by initiating showers in the calorimeters. Thus, it is possible to reconstruct different objects from the detector signals, like tracks and energy deposits, 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. Next, the reconstruction of jets is discussed in Sec. 5.2. In Sec. 5.3 and 5.4, the identification of decays from B hadrons and boosted top quarks is reviewed.

5.1 Global Event Description with the Particle-Flow Algorithm at CMS

The CMS experiment introduced the Particle-Flow algorithm [123] for the reconstruction of collision events which 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, as discussed in Sec. 3.2.3. Furthermore, the strength of the magnetic field together with a high ECAL granularity allows photons to be separated from charged-particle energy deposits.
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 [124]. 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.
Furthermore, calorimeter clusters are formed based on adjacent calorimeter cells in each sub-detector separately: ECAL barrel, ECAL endcap, HCAL barrel, HCAL endcap, PS first layer and PS second layer. First, seed clusters are defined as local calorimeter-cell energy maxima. Second, neighbouring cells are combined with seed clusters when their energy exceeds a pre-defined threshold representing two standard deviations of the electronics noise. In the HF however, no clustering is performed and each calorimeter cell gives rise to one cluster.
A particle traversing the detector gives typically rise to several of such elementary objects: one charged-particle track, and/or several calorimeter clusters and/or one muon track.

Consequently, 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 exceeds the measured momentum from the track beyond the expected calorimeter energy 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 jets as discussed in Sec. 5.2, missing transverse energy $\not{E}_{\mathrm{T}}$ or decay products of tau leptons. More detailed information on the specific quality criteria required for the identification of certain particles is given in Chap. 6, 7 and 8 for each analysis presented in this thesis.

5.2 Reconstruction of Jets

As stated earlier, the searches for supersymmetry presented in this thesis are performed in final states containing missing transverse energy and several jets. 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.

5.2.1 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 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 cross sections calculated at fixed order perturbation theory would diverge.

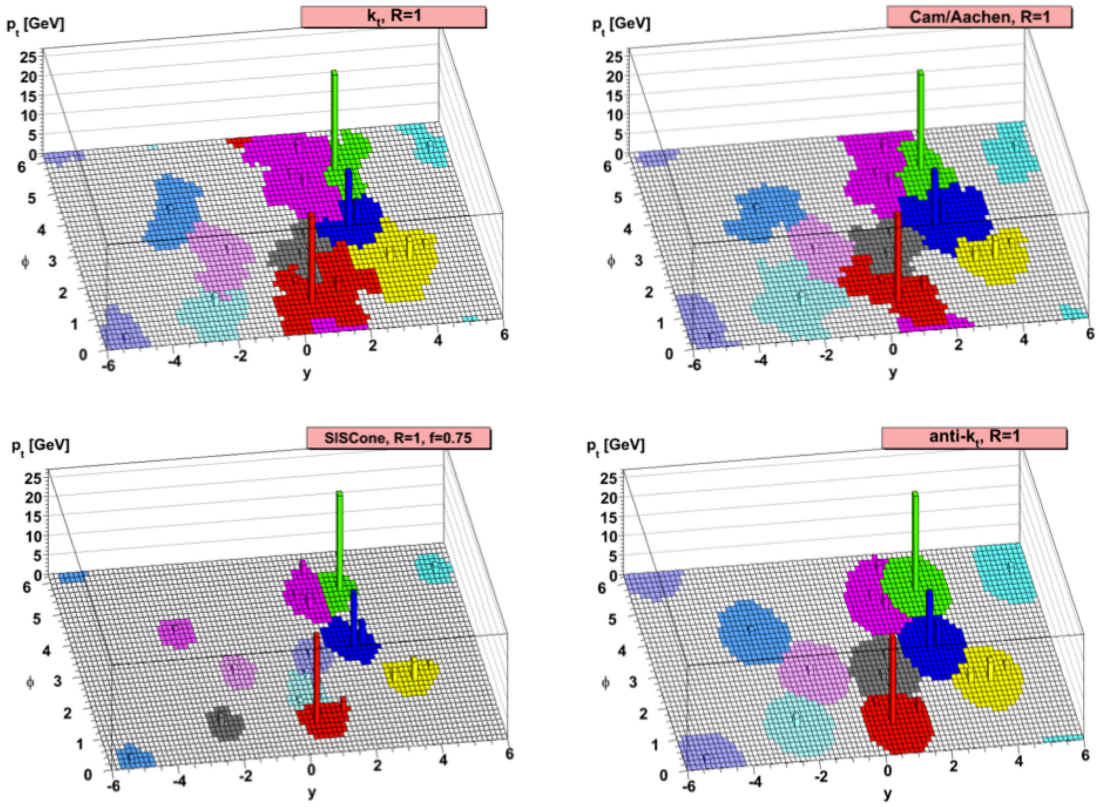

Figure 5.1: Example parton-level event generated with HERWIG++, adding many random soft particles, that is clustered with the k_T algorithm (*top left*), the CA algorithm (*top right*), the SISCone algorithm (*bottom left*) and the anti-k_T algorithm (*bottom right*). Colours illustrate all particles clustered to the resulting hard jets [125].

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 [125]. Technical implementations of various jet algorithms are provided by the *FastJet* package [126, 127].

Cone Algorithms: Cone algorithms 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. Then, momenta of all constituents j within a cone defined by

$$\Delta R_{i,j}^2 < R^2 \tag{5.1}$$

are added to the momentum of the seed, with $\Delta R_{i,j}$ calculated as introduced in Sec. 3.2.1. The resulting direction is used as new seed direction and the whole pro-

cedure 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 an $\mathcal{O}(N \log(N))$ time-dependence is given by the SISCone algorithm [128]. 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 based on variables invariant under longitudinal boosts is

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

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

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, i. e. 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. 5.1. The *k_T algorithm* uses $p = 1$ [129] and thus clusters soft particles first. This results mainly in irregularly shaped jets (cf. Fig. 5.1) and makes them sensitive to radiation in the event. Consequently, they are difficult to calibrate and therefore impractical to use at hadron colliders. The *Cambridge-Aachen algorithm* or short *CA algorithm* [130,131] 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 to be a jet algorithm especially suited for analyses of jet substructure as will be discussed later in Sec. 5.4. These jets still

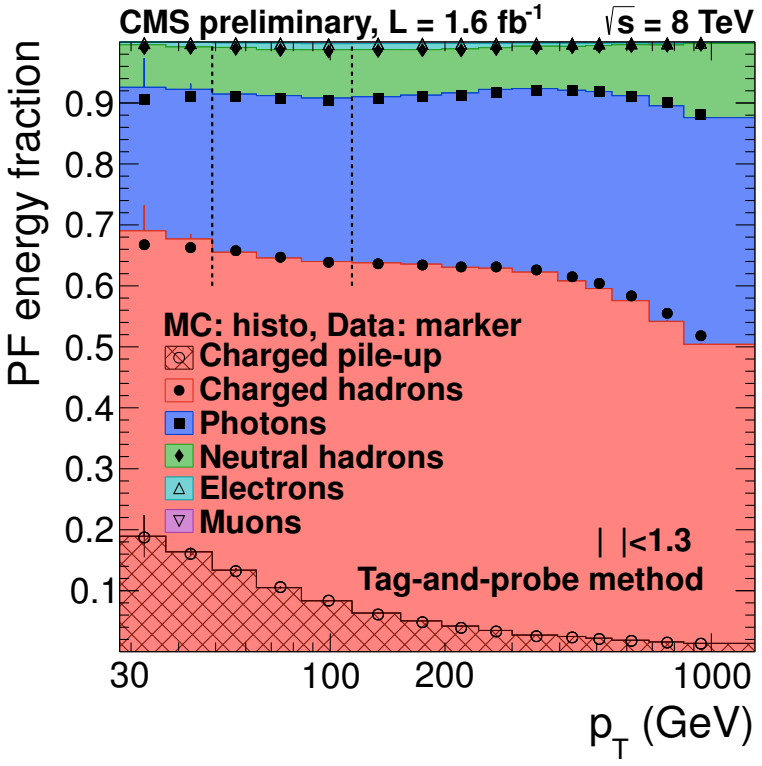

Figure 5.2: Composition of the PF jet energy versus jet p_{T} in the barrel detector region $|\eta| < 1.3$ in simulated events (coloured histograms) and data (solid markers) [134].

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 $p = -1$ [132] 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 easier calibration of anti-k_{T} jets than, e. g. of k_{T} jets.

5.2.2 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 algorithm 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.
Based on the information used from the various CMS subdetectors for the jet clustering, different types of detector-level jets are distinguished [133]:

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 [135]. 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 Sec. 5.1. 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. 5.2) 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 pileup, charged hadrons unambigously 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. The respective jets are denoted PFCHS jets.

5.2.3 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. This effect is quantified by the *jet transverse-momentum response* $\mathcal{R}$ which is defined as

$$\mathcal{R} = \frac{p_{\mathrm{T}}}{p_{\mathrm{T}}^{\mathrm{particle}}} \tag{5.2}$$

with the transverse momentum p_{T} of the jet measured at detector level and the transverse momentum $p_{\mathrm{T}}^{\mathrm{particle}}$ 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 jet response usually depends on the jet momentum as well as on the pseudorapidity which 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 non-linear response of the calorimeters, the specific track-reconstruction efficiency, the individual amount of detector material, cracks in the detector layout or different instrumentation.
In this thesis, the average response $\langle \mathcal{R} \rangle$ is referred to as *jet energy scale* (JES) while the width of the response distribution is denoted as relative *jet transverse-momentum resolution* (JER).

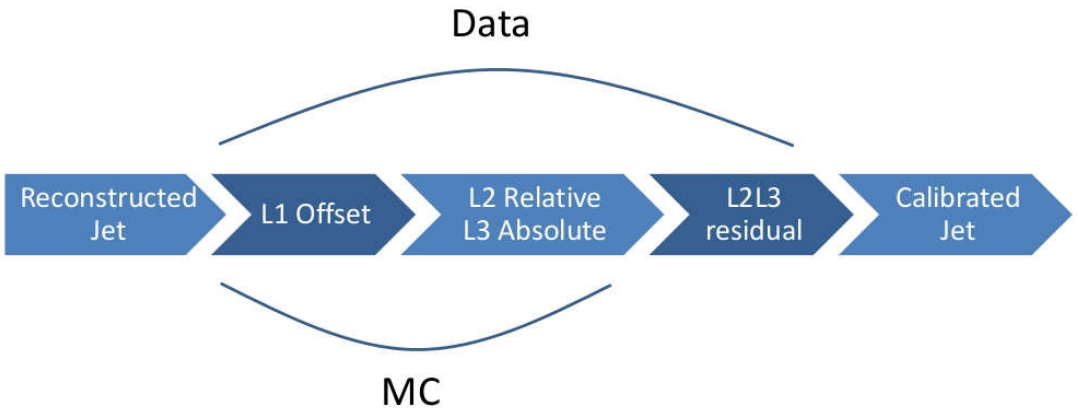

Figure 5.3: Sketch of the factorized approach used for jet energy corrections at the CMS experiment.

5.2.4 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. This compensates for the non-linear response of the calorimeters and ensures that $\langle \mathcal{R} \rangle = 1$. Within the CMS experiment, a factorized approach is utilized which is described in detail in [133] and illustrated in Fig. 5.3. The actual set of correction factors used in this thesis, if not stated otherwise, is documented in [136].
The calibrated jet four-momentum vector p^{cor} is obtained from the raw jet four-momentum vector p^{raw} by scaling the raw momentum with a correction factor C according to

$$p^{\text{cor}} = C \cdot p^{\text{raw}} = C_{\text{offset}}(p_{\text{T}}^{\text{raw}}, \eta) \cdot C_{\text{rel}}(\eta) \cdot C_{\text{abs}}(p'_{\text{T}}) \cdot C_{\text{res}}(p''_{\text{T}}, \eta) \cdot p^{\text{raw}} \tag{5.3}$$

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}. While C_{offset}, C_{rel} and C_{abs} are applied to both data and simulation, the residual correction factors C_{res} are applied to data only. 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 pileup 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* [137]).
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. 5.4.

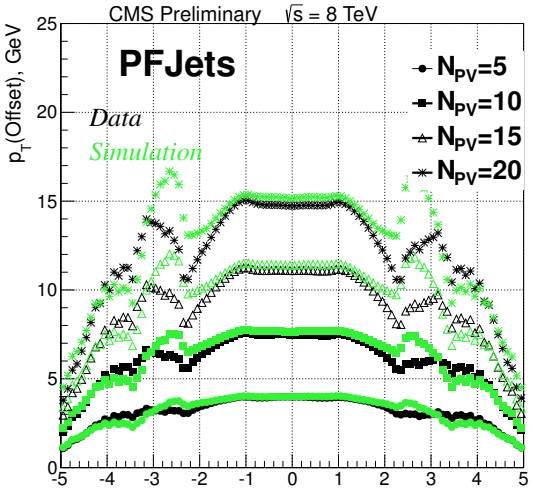

Figure 5.4: L1 offset transverse momentum correction for AK5 PF jets as a function of jet pseudorapidity in data (black) and simulation (green). Different intervals of reconstructed primary vertices (N_{PV}) are shown with different markers [136].

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. 5.5.

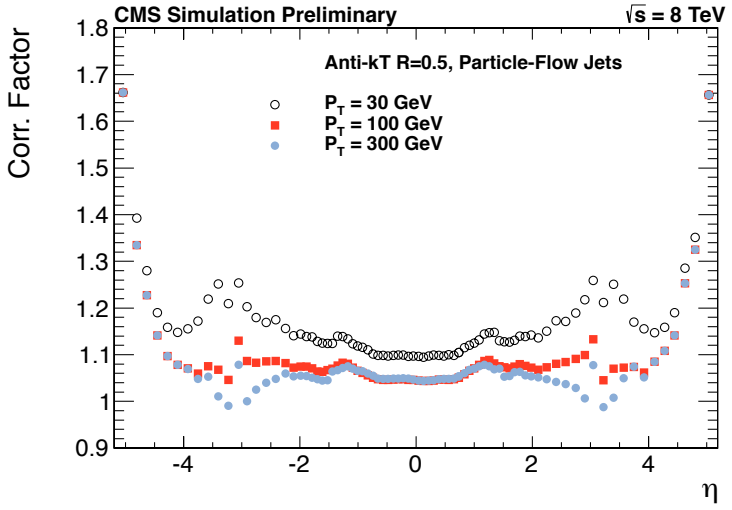

Figure 5.5: MC-Truth corrections for AK5 PF jets as a function of jet pseudorapidity for three reference transverse momentum values: 30 GeV (white hollow circles), 100 GeV (red squares) and 300 GeV (blue circles) [136].

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. Residual corrections can be derived from events that have momentum balance in the transverse plane, like dijet events (used for the determination of the L2 residual correction) or $Z/\gamma + \text{jet}$ events (used for the measurement of the L3 residual correction). The L2 residual correction is illustrated in Fig. 5.6.

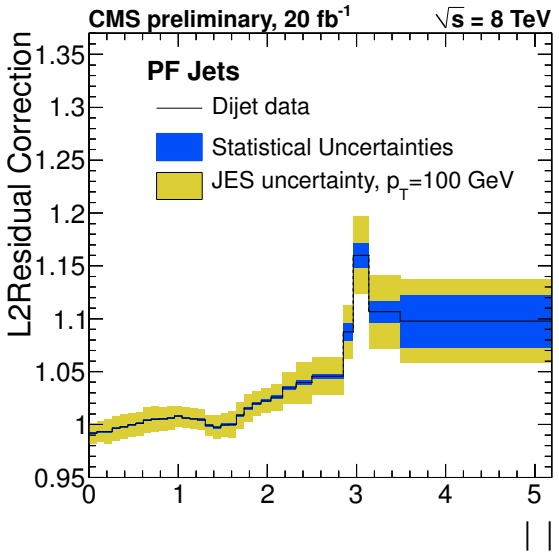

Figure 5.6: L2 residual corrections for AK5 PF jets as a function of jet pseudorapidity, obtained from dijet events, shown with JES uncertainty (yellow band) and the statistical uncertainty (blue band) [136].

The calibration factors are obtained with respect to the average flavour composition of a QCD mulitjet 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.

5.3 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 have a lifetime of around $c\tau \approx 500\,\mu\text{m}$, so that they 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 mesons.

The CMS experiment exploits the specific b-jet properties in dedicated b-tagging algorithms for an efficient b-jet identification. 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 comissioned within the CMS experiment are [138]:

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 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). The SSVHE is based on vertices with at least two associated tracks, while 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 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 an efficient identification of b jets, also in cases where no secondary vertex can 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. Important variables used in the CSV algorithm are the flight distance significance, vertex mass, number of tracks at the vertex, number of tracks in the jet and the IP sigificances 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.

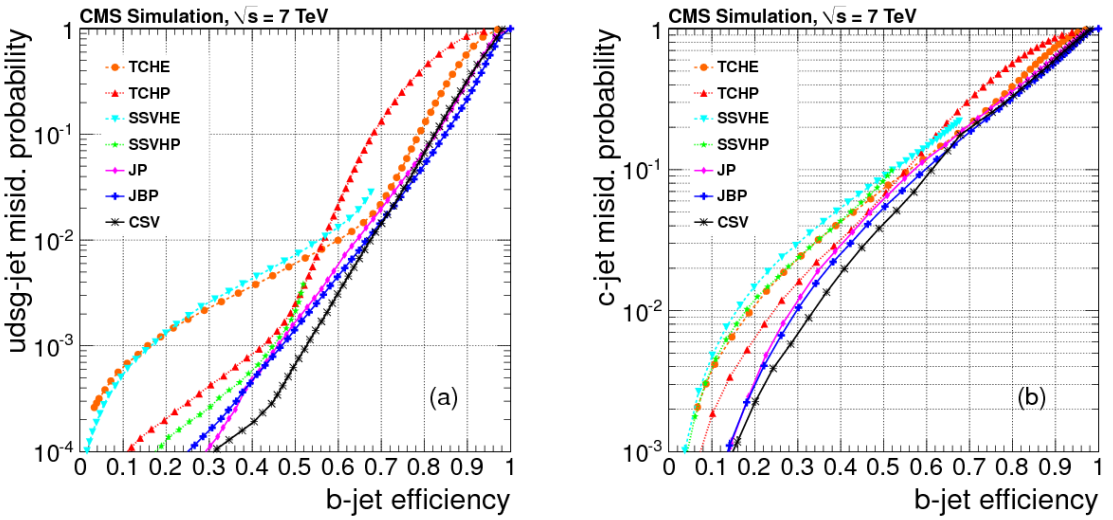

Figure 5.7: Performance curves obtained from simulation for the algorithms described in the text. (a) light-parton and (b) c-jet misidentification probabilities as a function of the b-jet efficiency. Taken from [138].

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. 5.7 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 $p_{\mathrm{T}} > 60$ 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 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 $\sqrt{s} = 8\,\mathrm{TeV}$ in 2012 were the TCHP, JP and CSV algorithms [139].

5.4 Identification of Boosted t-Quark Jets

Supersymmetric models or other scenarios describing physics beyond the standard model predict the extistence 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. 2.3.3, 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 $p_{\mathrm{T}} \ll m_t$, the top quark decay products show up as distinct objects in the detector. In the case of hadronic tops, these are well separated jets. However, if the top transverse momentum is high, the decay products are *boosted*, i. e. they have large Lorentz boost, and

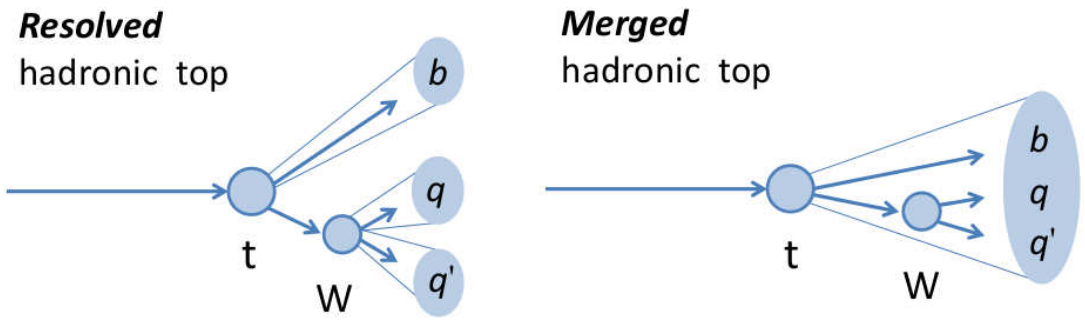

Figure 5.8: Schematic diagrams of the decay of a top-quark in the resolved case (*left*) and the boosted scenario (*right*).

thus are 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

$$\Delta R \approx 2\, m_t / p_{\mathrm{T}} \tag{5.4}$$

with the mass m_t and the transverse momentum p_{T} of the decaying particle. Schematic diagrams of resolved and boosted top quark decays are illustrated in Fig. 5.8.
In order to identify decays of boosted hadronic top quarks, several *top-tagging* algorithms (or short *taggers*) are commissioned within the CMS experiment [140, 141]. Two of them are used within this thesis:

CMS Top Tagger: The CMS Top Tagger [142] is based on the top tagger developed by Kaplan et al. [143] and acts on jets clustered by the Cambrigde-Aachen algorithm with distance parameter $R = 0.8$ (CA8 jet). These jets, used as input for the algorithm, are 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 $p_{\mathrm{T}}^{\mathrm{jet}} > 350\,\mathrm{GeV}$ 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. 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 [140]. 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 $140 < m_{\mathrm{jet}} < 250\,\mathrm{GeV}$.
- 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 $> 50\,\mathrm{GeV}$.

HEP Top Tagger: The HEP Top Tagger [144] is also based on jets clustered with the Cambrigde-Aachen algorithm, but with a larger distance parameter than the CMS Top Tagger of $R = 1.5$ (CA15 jets). This makes the HEP top-tagging algorithm especially suited for top quark decays with moderate boost and thus uses fat jets with a transverse momentum greater than 200 GeV as input. 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. Details on the applied decomposition and reclustering criteria can be found in [140]. In the end, the combination with the mass, determined from identified 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 top quark decays. 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 of $140 < m_{123} < 250\,\text{GeV}$.
- In order to select the W boson mass, the jet has to satisfy at least one of the following conditions based on the subjet pairwise masses

$$0.2 < \text{atan}\frac{m_{13}}{m_{12}} < 1.3 \;\text{ and }\; R_{\text{min}} < \frac{m_{23}}{m_{123}} < R_{\text{max}} \tag{5.5}$$

$$R^2_{\text{min}}\left(1+\left(\frac{m_{13}}{m_{12}}\right)^2\right) < 1-\left(\frac{m_{23}}{m_{123}}\right)^2 < R^2_{\text{max}}\left(1+\left(\frac{m_{13}}{m_{12}}\right)^2\right) \;\text{ and }\; \frac{m_{23}}{m_{123}} > 0.35 \tag{5.6}$$

$$R^2_{\text{min}}\left(1+\left(\frac{m_{12}}{m_{13}}\right)^2\right) < 1-\left(\frac{m_{23}}{m_{123}}\right)^2 < R^2_{\text{max}}\left(1+\left(\frac{m_{12}}{m_{13}}\right)^2\right) \;\text{ and }\; \frac{m_{23}}{m_{123}} > 0.35 \tag{5.7}$$

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

In addition, further selection criteria can be added to these taggers to achieve a potential performance improvement. These criteria are subjet-b tagging, as introduced in [139], and selections on the *n-subjettiness* ratio τ_3/τ_2 [145, 146] that denotes how likely a fat jet is composed of three subjets rather than two.

In analogy to the b-tagging algorithms, different working points for each top tagger can be defined characterized by a specific top-tagging efficiency and misidentification rate. Typically, the working points are chosen such that they have a minimum mistag rate for a given signal efficiency. The performance of various taggers is compared by determining the top-tagging efficiency versus the misidentification rate. Commonly, the top-tagging 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. 5.9, the mistag rate versus the top-tagging efficiency is illustrated for various taggers and parton momenta. The top-tagging 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 $> 200\,\text{GeV}$ for the

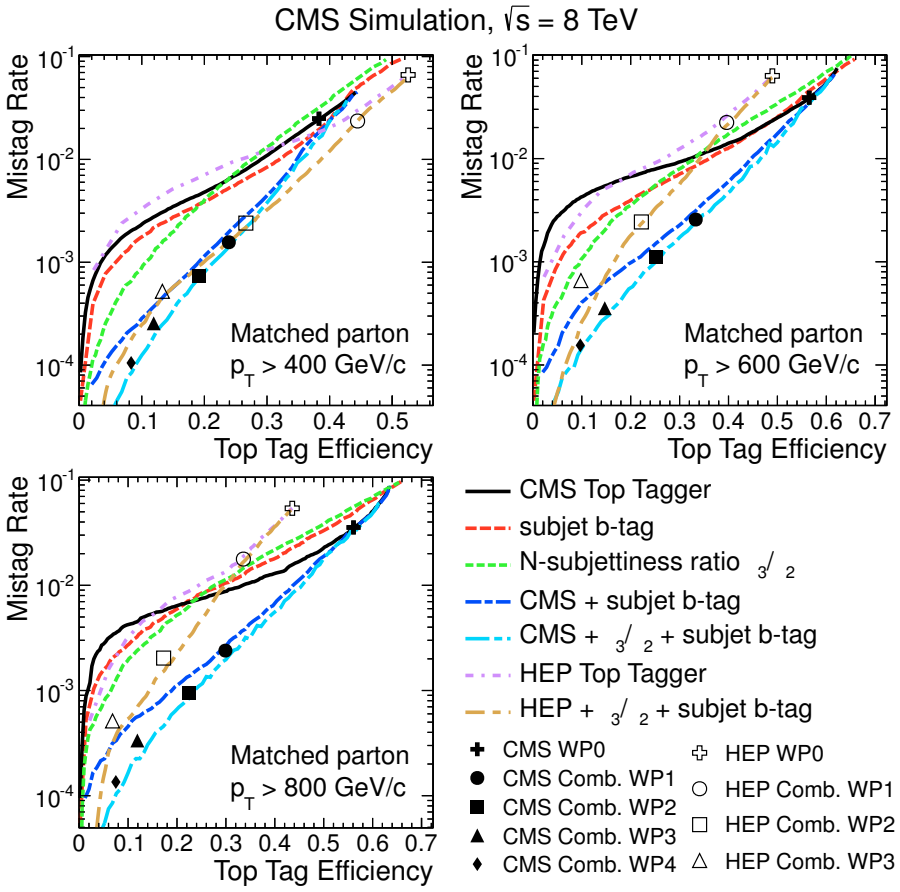

Figure 5.9: Mistag rate versus top-tagging efficiency as measured from QCD PYTHIA6 and POWHEG $t\bar{t}$ simulated events, respectively. In the cases where a jet mass cut is applied, the cut is not varied, but fixed at $140 < m_{\text{jet}} < 250\,\text{GeV}$. N-subjettiness is calculated using $R = 0.8$ 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 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 [140].

HEP Top Tagger (HEP WP0) and 38.3% for a matched parton-p_{T} of $> 400\,\text{GeV}$ 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 pileup such that in high pileup environments a performance degradation is expected. For instance, in the case of the CMS Top Tagger the mistag rate increases with a slope of $0.095\% \pm 0.006\%$ as a function of the number of primary vertices.

6 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 is of major importance and a crucial ingredient for such kind of analyses. One of such properties is the jet transverse-momentum resolution as introduced in Sec. 5.2.3.
Many new physics searches are carried out based on final states containing missing transverse momentum, several jets and no isolated leptons. 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 thus a keypoint in the prediction of such background contributions as discussed later in Chap. 7.
In this chapter, an analysis is presented in which 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 $\sqrt{s} = 7\,\mathrm{TeV}$ [133, 147], while a complementary approach utilizes γ + jet events [133, 148–150]. The measurement shown here is based on collision data corresponding to an integrated luminosity of $19.7\,\mathrm{fb}^{-1}$ recorded at $\sqrt{s} = 8\,\mathrm{TeV}$ in 2012. Parts of this chapter are taken from [151], having been written by the author.

6.1 Components of the Jet Response

As introduced in Sec. 5.2.3, the relation of the transverse momentum of a jet at detector level to the momentum of the corresponding particle-level jet is expressed by the jet response. In simulated events, the transverse momentum of the particle-level jet is known and corresponds to the p_{T} of the generated jet which is clustered from all stable particles after hadronisation and decay including neutrinos. Thus, the *MC-truth response* can be determined as

$$\mathcal{R} = \frac{p_{\mathrm{T}}}{p_{\mathrm{T}}^{\mathrm{gen}}} \tag{6.1}$$

for matched pairs of reconstructed and generated jets. In Fig. 6.1, an example for a jet response distribution derived from simulation is shown. It is obtained from a QCD multijet sample generated with Pythia6 tune Z2 [114] using the CTEQ6L1 PDF set [152] and processed with the full detector simulation. 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 jet within a distance of $\Delta R < 0.25$. Otherwise, the response is calculated in intervals of $p_{\mathrm{T}}^{\mathrm{gen}}$ and $|\eta^{\mathrm{gen}}|$. This is done since a dependence on the momentum and the respective

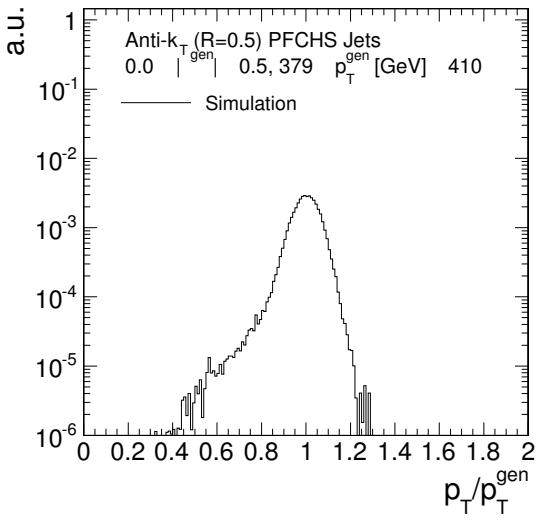

Figure 6.1: Jet response distribution for one example $|\eta^{\text{gen}}|$ and $p_{\text{T}}^{\text{gen}}$ interval derived from simulation.

detector region is expected, as discussed in Sec. 5.2.3. The calculation of the jet response is performed after applying the jet energy corrections discussed in Sec. 5.2.4 to the detector-level jet momenta, so that the mean of the response distribution is located at one.
Apparently, the jet response distribution consists 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 performance of the PF algorithm. Since the jet momenta obtained from the PF constituents are based on tracking- as well as calorimeter-based measurements, the jet resolution is closely connected to the evolution of the tracking and calorimeter resolution with p_{T}. In the tracking system, the intrinsic resolution is mainly caused by the uncertainty on the track curvature. This is limited by multiple scattering at low p_{T} and by the finite hit-position resolution at high p_{T}. In total, the track-p_{T} resolution degrades for increasing transverse momentum. However, the evolution of the resolution in the calorimeters behaves the other way round and improves with increasing momentum. At low momenta, the calorimeter resolution is mainly dominated by electronic noise and pileup 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 detector response due to the non-linearity of the calorimeters.
The response tails are caused by severe jet-momentum mismeasurements. These can be caused by detector effects or physics processes. Effects from the detector design, which lead to response tails, are mainly *shower leakage* and *punch-through*. Shower leakage means that not the whole shower is deposited in the calorimeters due to dead material, cracks in the detector or non-instrumented regions, whereas punch-through refers to events in which some particles cross the whole apparatus because of the limited depth of the calorimeters.

In addition, also malfunctioning detector components can lead to high or low reponse tails by creating artifical signals. Beyond the detector effects, the lower response tails get for instance 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.

6.2 Basic Concept of the Dijet Asymmetry Method

As introduced in Sec. 5.2.3 and 6.1, the jet transverse momentum resolution corresponds to the width of the Gaussian-shaped core part of the jet response distribution. 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

$$\mathrm{A} = \frac{p_{\mathrm{T},1} - p_{\mathrm{T},2}}{p_{\mathrm{T},1} + p_{\mathrm{T},2}} \,. \tag{6.2}$$

In this equation $p_{\mathrm{T},1}$ and $p_{\mathrm{T},2}$ correspond to the randomly ordered transverse momenta of the two leading jets.
Neglecting tails, the asymmetry is approximately normally distributed, with mean = 0, and the standard deviation is given as

$$\sigma_{\mathrm{A}} = \left|\frac{\partial \mathrm{A}}{\partial p_{\mathrm{T},1}}\right| \cdot \sigma(p_{\mathrm{T},1}) \oplus \left|\frac{\partial \mathrm{A}}{\partial p_{\mathrm{T},2}}\right| \cdot \sigma(p_{\mathrm{T},2}) \,. \tag{6.3}$$

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 $\langle p_{\mathrm{T},1}\rangle = \langle p_{\mathrm{T},2}\rangle = \langle p_{\mathrm{T}}\rangle$ and $\sigma(p_{\mathrm{T},1}) = \sigma(p_{\mathrm{T},2}) = \sigma(p_{\mathrm{T}})$. This allows the simplification of Eq. 6.3 and provides the following important relation between the width of the asymmetry σ_{A} and the jet-p_{T} resolution $\sigma(p_{\mathrm{T}})$

$$\frac{\sigma(p_{\mathrm{T}})}{\langle p_{\mathrm{T}}\rangle} = \sqrt{2} \cdot \sigma_{\mathrm{A}} \,. \tag{6.4}$$

This relationship was already used at the Tevatron experiments [153, 154], the ATLAS experiment [155] or in previous CMS analyses [133, 147] to measure the jet resolution from dijet events.

6.3 Application to Realistic Collision Events

The measurement of the jet transverse-momentum resolution in collision events is based on Eq. 6.4. As discussed in Sec. 5.2.3, the resolution is a function of p_{T} and η. In order to account for this dependence, the asymmetry has to be recorded in intervals of pseudorapidity and a measure for the transverse-momentum scale of the event, e. g. the p_{T} of the leading jet. However, the jet-p_{T} spectrum is affected by migration effects such that a binning based on single jet momenta is not suitable. Because of the limited jet resolution, a particular interval of reconstructed jet momenta is populated not only by jets whose

particle-level jet momentum belongs to that bin. In case of a steeply falling spectrum, as it is the case for jet momenta, more jets with low $p_{\mathrm{T}}^{\mathrm{gen}}$ migrate into a specific interval than jets with high $p_{\mathrm{T}}^{\mathrm{gen}}$. Consequently, the measured response is systematically higher and the measured relative response is biased towards the object with 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

$$p_{\mathrm{T}}^{\mathrm{ave}} = \frac{1}{2}(p_{\mathrm{T},1} + p_{\mathrm{T},2}) \tag{6.5}$$

such that the bias is cancelled on average. By definition, the $p_{\mathrm{T}}^{\mathrm{ave}}$ resolution is better by a factor of $\sqrt{2}$ than the single-jet resolution.

Beyond these aspects, the ideal dijet topology with exactly two jets that are 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 dijet system. This additional jet activity can be described by the variable α which is defined as the ratio of the transverse momentum of the third jet to the average momentum:

$$\alpha = \frac{p_{\mathrm{T},3}}{p_{\mathrm{T}}^{\mathrm{ave}}}\,. \tag{6.6}$$

The presence of jets beyond the third jet is neglected in the parametrization of the additional activity in the event as these have consecutively less momentum due to the decreasing jet production cross section versus jet-p_{T} [156]. The presence of additional jets and the thereby introduced imbalance leads to a broadening of the observed asymmetry distribution. In order to determine the intrinsic resolution, the measured resolution has to be corrected for this effect.
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. Typically, some particles might be too soft to be included in the clustered jet. Furthermore, additional contributions from the underlying event might be wrongly asscociated to a jet. In general, such effects have to be considered in the resolution measurement as well. The actual procedure how to determine and apply the required corrections is discussed in Sec. 6.6.

6.4 Samples and Event Selection

6.4.1 Datasets and Triggers

In this analysis, multijet events from pp collisions are considered which have been recorded in 2012 with the CMS detector at $\sqrt{s} = 8\,\mathrm{TeV}$. From these data, only those are considered in which 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.) $\pm$ 0.5% (stat.) [157].
Multijet events are preselected 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 $p_{\mathrm{T}}^{\mathrm{ave}}$ spectrum, different trigger paths are combined. Since they have different minimum $p_{\mathrm{T}}^{\mathrm{ave}}$

Table 6.1: Trigger paths with $p_{\mathrm{T}}^{\mathrm{ave}}$ thresholds at which the trigger efficiency reaches the 99% efficiency plateau. Thresholds are given for PFCHS jets.

Trigger	$p_{\mathrm{T}}^{\mathrm{ave}}$ threshold [GeV]
HLT_DiPFJetAve40	62
HLT_DiPFJetAve80	107
HLT_DiPFJetAve140	175
HLT_DiPFJetAve200	242
HLT_DiPFJetAve260	310
HLT_DiPFJetAve320	379
HLT_DiPFJetAve400	467

thresholds, a broad range in $p_{\mathrm{T}}^{\mathrm{ave}}$ is considered. In Tab. 6.1 the different trigger paths used for this analysis are listed and the particular offline $p_{\mathrm{T}}^{\mathrm{ave}}$ value, for which the respective trigger reaches 99% of the efficiency plateau, is given [158]. The low $p_{\mathrm{T}}^{\mathrm{ave}}$ triggers have been operated with prescale factors applied and only the trigger with the highest $p_{\mathrm{T}}^{\mathrm{ave}}$ threshold has been employed without the usage of prescale factors.
The simulated QCD sample that is used in the analysis is generated with Pythia, as discussed in Sec. 6.1. Since the cross section of the process has been scaled by $\hat{p}_{\mathrm{T}}^{4.5}$, with the scale parameter $\hat{p}_{\mathrm{T}}$ describing the momentum transfer in the hard process, the sample is reweighted with the inverse in order to regain the physical spectrum.

6.4.2 Selection Criteria

The physics objects used in the analysis are reconstructed with the PF algorithm including charged-hadron subtraction, as described in Sec. 5.1. Jets are clustered with the anti-k_{T} algorithm using a distance parameter of $R = 0.5$. They are calibrated in data and simulation following the procedure introduced in Sec. 5.2.4.
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 with respect to the azimuthal angle. 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 [159, 160]. In order to mitigate effects from pileup, only jets with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ are considered for the analysis. As discussed in Sec. 6.3, events with a topology of exactly two high-p_{T} jets and no further jet activity are very rare 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 pileup, it is required that each event

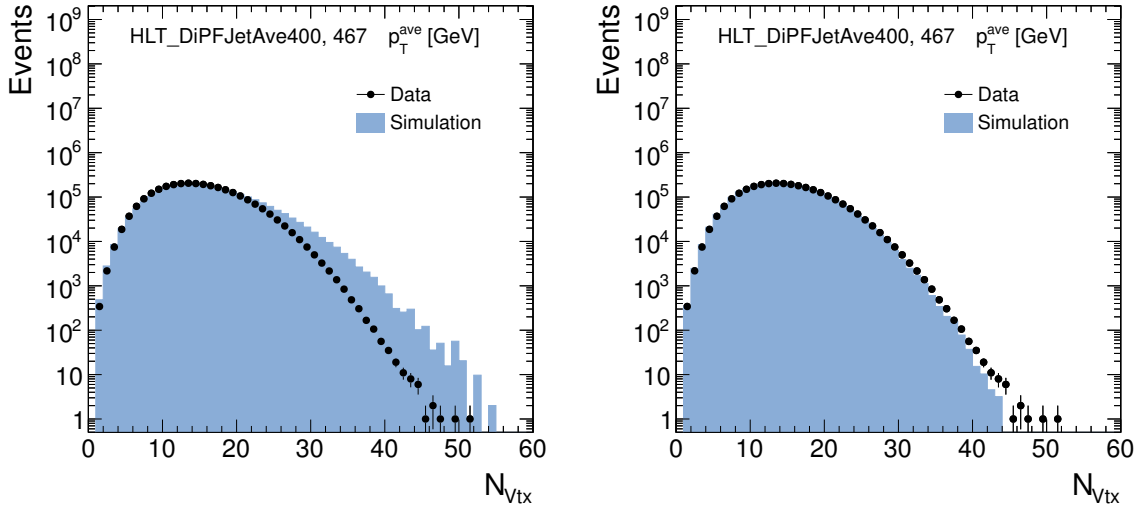

Figure 6.2: Distribution of number of primary vertices in data (black dots) and simulation (blue histogram) before (*left*) and after (*right*) reweighting of the pileup scenario in simulation for trigger path HLT_DiPFJetAve400.

has a third jet passing the p_T threshold of 10 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

$$\alpha < 0.25$$

is introduced. In order to enrich the sample with events close to the ideal dijet topology, in which two jets point into opposite directions in the transverse plane, the two leading jets have to fulfill

$$|\Delta\phi| > 2.7 \text{ with } \Delta\phi = \Delta\phi(\vec{p}_{T,1}, \vec{p}_{T,2})\,. \tag{6.7}$$

The selection criteria described above are applied to data and simulation in an identical manner. For the simulated sample a further adjustment is necessary. Typically, the simulation is performed before the actual data taking takes place and it is unknown which specific pileup conditions will be present in data. Thus, the simulation is performed with an estimated pileup scenario and it is necessary to adjust the simulation to the actual pileup scenario in data at analysis level. This is done by reweighting the simulated events to match the mean pileup distribution in data. Since the instantaneous luminosity changed throughout the data taking in 2012, the pileup conditions changed accordingly and the pileup distribution in data differs for each individual trigger. As stated above, the trigger paths utilized in this analysis have been mainly operated with prescale factors applied. Thus, each trigger path collected different amounts of data and consequently also the pileup distribution differs per trigger such that the reweighting of the pileup scheme has to be done for each trigger path individually. Depending on the offline p_T^{ave}, a simulated event can be unambigously assigned to the corresponding trigger path, which is fully efficient for that particular p_T range, and reweighted to that particular pileup scenario. The number of primary vertices in the event is a measure for the pileup activity so that the

Table 6.2: Overview of the $|\eta|$ and $p_{\mathrm{T}}^{\mathrm{ave}}$ interval boundaries used for the resolution measurement.

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

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. 6.2 for the trigger path with the highest $p_{\mathrm{T}}^{\mathrm{ave}}$ 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 App. 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 $p_{\mathrm{T}}^{\mathrm{ave}}$. These are summarized in Tab. 6.2. The $p_{\mathrm{T}}^{\mathrm{ave}}$ intervals are chosen in correspondence to the $p_{\mathrm{T}}^{\mathrm{ave}}$ values for which the different triggers become fully efficienct. If a trigger path provides enough statistical precision, the interval is separated into two. This definition of $p_{\mathrm{T}}^{\mathrm{ave}}$ intervals ensures that all events in one $p_{\mathrm{T}}^{\mathrm{ave}}$ 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 $0.0 < |\eta| < 0.5$ and $0.5 < |\eta| < 1.1$. The transition region, where the ECAL ends, is separated into an individiual bin $2.8 < |\eta| < 3.2$ and the forward detector is covered by one large bin $|\eta| = 3.2 - 5.0$, mainly due to the small number of available events there. In order to allow the application of Eq. 6.4 for the determination of the resolution, both leading jets are required to belong to the same $|\eta|$ interval $\Delta|\eta|$

$$\Delta|\eta|_{\mathrm{jet},1} = \Delta|\eta|_{\mathrm{jet},2} \tag{6.8}$$

To account for the α-dependence of the measured asymmetry distributions, the $p_{\mathrm{T}}^{\mathrm{ave}}$ and $|\eta|$ 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 $0.1, 0.125, 0.15, 0.175, 0.2, 0.225$ and 0.25. This inclusive definition of the α intervals implies that one specific event might be assigned to more than one α bin.
The resulting inclusive $p_{\mathrm{T}}^{\mathrm{ave}}$ spectrum after applying the described selection is shown for data and simulation in Fig. 6.3. The p_{T} spectra of the first three leading jets are shown as well. The number of simulated events in each trigger $p_{\mathrm{T}}^{\mathrm{ave}}$ interval is normalized to the respective integral in data. The shapes of the spectra in data and simulation agree well and also the effect of the prescales applied to the low-$p_{\mathrm{T}}^{\mathrm{ave}}$-threshold triggers is nicely visible resulting in the sawtooth-shaped $p_{\mathrm{T}}^{\mathrm{ave}}$ spectrum.

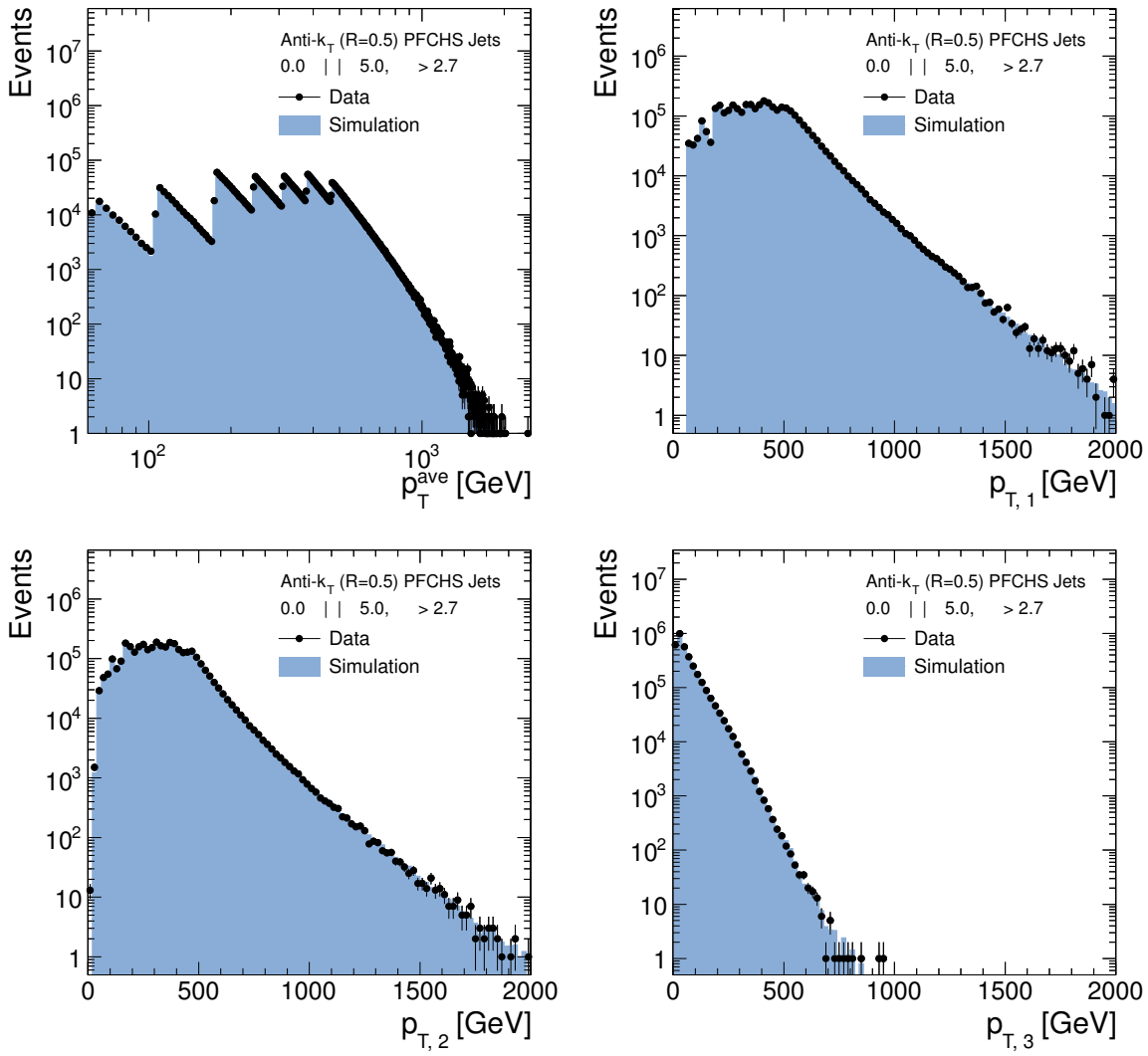

Figure 6.3: Inclusive p_T^{ave} spectrum of events after the total selection in data and in simulation (*top left*), leading jet p_T (*top right*), subleading jet p_T (*bottom left*) and third jet p_T (*bottom right*).

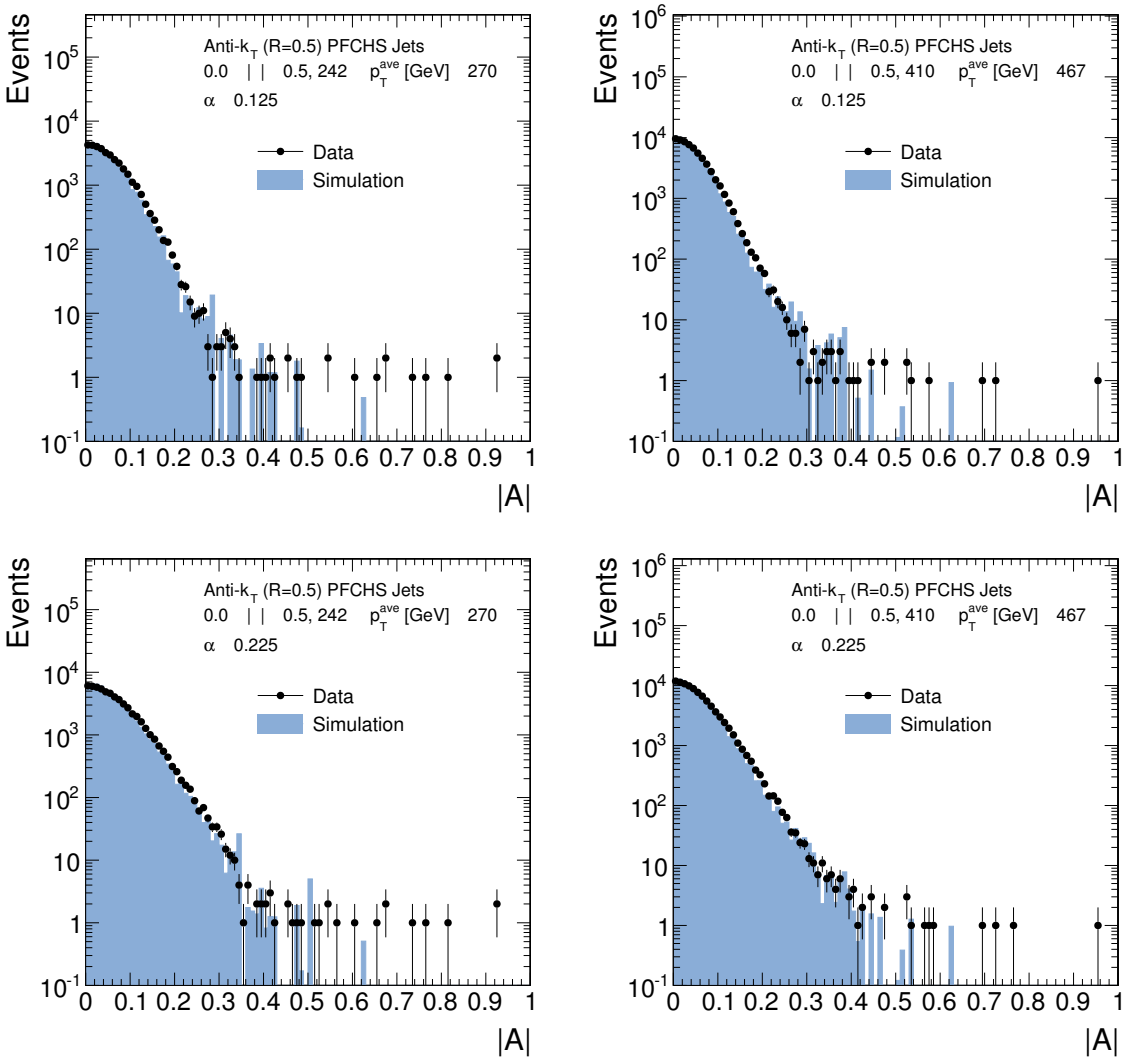

Figure 6.4: Some example asymmetry distributions for the lowest $|\eta|$ region, two medium $p_{\mathrm{T}}^{\mathrm{ave}}$ intervals and for a low (*top*) and a high (*bottom*) α interval.

Some example asymmetry distributions obtained after the described selection are depicted in Fig. 6.4 for data and simulated events. They have been obtained using Eq. 6.2. Instead of randomly assigning the first and second jet, the absolute value of the asymmetry is used. The effect that the asymmetry distribution gets broadened for larger values of α is visible when comparing the distributions in the top and the bottom row. In order to get a robust estimate of the asymmetry width, only asymmetry distributions with at least 100 events in the specified $p_{\mathrm{T}}^{\mathrm{ave}}$, η and α intervals are considered for the analysis.

6.5 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 needed.

From the tails of the jet 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 core part of the distribution is described. 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. This comparison is performed by calculating the width of the asymmetry with a certain definition and choosing the standard deviation of a Gaussian function according to the obtained asymmetry width.
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

$$\sigma_{\mathrm{A}} = \mathrm{RMS}_{t\%} = \sqrt{\frac{1}{\sum_i y_i} \sum_i y_i \cdot A_i^2} \tag{6.9}$$

with the frequency y_i of the asymmetry value A_i. The sum over i includes all values such that the total asymmetry distribution is covered from zero up to a percentage t and the asymmetry tail is neglected. Assuming a normal distribution, the statistical uncertainty is given by

$$\Delta\sigma_{\mathrm{A}} = \Delta\mathrm{RMS}_{t\%} = \frac{\mathrm{RMS}_{t\%}}{\sqrt{2 \cdot n_{\mathrm{eff}}}} \tag{6.10}$$

with the number of effective entries n_{eff}[1] in the specified ($p_{\mathrm{T}}^{\mathrm{ave}}$, $|\eta|$, α)-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. 6.5 for a certain $|\eta|$, $p_{\mathrm{T}}^{\mathrm{ave}}$ and α interval. The core part of the asymmetry distribution can in good approximation be described by a Gaussian function when choosing $\mathrm{RMS}_{98.5\%}$ as the asymmetry width. This is illustrated for one example interval in Fig. 6.5 and holds for other intervals as well. 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.

6.6 Corrections to the Dijet Asymmetry

The fundamental relation between the width of the asymmetry distribution and the jet energy resolution, as expressed in Eq. 6.4, holds in this form only for the case of an ideal dijet topology. In real collision events, various effects occur that disturb the exact balance of the two jets as discussed in Sec. 6.3. 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 measured asymmetry distribution which consequently also results in an increased measured jet resolution. In order to determine the intrinsic resolution, the measured asymmetry width has to be corrected for such effects.

[1] In case of an unweighted histogram, n_{eff} is equivalent to the number of histogram entries. However, for a weighted histogram, e. g. in simulation, n_{eff} corresponds to the hypothetical number of unweighted entries a histogram would need in order to have the same statistical power as the weighted histogram.

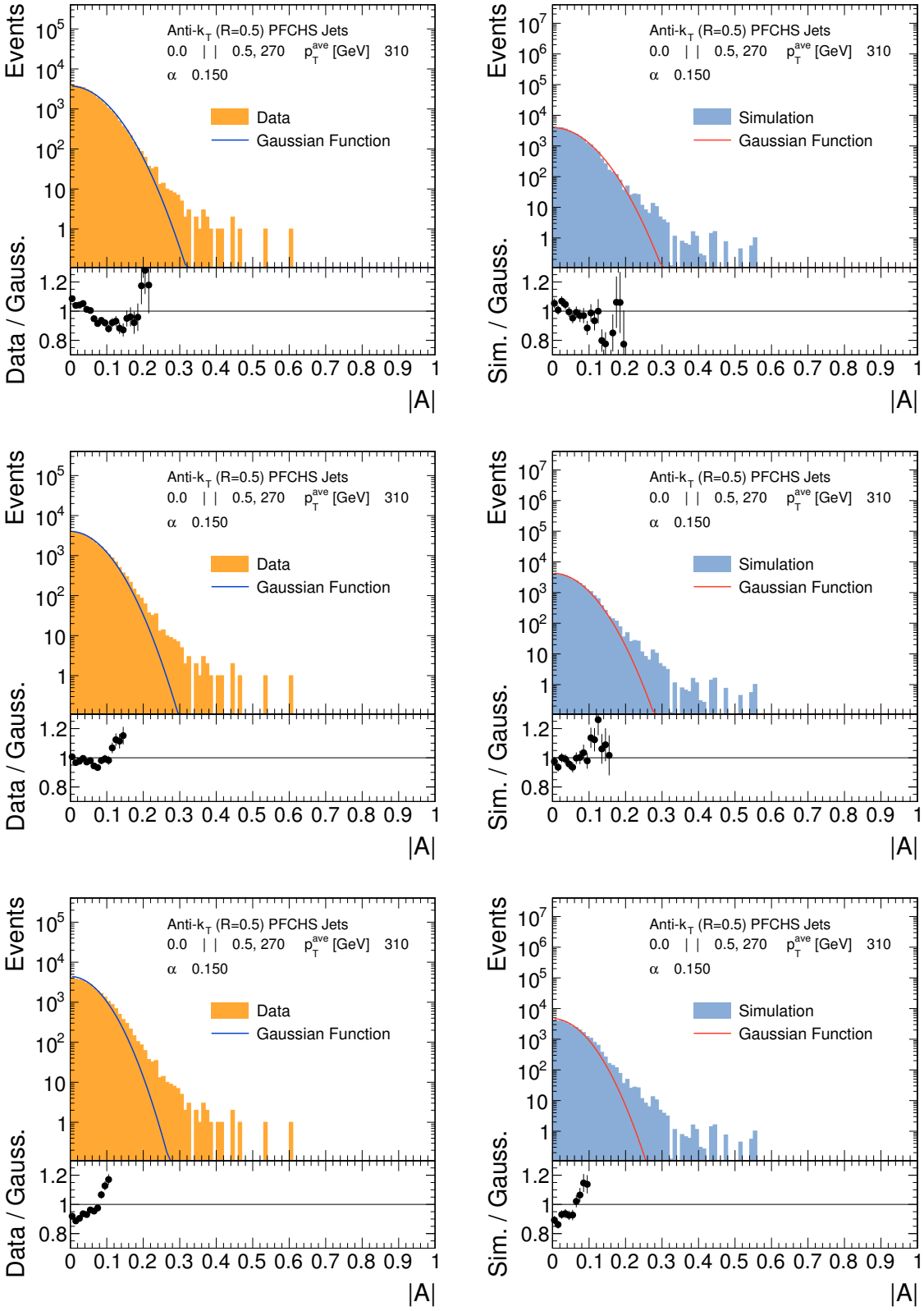

Figure 6.5: Asymmetry distributions for data (*left*) and simulation (*right*) compared to Gaussian distributions obtained with standard deviations corresponding to 0% truncation (*top*), 1.5% truncation (*middle*) and 5% truncation (*bottom*).

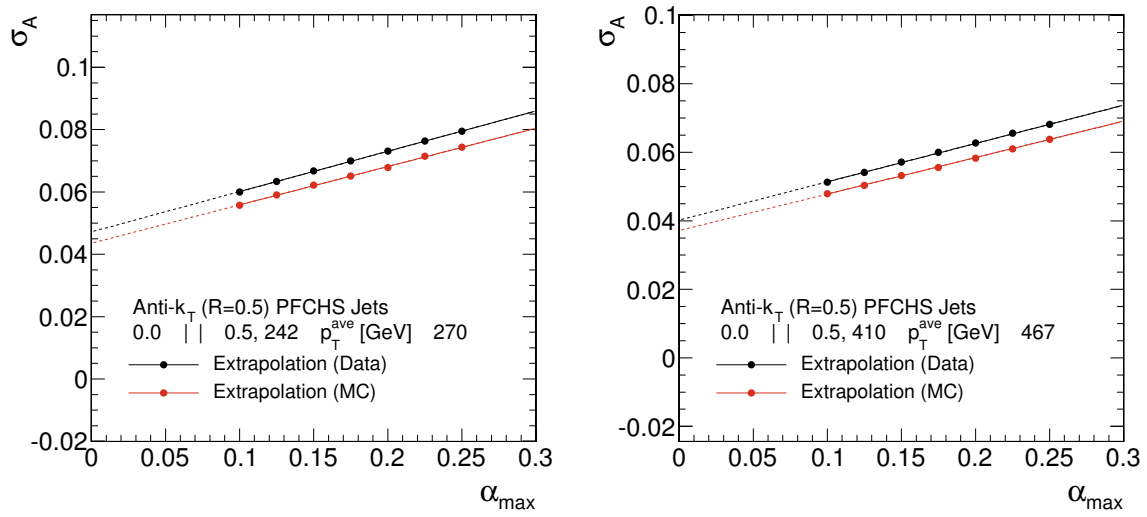

Figure 6.6: Two example extrapolations of measured values for σ_{A} in data (black) and simulation (red) to obtain the asymmetry width for zero additional jet activity.

6.6.1 Correction for Additional Jet Activity

The imbalance contribution arising from additional jet activity is removed by an extrapolation procedure. As described in Sec. 6.4.2, the asymmetry distribution is calculated for several intervals in $|\eta|$ and $p_{\mathrm{T}}^{\mathrm{ave}}$ with different selections on the maximum value of α ($= \alpha_{\mathrm{max}}$). For each of these selections, the width of the asymmetry is determined according to Eq. 6.9. The measured values of $\sigma_A(\alpha_{\mathrm{max}})$ are extrapolated to $\alpha_{\mathrm{max}} \to 0$ assuming a linear behaviour. The measured asymmetry widths in one particular ($p_{\mathrm{T}}^{\mathrm{ave}}$, $|\eta|$, α)-interval are fitted with a linear function and the y-intercept of the fitted linear function represents the resolution without further jet activity in the event. The statistical uncertainty is given by the respective fit uncertainty of the intercept.
As stated in Sec. 6.4.2, the selection is performed in inclusive intervals of $\alpha < \alpha_{\mathrm{max}}$. This results in a correlation of the measured values of σ_{A} for a particular ($p_{\mathrm{T}}^{\mathrm{ave}}$, $|\eta|$)-interval. In order to obtain a proper estimate of the statistical uncertainty, the correlation is propagated to the extrapolation fit. This approach is new with respect to previous analyses in which such correlations have not been considered in the extrapolation procedure [133, 147, 155]. More specifically, this means that the measured data points are described by a linear function

$$a \cdot \alpha_{\mathrm{max}} + b = \sigma_{\mathrm{A}}(\alpha_{\mathrm{max}}) \tag{6.11}$$

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

$$\chi^2 = dy^T \cdot C^{-1} \cdot dy \tag{6.12}$$

with $dy = y_{\mathrm{measured}} - y_{\mathrm{predicted}}$ and the covariance matrix C. Consequently, the asymmetry width for vanishing additional jet activity is given by $b = \sigma_{\mathrm{A}}(\alpha_{\mathrm{max}} \to 0)$. With the assumption that all events that belong to α interval i are also completely included in the

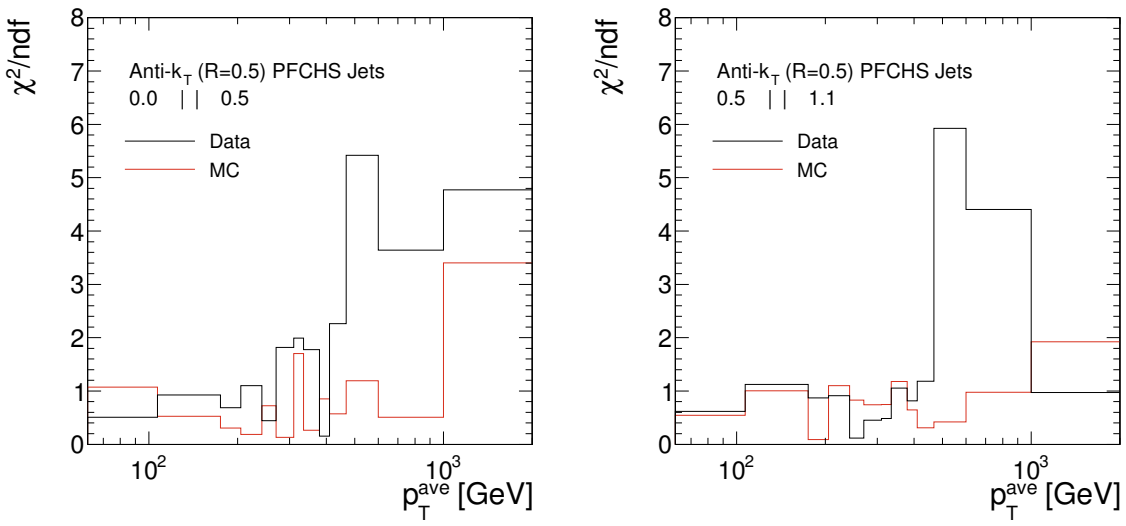

Figure 6.7: Goodness-of-fit test for the extrapolation fits in data and simulation in the two central $|\eta|$ regions as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$.

next higher α interval j, the covariance matrix is given by

$$C_{i,j}(\sigma_{A_i}, \sigma_{A_j}) = (\Delta\sigma_{A_i})^2 \cdot \frac{\sigma_{A_i}}{\sigma_{A_j}} \cdot \frac{n_i}{n_j} \tag{6.13}$$

with the number of events n in that particular α interval.[2] A discussion regarding the derivation of that expression is summarized in App. A.3 [161]. The function minimization itself is performed with the Minuit package [162]. The extrapolation procedure for two exemplative $|\eta|$ and $p_{\mathrm{T}}^{\mathrm{ave}}$ intervals is illustrated in Fig. 6.6. The performed extrapolation fits for all other non-empty intervals are shown in App. A.2.
In Fig. 6.7, an overview of a goodness-of-fit test is shown for the extrapolation fits in data and simulation. The resulting values for χ^2 over the number of degrees of freedom are shown as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$ for the two central $|\eta|$ intervals. The fit quality is in general quite good, but worsens for larger $p_{\mathrm{T}}^{\mathrm{ave}}$ values. Since especially for higher $p_{\mathrm{T}}^{\mathrm{ave}}$ intervals in the central $|\eta|$ regions, the statistical uncertainty is small, the fit is very sensitive to even small deviations from a linear behaviour which results in a minor fit quality. A similar behaviour is observed also for other $|\eta|$ intervals. However, a possible non-linearity is considered in the systematic uncertainties of the measurement as discussed in Sec. 6.9.

6.6.2 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 for instance from out-of-cone showering. This additional imbalance contribution is referred to as *particle-level imbalance* (PLI) and estimated from simulation.

[2] The assumption that all events from α interval i are also included in α interval j is only almost true since the asymmetry distributions are truncated to reject non-Gaussian components and some events might not fulfill this criterion.

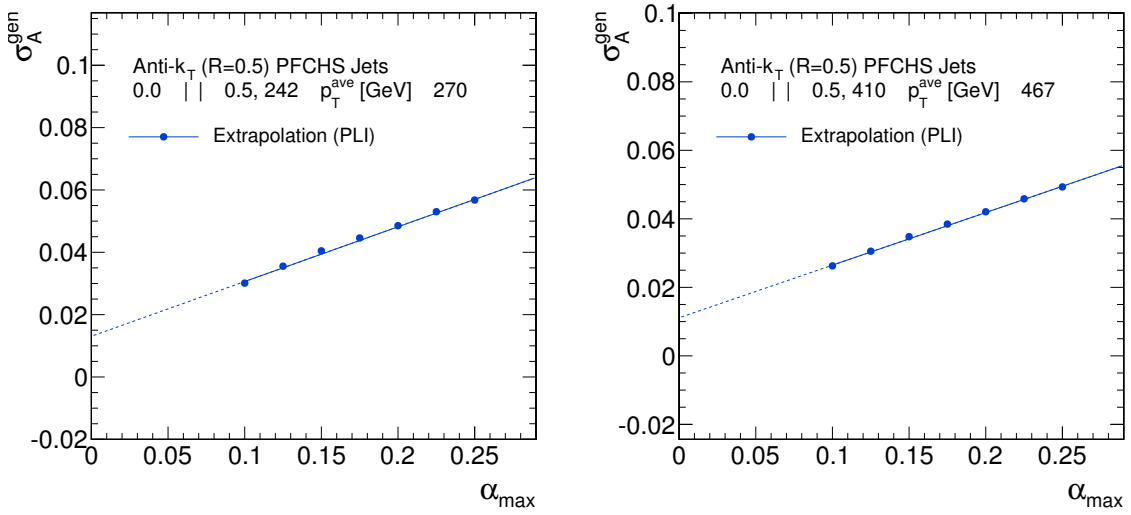

Figure 6.8: Two example extrapolations of measured values for $\sigma_{\mathrm{A}}^{\mathrm{gen}}$ in simulation to determine the contribution from PLI.

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

$$\mathrm{A}^{\mathrm{gen}} = \frac{p_{\mathrm{T},1}^{\mathrm{gen}} - p_{\mathrm{T},2}^{\mathrm{gen}}}{p_{\mathrm{T},1}^{\mathrm{gen}} + p_{\mathrm{T},2}^{\mathrm{gen}}} \tag{6.14}$$

with $p_{\mathrm{T},1}^{\mathrm{gen}}$ and $p_{\mathrm{T},2}^{\mathrm{gen}}$ referring to the momenta of the two leading generated jets. This distribution is affected by additional parton radiation as well so that the same procedure to obtain $\sigma_{\mathrm{A}}^{\mathrm{gen}}(\alpha_{\mathrm{max}} \to 0)$ is used as for the asymmetry at detector level. The generator asymmetry is calculated in the same ($p_{\mathrm{T}}^{\mathrm{ave}}$, $|\eta|$, α)-intervals as the detector-level asymmetry in order to derive the size of the particle-level imbalance for the exact same events. The asymmetry width of the generator asymmetry is also calculated as $\mathrm{RMS}_{98.5\%}$. In order to obtain the values of $\sigma_{\mathrm{A}}^{\mathrm{gen}}$ for zero additional jet activity, an analogue extrapolation procedure is performed as described in Sec. 6.6.1. Some example extrapolations are illustrated in Fig. 6.8.
The results of the extrapolated values for $\sigma_{\mathrm{A}}^{\mathrm{gen}}$ can be used to quantify the size of the particle-level imbalance. This is given by

$$\sigma_{\mathrm{PLI}} = \sqrt{2} \cdot \sigma_{\mathrm{A}}^{\mathrm{gen}}(\alpha_{\mathrm{max}} \to 0)\,. \tag{6.15}$$

6.6.3 Results of the Corrections to the Asymmetry

The results of the various extrapolation fits to determine $\sigma_{\mathrm{A}}(\alpha_{\mathrm{max}} \to 0)$ in data and simulation as well as for the particle-level imbalance are shown together in Fig. 6.9 as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$ for two different $|\eta|$ regions.
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

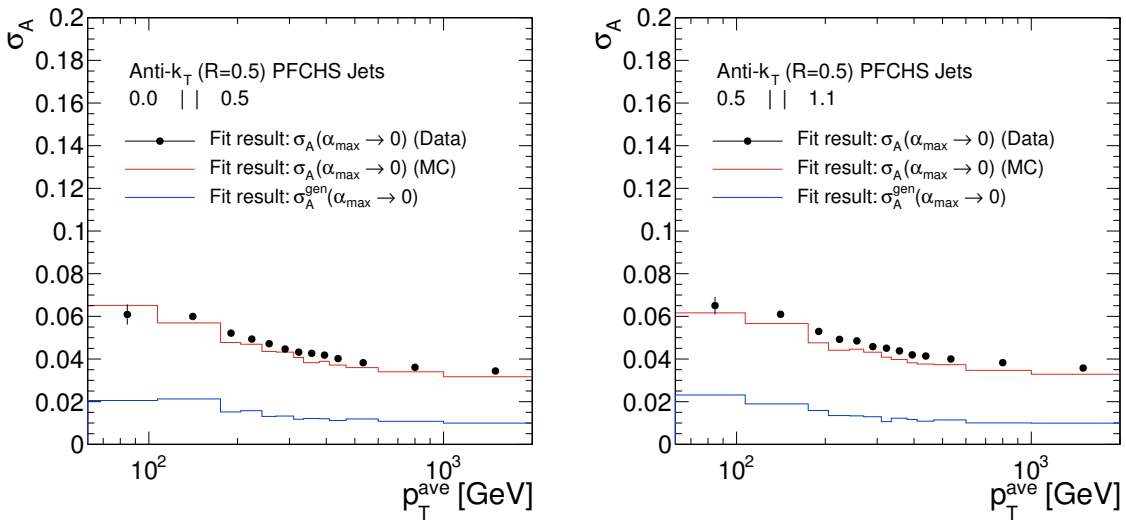

Figure 6.9: Results of extrapolation fits in the two central $|\eta|$ regions as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$ for data (black), simulation (red) and particle-level imbalance (blue).

by subtracting the PLI correction in quadrature from the measured σ_{A} after the correction for additional jet activity

$$\sigma_{\mathrm{JER}} = \sqrt{2} \cdot \sigma_{\mathrm{A}}(\alpha_{\mathrm{max}} \to 0) \ominus \sigma_{\mathrm{PLI}} \,. \tag{6.16}$$

The same PLI correction is substracted from the data fit results as well as from the fit results obtained in simulated events. As can be seen in Fig. 6.9, the correction due to particle-level imbalance is small compared to the measured asymmetry widths. In order to account for a possible imprecise modelling of the particle-level imbalance contribution, a systematic uncertainty will be considered as discussed in Sec. 6.9.

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

After applying all corrections described in Sec. 6.6, the measured resolution in data and simulation can be compared by calculating the data-to-simulation ratio

$$c(\mathrm{Data/MC}) = \frac{\sigma_{\mathrm{JER}}^{\mathrm{Data}}}{\sigma_{\mathrm{JER}}^{\mathrm{MC}}} \,. \tag{6.17}$$

The resulting distributions as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$ for the different $|\eta|$ regions are shown in Fig. 6.10. As no significant p_{T} dependence is observed, the ratio is parametrized by a constant fit. In each $|\eta|$ region, the fit result shows a value larger than one. This means that the resolution in data is in general worse than in simulation. The constant fit is visualized in Fig. 6.10 as a red line, together with the statistical uncertainty resulting from the fit (grey shaded area). The description with a constant seems justified as the resulting values of χ^2/ndf (also displayed in Fig. 6.10) are close to unity. A potential trend of the data-to-simulation ratio as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$ is considered as systematic uncertainty of the

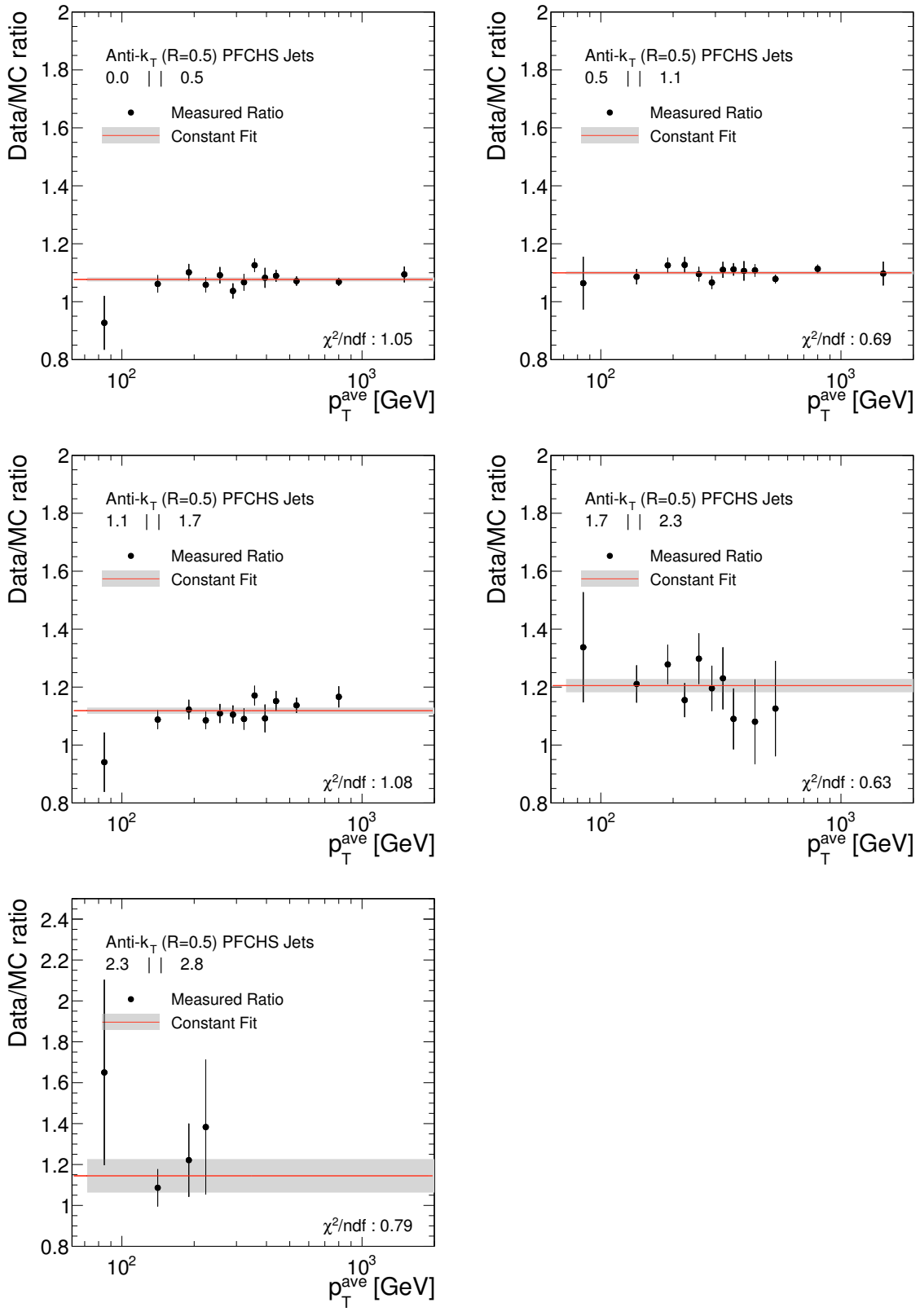

Figure 6.10: Ratio of the resolution in data to the resolution in the PYTHIA QCD sample in non-empty $|\eta|$ regions as a function of p_T^{ave} shown together with a constant fit (red line) and corresponding statistical uncertainty (grey shaded area).

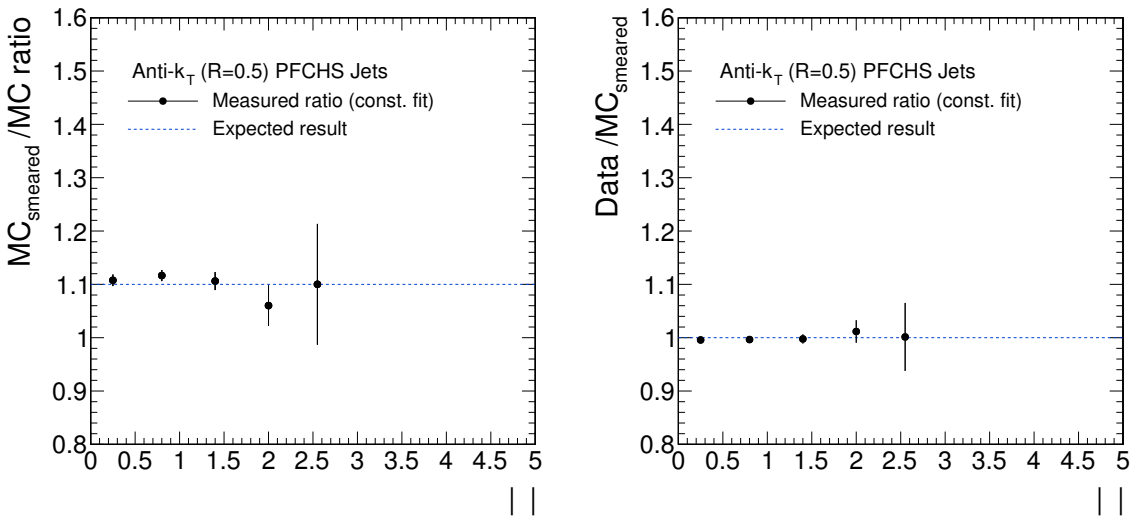

Figure 6.11: Ratio $c(\mathrm{MC_{smeared}}/\mathrm{MC})$ of the resolution in the smeared QCD sample to the resolution of the unsmeared QCD sample as a function of $|\eta|$ with statistical uncertainty (*left*) and ratio $c(\mathrm{Data}/\mathrm{MC_{smeared}})$ of the resolution in data to the resolution of the QCD sample smeared with the measured data-to-simulation ratio as a function of $|\eta|$ with statistical uncertainty (*right*). The blue dashed line indicates the target value of each validation test.

measurement. The determined data-to-simulation ratios are summarized below, together with statistical and systematic uncertainties, in Tab. 6.3. Because of 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 prescales of the triggers with low momentum thresholds.

6.8 Validation of the Method

6.8.1 Validation in Simulated Events

In order to test the quality of the method to predict the correct data-to-simulation ratio, a validation test based 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 jets in p_T in each event are smeared with a smearing factor c which increases the p_T resolution.[3] This smearing 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

$$p'_\mathrm{T} = p_\mathrm{T}^\mathrm{gen} + c \cdot (p_\mathrm{T} - p_\mathrm{T}^\mathrm{gen}) \,. \tag{6.18}$$

[3] It is necessary to smear all jets which might get relevant for the analysis, i. e. become one of the three leading jets. In this analysis, the five leading jets in p_T are smeared.

resulting in the smeared momentum p'_{T}. After the smearing procedure, the jet momenta are reordered descendent in p_{T}.
For this test, the smearing factor has been chosen to be $c = 1.1$ for each $|\eta|$ interval. 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 Sec. 6.6, it is expected that the resulting ratio resembles the input smearing factor c. Consequently, the test is passed when the measured ratio recovers the input value according to $c(\mathrm{MC_{smeared}}/\mathrm{MC}) = 1.1$. The resulting values obtained from a constant fit of the ratios as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$ are shown in each $|\eta|$ region with statistical uncertainties in Fig. 6.11 (left). It can be seen that the input scaling factor of 1.1 is recovered within the statistical uncertainties.

6.8.2 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 the previous section.
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 $c(\mathrm{Data}/\mathrm{MC})$ according to Eq. 6.18. Thus, jets are corrected with different scale factors depending on the respective $|\eta|$ region they belong to. After applying the smearing procedure to the jet momenta in simulation, the whole measurement procedure of the data-to-simulation ratio including the corrections for further jet activity and PLI is performed. Since the measured resolution differences between data and simulation have been compensated before the actual redetermination of the data-to-simulation ratio, it is expected to obtain ratios of $c(\mathrm{Data}/\mathrm{MC_{smeared}}) = 1$. The result of this test is illustrated in Fig. 6.11 (right) with statistical uncertainties. Good agreement with the expected value of 1 is achieved, even without the consideration of systematic uncertainties in this consistency test.

6.9 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 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

$$\delta c(\mathrm{Data}/\mathrm{MC}) = c(\mathrm{Data}/\mathrm{MC})_{\Delta} - c(\mathrm{Data}/\mathrm{MC}) \tag{6.19}$$

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

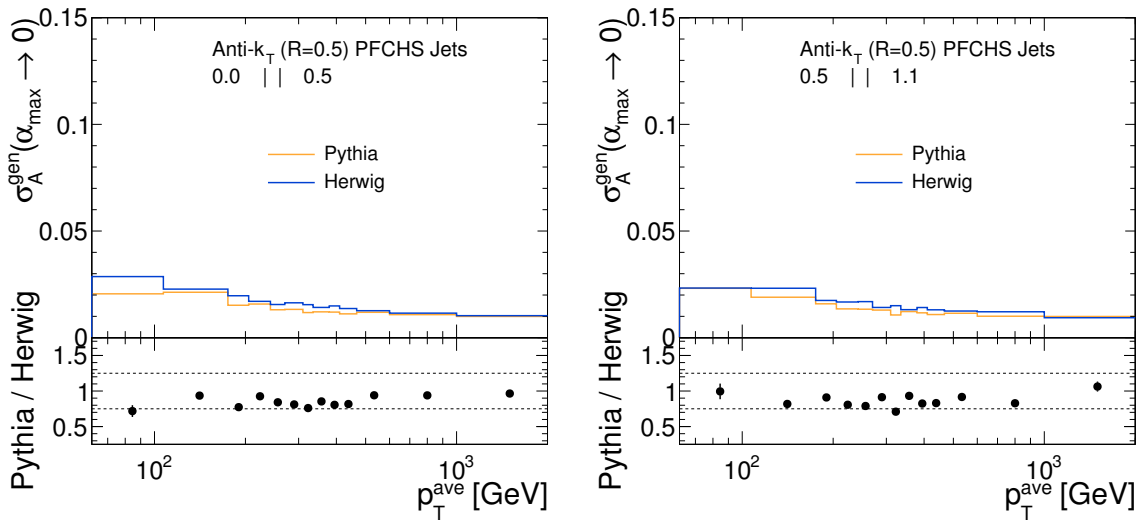

Figure 6.12: Comparison of the contributions to the asymmetry width due to particle-level imbalance in QCD multijet events generated with PYTHIA (orange) to the same quantity derived with the HERWIG++ generator (blue) in the two central $|\eta|$ regions as a function of p_T^{ave}. The two dashed lines in the ratio display a deviation of $\pm 25\%$.

case only an up- or downward variation is performed, the systematic uncertainty is taken symmetrically as the absolute shift.

PU reweighting: The trigger dependent pileup distributions in data, which are used to reweight the simulated sample to match the observed pileup 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 pileup 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.

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

$$c(\mathrm{Data/MC})_{\mathrm{PLI}} = \frac{\sqrt{2} \cdot \sigma_{\mathrm{A}}^{\mathrm{Data}}(\alpha_{\max} \to 0) \ominus f \cdot \sigma_{\mathrm{PLI}}}{\sqrt{2} \cdot \sigma_{\mathrm{A}}^{\mathrm{MC}}(\alpha_{\max} \to 0) \ominus f \cdot \sigma_{\mathrm{PLI}}} \tag{6.20}$$

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++

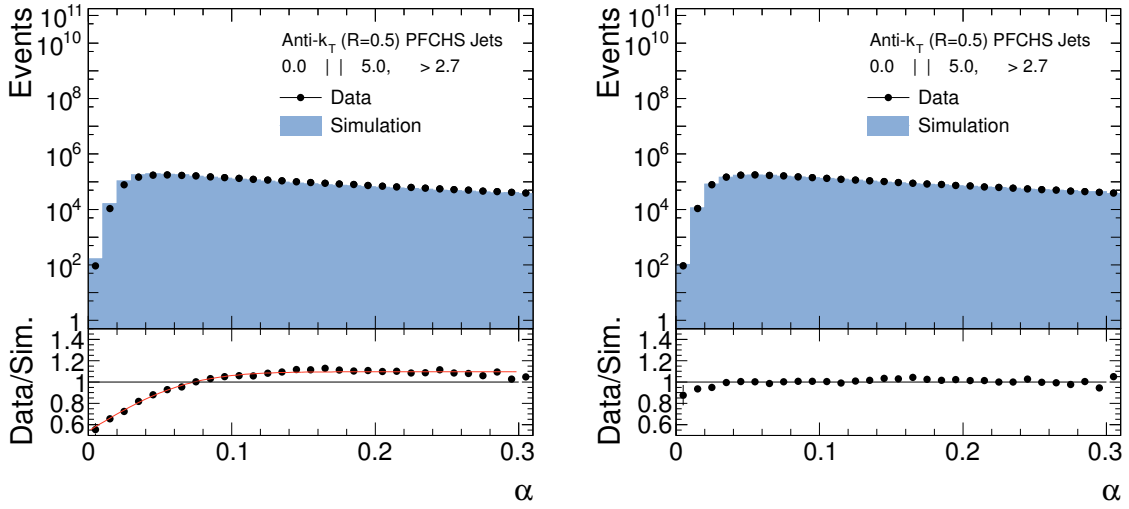

Figure 6.13: Inclusive α-spectrum before (*left*) and after (*right*) reweighting the α-spectrum of the simulated sample. The red curve in the bottom of the left figure illustrates the function used for the α-spectrum reweighting.

generator with tune EE3C. This comparison is illustrated in Fig. 6.12 for the two central $|\eta|$ regions. The size of the PLI correction agrees within 25% between both generators.

Jet energy scale: The jet energy scale has been corrected to particle level by the application of dedicated calibration factors. In order to propagate the uncertainty of the jet energy scale corrections 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 redetermined based on the varied jet momenta.

α-spectrum: The observed inclusive α-spectrum in the simulated sample is compared to the spectrum in data and shown in Fig. 6.13 (left). The bottom part displays the ratio Data/Sim. Since the ratio shows that the spectra do not agree, the influence of the α-spectrum on the data-to-simulation ratio is evaluated by reweighting the α-spectrum in simulated events to 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

$$w(\alpha) = 0.545 \cdot (\mathrm{erf}(13.5 \cdot \alpha - 0.02) + 1)) \tag{6.21}$$

with the error function erf[4]. This weight is considered as multiplicative factor onto the usual event weight in simulation. For comparison, the reweighted α-spectrum is shown in the right part of Fig. 6.13.

[4] $\mathrm{erf(x)} = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt$

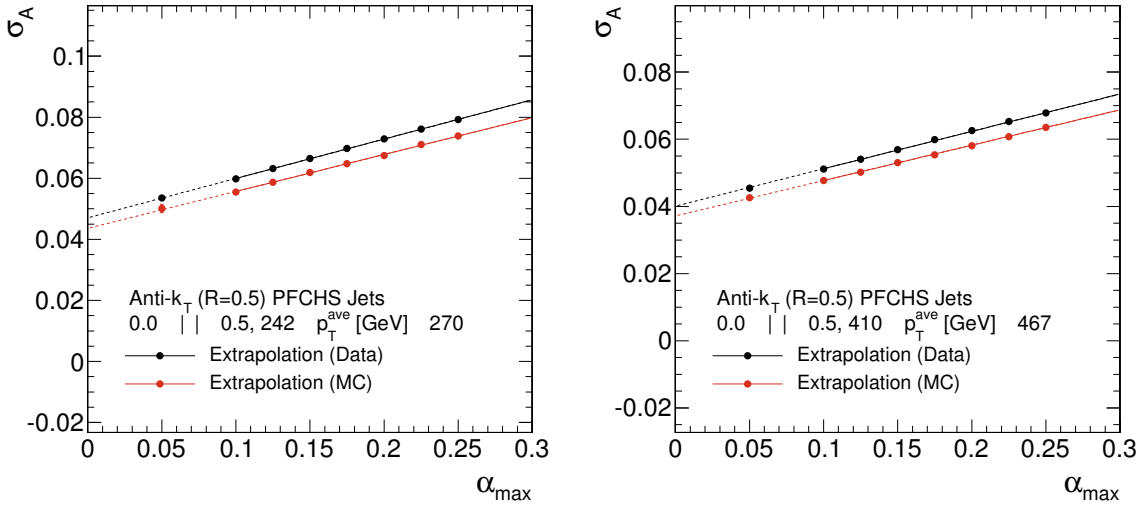

Figure 6.14: Two example extrapolations of measured values for σ_{A} in data and simulation to obtain the result at zero additional jet activity when adding one additional α interval 0.0–0.05.

α-extrapolation: 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. The choice of the linear function implies that the linear behaviour holds also for small values of α. However, especially for small values of α the linear behaviour could not be tested explicitly because of 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. 6.14, some example 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. 6.5, 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%.

Flavour uncertainty: As discussed in Sec. 6.1, the jet response can be quite different for light-flavour (u, d, s, g) and heavy-flavour jets (c, b). If the flavour composition is the same in data and simulation, this flavour difference cancels out in the data-to-simulation ratio. However, the rate of gluon splitting events with $g \to b\bar{b}$ is known to

be not well modelled [163]. 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-flavour 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. The event weight of all events identified as gluon splitting into heavy-flavour quarks is varied by +50%. 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: In general, the data-to-simulation ratio is determined by fitting the ratio of $\sigma_{\mathrm{JER}}^{\mathrm{Data}}$ to $\sigma_{\mathrm{JER}}^{\mathrm{MC}}$ with a constant. This assumes that the ratio is flat as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$. In order to test alternatives to this assumption, one can fit the $\sigma_{\mathrm{JER}}^{\mathrm{Data}}$ and $\sigma_{\mathrm{JER}}^{\mathrm{MC}}$ distributions with a more model dependent approach.
The following function, denoted *NSC-function*, is used to fit the resolution in simulation

$$f(p_{\mathrm{T}}) = \sqrt{\frac{N^2}{p_{\mathrm{T}}^2} + \frac{S^2}{p_{\mathrm{T}}} + C^2} \tag{6.22}$$

with $p_{\mathrm{T}} = p_{\mathrm{T}}^{\mathrm{ave}}$ 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 (cf. Sec. 3.2). If the resolution in simulation is fitted with this NSC-function, it is expected to find one common scale factor k_{NSC} for the N, S and C parameters to describe the data in case the data-to-simulation ratio is flat as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$. Significant differences between data and the actual modelling of the detector behaviour consequently can result in the necessity to use individual scale factors k_N, k_S and k_C for the parameters N, S and C in order to describe the data based on 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

$$f(p_{\mathrm{T}}) = \sqrt{\frac{(k_{NS} \cdot N)^2}{p_{\mathrm{T}}^2} + \frac{(k_{NS} \cdot S)^2}{p_{\mathrm{T}}} + (k_C \cdot C)^2} \tag{6.23}$$

with parameters N, S and C fixed to the fit results from simulation. This function makes use of 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 effects from pileup which are expected to mainly influence N.
Some example fits for the two most central $|\eta|$ intervals are summarized in Fig. 6.15. 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 a function of $p_{\mathrm{T}}^{\mathrm{ave}}$ 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. 6.16. The ratio derived from the NSC-fits is also compatible with the measured values for the ratio. The difference of the measured scale factors k_{NS} and k_C to the central value of the measurement from the constant

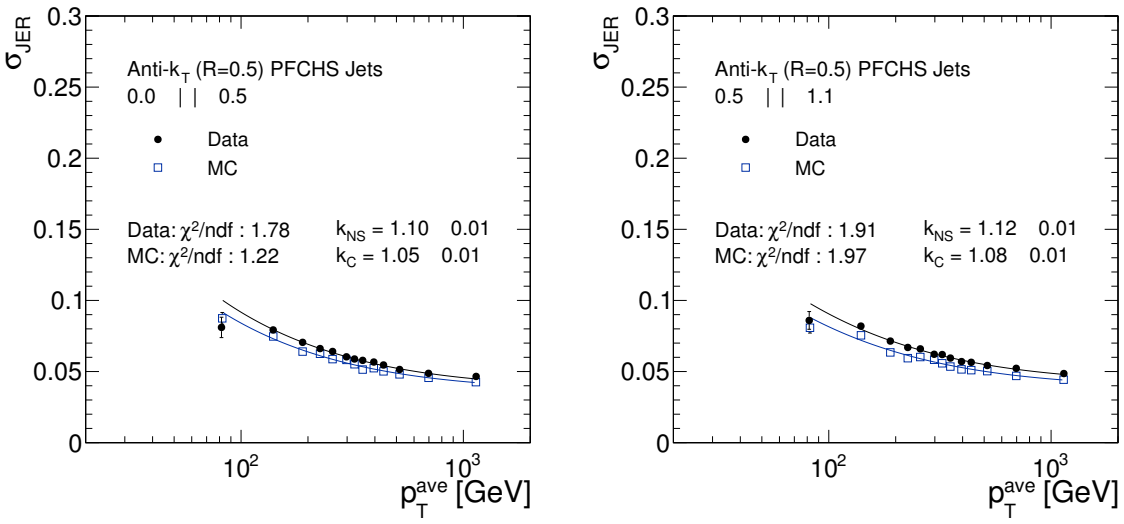

Figure 6.15: Results of fitting the resolution in data and simulation with the respective NSC-functions in the two central $|\eta|$ intervals.

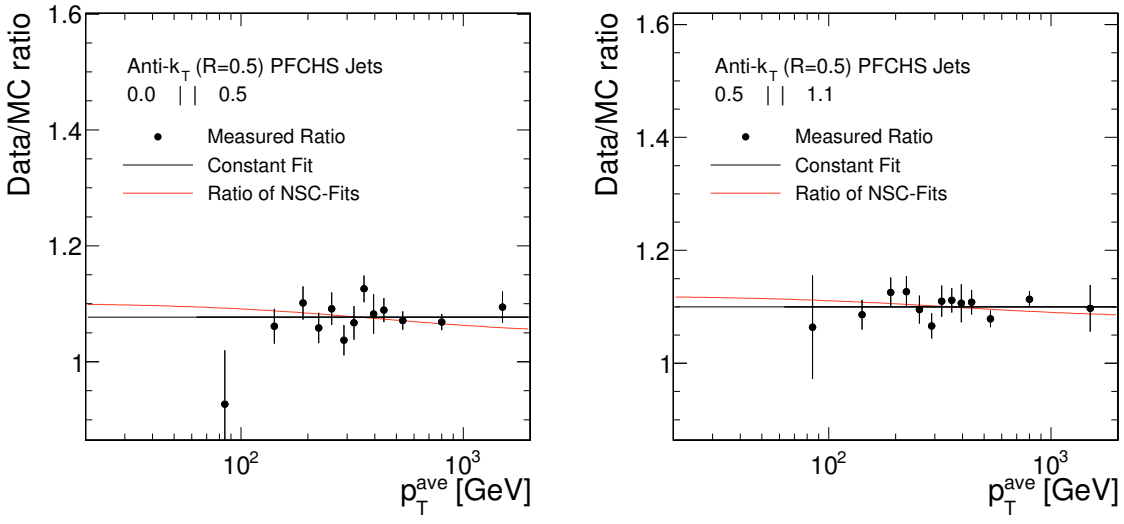

Figure 6.16: Ratio of the results fitting the resolution in data and simulation with the NSC-functions in the two central $|\eta|$ intervals together with the measured ratio and a constant fit.

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. 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. 6.3. Because of 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 $|\eta| \in [2.3, 2.8]$ region the measurement is limited by the statistical uncertainty.

Table 6.3: Summary of the measured data-to-simulation ratios c(Data/MC) with absolute 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(Data/MC)	1.077	1.100	1.119	1.205	1.145
Stat. uncertainty	± 0.007	± 0.006	± 0.010	± 0.027	± 0.078
PU	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-extrapolation	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

6.10 Extension of the Method to the Forward Detector Region

The measurement presented above is based on the requirement that the two leading jets in an event both belong to the same $|\eta|$ region. Consequently, the resolution of these two jets

is equal which results in the simple relation expressed in Eq. 6.4 between asymmetry width and jet resolution. The requirement that both leading jets have to belong to the same $|\eta|$ interval reduces the available number of events significantly. This poses a problem, especially in the forward region of the detector, as the jets in these events have low transverse momenta and consequently have to be triggered by the highly prescaled lowest $p_{\mathrm{T}}^{\mathrm{ave}}$ triggers. Hence, the available number of events is small and for the intervals $|\eta| > 2.8$ no data-to-simulation ratio could be determined with this approach.
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 both leading jets have to belong to the same $|\eta|$ interval has to be dropped. Instead, events are selected in which the two leading jets belong to different pseudorapidity regions. The jet resolution $\sigma(p_{\mathrm{T}}^{\mathrm{probe}})$ in a probe interval $|\eta_{\mathrm{probe}}|$ can be determined if the resolution $\sigma(p_{\mathrm{T}}^{\mathrm{ref}})$ in a reference interval $|\eta_{\mathrm{ref}}|$ is known.
Based on Eq. 6.3, it can be shown that for $\langle p_{\mathrm{T}}^{\mathrm{probe}} \rangle = \langle p_{\mathrm{T}}^{\mathrm{ref}} \rangle = \langle p_{\mathrm{T}} \rangle$ and $\sigma(p_{\mathrm{T}}^{\mathrm{probe}}) \neq \sigma(p_{\mathrm{T}}^{\mathrm{ref}})$ the relation

$$\frac{\sigma(p_{\mathrm{T}}^{\mathrm{probe}})}{\langle p_{\mathrm{T}} \rangle} = \sqrt{4 \cdot \sigma_{\mathrm{A}}^{\mathrm{ext}} - \left(\frac{\sigma(p_{\mathrm{T}}^{\mathrm{ref}})}{\langle p_{\mathrm{T}} \rangle} \right)^2} \tag{6.24}$$

is obtained. Here, $\sigma_{\mathrm{A}}^{\mathrm{ext}}$ 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 belongs to another $|\eta|$ region. The resolution of the jet in the reference interval has to be determined with the method imposing the same-$|\eta|$ requirement. The asymmetry width is meant to be corrected for additional jet activity and the particle-level imbalance determined from $\sigma_{\mathrm{A,gen}}^{\mathrm{ext}}$. The latter is the asymmetry width of the generator asymmetry calculated for events in which the two leading jets do not belong to the same $|\eta|$ interval. The reference interval is chosen to be in the central detector region where the statistical precision from the same-$|\eta|$ measurement already is sufficient. The statistical uncertainty of $\frac{\sigma(p_{\mathrm{T}}^{\mathrm{ref}})}{\langle p_{\mathrm{T}} \rangle}$ is propagated to the statistical uncertainty of $\frac{\sigma(p_{\mathrm{T}}^{\mathrm{probe}})}{\langle p_{\mathrm{T}} \rangle}$.
As long as both jets belong to the same $|\eta|$ region, residual effects from jet energy scale differences in data and simulation, where the mean of the asymmetry is shifted and not exactly located at zero, affect both jets the same way and should not have an impact on the resolution measurement. Since residual effects from the jet energy scale become more important if both jets belong to different probe and reference intervals, a slightly modified definition of the asymmetry of

$$\mathrm{A} = \frac{p_{\mathrm{T}}^{\mathrm{probe}} - p_{\mathrm{T}}^{\mathrm{ref}}}{p_{\mathrm{T}}^{\mathrm{probe}} + p_{\mathrm{T}}^{\mathrm{ref}}} \tag{6.25}$$

is used for the reference-and-probe-interval measurement. The transverse momenta $p_{\mathrm{T}}^{\mathrm{ref}}$ and $p_{\mathrm{T}}^{\mathrm{probe}}$ 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. The asymmetry width is calculated as

$$\sigma_{\mathrm{A}} = \mathrm{RMS}_{98.5\%} = \sqrt{\frac{1}{\sum_i y_i} \cdot \sum_i y_i (A_i - A_{\mathrm{mean}})^2} \tag{6.26}$$

Table 6.4: Summary of the forward extension measurements with different reference intervals showing the nominal data-to-simulation ratios c(Data/MC) with absolute statistical and total systematic uncertainty in different probe $|\eta|$ regions.

$\lvert\eta\rvert$	$\lvert\eta_{\text{ref}}\rvert \in [0.0, 0.5]$ c(Data/MC) $\pm$stat. $\pm$syst.	$\lvert\eta_{\text{ref}}\rvert \in [0.5, 1.1]$ c(Data/MC) $\pm$stat. $\pm$syst.	$\lvert\eta_{\text{ref}}\rvert \in [1.1, 1.7]$ c(Data/MC) $\pm$stat. $\pm$syst.
$0.0 - 0.5$	—	$1.081 \pm 0.008 \pm 0.026$	$1.084 \pm 0.012 \pm 0.033$
$0.5 - 1.1$	$1.106 \pm 0.008 \pm 0.028$	—	$1.082 \pm 0.012 \pm 0.036$
$1.1 - 1.7$	$1.133 \pm 0.009 \pm 0.030$	$1.111 \pm 0.009 \pm 0.030$	—
$1.7 - 2.3$	$1.227 \pm 0.025 \pm 0.058$	$1.206 \pm 0.023 \pm 0.039$	$1.189 \pm 0.031 \pm 0.063$
$2.3 - 2.8$	$1.253 \pm 0.047 \pm 0.112$	$1.300 \pm 0.047 \pm 0.082$	$1.250 \pm 0.051 \pm 0.065$
$2.8 - 3.2$	$1.410 \pm 0.068 \pm 0.067$	$1.356 \pm 0.058 \pm 0.105$	$1.432 \pm 0.066 \pm 0.077$
$3.2 - 5.0$	$1.171 \pm 0.116 \pm 0.079$	$0.829 \pm 0.082 \pm 0.149$	$1.137 \pm 0.105 \pm 0.096$

with the frequency y_i of the individual asymmetry values A_i. The sum over i includes all values such that 98.5% of the total asymmetry distribution are covered symmetrically 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. This allows a measurement of the data-to-simulation ratio also in the $|\eta|$ intervals 2.8–3.2 and 3.2–5.0. In addition, it provides an enhancement of the available number of events for the innermost $|\eta|$ intervals and thus a further reduction of the statistical uncertainties.
The systematic uncertainties of the two new measurements are evaluated according to the same procedure as described above 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 σ_A^{ext} value, as well as for $\frac{\sigma(p_{\text{T}}^{\text{ref}})}{\langle p_{\text{T}} \rangle}$, for each uncertainty source. In general, the resulting uncertainties are of the same order as for the same-$|\eta|$ measurement.
A summary of the resulting data-to-simulation ratios determined in the measurements with the different reference intervals and the corresponding absolute statistical and total systematic uncertainties is shown in Tab. 6.4. A detailed overview of these measurements, listing all individual components of systematic uncertainties, is given in App. A.4.

6.11 Measurement for Simulated Events Obtained with the Herwig++ Generator

The data-to-simulation ratio has been determined with respect to simulated events obtained from the Pythia generator. However, other generators use for instance different fragmentation models and it is interesting to study if the determined data-to-simulation ratio is generator independent. For this study, the data-to-simulation ratio is derived for

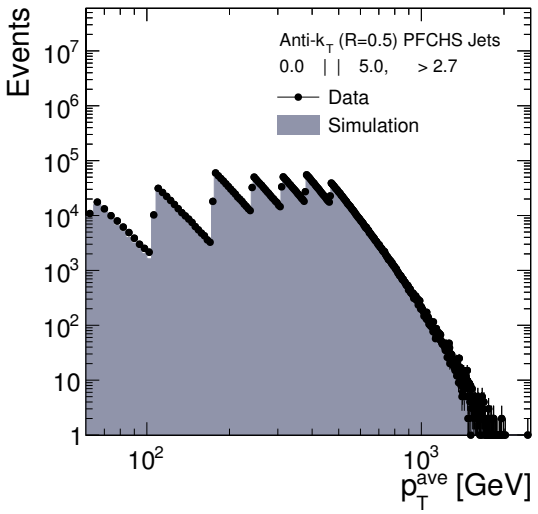

Figure 6.17: Inclusive $p_{\mathrm{T}}^{\mathrm{ave}}$ spectrum of events obtained after applying the event selection in data and in simulated events generated with HERWIG++.

Table 6.5: Measured data-to-simulation ratio in various probe $|\eta|$ regions with statistical uncertainty for simulated events obtained with the HERWIG++ generator.

$\lvert\eta\rvert$	$\lvert\eta_1\rvert = \lvert\eta_2\rvert$ c(Data/MC)	$\lvert\eta_{\mathrm{ref}}\rvert \in [0.0, 0.5]$ c(Data/MC)	$\lvert\eta_{\mathrm{ref}}\rvert \in [0.5, 1.1]$ c(Data/MC)	$\lvert\eta_{\mathrm{ref}}\rvert \in [1.1, 1.7]$ c(Data/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

simulated events not taken from the PYTHIA generator, but for events obtained from HERWIG++ tune EE3C. This is the same sample as introduced in Sec. 6.9 for the evaluation of the systematic PLI uncertainty.
The measurement is performed applying the selection criteria described in Sec. 6.4.2 and also the pileup reweighting for the pileup scenario of the HERWIG++ sample is performed following the procedure described in the same section. In Fig. 6.17, the inclusive $p_{\mathrm{T}}^{\mathrm{ave}}$ spectrum is compared in data and simulation for simulated events taken from HERWIG++. The agreement of the $p_{\mathrm{T}}^{\mathrm{ave}}$ spectrum between data and simulation looks reasonable and exhibits a similar trend as the spectrum in simulated events from PYTHIA.
The nominal values of the data-to-simulation ratio have been determined with statistical uncertainties with the same-$|\eta|$ requirement as well as for the three different reference-and-probe jet combinations. Since this study is only used as cross-check, no systematic uncertainties are 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. 6.5. In general, the obtained data-to-simulation ratios for HERWIG++ are quite similar to the values obtained from PYTHIA for most $|\eta|$ regions.

6.12 Results

6.12.1 Determination of a Combined Result

The data-to-simulation ratio has been measured with different approaches for various $|\eta|$ regions. Since these are in general in good agreement, the results obtained from the individual measurements are combined into one single scale factor for each $|\eta|$ region. This combination is derived from the measurements done with simulated events from PYTHIA, while the HERWIG++ results are only used as cross check. The only exception is the last $|\eta|$ region, ranging from 3.2–5.0, which is treated separately.
The combined data-to-simulation ratios are calculated as weighted mean from the individual measurements with the same-$|\eta|$ 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 the shifted ratios are combined as the weighted mean according to their statistical uncertainty and the difference of each systematically shifted weighted mean to the nominal weighted mean is the respective systematic uncertainty for the combined data-to-simulation ratio.
As stated above, the $|\eta|$ region covering the most forward region of the detector from 3.2–5.0, i. e. the hadronic forward, is treated differently than the other $|\eta|$ intervals. In contrast to 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 detector part. In order to find one common scale factor for this $|\eta|$ region, respresenting the various measured values, all six individual measurements from PYTHIA and HERWIG++ are combined. This is done following the procedure described above for the other intervals considering all nominal values obtained with PYTHIA and HERWIG++ and their statistical uncertainty. Since the systematic uncertainties for the HERWIG++ measurements have not been determined, the combined systematic uncertainty from the PYTHIA results is considered and a total scale factor of $c_{\mathrm{comb.}} = 1.056 \pm 0.048$ (stat. unc.) $\pm$ 0.079 (syst. unc.) is obtained. This

Table 6.6: Measured data-to-simulation ratio in various $|\eta|$ regions with statistical and systematic uncertainty as well as the total uncertainty.

$\|\eta\|$	c(Data/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

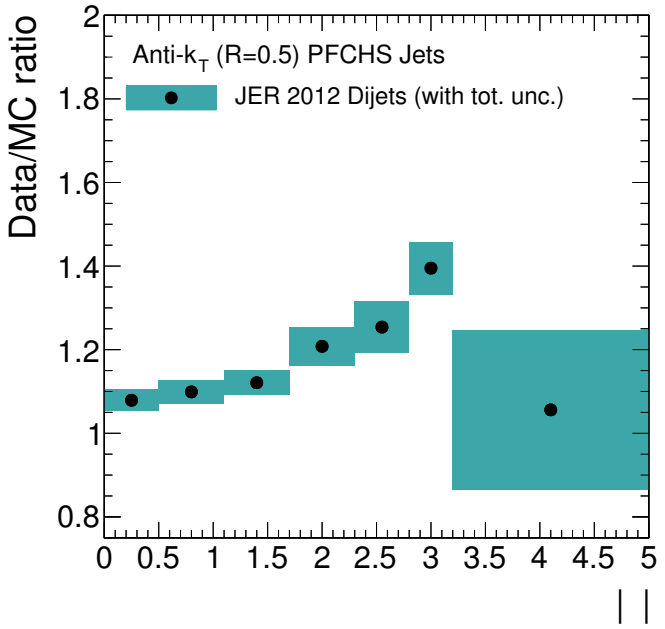

Figure 6.18: Measured data-to-simulation ratio in various $|\eta|$ regions displayed with total uncertainty.

corresponds to a total uncertainty of $\sigma^{\text{tot.}}_{c_{\text{comb.}}} = 0.092$. In order to test the consistency of this combined value with the single measurements, a χ^2/ndf is calculated according to

$$\chi^2/\text{ndf} = \left[\sum_i \left(\frac{c_i - c_{\text{comb.}}}{\sigma^{\text{tot.}}_{c_{\text{comb.}}}}\right)^2\right] / 5 \tag{6.27}$$

with the scale factors c_i of the six individual measurements i. This results in $\chi^2/\text{ndf} = 4.3$ and shows that the obtained total uncertainty is too small to reasonably cover the spread of the individual results. In order to obtain $\chi^2/\text{ndf} = 1.0$, the systematic uncertainty is increased to a value of $\pm$ 0.185. This gives a total uncertainty of $\pm$ 0.191 and ensures that the single measurements are well represented by the combined result as the observed differences are covered by the assigned uncertainty.
The obtained data-to-simulation ratios for the various $|\eta|$ regions are summarized in Tab. 6.6, together with the statistical and systematic uncertainties. In addition, Fig. 6.18 shows the measured ratios together with the total uncertainty obtained as the quadratic sum of the statistical and systematic uncertainties.
The determined ratios vary from 1.06 to 1.40 and exceed the value of one in all detector parts. Possible sources of this difference can be, e. g. mismodelled noise effects or inhomogenities of the detector. However, no distinct reason for the discrepancy of the resolution in data and simulation has been identified yet.

6.12.2 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 γ + 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 $\sqrt{s} = 7\,\text{TeV}$ [147] and results for γ + jet measurements at $\sqrt{s} = 8\,\text{TeV}$ [150] are shown in Fig. 6.19.
The results obtained from different methods and for the different centre-of-mass energies are well compatible. The main advantage of the dijet measurement presented in this thesis compared to the γ+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 better.
In comparison to the $\sqrt{s} = 7\,\text{TeV}$ results from dijet events, the total uncertainty could be significantly reduced in the measurement presented in this thesis 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 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.

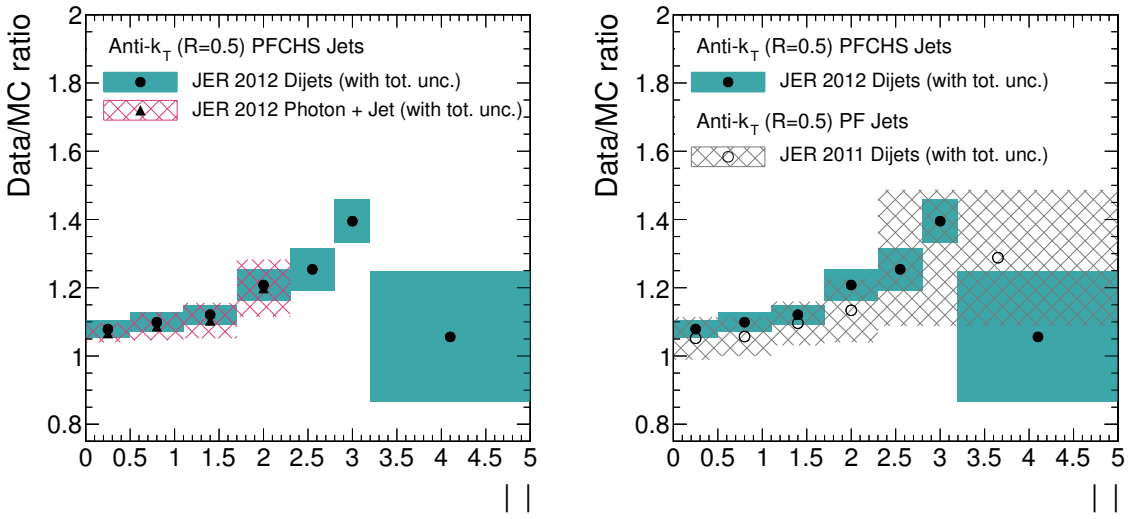

Figure 6.19: Measured data-to-simulation ratio in various $|\eta|$ regions displayed with total uncertainty. Comparison to results obtained for $\sqrt{s} = 8\,\text{TeV}$ data from γ+jet events (*left*) and to $\sqrt{s} = 7\,\text{TeV}$ data from dijet events (*right*).

6.12.3 Impact of the Improved JER Measurement

As mentioned in the introduction of this chapter, the jet transverse-momentum resolution is a key ingredient for several analyses. In this section, the impact of the improved JER measurement is discussed for two use cases.

First, the influence of the JER measurement on the jet energy corrections is examined. As the derived data-to-simulation factors for the resolution feature a dependence on $|\eta|$, it is important to propagate their uncertainty to the L2 residual correction of the jet energy scale. The impact of the JER uncertainty on the L2 residual correction is illustrated in Fig. 6.20.

The blue-shaded band denotes the uncertainty on the L2 residual correction caused by the uncertainty of the JER measurement from 2011 dijet data, as illustrated in Fig. 6.19 (right). Red and green dots illustrate the resulting up and down variation of the L2 residual correction when propagating the JER uncertainty as determined in the analysis presented in this thesis. The open and closed symbols represent two different methods used to derive the L2 residual correction, while the method denoted 'MPF' is the nominal one. Especially in the region up to around $|\eta| = 3.0$, the uncertainty is reduced by about 50%.

The improved jet resolution measurement has also a major impact on the determination of the top-quark mass. The currently most precise single measurement of the top-quark mass from CMS is performed in the lepton+jets channel and yields a top quark mass of $m_t = 172.04 \pm 0.19\ (\text{stat.} + \text{JSF}) \pm 0.75\ (\text{syst.})\,\text{GeV}$ [164]. The propagated uncertainty of the jet transverse-momentum resolution is among the dominant systematic uncertainties and amounts to 0.26 GeV when considering the data-to-simulation scale factors as measured with the 2011 dijet data. The improved uncertainty on the JER scale factors decreases the contribution to the systematic uncertainties of the top-quark mass measurement from 0.26 GeV to 0.11 GeV. Accordingly, the total systematic uncertainty drops from 0.75 GeV

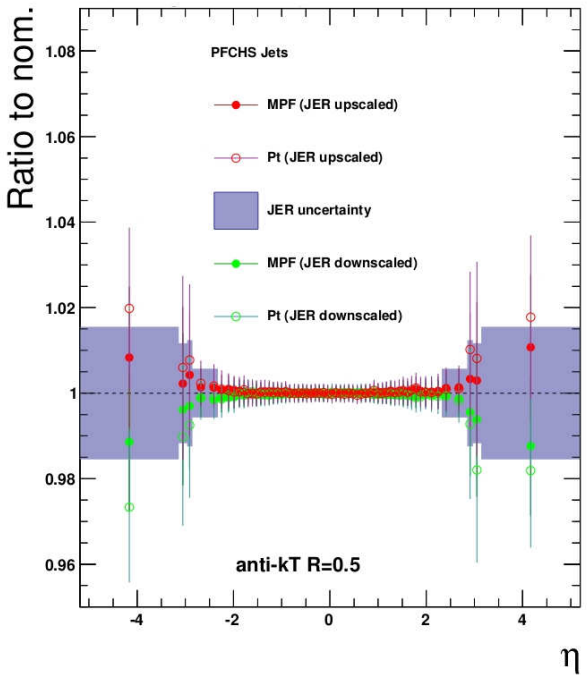

Figure 6.20: Impact of the JER uncertainty on the L2 residual correction. The variation of the correction with respect to the nominal correction factor is shown as a function of the pseudorapidity [158]. More information is given in the text.

to 0.71 GeV [165]. Consequently, the precision of the top-quark mass measurement in the above mentioned lepton+jets channel reaches the accuracy of the currently overall most precise single measurement of the top-quark mass which yields a total uncertainty of 0.76 GeV as performed by the DØ experiment [166].

6.13 Adjustment of the MC-Resolution to Data

The determined data-to-simulation ratios as given in Tab. 6.18 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. 6.8.2), 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 in Eq. 6.18.
Alternatively, the resolution in simulation can be adjusted by convolution with a Gaussian. Taking a Gaussian function with appropiate width σ_c, the width of the response function after convolution corresponds to the resolution in data according to

$$\sigma_c = \sqrt{c^2 - 1} \cdot \left(\frac{\sigma_{MC}}{p_{\mathrm{T}}} \right) . \tag{6.28}$$

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 Sec. 7.3 where this is applied in the estimation of background contributions arising from QCD multijet events to a search for new physics.

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

The search for supersymmetric particles is among the most important goals of the LHC physics program. As discussed in Sec. 2.3.3, 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. 7.1, the theory cross section for the production of supersymmetric particles is shown as a 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.

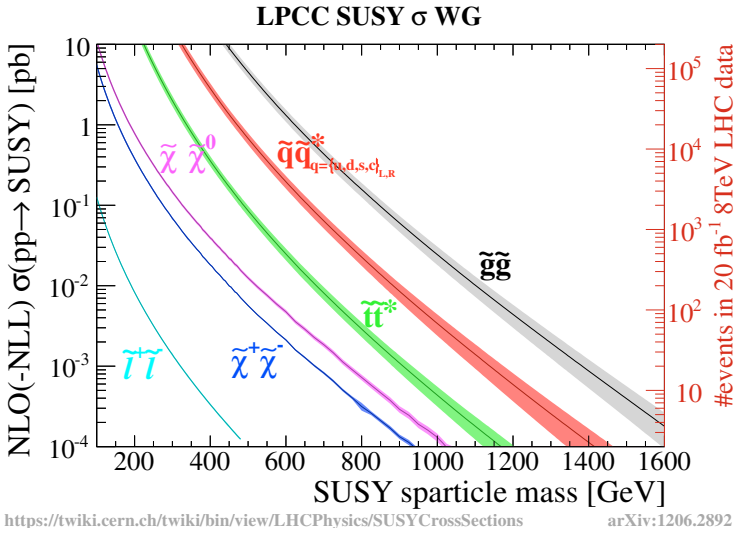

Figure 7.1: Theory cross sections for selected SUSY processes as a function of the sparticle mass. The y-axis on the right indicates the expected number of events in 20 fb^{-1} of pp collision data at the LHC at $\sqrt{s} = 8$ TeV [74].

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 searches for supersymmetric cascade decays arising from strongly produced light-flavour squarks or gluinos. As introduced in Sec. 2.3.3, the targeted experimental signature contains several hard jets, a certain amount of missing transverse energy and no isolated leptons. Thus, events are selected based on the number of jets (N_{Jets}), the scalar sum of the jet transverse momenta (H_{T}) and the missing transverse momentum calculated from the jet momenta ($\not{H}_{\text{T}}$). 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 in case the cross section of the process and the acceptance of the selection is large enough.
After the description of the event selection (Sec. 7.1), 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. 7.3) as this is performed in the context of this thesis. Finally, results are presented and interpreted in various simplified supersymmetric models (Sec. 7.4). Parts of this chapter are taken from [167], written by the author. This analysis follows previous inclusive searches [77, 78] and is published in [168].

7.1 Event Selection

7.1.1 Data Samples

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 $19.5\,\text{fb}^{-1}$ for all sub-detectors fully functional.
In addition, it is made use of several simulated samples describing SM background processes. These are especially employed in the validation of the background estimation methods described in Sec. 7.2 and 7.3. 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 MadGraph [117] generator at leading order and are interfaced with the parton-shower model in Pythia 6.4.24 [115]. While $t\bar{t}$ events are generated with up to three additional jets, other background samples contain up to four additional jets. The samples are scaled to cross-section predictions at next-to-leading order or next-to-next-to-leading order, when available [169, 170]. The events are processed with the full detector simulation.
Furthermore, SUSY signal samples are obtained from simulation. They are generated with MadGraph [117] (with up to two additional partons), the CTEQ6L parton distribution functions [152] and are 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 [74, 171–175]. 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.

7.1.2 Trigger

The data have been collected by triggering on H_T, the scalar sum of the jet transverse momenta, and $\not\!\!E_\text{T}$, 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. 7.1. H_T and $\not\!\!E_\text{T}$ 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 runs, also charged-hadron subtraction was applied to jets at trigger level.
In order to determine the offline values for H_T and $\not\!\!H_\text{T}$ (calculated according to the definition following in Sec. 7.1.4) for which the triggers reach the plateau efficiency, the trigger efficiencies are measured with respect to a single electron trigger (HLT_Ele27_WP80), i. e.

Table 7.1: Signal trigger paths used in different 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

it is tested how many events that are triggered by the reference electron trigger also pass the trigger under study. In principle, an independent trigger is desirable in order to get an unbiased estimate of the trigger efficiency. However, due to the PF algorithm all subdetectors are used simultaneously to reconstruct the particles in an event. Hence, no independent trigger path providing enough statistical precision is available. Consequently, only the reach and position of a plateau efficiency for a certain trigger path can be determined.
The determination of the relative trigger plateau efficiency is performed for different jet multiplicity intervals ($3 \leq N_{\mathrm{Jets}} \leq 5$, $6 \leq N_{\mathrm{Jets}} \leq 7$ and $N_{\mathrm{Jets}} \geq 8$). The obtained trigger turn-on curves for the two different HLT paths are shown as a function of H_{T} and $\not\!\!H_{\mathrm{T}}$, as used in the analysis, for jet multiplicity $3 \leq N_{\mathrm{Jets}} \leq 5$ in Fig. 7.2. The respective turn-on curves for jet multiplicities $6 \leq N_{\mathrm{Jets}} \leq 7$ and $N_{\mathrm{Jets}} \geq 8$ are shown in App. B.1 and App. B.2. For both trigger paths, the efficiency plateau is reached around values of $H_{\mathrm{T}} = 500\,\mathrm{GeV}$ and $\not\!\!H_{\mathrm{T}} = 200\,\mathrm{GeV}$.
The integrated trigger efficiencies for these particular values in different jet multiplicity intervals are summarized with statistical uncertainties in Tab. 7.2. In general, they are close to 100% with small uncertainties below 1%. However, for the highest jet multiplicity selection of $N_{\mathrm{Jets}} \geq 8$ only few events where selected such that statistical uncertainties are a 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 values of $H_{\mathrm{T}} > 500\,\mathrm{GeV}$ and $\not\!\!H_{\mathrm{T}} > 200\,\mathrm{GeV}$ independent of the jet multiplicity.

7.1.3 Event Cleaning

The analysis presented in this chapter 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 detect events in which protons from the beam interact with residual gas molecules in the

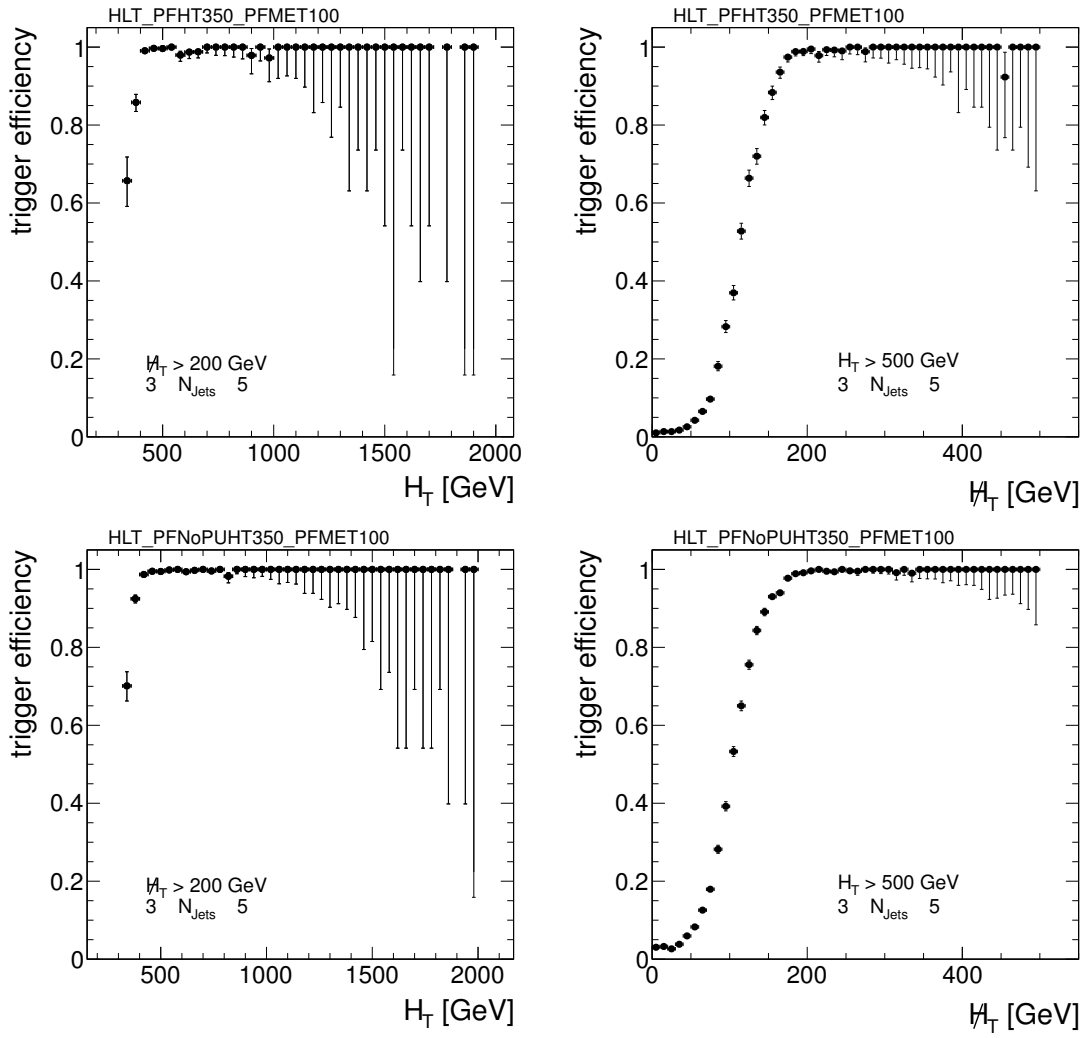

Figure 7.2: Measured relative trigger efficiency for paths HLT_PFHT350_PFMET100 (*top*) and HLT_PFNoPUHT350_PFMET100 (*bottom*) as a function of H_{T} (*left*) and $\not{H}_{\mathrm{T}}$ (*right*) shown for $3 \leq N_{\mathrm{Jets}} \leq 5$.

Table 7.2: Summary of total trigger efficiencies of the signal triggers for selections of $H_{\mathrm{T}} > 500\,\mathrm{GeV}$ and $\not{H}_{\mathrm{T}} > 200\,\mathrm{GeV}$ in different jet multiplicity intervals determined with respect to the reference trigger path.

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	$100.0^{+0.0}_{-36.9}$	$100.0^{+0.0}_{-10.9}$

beam pipe, the CSC subdetector is used to identify muons moving parallel to the beam and removing them accordingly (*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 artifical 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 result in fake $\not{H}_\mathrm{T}$ and 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 malfunctioning. These dead ECAL cells make up around 1% in total and can be responsible for energy losses resulting in large values of fake-$\not{H}_\mathrm{T}$. Such events can be identified by using the separate trigger primitive information of the L1 trigger to determine how much energy was lost (*TP filter*) or by using the energy of the cells surrounding the masked cells (*BE filter*).

Tracking Failure: In some events, the track reconstruction is observed to fail which manifests 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 centre 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 in which energy from energetic HCAL towers traversed by soft muons is wrongly associated to the muon momentum (*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, in which 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. Thus, 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 arising for instance from detector noise, events are discarded if the energy of a jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ is composed of more than 95% from PF photon candidates or more than 90% from PF neutral hadron energy (*PBNR filter*).

7.1.4 Baseline Selection

The physics objects used in the analysis are reconstructed with the PF algorithm. Jets are clustered from Particle-Flow objects with the anti-k_{T} algorithm using $R = 0.5$ including charged-hadron subtraction. Furthermore, they are calibrated, as discussed in Sec. 5.2.4, including residual correction factors for data.
The following *baseline selection* criteria are used to define the event sample used for the analysis. This selection defines a validation region and provides a basis for tighter criteria.

- The number of jets (N_{Jets}) is required to be ≥ 3. N_{Jets} is defined as the number of jets with $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ 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 $\geq 500\,\mathrm{GeV}$ with

$$H_{\mathrm{T}} = \sum_{\mathrm{jets}} p_{\mathrm{T}}$$

  for all jets that have $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ 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 ($\not{H}_{\mathrm{T}}$) is required to be $\geq 200\,\mathrm{GeV}$ with

$$\not{H}_{\mathrm{T}} = |\vec{\not{H}}_{\mathrm{T}}| = |-\sum_{\mathrm{jets}} \vec{p}_{\mathrm{T}}|$$

  for all jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 5.0$. This selection reduces contributions from standard model processes in which missing transverse momentum is expected to be small. In particular, QCD multijet background is suppressed.

- In order to suppress events in which missing transverse energy is mainly arising from jet mismeasurements, as for QCD multijet events, it is required that $\not{H}_{\mathrm{T}}$ is not aligned with any of the leading three jets and events with

$$\Delta\phi(\mathrm{jet}_n, \not{H}_{\mathrm{T}}) > 0.5 \;\text{ for }\; n = 1, 2 \;\text{ and }\; \Delta\phi(\mathrm{jet}_3, \not{H}_{\mathrm{T}}) > 0.3$$

  are selected. The value of 0.5 is chosen according to the jet distance 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+jets events are reduced by rejecting events containing isolated electrons or muons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$. These are required to have a good quality track that can be associated with the primary interaction vertex [176, 177]. The isolation is given as the scalar sum of transverse momenta of PF particles (except for the lepton itself) within a cone of radius $\Delta R = 0.3$ for

the electron and $\Delta R = 0.4$ for the muon, respectively. It 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.

After the application of the baseline selection, 26909 events are selected in data without the application of event cleaning filters. When imposing in addition the cleaning sequence, 11753 events constitute the event sample in data used for the analysis. This corresponds to a cleaning efficiency of around 56%. In simulated events, around 28% of the QCD multijet sample are rejected by the filters after the baseline criteria while it is less than 4% for other background contributions.
A comparison of the events selected by the baseline criteria, including cleaning filters, in data and simulation is shown in Fig. 7.3. In general, a reasonable agreement between data and MC distributions is observed. Especially in the bulk of the distributions, deviations are only at the order of 10–20%. However, 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 extreme tails of the H_{T}, $\not{H}_{\mathrm{T}}$ and N_{Jets} phase space with only few events and large uncertainties in the simulation, SM background contributions are estimated solely with data-based methods.

7.1.5 Exclusive Search Regions

As mentioned in the beginning, the analysis presented in this chapter is an extension of previous inclusive searches. These were based on the requirement of at least three jets in the final state. In this analysis, the data are further subdivided into three exclusive jet multiplicity categories: $3 \leq N_{\mathrm{Jets}} \leq 5, 6 \leq N_{\mathrm{Jets}} \leq 7$ and $N_{\mathrm{Jets}} \geq 8$. 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 improves 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 [178–181].
In order to gain sensitivity to a variety of models, the jet categories are further classified according to H_{T} and $\not{H}_{\mathrm{T}}$. 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 $\not{H}_{\mathrm{T}}$ is given in Tab. 7.3.

7.2 Estimation of Non-QCD Backgrounds

As discussed in Sec. 2.3.3, events from the SM processes $Z(\nu\bar{\nu}) + \mathrm{jets}$, $W + \mathrm{jets}$ or semi-leptonic $t\bar{t}$ events (in which either the lepton is lost or a hadronically decaying τ lepton) and mismeasured QCD multijet events constitute important background contributions to hadronic final states. In this analysis, dedicated data-based methods are employed to estimate their size in the selected data. The prediction of the QCD background has been performed as part of the work presented in this thesis. For completeness, also the techniques to estimate other backgrounds are introduced briefly.

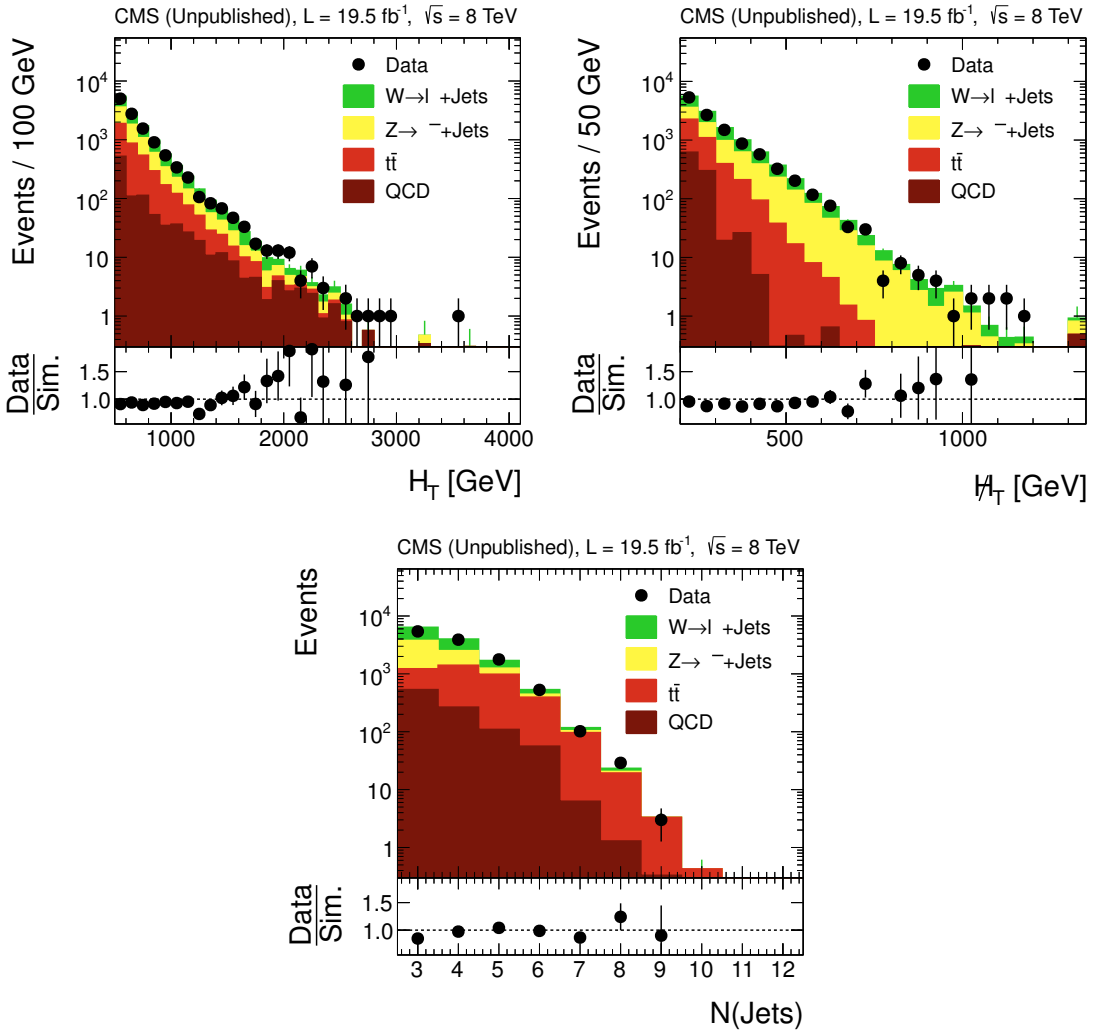

Figure 7.3: Comparison of selected H_{T} (*top left*), $\not{H}_{\mathrm{T}}$ (*top right*) and N_{Jets} (*bottom*) distributions in data (*black dots*) and simulated events (*shaded curve*) found from applying the event cleaning and baseline selection criteria described in Sec. 7.1.3 and 7.1.4. Only statistical uncertainties are shown. Taken from [168].

7.2.1 Invisible Z Background

The irreducible background contributions arising from $Z(\nu\bar{\nu}) + \text{jets}$ events are estimated using γ + jets events. This is a well-suited method as for high transverse momenta of the vector boson the event kinematics are the same and the cross sections differ mainly according to the different boson-quark couplings [182–184].
The γ+jets sample is collected by triggering on a γ candidate and large values of H_{T}. Photon candidates are selected if they satisfy $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ 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,

Table 7.3: Exclusive search regions used in the analysis binned in H_T, $\not{H}_T$ and N_{Jets}.

N_{Jets}	[3-5]	[6-7]	[≥8]
	$\not{H}_T$ [GeV]	$\not{H}_T$ [GeV]	$\not{H}_T$ [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	

photon candidates are required to be isolated meaning that in a cone of radius $\Delta R = 0.3$ 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 $\gamma +$ jets events is corrected for photon acceptance, reconstruction and isolation efficiency. Furthermore, the purity of the $\gamma +$ 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 for instance by misidentified jet fragments, is estimated by exploiting the difference between the shower profile of prompt and background photons. The average purity of the $\gamma +$ jets sample is measured to be 93%.
Subsequently, the number of $Z(\nu\bar{\nu}) +$ jets events in data $N^{data}_{Z(\nu\bar{\nu})+jets}(H_T, \not{H}_T, N_{Jets})$ using the number of selected $\gamma +$ jets events $N^{data}_{\gamma+jets}(H_T, \not{H}_T, N_{Jets})$ is obtained according to

$$N^{data}_{Z(\nu\bar{\nu})+jets}(H_T, \not{H}_T, N_{Jets}) = R^{MC}_{Z(\nu\bar{\nu})/\gamma}(H_T, \not{H}_T, N_{Jets}) \times N^{data}_{\gamma+jets}(H_T, \not{H}_T, N_{Jets}) \times \frac{R^{data}_{Z(\mu\mu)/\gamma}}{R^{MC}_{Z(\mu\mu)/\gamma}}$$

with the ratio relating the production cross section of $Z(\nu\bar{\nu}) +$ jets and $\gamma +$ jets events $R^{MC}_{Z(\nu\bar{\nu})/\gamma}(H_T, \not{H}_T, N_{Jets})$ determined in simulation and the double ratio $\frac{R^{data}_{Z(\mu\mu)/\gamma}}{R^{MC}_{Z(\mu\mu)/\gamma}}$ selected in data and MC to account for theoretical uncertainties on $R_{Z/\gamma}$ especially for high jet multiplicities [183, 184]. The missing momentum in the event is emulated by ignoring the

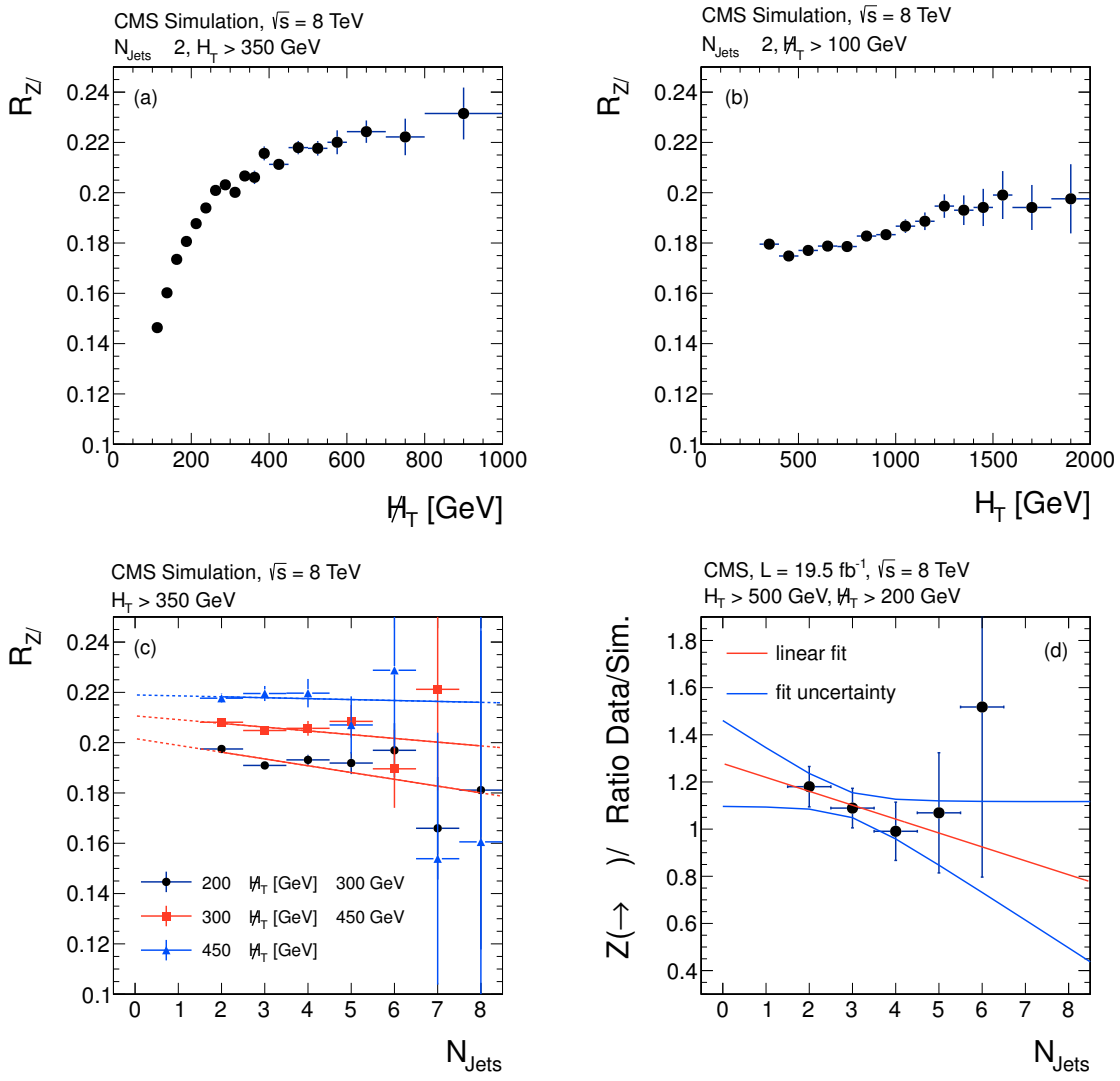

Figure 7.4: The simulated ratio $R_{Z/\gamma}$ as a function of (a) $\not{H}_{\mathrm{T}}$, (b) H_{T}, (c) N_{Jets}, where the values for three $\not{H}_{\mathrm{T}}$ selections are shown with linear fits, and (d) the double ratio of $R_{Z(\mu\mu)/\gamma}$, using events from data to those from simulation. The linear fit and its uncertainty band are overlaid [168].

momentum of the photon candidate in the calculation of $\not{H}_{\mathrm{T}}$.
The behaviour of $R_{Z/\gamma}$ is examined in simulated events as a function of $\not{H}_{\mathrm{T}}$, H_{T} and N_{Jets}. The obtained distributions are shown in Fig. 7.4 (a)–(c). While a strong dependence on $\not{H}_{\mathrm{T}}$ for small values ($\lesssim 500\,\mathrm{GeV}$) is observed, the variation as a function of H_{T} amounts to only $(12 \pm 5)\%$ in the relevant range for this analysis of $500 < H_{\mathrm{T}} < 1500\,\mathrm{GeV}$. The ratio as a function of the jet multiplicity is determined for different $\not{H}_{\mathrm{T}}$ ranges, which are $200 < \not{H}_{\mathrm{T}} < 300\,\mathrm{GeV}$, $300 < \not{H}_{\mathrm{T}} < 450\,\mathrm{GeV}$ and $\not{H}_{\mathrm{T}} > 450\,\mathrm{GeV}$. The behaviour in each of these $\not{H}_{\mathrm{T}}$ ranges is described by a linear function also displayed in Fig. 7.4 (c). It is found that the ratio decreases slightly with increasing jet multiplicity which is consistent with findings from theory [183, 184]. In order to take the theoretical uncertainty on $R_{Z/\gamma}$

obtained from simulated events into account, differences of this phenomenological ratio in data and simulated events are studied using $Z(\mu\mu)$ events. The double ratio of $R^{\text{data}}_{Z(\mu\mu)/\gamma}$ to $R^{\text{MC}}_{Z(\mu\mu)/\gamma}$ is derived as a function of the jet multiplicity and shown in Fig. 7.2.1 (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 three 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.

7.2.2 Hadronic-tau 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 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 isolated muon or a muon accompanied by at least two jets. Furthermore, instead of applying a lepton veto, exactly one μ with $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.1$ is required. In order to prevent the control sample from signal contamination, a selection on the transverse mass $m_{\text{T}} = \sqrt{2p^{\mu}_{\text{T}}\not{E}_{\text{T}}[1-\cos(\Delta\phi)]}$ of $m_{\text{T}} \leq 100\,\text{GeV}$ is imposed with the azimuthal angle $\Delta\phi$ between the direction of the muon four-momentum and the $\not{E}_{\text{T}}$ 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 $p^{\text{jet}}_{\text{T}}/p^{\tau}_{\text{T}}$, obtained from simulation, of a reconstructed jet with $p^{\text{jet}}_{\text{T}}$ matched to a generated hadronically decaying τ lepton with p^{τ}_{T}. Technically, sampling means that the four-momentum of the muon is scaled with the proper value from the τ_h response. However, if an event in the control sample is obtained from a prescaled trigger, the sampling is performed not only once but increased according to the prescale factor. For the calculation of the response, the generated τ lepton has to fulfill $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.1$ and the distance for the matching is chosen to be $\Delta R < 0.2$ for tau-$p_{\text{T}} < 50\,\text{GeV}$ 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 sample the complete response function, the random sampling of the response is repeated one hundred times for each event following a bootstrap method [185]. The actual prediction is given by the mean of the set of predictions from the bootstrapping and the statistical uncertainty is obtained from the standard deviation. Furthermore, also the statistical uncertainty of the seed sample is taken into account by considering the number of occurences of a seed event in the signal region.
In the following, $H_{\text{T}}, \not{H}_{\text{T}}$ 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. 7.1.4, are applied to the sample. The background contribution due to hadronic-tau 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 $W \to \tau_h\nu$ to $W \to \mu\nu$ events.

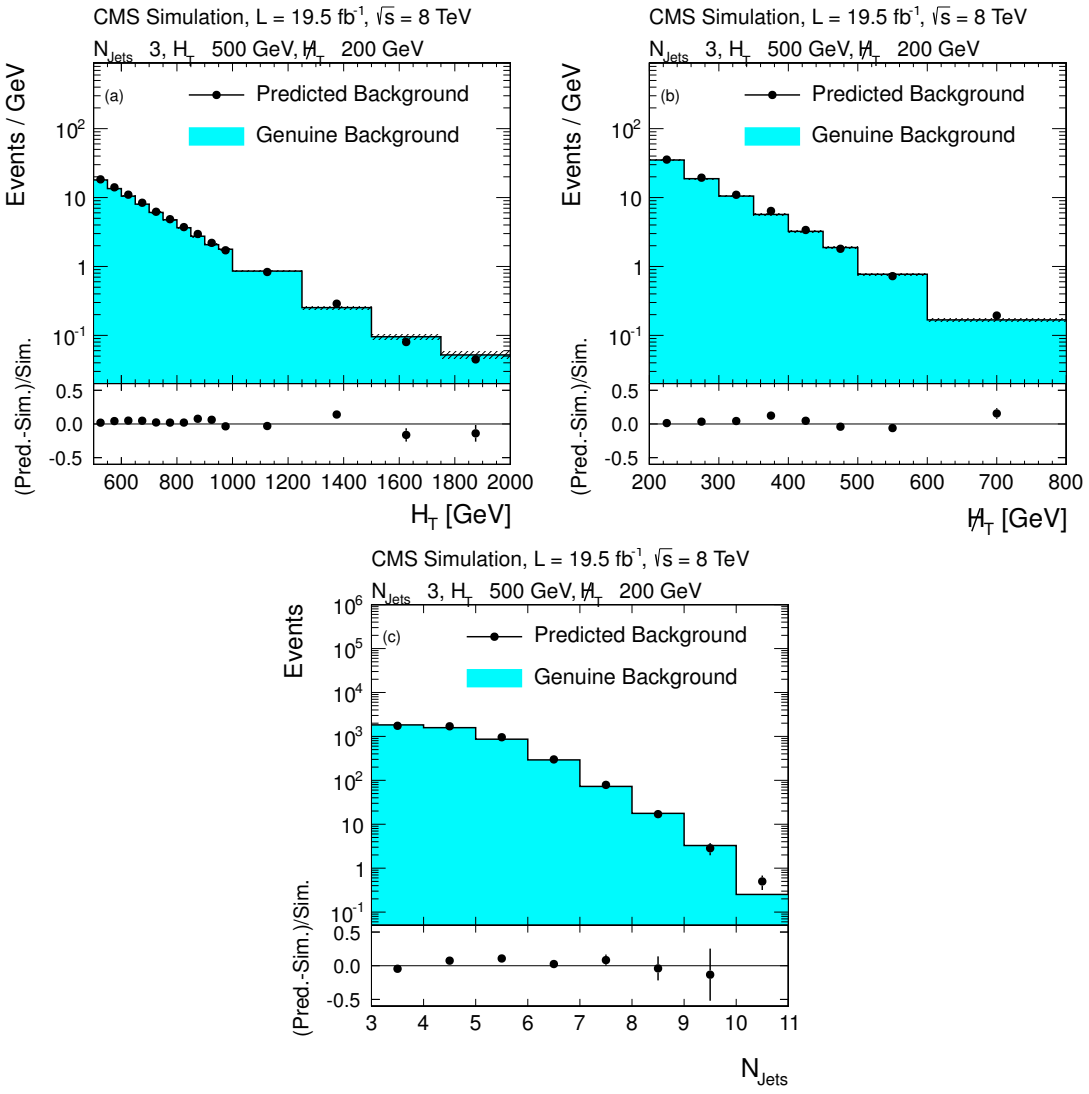

Figure 7.5: Predicted (a) H_{T}, (b) $\not\!\!H_{\mathrm{T}}$ and (c) N_{Jets} distributions found from applying the hadronic-tau background evaluation method to simulated $t\bar{t}$ and $W+\mathrm{jets}$ events (solid points) in comparison to the genuine $t\bar{t}$ and $W+\mathrm{jets}$ background from simulation (shaded curve). Only statistical uncertainties are shown [168].

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 $W+\mathrm{jets}$ events to the respective genuine background obtained from simulation. This comparison is shown as a function of H_{T}, $\not\!\!H_{\mathrm{T}}$ and N_{Jets} in Fig. 7.5 after the baseline selection. Although the agreement is quite reasonable and hence the method is observed to work reliable, uncertainties of 10% are considered for $3 \leq N_{\mathrm{Jets}} \leq 5$ and 20% for jet multiplicities $6 \leq N_{\mathrm{Jets}} \leq 7$ and $N_{\mathrm{Jets}} \geq 8$. These uncertainties mainly reflect the statistical precision of the validation test.
Further systematic uncertainties taken into account for the hadronic-tau background prediction cover differences between data and MC for the muon isolation and reconstruction

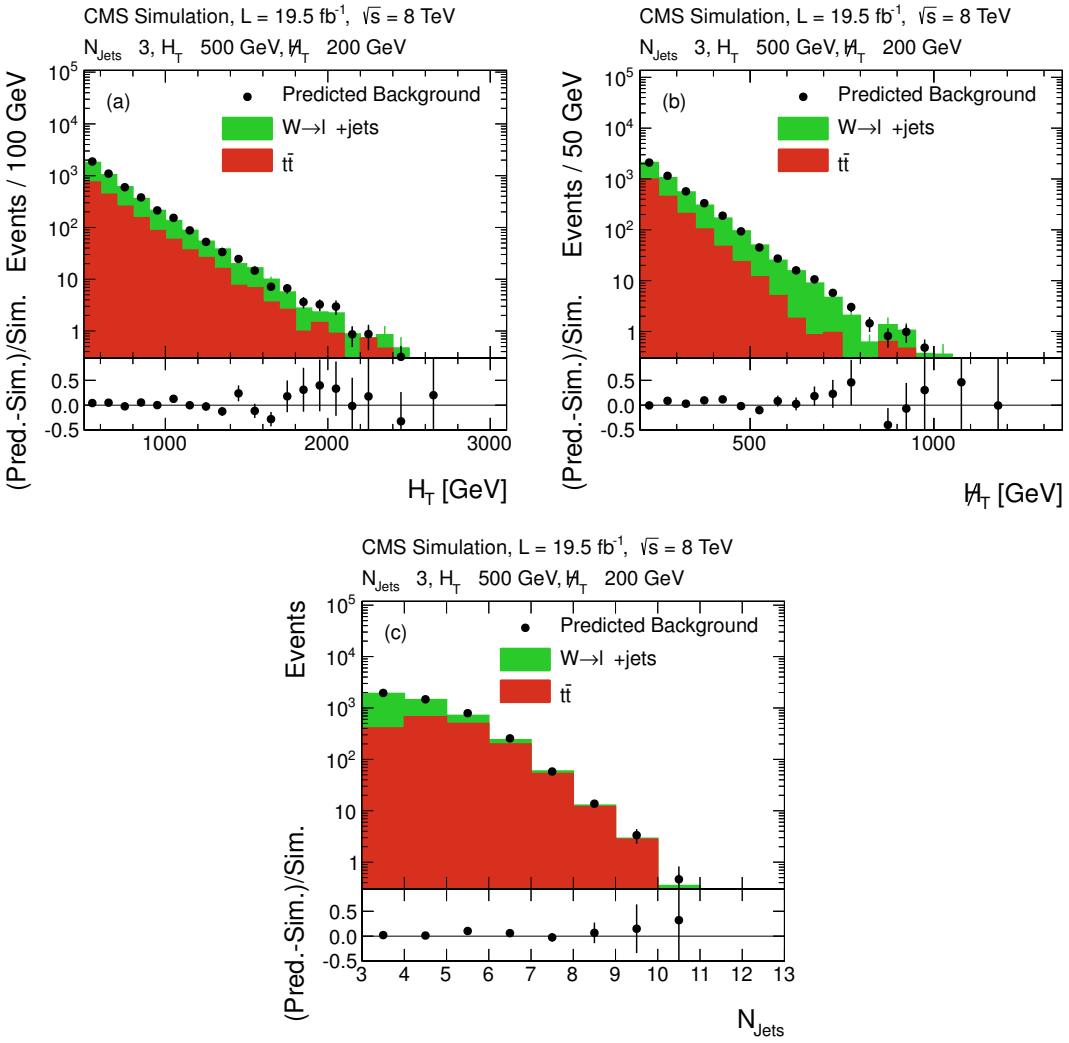

Figure 7.6: Predicted (a) H_{T}, (b) $\not\!\!H_{\mathrm{T}}$ and (c) N_{Jets} distributions found from applying the lost-lepton background evaluation method to simulated $t\bar{t}$ and $W+$jets events (solid points) in comparison to the genuine $t\bar{t}$ and $W+$jets background from simulation (shaded curves). Only statistical uncertainties are shown [168].

efficiencies as well as uncertainties on the kinematic and geometric acceptance, the τ_h jet response and the acceptance of the transverse mass cut.

7.2.3 Lost-Lepton Background

Similarly to background events from hadronic-tau decays, the background contribution due to a failed veto of a light lepton is estimated from a $\mu+$jets control sample. This is selected with the same trigger as used as signal trigger for the search. The sample is selected by requiring exactly one well-reconstructed and isolated muon with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$. Furthermore, the same transverse mass requirement of $m_{\mathrm{T}} < 100\,\mathrm{GeV}$, 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 $\epsilon^{e,\mu}_{\text{reco}}$ and isolation $\epsilon^{e,\mu}_{\text{iso}}$ inefficiencies 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 + jets events and determined in intervals of H_{T}, $\not{H}_{\text{T}}$ and N_{Jets}.
The number of events due to unidentified leptons is determined by weighting the events in the control sample according to

$$\frac{1}{\epsilon^{\mu}_{\text{iso}}} \times \frac{1-\epsilon^{e,\mu}_{\text{reco}}}{\epsilon^{\mu}_{\text{reco}}} \,.$$

To account for non-isolated leptons, events in the control sample are weighted with

$$\frac{\epsilon^{e,\mu}_{\text{reco}}}{\epsilon^{\mu}_{\text{reco}}} \times \frac{1-\epsilon^{e,\mu}_{\text{iso}}}{\epsilon^{\mu}_{\text{iso}}} \,.$$

The method is validated in simulation by comparing the predicted event yields for lost-lepton events in $t\bar{t}$ and W + jets from a single-muon control sample after the baseline selection to the simulated genuine background. This comparison is illustrated in Fig. 7.6 as a function of H_{T}, $\not{H}_{\text{T}}$ and N_{Jets} and shows a good overall agreement. An uncertainty of 15% is assigned to jet multiplicities 3–5 and 40% to other jet multiplicity selections in order to account for the statistical precision of this validation test.
Other uncertainties of the lost-lepton background prediction arise from a lack of events in the control sample in some search regions, 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.

7.3 QCD Background Estimation with the Rebalance-And-Smear Method

The background contribution which is most difficult to model in hadronic SUSY searches, is typically QCD background. This is caused by the fact that a precise description of the underlying particle-level jet spectrum and its manifestation in the detector is needed. 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 *Rebalance-and-Smear* (R+S) method was developed. It is based on the assumption that if the momenta of particle-level jets in an event are known, the reconstructed jet momenta can be modelled by a per-jet resolution function. This approach has been successfully used already in previous analyses [77, 78]. In this thesis, essentially improvements of the R+S method are discussed that became necessary to handle the main changes of this analysis when extending the search regions in several jet mulitplicity intervals at $\sqrt{s} = 8\,\text{TeV}$.
After the discussion of the general concept of the R+S method, the adjustment of the procedure to the actual conditions for 8 TeV data is introduced. This is followed by a description of the validation procedure of the method. Furthermore, the application of the R+S method to data is presented and systematic uncertainties are discussed. This section concludes with a presentation of the results of the QCD background prediction.

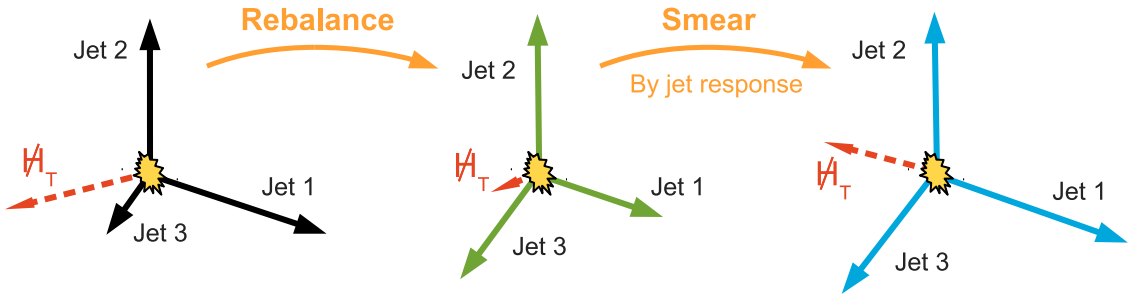

Figure 7.7: Outline of the two steps performed in the R+S method for estimation of QCD background events. Sketch from [147].

7.3.1 General concept of the Rebalance-and-Smear method

In general, QCD background contributions arise from jet mismeasurements in the detector. The main contributions stem from QCD multijet events which have in principle no intrinsic missing energy, except for neutrinos in jets arising for instance from electroweak decays of heavy-flavour quarks. Minor contributions originate also from fully-hadronic decaying $t\bar{t}$, $W+\text{jets}$ and $Z+\text{jets}$ events. Since QCD background is caused by jet mismeasurements, the general idea is to estimate this background contribution by emulating 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 idealised multijet events denoted *seed events* are obtained. These seed events are 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 to model the interaction of the multijet state with the detector. This means that all jet momenta are scaled with a factor randomly drawn according to the jet response distribution. Smeared events contain the whole information about event kinematics of the studied background contribution such that these can be used to derive contributions to various kinematic distributions, like H_T or $\not{H}_\mathrm{T}$. This is a main advantage of the R+S method compared to other QCD background estimation methods which typically predict event rates rather than event kinematics. The general outline of the R+S method is illustrated in Fig. 7.7.

Response Templates

As indicated above, the R+S method 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, including the full detector simulation, in intervals of $p_\mathrm{T}^\mathrm{gen}$ and $|\eta^\mathrm{gen}|$ as summarized in Tab. 7.4. Reconstructed jets at detector-level are Particle-Flow jets clustered with the anti-k_T algorithm using a distance parameter of $R = 0.5$ including charged-hadron subtraction. These jets are calibrated according to the description in Sec. 5.2.4. Furthermore, the pileup scenario of the simulated sample is reweighted to match the one observed in data as explained in Sec. 6.4.2. Although a fine $p_\mathrm{T}^\mathrm{gen}$ and $|\eta^\mathrm{gen}|$ binning is chosen, the response is averaged over a certain part of the $p_\mathrm{T}^\mathrm{gen}$ spectrum in each interval. Thus, the response distribution tends to overestimate the resolution for high-p_T jets while it behaves oppositely for low-p_T jets.

Table 7.4: Overview of the $|\eta^{\text{gen}}|$ and $p_{\text{T}}^{\text{gen}}$ interval boundaries for the MC-truth response determination used as input for the R+S method.

$\lvert\eta^{\text{gen}}\rvert$
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
$p_{\text{T}}^{\text{gen}}$ [GeV]
0, 20, 30, 50, 80, 120, 170, 230, 300, 380, 470, 570, 680, 800, 1000, 1300, 1700, 2200, 2800, 3500

The truth jet response is derived by performing an unambigous one-to-one matching of reconstructed jet i to generated jet i using a distance criterion of $\Delta R < 0.1$. 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 jet pair is vetoed in a cone of size $R < 0.7$ by requiring

$$p_{\text{T}}^{\text{GenJet}_\text{j}} / p_{\text{T}}^{\text{GenJet}_\text{i}} < 0.05 \tag{7.1}$$

and

$$p_{\text{T}}^{\text{Jet}_\text{j}} < 30\,\text{GeV} \;\text{ and }\; p_{\text{T}}^{\text{Jet}_\text{j}} / p_{\text{T}}^{\text{Jet}_\text{i}} < 0.05\,. \tag{7.2}$$

The obtained jet response distributions are averaged over all jets in an event not separating them according to their rank, i. e. their position in a descending p_{T} order, or according to the jet flavour. Thus, they reflect the flavour composition of an average QCD multijet sample. One example response distribution is illustrated in Fig. 7.8. Here, the blue line indicates the response as used in the R+S method, while the red line shows the response distribution for b jets. The latter illustrates that the lower tail of the response is mainly caused by the decay of heavy-flavour quarks.

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. 6, 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. For that purpose, 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 $\pm$ 1 RMS around the mean which is then subtracted from the total response distribution in order to obtain the tail parts. 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. 6.28. 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 $\sqrt{s} = 7\,\text{TeV}$ as illustrated in Fig. 6.19 (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 ($\text{A} > 2\sigma_c$) [147].

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

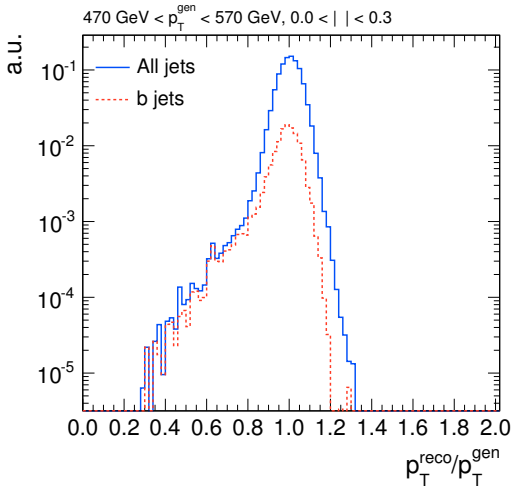

Figure 7.8: Example truth response template as used in the R+S method for one particular $p_{\mathrm{T}}^{\mathrm{gen}}$ and $|\eta^{\mathrm{gen}}|$ interval averaged over all jets (*blue line*) and for b jets only (*red line*).

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* [186].
This fit is based on the assumption 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 the measured quantities, these 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. These Lagrange multipliers provide a general method to determine local extrema of non-linear functions of many variables. Mathematically, the likelihood function

$$-2\ln[L(\vec{y}_{\mathrm{true}})] = d\vec{y}^{\,T} C^{-1} d\vec{y} \tag{7.3}$$

with $dy^i = y^i_{\mathrm{true}} - y^i_{\mathrm{measured}}$ and covariance matrix C is minimised. In this particular case, the measured jet four-momenta correspond to the values of y^i_{measured} and are fitted using the constraint of transverse momentum balance. The covariance matrix is given by the jet resolution. However, since the resolution for the angular components is not explicitly determined, in fact no proper angular fit is performed. Thus, the imbalance in each multijet event is removed by actually scaling the jet transverse momenta within the range of the respective resolutions which are approximated by the Gaussian MC-truth resolution. For each response template, the Gaussian core is extracted, as described above, by performing a fit with a Gaussian function within $\pm$ 1 RMS around the mean. The obtained resolutions with uncertainties from the Gaussian fit are illustrated as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$ for two example $|\eta^{\mathrm{gen}}|$ intervals in Fig. 7.9. Subsequently, the obtained resolution

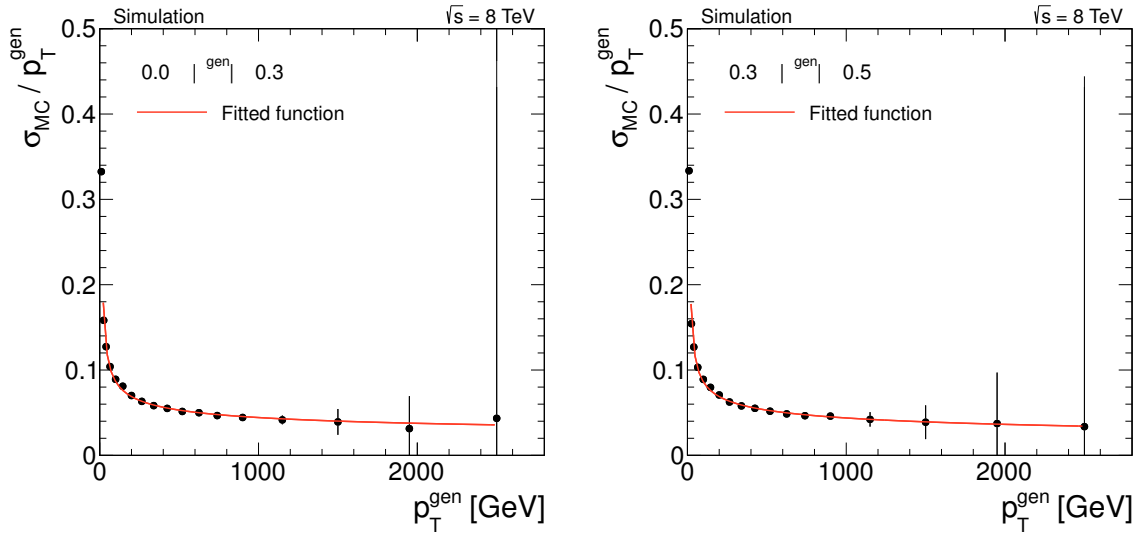

Figure 7.9: Relative truth-p_{T} resolution derived from simulated events shown as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$. The distribution is fitted with a function as described in the text used as input for the kinematic fit employed to gain a balanced seed sample.

values are fitted with the function

$$\frac{\sigma_{\mathrm{MC}}(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \sqrt{\mathrm{sgn}(N) \cdot \left(\frac{N}{p_{\mathrm{T}}}\right)^2 + S^2 \cdot p_{\mathrm{T}}^{m-1} + C^2} \tag{7.4}$$

and free parameters N, S, C and m. This function is a modified version of Eq. 6.22 introduced in Chap. 6 for the characterization of the resolution. It is adjusted here to better describe the resolution of Particle-Flow jets. The term $\mathrm{sgn}(N)$ 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 illustrated in Fig. 7.9 as red curve and used as input for the kinematic fit. The whole set of truth resolution histograms displayed with the fitted resolution functions for all $|\eta^{\mathrm{gen}}|$ intervals used as input for the kinematic fit are illustrated in Fig. B.3 and Fig. B.4 in the appendix.
Finally, all events for which the fit converged within ($|\not{H}_{\mathrm{T}}^{x}| + |\not{H}_{\mathrm{T}}^{y}| < 0.02\,\mathrm{GeV}$) are kept as seed events. Contributions to the seed sample from non-QCD multijet SM processes or even signal events in data do not have to be treated specially. Although these events might get successfully balanced, their contribution to the seed sample is negligible since their production cross section is orders of magnitudes smaller than the QCD multijet production cross section (cf. Fig. 3.2).
Furthermore, the kinematic fit allows to generate a seed sample from an inclusive sample. This is an advantage compared to for instance defining a seed sample by selecting data events with low values of $\not{H}_{\mathrm{T}}$ since selections suppressing high tails of missing energy often tend to bias the QCD kinematics. Typically, H_{T} and $\not{H}_{\mathrm{T}}$ are correlated quantities in QCD events as high values of $\not{H}_{\mathrm{T}}$ caused by severe jet mismeasurements can only occur if there is a certain amount of energy in the event. Consequently, selection cuts removing high-$\not{H}_{\mathrm{T}}$

tails also remove parts of the H_T spectrum which results in an overall underestimation of QCD contributions to the high H_T and $\not{H}_T$ tails. Thus, the rebalancing of an inclusive sample with a kinematic fit allows to generate an unbiased seed sample.

Response Smearing

The second step of the R+S method is the smearing procedure. Here, all jets of a seed event are smeared with the full jet response distributions including non-Gaussian tails. This means that for each event the magnitudes of the transverse momenta of the jets are scaled with a factor that is randomly obtained from the jet response distribution histogram 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.
However, this procedure typically results in predictions with large statistical uncertainties since the probability that a seed event is smeared into the signal region is small. In order to obtain a more robust estimate of the prediction, each event is smeared not only once, but $N = 100$ times (bootstrap method). The mean of these N predictions is considered as final result, while the statistical uncertainty is obtained as the standard deviation of the mean estimate, i. e. the standard deviation of the set of predictions divided by $1/\sqrt{N}$. This definition of the statistical uncertainty of the prediction ignores the statistical fluctuations of the seed sample. However, since the seed sample is very large, this uncertainty 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 with loose preselection at detector-level of $N_{\text{Jets}} \geq 2$ and $H_T > 350\,\text{GeV}$.[2] The result of this generator-jet smearing is shown in Fig. 7.10. Overall the distributions derived from the smeared generator jets reproduce nicely the expectation from the simulation.
In general, disagreements between predicted and expected quantities in simulated events are in this thesis denoted as *non-closure* of the method and the respective tests of the agreement *closure tests*. Non-closure can occur for instance due to the limited granularity of the response template 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. 7.3.3) and is assigned as systematic uncertainty to the QCD background prediction (see Sec. 7.3.6).

7.3.2 Application to Collision Events

The successful performance of the smearing procedure using simulated events has been shown in the previous section. However, when applying the whole R+S method to collision events including the rebalancing step, further aspects have to be considered.
Jets reconstructed in an event, especially soft jets, do not necessarily originate from the hard interaction, but might arise for instance from pileup or the underlying event. Thus, it is necessary to discard jets below a certain p_T threshold in the rebalancing step in order to

[2] The same preselection is applied to all simulated and data samples studied in this chapter. This does not bias the QCD prediction since both cut values are sufficiently below the analysis selection criteria.

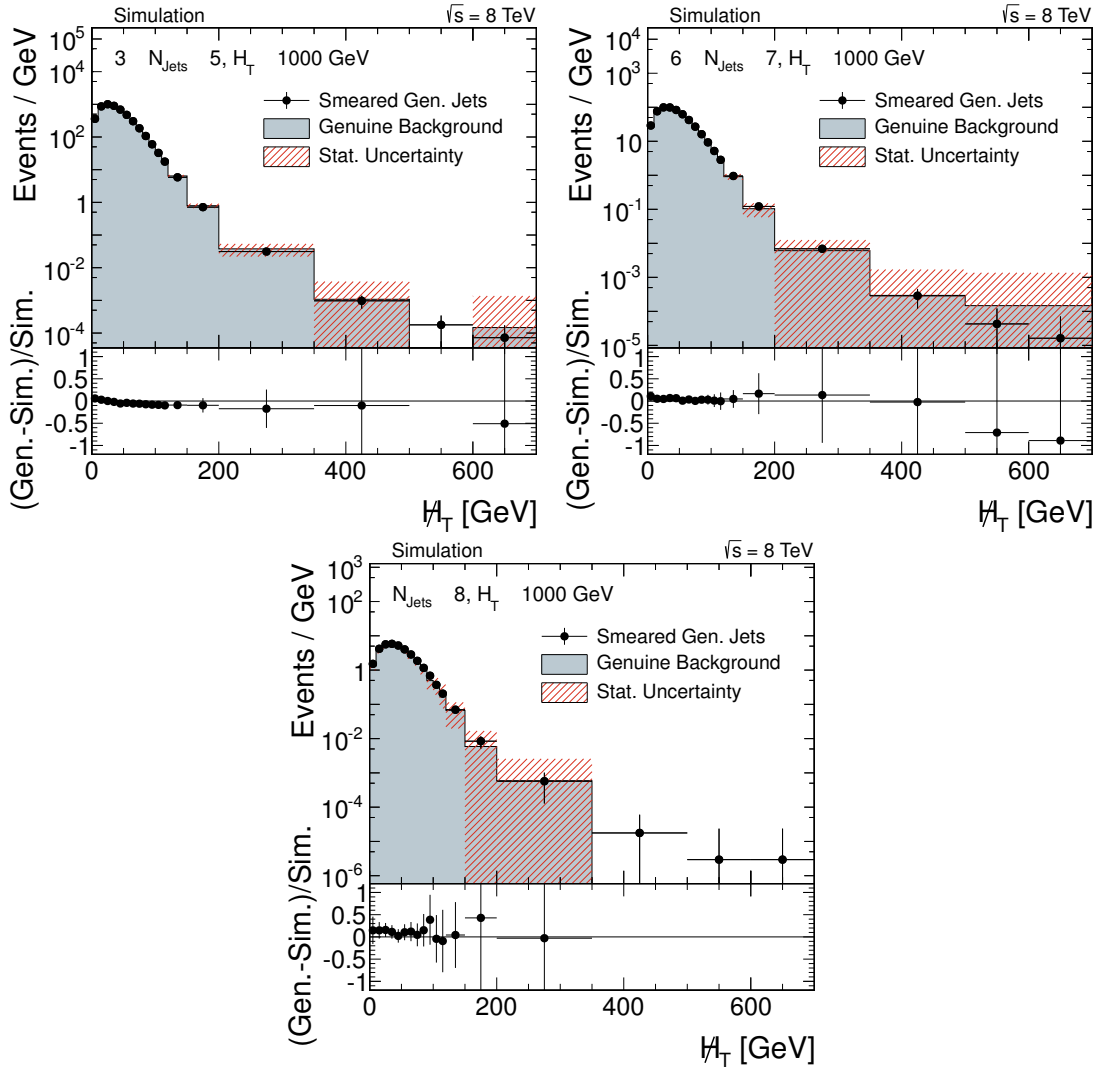

Figure 7.10: Particle-level jets obtained from a QCD multijet sample generated with MadGraph smeared with the truth response templates are compared to the expectation from full simulation. This comparison is shown for search regions with non-negligible QCD multijet background contributions as a function of $\not{H}_{\mathrm{T}}$ defined by $H_{\mathrm{T}} \geq 1000\,\mathrm{GeV}$ for $3 \leq N_{\mathrm{Jets}} \leq 5$ (*top left*), $6 \leq N_{\mathrm{Jets}} \leq 7$ (*top right*) and $N_{\mathrm{Jets}} \geq 8$ (*bottom*) after the application of the minimum $\Delta\phi$ selection.

not balance them against the hard process. This minimum p_{T} threshold for jets considered in the rebalancing step can for instance be chosen such that a good inclusive closure of the method in simulated events is observed. Here, only jets with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ are considered for the rebalancing procedure. A similar p_{T} threshold has also been imposed in earlier versions of the R+S method. There, it has been shown that the R+S method with this configuration provides reliable predictions of H_{T} and $\not{H}_{\mathrm{T}}$ distributions for an inclusive jet multiplicity selection of $N_{\mathrm{Jets}} \geq 3$. Since in the analysis presented in this thesis, the search regions have been extended to several jet multiplicity intervals, it is of particular interest

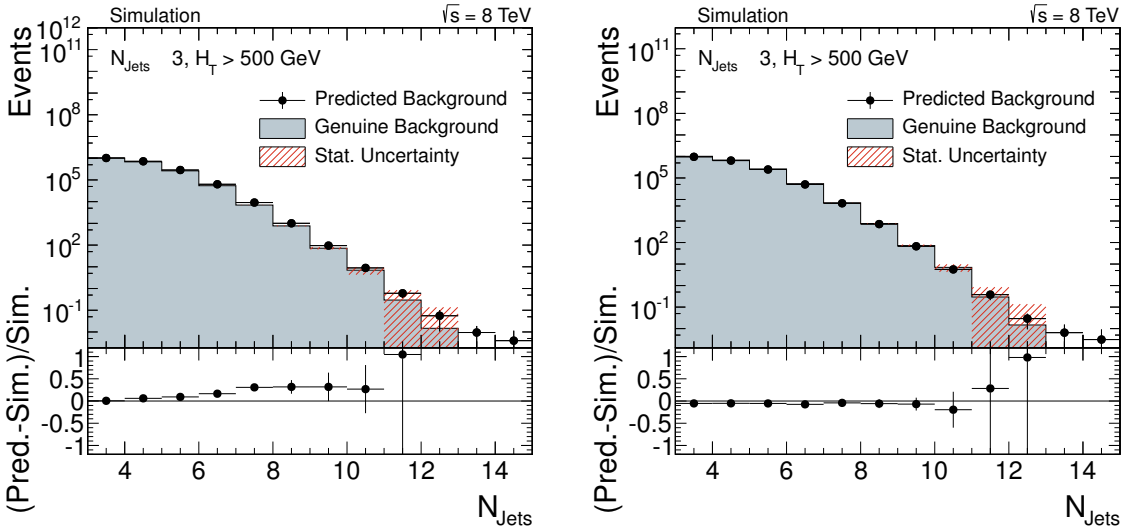

Figure 7.11: Prediction of QCD background on a QCD multijet sample generated with MadGraph compared to the expectation from full simulation with a p_{T} cut of 10 GeV imposed on jets considered in the rebalancing. The predicted N_{jets} distribution is obtained using the R+S method as in previous versions of the analysis (*left*) and using the R+S method with a correction of the rebalancing step as described in the text (*right*).

to study if the R+S method is able to correctly predict the jet multiplicity as well.
A comparison of the predicted N_{Jets} distribution with the R+S method, using a p_{T} cut of 10 GeV for jets considered in the rebalancing, to the expected genuine background from simulation is shown in Fig. 7.11 (left) after imposing baseline requirements without the $\not{H}_{\mathrm{T}}$ selection. The distributions exhibit that the R+S method tends to overpredict the expected number of events with increasing jet multiplicities by up to around 40%. The smearing procedure has been shown in Sec. 7.3.1 to provide a reliable performance so that this overprediction can be attributed to the rebalancing step.
Since the rebalancing step is supposed to provide an estimate of the particle-level jet spectrum, the quality of the rebalancing can be tested by comparing the p_{T} of the rebalanced jets to the p_{T} of matched generated jets after performing the rebalancing when excluding jets below 10 GeV from the rebalancing procedure. The mean of this ratio, denoted as rebalance factor f, as a function of the momentum of the matched reconstructed jets is shown in Fig. 7.12. It is observed that especially jets with small transverse momentum ($p_{\mathrm{T}} \leq 100$ GeV) are, on average, rebalanced to too high momenta. The observed too high momenta of rebalanced jets thus result in more jets passing the N_{Jets} threshold of 50 GeV than expected. As a consequence, the number of predicted QCD events in the higher jet multiplicity bins is too large.
This effect can be explained by the necessity to discard soft jets in the rebalancing procedure. By not taking into account all jets in the rebalancing, also soft jets that do belong to the actual hard interaction are not considered. This introduces an artifical additional imbalance in the event which has to be compensated for in the kinematic fit. In order to further study this threshold effect, a closure test is performed on a simulated QCD sample

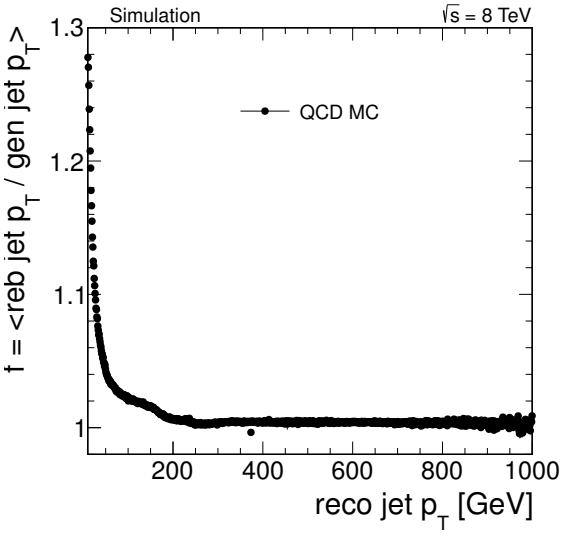

Figure 7.12: Mean of transverse momenta of rebalanced jets divided by transverse momenta of matched generated jets as a function of transverse momentum of matched reonstructed jets obtained from simulated events generated with MadGraph.

which is generated without pileup obtained from the Pythia6 [115] event generator with tune Z2* [187]. This offers the possibility to drop the requirement of $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ for jets considered in the rebalancing. For consistency, also the response templates are in this closure test taken from the same QCD sample. The results of this test are shown in Fig. 7.13. In general, the disagreement for the shown regions is statistically compatible with zero. Especially the higher jet multiplicities are in this case predicted correctly by the R+S method when there is no necessity to impose the minimum p_{T} threshold in the rebalancing.
In order to account for the observed threshold effect in events in which the minimum p_{T} selection in the rebalancing actually is needed, an empirical correction factor is introduced. The correction factor is employed such that in the rebalancing procedure each jet momentum is scaled by 1/f before the kinematic fit is performed. Thus, the effect of overpredicting the jet momenta of jets with small transverse momenta is taken care of by downscaling the jet momenta before the kinematic fit by a factor representing the average overprediction. The prediction of the N_{Jets} distribution with the adjusted R+S method compared to the expected background is shown in Fig. 7.11 (right) and illustrates that the regulated method leads to correct predictions of the QCD event yields also in the high jet multiplicity regions. The correction factor derived from simulation is later also applied to the rebalancing of data events since in data the same minimum jet-p_{T} threshold of 10 GeV is chosen.

7.3.3 Validation in Simulated Events

The quality of the R+S method to predict background contributions from QCD multijet events is tested on simulated samples by several closure tests in different kinematic regions. In these tests, the data-based prediction is applied to simulated events and compared to

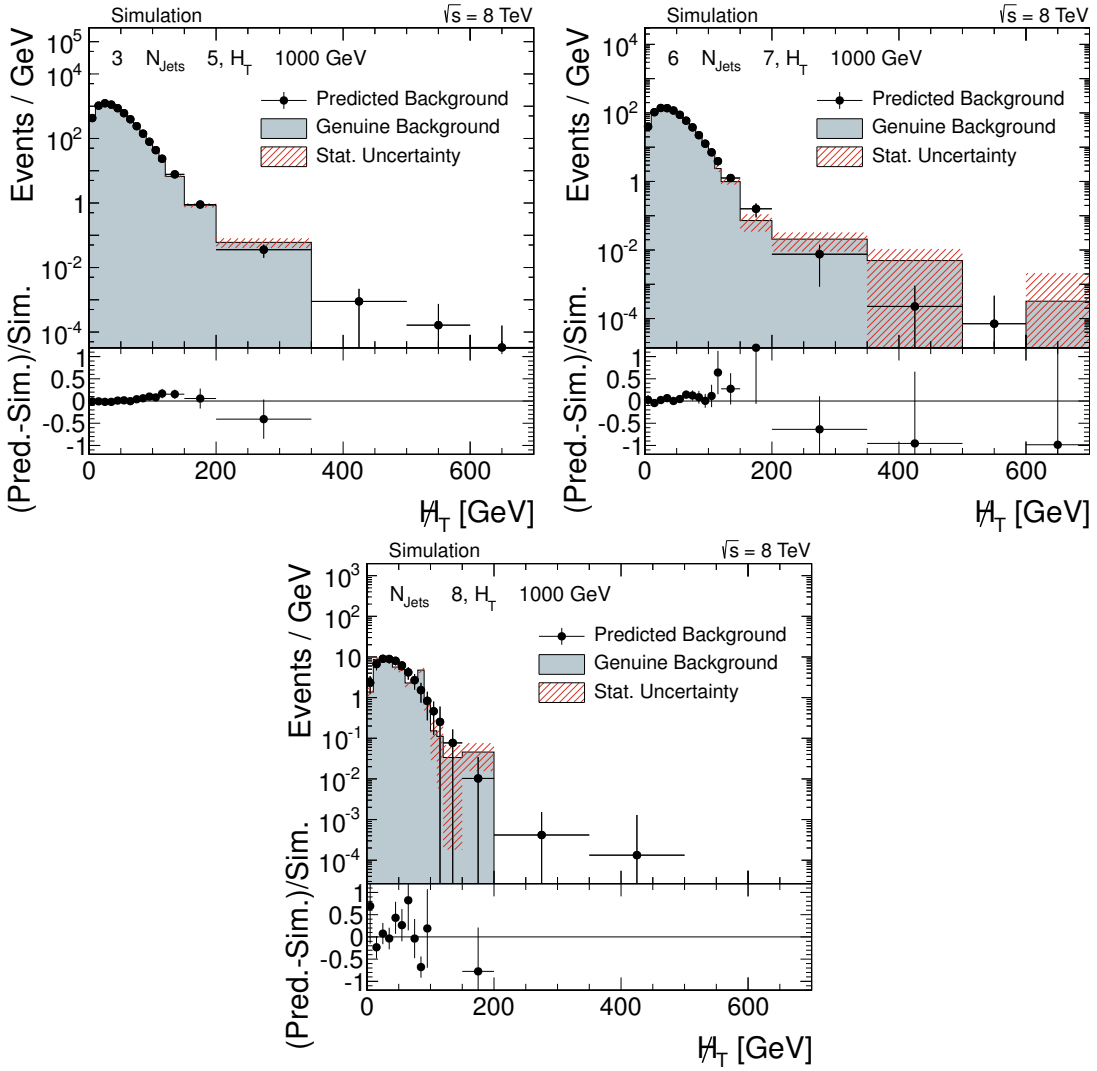

Figure 7.13: Prediction of QCD background on a QCD multijet sample generated with Pythia without the simulation of pileup compared to the expectation from full simulation. This comparison is shown as a function of $\not{H}_{\mathrm{T}}$ for regions defined by $H_{\mathrm{T}} \geq 1000\,\mathrm{GeV}$ and $3 \leq N_{\mathrm{Jets}} \leq 5$ (*top left*), $6 \leq N_{\mathrm{Jets}} \leq 7$ (*top right*) or $N_{\mathrm{Jets}} \geq 8$ (*bottom*) after applying the minimum $\Delta\phi$ selection.

the results from full simulation, as explained in Sec. 7.3.2. The different closure tests as a function of $\not{H}_{\mathrm{T}}$ for various jet multiplicity bins for a low $H_{\mathrm{T}} = [500, 1000]\,\mathrm{GeV}$ and a high $H_{\mathrm{T}} \geq 1000\,\mathrm{GeV}$ selection are shown in Fig. 7.14 for simulated events obtained from MadGraph. In general, the prediction shows a good agreement with the expected QCD background contributions. Nonetheless, deviations between prediction and expectation occur. These are considered as remaining bias of the R+S method and treated as systematic uncertainty. Due to the limited statistics of the simulated sample, the bias uncertainty is not evaluated for each search region individually, but for the different jet multiplicity intervals and a low and a high H_{T} selection corresponding to the kinematic regions defined in Fig. 7.14.

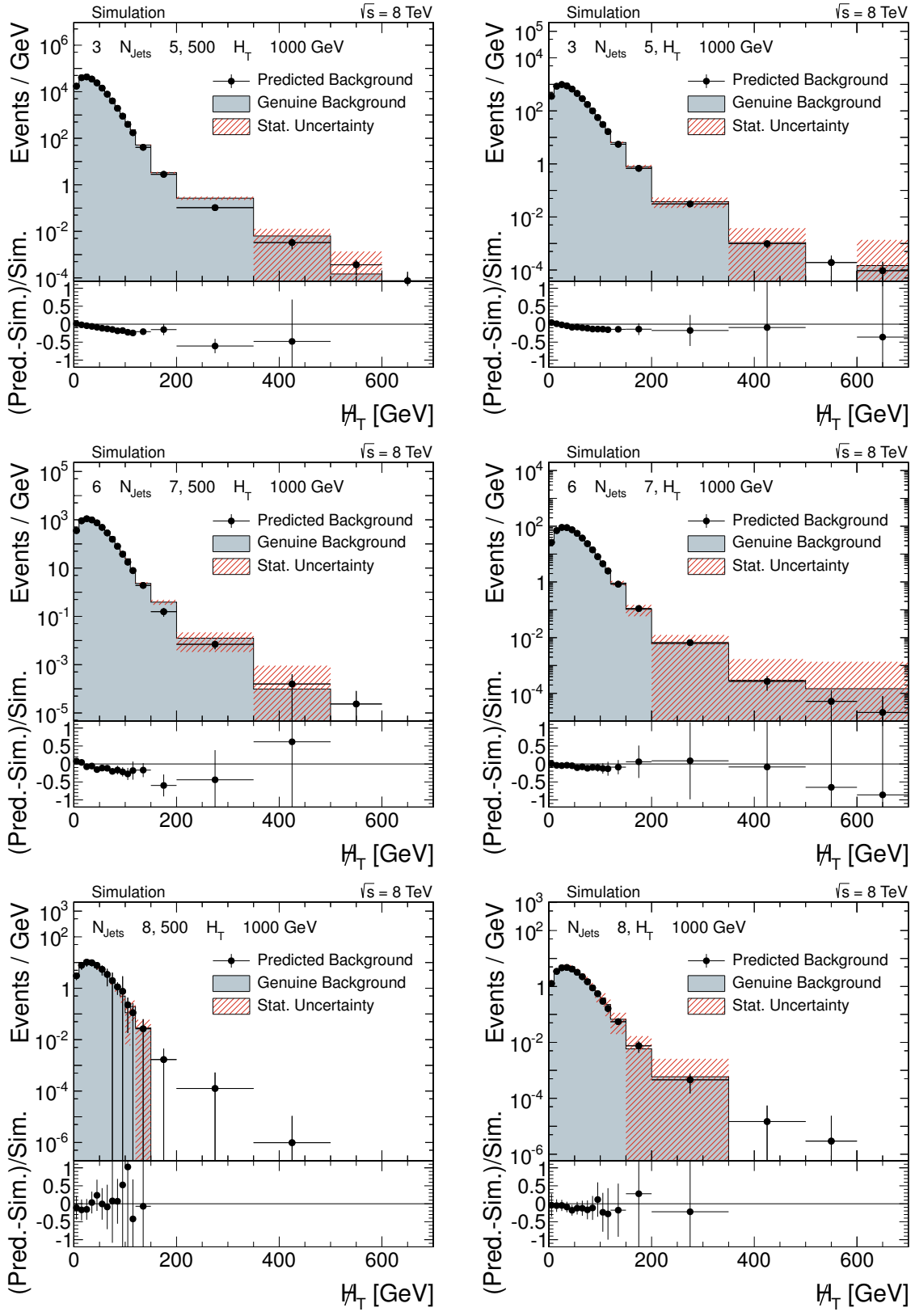

Figure 7.14: Prediction of QCD background on a QCD multijet sample generated with MadGraph compared to the expectation from full simulation. The closure test is shown for various jet multiplicity bins and low (*left*) or high (*right*) H_T selections.

Table 7.5: Summary of non-closure uncertainties with statistical uncertainties derived from the MadGraph QCD sample for the signal region (*first column*) and two control regions which are defined by $100 < \not{H}_T < 200\,\text{GeV}$ (*second column*) and inverted $\Delta\phi$ criterion (*third column*). The fourth column is used as additional cross check region as described in the text. The numbers marked in bold letters are considered 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)	$\not{H}_T > 200$ GeV $\Delta\phi$ cut	$\not{H}_T = 100 - 200$ GeV $\Delta\phi$ cut	$\not{H}_T > 200$ GeV $\Delta\phi$ cut inverted	$\not{H}_T = 100 - 200$ GeV $\Delta\phi$ cut inverted
3 – 5	500 – 1000	(**60.4** ± 9.8)%	(22.6 ± 1.6)%	(20.1 ± 6.0)%	(2.8 ± 1.3)%
6 – 7	500 – 1000	(43.1 ± 46.5)%	(**25.4** ± 11.1)%	(59.3 ± 96.0)%	(4.6 ± 20.0)%
≥ 8	500 – 1000	–	(8.9 ± 90.1)%	(**86.0** ± 38.2)%	(12.2 ± 66.4)%
3 – 5	≥ 1000	(17.1 ± 35.0)%	(14.4 ± 3.1)%	(**14.5** ± 8.9)%	(5.1 ± 1.7)%
6 – 7	≥ 1000	(5.5 ± 108.0)%	(**10.9** ± 8.8)%	(14.5 ± 42.9)%	(3.0 ± 7.0)%
≥ 8	≥ 1000	(19.4 ± 276.0)%	(21.8 ± 28.6)%	(40.4 ± 293.5)%	(21.1 ± **42.6**)%

The first choice for the determination of remaining biases is to calculate for the signal region, defined by $\not{H}_T > 200$ GeV and the application of the $\Delta\phi$ cut, the relative difference between prediction and expectation normalized to the expectation. Then, this relative difference can be considered as systematic uncertainty. In case it is statistically significant within $\pm 1\sigma$ uncertainty, the prediction could also be corrected for the non-closure. The calculated relative differences with their statistical uncertainties are summarized in Tab. 7.5 (first column) for the signal region. The table shows that there is one bin ($3 \leq N_{\text{Jets}} \leq 5$ and $500 \leq H_T \leq 1000\,\text{GeV}$) in which the signal region shows a statistically significant non-closure. In Fig. 7.15, this particular distribution is shown for the MadGraph QCD sample (left) and alternatively for a QCD sample generated with Pythia (right) as used for the studies in Chap. 6. In the region for $200 < \not{H}_T < 350\,\text{GeV}$, the MadGraph sample exhibits an underprediction of $\approx 60\%$, while the Pythia sample tends to a statistically non-significant overprediction. Thus, it is difficult to judge if the observed underprediction in the test performed with MadGraph is a systematic effect or just a statistical fluctuation. In order to treat this observed difference conservatively, the predicted result is not corrected for this potential deviation, but the total 60% deviation observed in the MadGraph sample is considered as systematic uncertainty. Since QCD is not the dominant background contribution in the search regions defined by $3 \leq N_{\text{Jets}} \leq 5$ and $500 \leq H_T \leq 1000\,\text{GeV}$, this rather large uncertainty has hardly an impact on the final result of the analysis.
Since the number of events in the signal region with $\not{H}_T \geq 200\,\text{GeV}$ is low for all bins except for the one discussed above, the respective statistical uncertainties are large. Hence, these bins do not allow reasonable conclusions concerning the closure of the method. Thus, also two sidebands of the signal region are studied and the prediction is compared to the full simulation either for control region 1 defined by $100 \leq \not{H}_T \leq 200\,\text{GeV}$ (second column in Tab. 7.5) or for control region 2 defined by an inverted $\Delta\phi$ criterion (third column in Tab. 7.5). The evaluation of the remaining bias aiming at a conservative treatment, proceeds as follows:

- If the differences in both control regions are statistically significant, the larger one

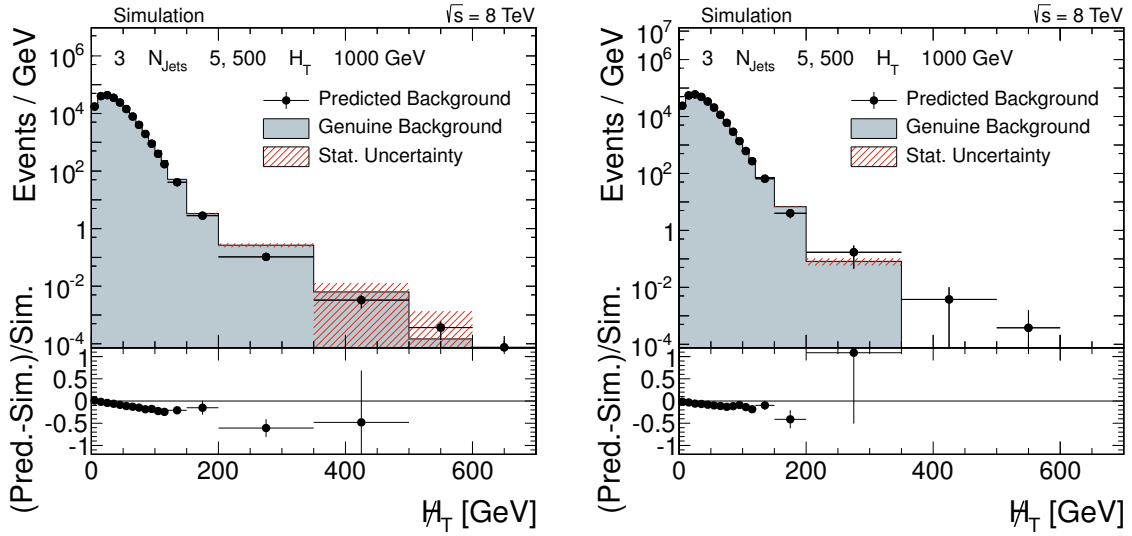

Figure 7.15: Prediction of QCD background on a simulated QCD multijet sample compared to the expectation from full simulation. The closure test is shown for 3–5 jets and $500 \leq H_{\mathrm{T}} \leq 1000$ GeV obtained from the MadGraph QCD sample (*left*) and alternatively from a Pythia QCD sample (*right*).

is considered as systematic uncertainty.

- If only one of the two numbers in the control regions 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. Thus, the value is compared to the value and its uncertainty in the corresponding cross check region bin (right column of Tab. 7.5). If the value and its uncertainty in the cross check region are smaller than the chosen value from the control region, the number from the control region is considered as systematic error. Otherwise the largest number, i. e. the deviation or its uncertainty, from the cross check region is taken.

- If none of the numbers in the control regions is statistically significant, the number with higher precision is taken and it is proceeded as above by comparing this value to the numbers in the cross check region. If the cross check region does not show larger values, the number with highest precision is taken and otherwise the largest number from the cross check region is chosen.

The uncertainty which is finally considered by the procedure described above as the uncertainty quantifying the remaining bias of the R+S method, is printed in bold letters in Tab. 7.5.

7.3.4 Application to Data Events

After the successful validation of the R+S method in simulated events and a quantification of a possible remaining bias, the procedure can finally be applied to data in order to estimate the QCD background contributions.

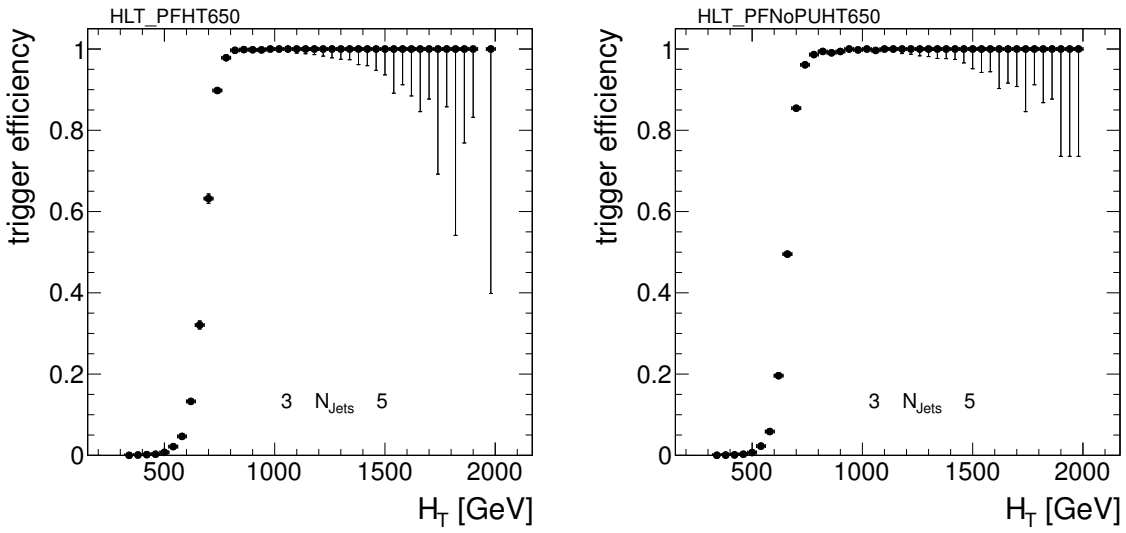

Figure 7.16: Measured trigger efficiency for paths HLT_PFHT650 (*left*) and HLT_PFNoPUHT650 (*right*) as a function of H_{T}, shown for $3 \leq N_{\mathrm{Jets}} \leq 5$.

The QCD background prediction is performed on a QCD multijet data control sample. This is collected by two triggers based on H_{T} calculated from PF jets. The nominal H_{T} thresholds of the two triggers are 350 GeV and 650 GeV, respectively. The trigger efficiency as a function of H_{T} for a nominal threshold of 350 GeV has been evaluated already for the signal trigger in Sec. 7.1.1 and was observed to be fully efficient for the baseline H_{T} cut. The trigger efficiencies for the respective trigger with $H_{\mathrm{T}} = 650\,\mathrm{GeV}$ are shown in Fig. 7.16 for a jet multiplicity of 3–5 and for jet multiplicities $6 \leq N_{\mathrm{Jets}} \leq 7$ and $N_{\mathrm{Jets}} \geq 8$ in App. B.5 and App. B.6, respectively. For these jet multiplicities, the trigger paths are fully efficient for a H_{T} selection of 800 GeV. Since two different H_{T} threshold triggers are used, each event of the multijet control sample has to be unambigously assigned to one trigger to avoid double-counting of certain events and to gain a smooth H_{T} spectrum which allows an unbiased background prediction. Taking into account that the trigger with the lower H_{T} threshold has been prescaled during operation, this is done as follows: For each event, the trigger which fired and has the lowest prescale factor is determined. Then the event is weighted according to the prescale factor. Since only one prescaled and one unprescaled trigger is considered, this assignment is unambigious and leads to a smooth H_{T} spectrum which starts at the lowest trigger threshold. The prescale weighted seed H_{T} distribution is illustrated in Fig. 7.17. The usage of the prescaled trigger is necessary, in order to collect also events with $H_{\mathrm{T}} < 800\,\mathrm{GeV}$ and in particular also events with H_{T} values below the minimum H_{T} threshold of the signal region. The latter events have to be considered since they can pass $H_{\mathrm{T}} > 500\,\mathrm{GeV}$ through a fluctuation to large response values. However, the events collected by the prescaled trigger have high event weights and spoil artificially the prediction when they enter the signal region as they lead to a substantially higher uncertainty. As described for the hadronic-tau background, this is solved by smearing the events obtained from the prescaled trigger not only once, but M times (with M = prescale factor) and weighting them accordingly with one.

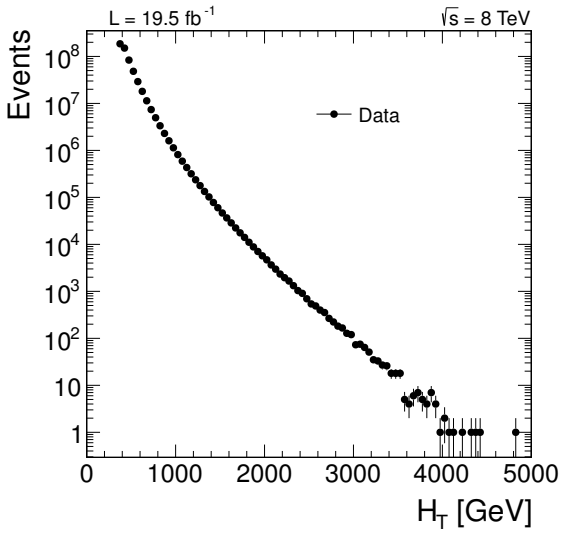

Figure 7.17: Seed H_{T} spectrum of full 2012 dataset used as input for the R+S method after correcting for trigger prescales.

7.3.5 Validation in Data Events

Similar to the validation in simulated events, also the application of the R+S method to data can be validated. This is done by comparing the prediction of the R+S method from data to selected events in a QCD enriched data control sample.
The control sample is defined by at least three jets, $H_{\mathrm{T}} > 1000\,\mathrm{GeV}$, an inverted $\Delta\phi$ criterion and $100 \leq \not{H}_{\mathrm{T}} \leq 200\,\mathrm{GeV}$. The resulting comparison between predicted QCD multijet background events obtained from the R+S method and selected data events in the control region is shown in Fig. 7.18 and exhibits reasonable agreement within 10–20%. This also justifies the approach to consider the correction factor for the rebalancing derived from simulation for data events as well. In general, no perfect agreement is expected in this test since contaminations from other backgrounds are still present in the data control sample. This is illustrated in the lower row of Fig. 7.18 in which the background composition is studied in simulated events. The ratios in the bottom display the comparison of the QCD background to the total background and thus provide an estimate of the purity of the control sample. The ratios exhibit a similar trend as seen in the upper row for the comparison of the QCD prediction to the data control sample and confirm that the deviations observed in the data closure-test can essentially be attributed to the contamination of the control region by other backgrounds. The residual contribution of non-QCD background to the chosen QCD control region amounts to 5% in total and is primarily present for H_{T} values up to 1500 GeV and high jet multiplicities. Deviations observed in the data closure test beyond the contamination of the control sample from other backgrounds are covered by the bias uncertainty studied in Sec. 7.3.3.
Overall, the application of the R+S method to predict the QCD background contributions in data is expected to provide reliable results since the validation tests in simulation as well as in data have a positive outcome. Systematic uncertainties that have to be considered for the prediction of QCD background are discussed in the next section.

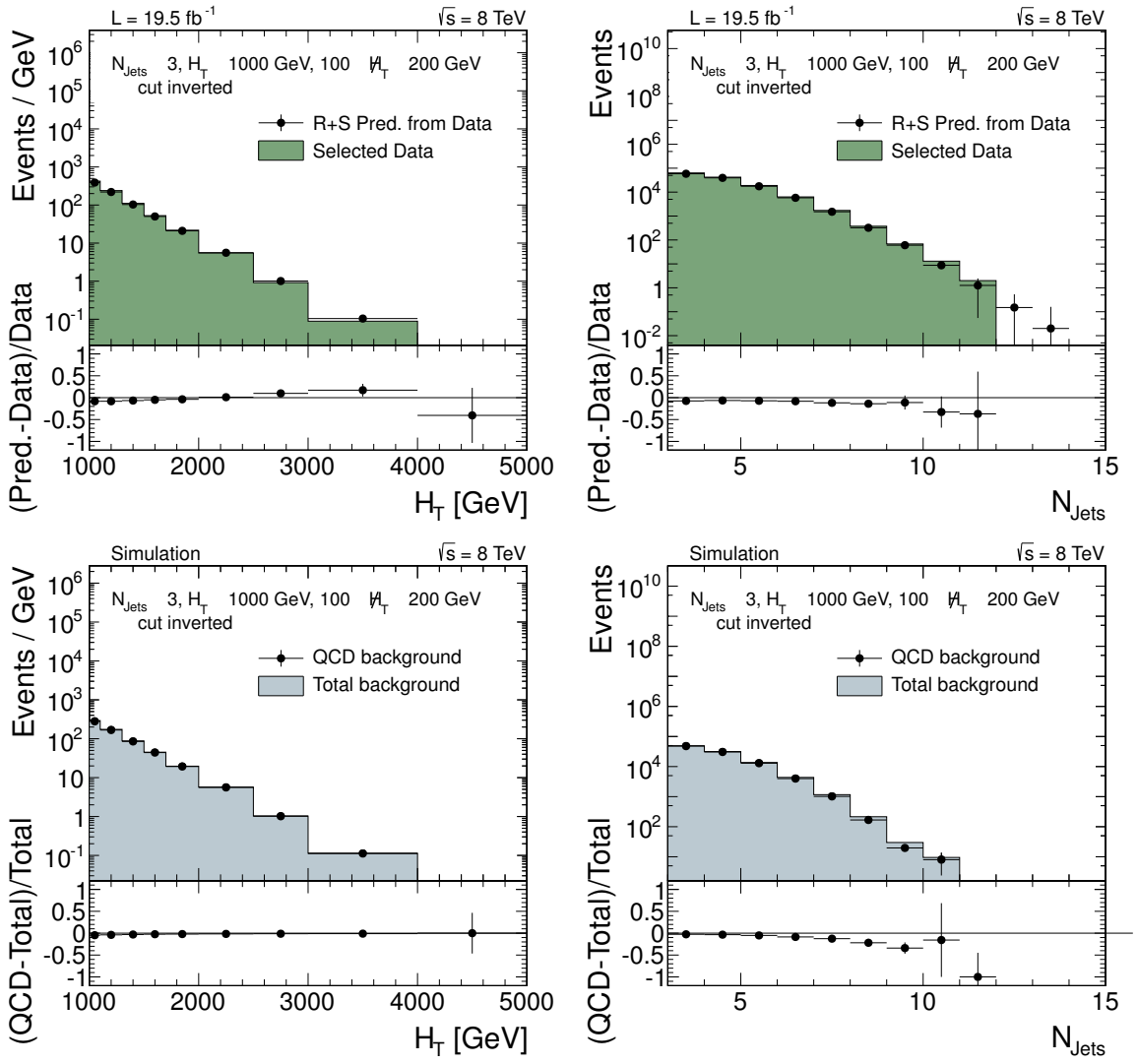

Figure 7.18: Prediction of QCD background on data compared to the expectation from data for a QCD enriched control region with $H_{\mathrm{T}} > 1000$ GeV, inverted $\Delta\phi$ cut, $N_{\mathrm{Jets}} \geq 3$ and $100 \leq \not{H}_{\mathrm{T}} \leq 200\,\mathrm{GeV}$ (*upper row*) and selected QCD events in simulation compared to the total number of selected background events in simulation in the same control region (*lower row*).

7.3.6 Systematic Uncertainties

Core of response functions

The uncertainties on the factors of the Gaussian core resolution accounting for the differences in data and simulation denoted in Tab. B.1 are propagated to the prediction. This is done by shifting the scaling factors by $\pm 1\sigma$ up and down.

Tail of response functions

The uncertainties on the scaling factors of the non-Gaussian tails listed in Tab. B.2 are also propagated to the prediction by varying them within $\pm 1\sigma$ up and down.

Bias uncertainty

The systematic uncertainty caused by a potential non-closure and remaining biases of the R+S method is evaluated as described in Sec. 7.3.3. This uncertainty also covers uncertainties on the rebalance correction factor and the jet-p_{T} cut value chosen for the jets considered in the rebalancing such that no additional uncertainties are considered for those. The bias uncertainty is evaluated for each jet multiplicity bin separated into a low $500 \leq H_{\mathrm{T}} \leq 1000\,\mathrm{GeV}$ and a high $H_{\mathrm{T}} > 1000\,\mathrm{GeV}$ region.

Pileup

In general, the R+S method is protected against influences from pileup by applying L1 jet energy corrections, making use of charged-hadron subtraction and neglecting soft jets in the rebalancing procedure. Furthermore, pileup, which is an issue especially for soft jets, does in general not contribute significantly to the high $\not{H}_{\mathrm{T}}$ search bins as these are mainly populated due to heavily mismeasured hard jets.
The residual pileup influence is evaluated in a low $\not{H}_{\mathrm{T}}$ control region with $\not{H}_{\mathrm{T}} > 100$ GeV since residual pileup effects are expected to contribute mostly in such low $\not{H}_{\mathrm{T}}$ regions. Pileup contributions to the high $\not{H}_{\mathrm{T}}$ signal regions might be even less, so that this approach provides a conservative estimate of the pileup uncertainty. Furthermore, the usage of a low $\not{H}_{\mathrm{T}}$ control region ensures that statistical limitations are reduced.
The general approach is to estimate the pileup-dependent fraction of the nominal QCD prediction by studying the difference in the behaviour of the prediction in data and simulation (taken from MadGraph) for different pileup conditions. In order to do this, the sample is divided into three different bins of primary vertices $N_{\mathrm{Vtx}} = [0, 10]$, $N_{\mathrm{Vtx}} = [11, 20]$ and $N_{\mathrm{Vtx}} > 20$. Then, the QCD prediction is calculated for each vertex bin when applying baseline selection criteria with the relaxed $\not{H}_{\mathrm{T}}$ requirement mentioned above. In order to compensate for the different number of seed events contributing to each N_{Vtx} interval, the prediction in each N_{Vtx} bin is normalized to the respective number of seed events contributing to that particular vertex bin. Furthermore, it is assumed that pileup effects are negligible in the lowest primary vertex bin. Hence, the predictions corrected for the number of seed events in data and simulation are normalized to each other such that they have the same yield in the first primary vertex bin in order to study the difference between data and MC prediction in the two higher primary vertex intervals. The distributions obtained with this procedure are illustrated for jet multiplicity 3–5 in Fig. 7.19. In order to relate the observed difference between the prediction in data and simulation for different pileup conditions to the nominal QCD prediction in data, the absolute difference between data and Monte Carlo prediction for the second and the third vertex bin is taken, multiplied each with the seed events in data for that vertex bin and summed up. Finally, the ratio of this pileup dependent fraction of the prediction to the nominal QCD prediction is considered as pileup uncertainty. Due to statistical limitations it is not evaluated for the various H_{T} regions, but for the three jet multiplicity selections only.

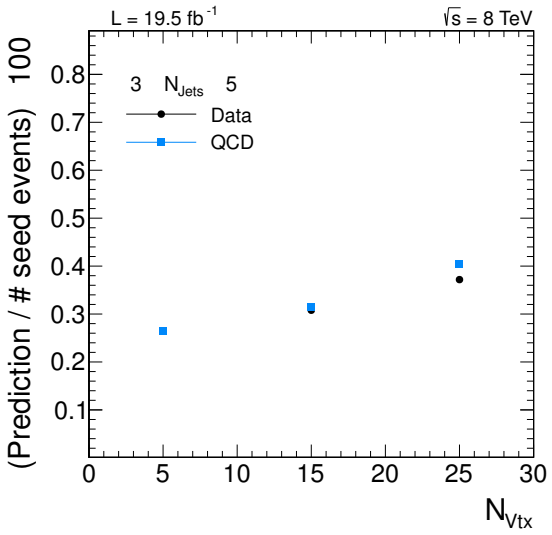

Figure 7.19: Prediction of QCD background in data and Monte Carlo as a function of primary vertices (N_{Vtx}) normalized to the number of contributing seed events used for the determination of the pileup uncertainty as described in the text.

7.3.7 QCD Background Prediction

The final prediction for QCD background contributions derived with the R+S method from a multijet control sample in data is summarized in Tab. 7.6 for jet multiplicity bins 3–5, 6–7 and $\geq$ 8. The quoted total systematic uncertainty is obtained by adding the single contributions in quadrature.
Some search bins show very large statistical uncertainties of $\geq 100\%$. This then also affects the evaluation of systematic effects, e. g. the core and tail scaling uncertainties. However, since this happens only in bins in which QCD background is almost negligible, it does not impact the final result of the analysis significantly. For the affected systematic variations, the largest observed variation is considered as systematic uncertainty, as a conservative estimate. Affected search bins are mainly the high $\not{H}_{\mathrm{T}}$ and the highest jet multiplicity search bins.
However, search regions with non-negligible contributions from QCD background, like high H_{T} ($\geq 1000\,\mathrm{GeV}$) and low $\not{H}_{\mathrm{T}}$ bins, show in general quite moderate total uncertainties. Typical values lie around 50% which is a remarkable precision for a prediction of a background that is so difficult to model as discussed in the beginning of this section. The main contributions to the systematic uncertainty arise from the propagated uncertainties of the core and tail scaling factors. The resulting variations in the prediction from the variation of the core scaling factors are typically between 10–30 % in the search bins with non-negligible QCD contributions while it is between 20–35 % from the tail scaling factors. This emphasizes the importance to precisely measure the resolution data-to-simulation scale factors as described in Chap. 6 for the scaling factors of the response core.

Table 7.6: Predicted event yields for the QCD background in the search regions defined by $H_T, \not{H}_T$ 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]	$\not{H}_T$ [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	$^{+4.4}_{-4.8}$
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	$^{+0.95}_{-0.89}$

7.4 Results and Interpretation

The selected number of events in $19.5\,\text{fb}^{-1}$ of data together with the predicted event yields for the various SM background contributions estimated as discussed in Sec. 7.2 and Sec. 7.3 are listed in Tab. 7.7 for all 36 exclusive search regions. The displayed uncertainties for the background predictions are the total uncertainties. Furthermore, the obtained yields in data and the predicted background are visualized in Fig. 7.20. The ratio in the bottom shows 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 $6 \leq N_{\text{Jets}} \leq 7$, $500 < H_{\text{T}} < 800\,\text{GeV}$ and $\not{H}_{\text{T}} \geq 450\,\text{GeV}$ with a local p-value of 0.05. However, this is insignificant when including the probability to observe a statistical fluctuation as large or larger in any of the search regions corresponding to a global p-value of 0.78.

Furthermore, 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. These are

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

in case of light-flavour squarks and

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

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

(d) $\tilde{g} \to q\bar{q} + \tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ where $\tilde{\chi}_1^{\pm} \to W + \tilde{\chi}_1^0$, $\tilde{\chi}_2^0 \to Z + \tilde{\chi}_1^0$ and $m_{\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0} = 0.5(m_{\tilde{\chi}_1^0} + m_{\tilde{g}})$

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 CL_s [188–190] approach and denote the 95% confidence level (C.L.) 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 calculated for all 36 search regions considering the uncertainties of the acceptance, efficiencies and uncertainties for the signal as well as the background predictions. The statistical and systematic uncertainties of the QCD background are considered to be fully correlated, while for other backgrounds and signal only each systematic uncertainty is treated as fully correlated among all the search regions. Furthermore, the statistical uncertainty of the lost-lepton and hadronic-tau background is assumed to be fully correlated for each bin since both make use of a single muon control sample. The uncertainties considered for the signal acceptance and efficiency in the limit setting procedure are:

- 2.6% uncertainty on the integrated luminosity [157]
- 2% uncertainty for a possible trigger inefficiency (cf. Sec. 7.1.1)
- 3% uncertainty due to a possible mismodelling of the event cleaning efficiency

Table 7.7: Predicted event yields for the different background components in the search regions defined by H_T, $\not{H}_T$ and N_{Jets}. The uncertainties of the different background sources are added in quadrature to obtain the total uncertainties. Taken from [168].

Selection			$Z \to \nu\bar{\nu}$	$t\bar{t}/W$	$t\bar{t}/W$	QCD	Total	Data
N_{Jets}	H_T [GeV]	$\not{H}_T$ [GeV]		$\to$ e, μ+X	$\to \tau_h$+X		background	
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

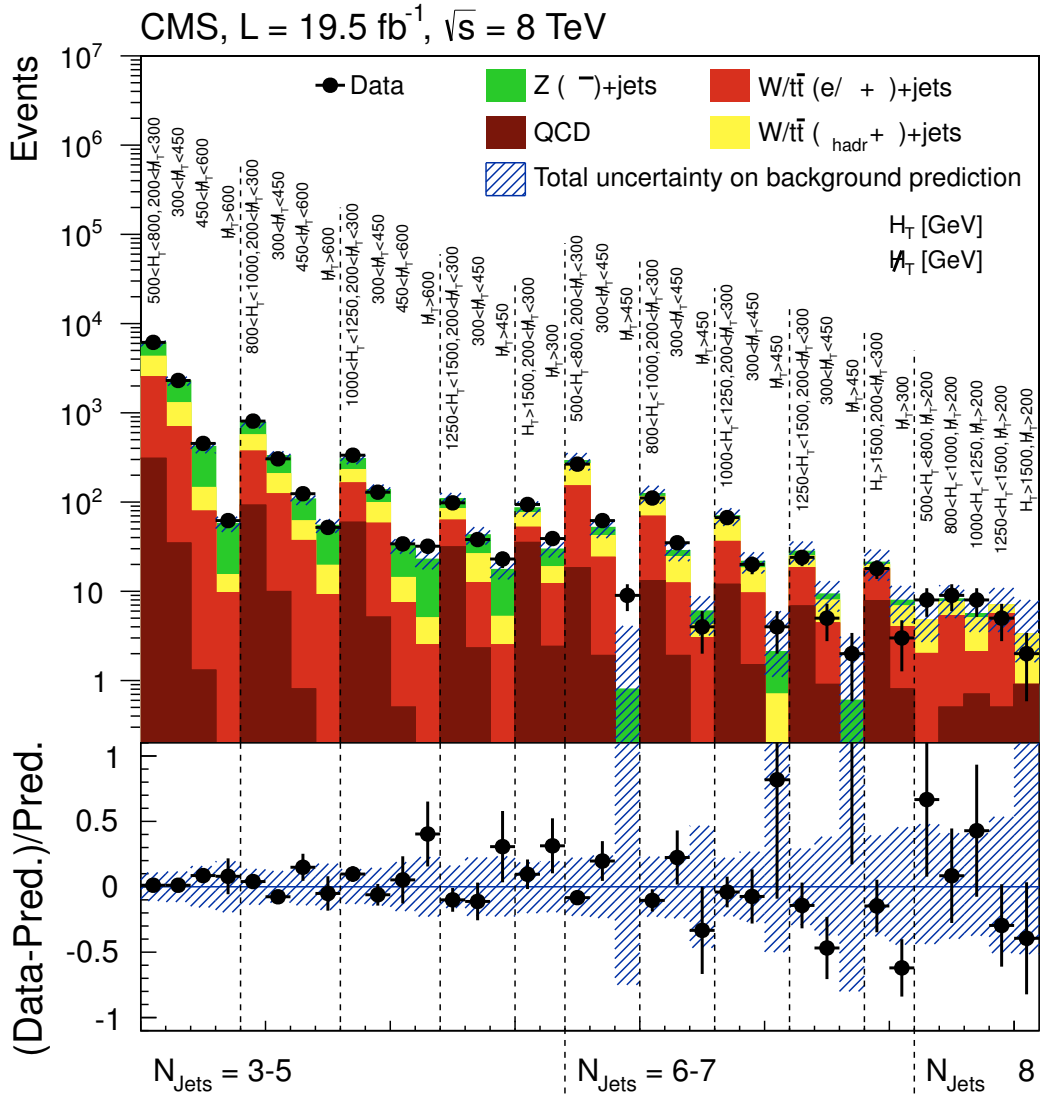

Figure 7.20: Summary of the observed number of events in each of the 36 search regions in comparison to the corresponding background prediction. The hatched region shows the total uncertainty of the background prediction. Taken from [168].

- 2–8% and 1–2% in the signal acceptance from the propagation of the respective uncertainties on the jet energy calibration and resolution
- 1–8% in the signal acceptance from the systematic variation of PDFs [191]
- an uncertainty considering the adjustment of the rate of initial-state radiation in simulation to match the measured rate in data [192] resulting for model parameter points with small differences between the LSP and the gluino or squark mass in an uncertainty of up to 22% and typically less than a few percent for others

Finally, the resulting exclusion limits for the above described processes (a)–(d) are shown in Fig. 7.21 (a)–(d), respectively. The expected (*dashed*) and observed (*solid*) 95% C.L. upper limits are shown in the gluino-LSP and squark-LSP mass plane for the signal production cross sections, respectively. 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 [193] and MSTW2008 [194] 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.
In general, the analysis provides good sensitivity to signal points with large mass differences between squark/gluino and LSP mass, nearly independent of the LSP mass. This is due to the fact that for such mass scenarios the selection based on large values of H_{T} and $\not{H}_{\mathrm{T}}$ is most efficient. For signal scenarios in which the mass difference between squark/gluino and LSP is small, softer jets and smaller values of $\not{H}_{\mathrm{T}}$ are expected and consequently the analysis sensitivity drops resulting in the weaker cross section limits.

7.4.1 Comparison to Other Measurements

The exclusion limits obtained with the analysis presented in this thesis 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 provides a good sensitivity towards the gluino-mediated production of third generation squarks and to decays involving W and Z bosons which could not be explored before.
Furthermore, studies using the $\sqrt{s} = 8$ TeV data targeting the same simplified models, but based on different analysis techniques or final states, have been performed within CMS as well. A comparison of the analysis presented here (SUS-13-012) to other CMS analyses is illustrated in Fig. 7.22 for models (a)–(c) introduced in Sec. 7.4.
This comparison exhibits that the expected sensitivity to the models $\tilde{g} \to q\bar{q} + \tilde{\chi}_1^0$ and $\tilde{q} \to q + \tilde{\chi}_1^0$ is similar for the different analyses. This is not unexpected as the compared analyses all make use of the hadronic final state. However, the analysis published in [180] performs better especially in case of squark production as here also search regions based on dijet events are employed.
Concerning gluino-mediated top squark production it is interesting to note that the analysis presented in this thesis is also similarly sensitive to this model as other hadronic

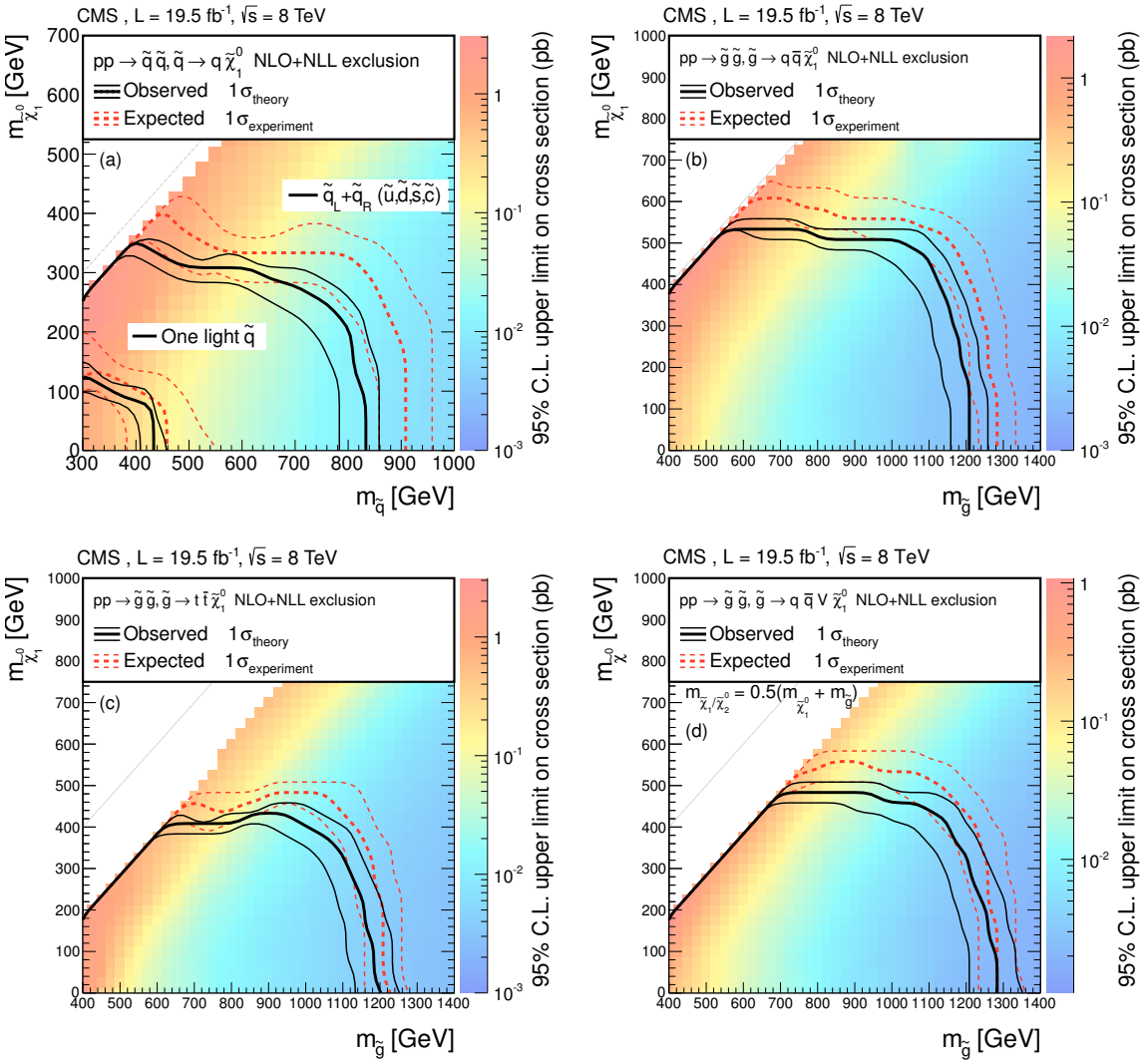

Figure 7.21: The observed and expected 95% C.L. upper limits on the (a) squark-squark and (b–d) gluino-gluino production cross sections in either the m(squark)-m(LSP) or the m(gluino)-m(LSP) plane obtained with the simplified models. For the squark-squark production the upper set of curves corresponds 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 [168].

searches (cf. SUS-12-028, SUS-12-024 or SUS-13-019). These other hadronic analyses typically employ b-tagging information, while the analysis presented in this thesis followed a complementary approach by employing high jet multiplicity search regions. However, the best sensitivity to this respective model is achieved by an analysis which is based on a single lepton, multiple jets and b-tags (SUS-13-007).
Comparable analyses have also been performed by the ATLAS experiment [195–197]. The obtained exclusion limits on sparticle masses lie in a very similar mass region.

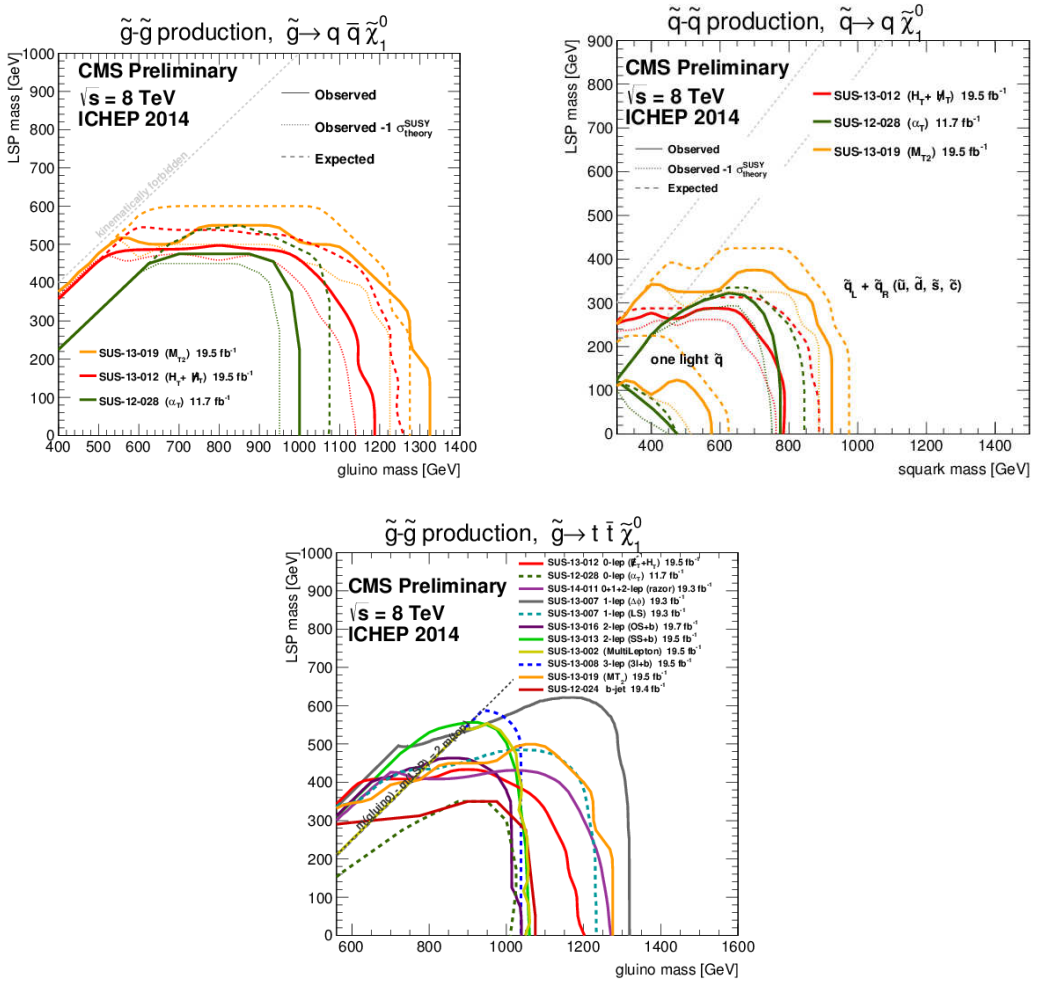

Figure 7.22: Comparison of various exclusion limits derived by different CMS analyses for the process $\tilde{g} \to q\bar{q} + \tilde{\chi}_1^0$ (*top left*), $\tilde{q} \to q + \tilde{\chi}_1^0$ (*top right*) and $\tilde{g} \to t\bar{t} + \tilde{\chi}_1^0$ (*bottom*). Taken from [76].

7.5 Status of supersymmetry after LHC Run I

In general, the SUSY search presented in this thesis as well as other measurements, as discussed in Sec. 7.4.1, pushed the mass limits of supersymmetric particles closer to the TeV range than the 7 TeV analyses or even beyond. However, most of the interpretations are in fact presented in simplified models assuming 100% branching fraction of that specific decay which is most probably not realised in nature. Scaling the respective branching ratios down, typically results in much weaker mass limits.

Therefore, it is also interesting to look at more realistic SUSY models, e. g. the CMSSM again. In addition to direct interpretations of searches in the CMSSM, which for instance in case of the ATLAS experiment, result in exclusions of $m_{1/2} \lesssim 800\,\text{GeV}$ for $m_0 \lesssim 1\,\text{TeV}$ and $m_{1/2} \lesssim 600\,\text{GeV}$ for $m_0 \lesssim 6\,\text{TeV}$ ($\tan\beta = 30$, $A_0 = -2m_0$, $\mu > 0$) [79], also global fits to constrain the respective CMSSM parameters are performed (cf. for instance [199, 200]). Typically, these fits do not only consider direct searches from the

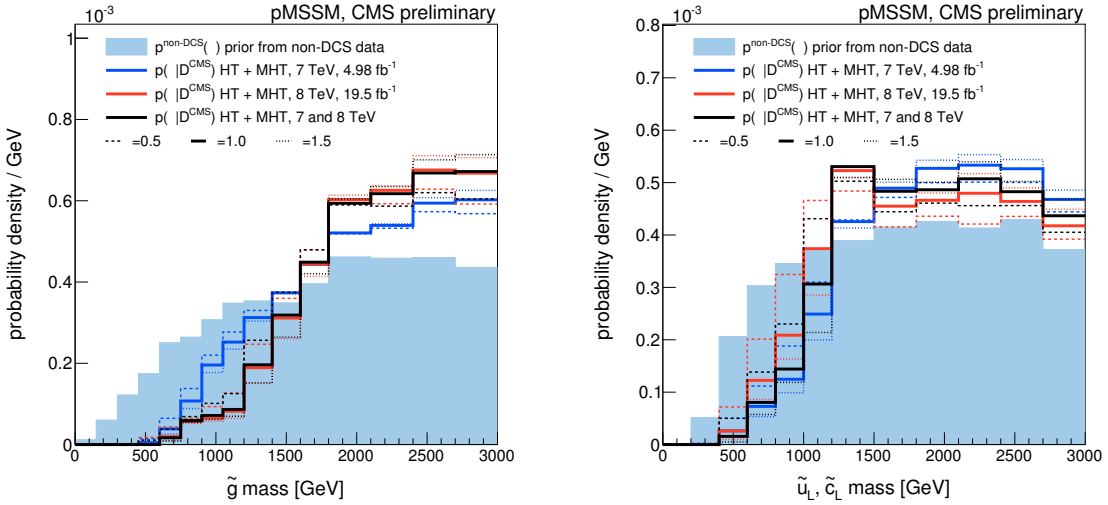

Figure 7.23: Marginalized distributions of gluino masses (*left*) and $\tilde{u}_L, \tilde{c}_L$ masses (*right*). Filled histograms show prior distributions, while line histograms illustrate posterior distributions based on the results of the analysis presented in this chapter and [78]. Solid curves denote the nominal curves while dashed lines represent systematic variations. Taken from [198].

LHC, but also use constraints from low-energy precision observables, flavour measurements or the cosmological cold dark matter density. In general, in these fits a growing tension between low-energy observables, the non-observation of SUSY at the LHC and the CMSSM is observed and consequently best fit values are pushed up to large values of $m_0/m_{1/2}$ which is in contradiction to expectations for natural SUSY.
However, as discussed in Sec. 2.3.3 the interpretation of search results only within the CMSSM and simplified models is not considered as sufficient anyway. Thus, also models beyond the CMSSM catch growing attention. One of these is the pMSSM which has been introduced in Sec. 2.3.3 as well. An interpretation of SUSY searches, comprising especially the analysis presented in this thesis, in the context of the pMSSM, has been performed by the CMS experiment [198]. In order to investigate the impact of direct searches at CMS, a global Bayesian analysis [201, 202] is performed that furthermore also includes pre-CMS data and indirect measurements. Here, probability distributions prior and posterior to the CMS searches are investigated. In Fig. 7.23, example distributions of such prior and posterior probability distributions for gluino and squark masses are illustrated. The distributions exhibit that the data disfavours pMSSM scenarios with $\tilde{g}$ masses below 1200 GeV and scenarios with $\tilde{u}_L, \tilde{c}_L$ masses below 1000 GeV. Consequently, also in this interpretation of the analysis presented in this chapter, excluded mass ranges reach or exceed already the 1 TeV mark such that the impression arises that natural SUSY is under increasing pressure.
However, as discussed in Sec. 2.2.2, especially the supersymmetric partner of the top quark should not be too heavy if SUSY is supposed to provide a solution to the hierarchy problem. This is still feasible since mass limits for third generation squarks are typically weaker than for light-flavour squarks. A summary of searches performed with the CMS

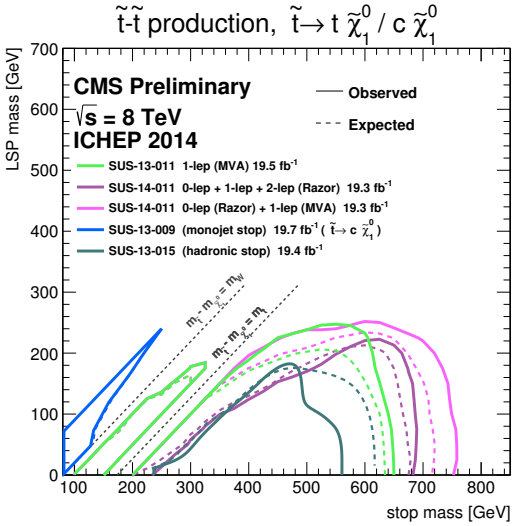

Figure 7.24: Exclusion limits derived by different CMS analyses for the process $\tilde{t} \to t\tilde{\chi}_1^0$ in the $m_{\tilde{t}}$ versus m_{LSP} plane. Taken from [76].

experiment at a centre-of-mass energy of $\sqrt{s} = 8\,\mathrm{TeV}$ for direct production of top squark pairs is illustrated in Fig. 7.24.
Here, top squarks with masses between 200–750 GeV have been excluded for LSP masses below around 200 GeV. This corresponds to a generic tuning (as introduced in Sec. 2.2.1) of $\Delta \lesssim 20$. Although this exceeds the traditional value of $\Delta \lesssim 10$ already, values up to $\Delta \lesssim 100$ are often considered acceptable to date [32]. Consequently, a top squark mass up to around 1 TeV is considered eligible and a wide range of the parameter space is still not investigated. Since the LHC starts a second run period in 2015 with an increased centre-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$, this mass region up to 1 TeV can be further studied.
A feasibility study to search for direct top squark pair production at $\sqrt{s} = 13\,\mathrm{TeV}$ is presented in the next chapter based on simulated events. Here, also techniques to identify boosted hadronically decaying top quarks, as introduced in Chap. 5, emerging from the top squark decays are employed.

8 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 $\sqrt{s} = 13\,\text{TeV}$, provides the excellent opportunity to further investigate the question if supersymmetry is realised in nature at the TeV scale. As discussed in Section 7.5, the current limits exclude top squarks with masses up to around 750 GeV for LSP masses below 100 GeV in case of direct top squark production. Since for natural supersymmetry the top squark mass is expected to not exceed 1 TeV significantly, it is of particular interest to probe the top squark mass range above 750 GeV during the next run period of the LHC.
The targeted process of this analysis has been introduced in Fig. 2.6 where the pair production of top squarks is shown with subsequent decay into a top quark and a neutralino LSP. Since each of the top quarks decays into a b quark and a W boson, the final state signature further depends on the decay modes of the W bosons. In this analysis, only final states with fully hadronic top quark decays are considered. Consequently, this analysis targets a jet final state accompanied by missing transverse energy caused by the LSPs and no isolated leptons. As discussed in Sec. 2.3.3, 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 $\sqrt{s} = 13\,\text{TeV}$ are discussed and compared in order to address mass scenarios with large mass differences between the top squark mass and the LSP. Furthermore, the performance of new selections is compared to the existing hadronic top squark search performed by the CMS experiment at $\sqrt{s} = 8\,\text{TeV}$ which is published in [203] and denoted *SUS-13-015* in the following. In addition, it is interesting to study if selections that are developed for a search for direct top squark production are also sensitive to gluino-mediated production of third generation quarks.

8.1 Samples of Simulated Events

The studies presented in this chapter are based on simulated samples at $\sqrt{s} = 13\,\text{TeV}$. These are processed with the fast detector simulation and consider the pileup scenario as it was present at $\sqrt{s} = 8\,\text{TeV}$ including out-of-time pileup according to a bunch spacing of 50 ns. Although these are not the pileup conditions expected for $\sqrt{s} = 13\,\text{TeV}$, the selections studied here typically are based on objects with high transverse momenta such that influences from pileup are negligible in good approximation. Moreover, the studies presented in this chapter are focusing on the identification of general analysis strategies comparing the relative performance of different selections such that also simplifications caused by the use of the fast detector simulation are not explicitly accounted for. Some discussion about the impact of these simplifications concerning the sensitivity of the analysis follows in Sec. 8.9.

Table 8.1: 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 + jets	$H_{\mathrm{T}} = [0, 50]$ GeV	99.92 [nb]	LO
	$H_{\mathrm{T}} = [50, 150]$ GeV	15.98 [nb]	LO
	$H_{\mathrm{T}} = [150, 300]$ GeV	1.328 [nb]	LO
	$H_{\mathrm{T}} = [300, \infty]$ GeV	0.169 [nb]	LO
Z + jets	$H_{\mathrm{T}} = [0, 100]$ GeV	22.0 [nb]	LO
	$H_{\mathrm{T}} = [100, 300]$ GeV	0.951 [nb]	LO
	$H_{\mathrm{T}} = [300, \infty]$ GeV	0.0396 [nb]	LO
QCD multijet	$H_{\mathrm{T}} = [100, 250]$ GeV	22 930 [nb]	LO
	$H_{\mathrm{T}} = [250, 500]$ GeV	465 [nb]	LO
	$H_{\mathrm{T}} = [500, 1000]$ GeV	18.66 [nb]	LO
	$H_{\mathrm{T}} = [1000, \infty]$ GeV	0.536 [nb]	LO
Top squark 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

The processes considered in the analysis are summarized in Tab. 8.1. W + jets, Z + jets and QCD multijet events are generated with MADGRAPH5 [117] using the PDF CTEQ6L1 [152], while $t\bar{t}$ events are generated with POWHEG [118] and the MSTW2008 PDF [194]. For all samples, the showering process is performed with PYTHIA6 [115] tune Z2* [187]. The process W+jets includes decays with $W \to l\nu$ with up to two jets modelled in the matrix element. Similarly, also the process Z + jets includes up to two jets. Here, decays $Z \to \nu\nu$ are modelled. For these two processes, cross sections at NLO are obtained by applying scale factors computated with POWHEG to the leading order cross sections. These scale factors amount to 1.04 and 1.07 for W + jets and Z + jets, respectively. The cross section calculation for $t\bar{t}$ events is obtained from HATHOR [204].
In order to study signal events, two different processes are generated with MADGRAPH5 and showered with PYTHIA6. On the one hand, direct top squark 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 for top squark masses listed in Tab. 8.1 for 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 $\tilde{g} \to t\bar{t}\tilde{\chi}_0^1$. For the top quark, only fully hadronic decays are simulated. The cross section given in Tab. 8.1 for gluino pair production is corrected for the hadronic branching fraction of the top quark. In this case, the neutralino mass is fixed to 50 GeV.
All samples are normalized to an integrated luminosity of $19.5\,\text{fb}^{-1}$ which corresponds to the same integrated luminosity as recorded at $\sqrt{s} = 8\,\text{TeV}$.

8.2 Sensitivity of a Basic Selection using H_T and $\not\!\!E_\text{T}$

The targeted signature of the direct pair production of top squarks in the hadronic channel is based on jets and missing transverse momentum similar to the search presented in Chap. 7. Thus, a very similar baseline selection is employed as a basis for further studies. If not stated otherwise, jets are clustered with the anti-k_T algorithm and a distance parameter $R = 0.5$ including charged-hadron subtraction. Since no dedicated jet energy corrections for $\sqrt{s} = 13\,\text{TeV}$ have been determined, jets are corrected with the respective correction factors for $\sqrt{s} = 8\,\text{TeV}$. The applied analysis criteria are:

- Background contributions arising from $t\bar{t}$ and W+jets events are reduced by rejecting events containing isolated electrons or muons with $p_\text{T} > 10\,\text{GeV}$ and $|\eta| < 2.5$. These are required to have a good quality track that can be associated with the primary interaction vertex [176, 177]. 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, and 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 $p_\text{T} > 50\,\text{GeV}$ and $|\eta| < 2.5$. This requirement is imposed in order to select multijet events as expected from the decays of the two top quarks.

- The scalar sum of jet momenta (H_T) is required to be $\geq 500\,\text{GeV}$ with

$$H_\text{T} = \sum_\text{jets} p_\text{T}$$

  for all jets with $p_\text{T} > 50\,\text{GeV}$ 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 missing transverse energy $\not\!\!E_\text{T}$ calculated from the PF candidates is required to be $\geq 200\,\text{GeV}$. This selection reduces contributions from standard model processes in which missing transverse momentum is expected to be small. Especially, QCD multijet background is suppressed.

- In order to suppress events in which missing transverse energy is mainly arising from jet mismeasurements, as for QCD multijet events, it is required that $\not\!\!E_\text{T}$ is not aligned with any of the leading three jets. Thus, events with

$$\Delta\phi(\text{jet}_n, \not\!\!E_\text{T}) > 0.5 \;\text{ for }\; n = 1, 2 \;\text{ and }\; \Delta\phi(\text{jet}_3, \not\!\!E_\text{T}) > 0.3$$

  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 7.1.3). The selection described above is again denoted *baseline selection* in the following. In Fig. 8.1, the obtained spectra of H_{T}, $\not{E}_{\mathrm{T}}$ and N_{Jets} after applying the baseline selection are shown for the SM backgrounds and two selected signal points normalized to an integrated luminosity of $19.5\,\mathrm{fb}^{-1}$. The signal points represent mass values of 600 GeV and 1100 GeV for the top squark 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 top squark masses. Typically, the H_{T} and $\not{E}_{\mathrm{T}}$ spectrum get harder for higher top squark masses while the shape of the N_{Jets} spectrum stays nearly unchanged. After applying baseline selection criteria, the background is composed almost equally of all four SM processes ($t\bar{t}$: 27%, W + jets: 21%, Z + jets: 19%, QCD: 33%).
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 selection efficiencies $\epsilon_{\mathrm{sig/bg}}$ are defined according to

$$\epsilon = \frac{\text{no. of selected events}}{\text{no. of all events}} \tag{8.1}$$

for the number of signal and background events, respectively. In Fig. 8.2, the evolution of the signal versus background efficiencies is shown for the top squark mass 600 GeV and 1100 GeV signal points and $t\bar{t}$ as well as total background events. The curves are 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}, $\not{E}_{\mathrm{T}}$ and N_{Jets} is compared. As can be seen in Fig. 8.2, variations of the $\not{E}_{\mathrm{T}}$ selection show in general the best performance when comparing these variables. This holds for low and high top squark masses as well as for the total background and when considering $t\bar{t}$ background only. Furthermore, also H_{T} provides a good separation power and is only in case of the top squark mass of 600 GeV inferior to N_{Jets} when considering the total background. 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, $\not{E}_{\mathrm{T}}$ 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 Chap. 7 is pursued. Events selected with the baseline selection are further categorized according to their H_{T} and $\not{E}_{\mathrm{T}}$ values in exclusive search regions shown in Tab. 8.2. The event yields of background and two signal processes obtained after applying baseline requirements are shown for these different signal regions as well as for the total selected sample in Tab. 8.3. Furthermore, also the signal over background ratios are displayed. In general, the signal to background ratio in SR1 is very low such that this bin is not expected to contribute to the search sensitivity significantly, but can be used to constrain the background estimate. As seen already from the signal versus background efficiency curves, the best discrimination between signal and background events can be obtained by selecting high values of $\not{E}_{\mathrm{T}}$ (SR2 and SR4), while the overall best sensitivity is given when combining the high $\not{E}_{\mathrm{T}}$ selection also with the high H_{T} selection (SR4). However, the signal to background ratios in all

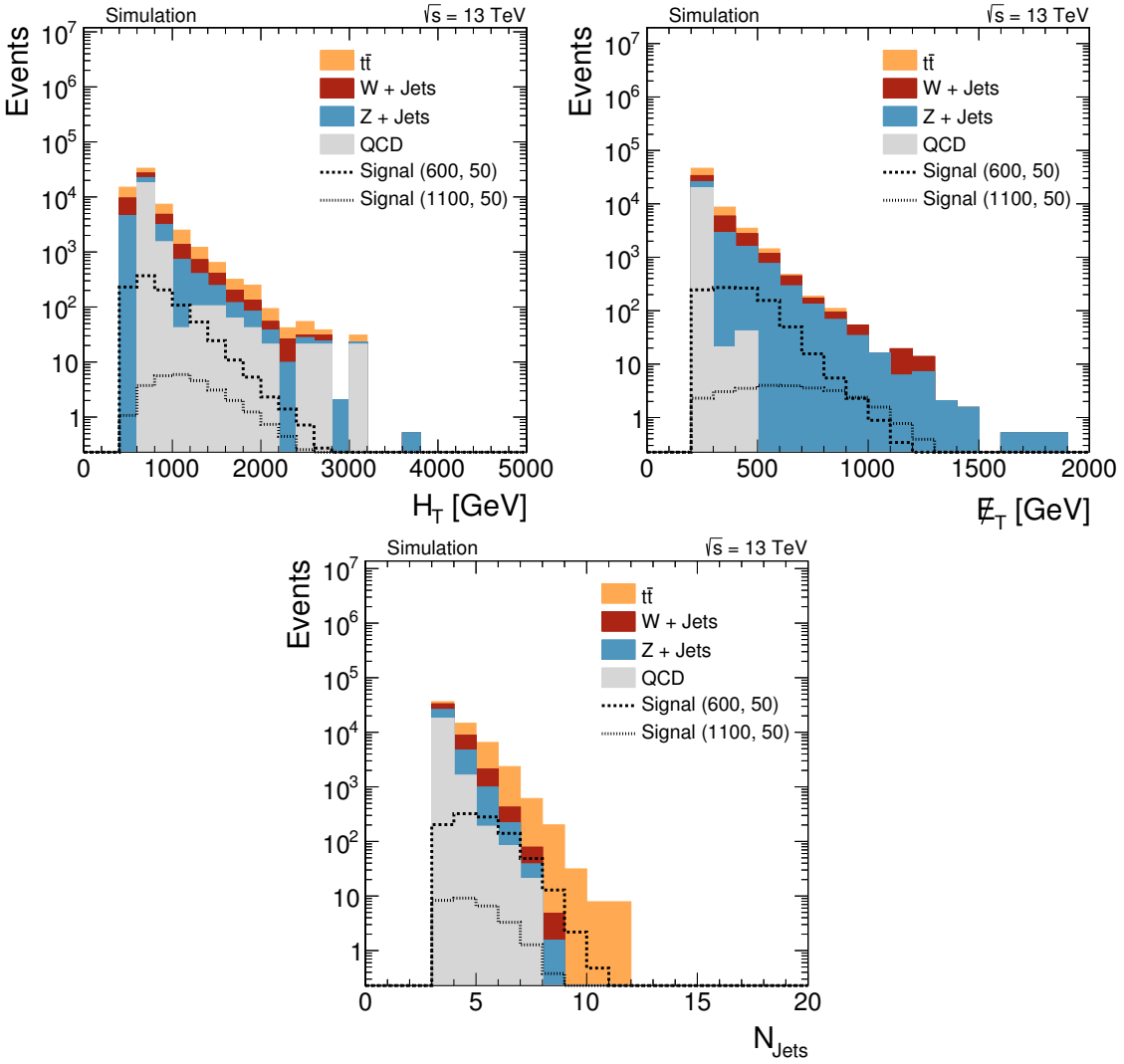

Figure 8.1: Comparison of selected H_{T} (*top left*), $\not{E}_{\mathrm{T}}$ (*top right*) and N_{Jets} (*bottom*) distributions in simulated events found from applying the baseline selection criteria. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

search regions are so low that it is not expected to be able to really probe the considered top squark mass scenarios with such a basic selection.

In order to get an even better estimate of the search sensitivity than from the signal to background ratios of the individual search regions, the sensitivity of a certain selection is quantified by determining the expected exclusion reach. While the signal to background ratios are useful to get a general impression of the selection quality, the expected exclusion reach combines the information from all search regions and thus provides a more accurate estimate of the search sensitivity. Thus, it is used in the following to quantify the quality of a certain selection and compare it to others.
In order to determine the exclusion reach, the uncertainties of the individual background

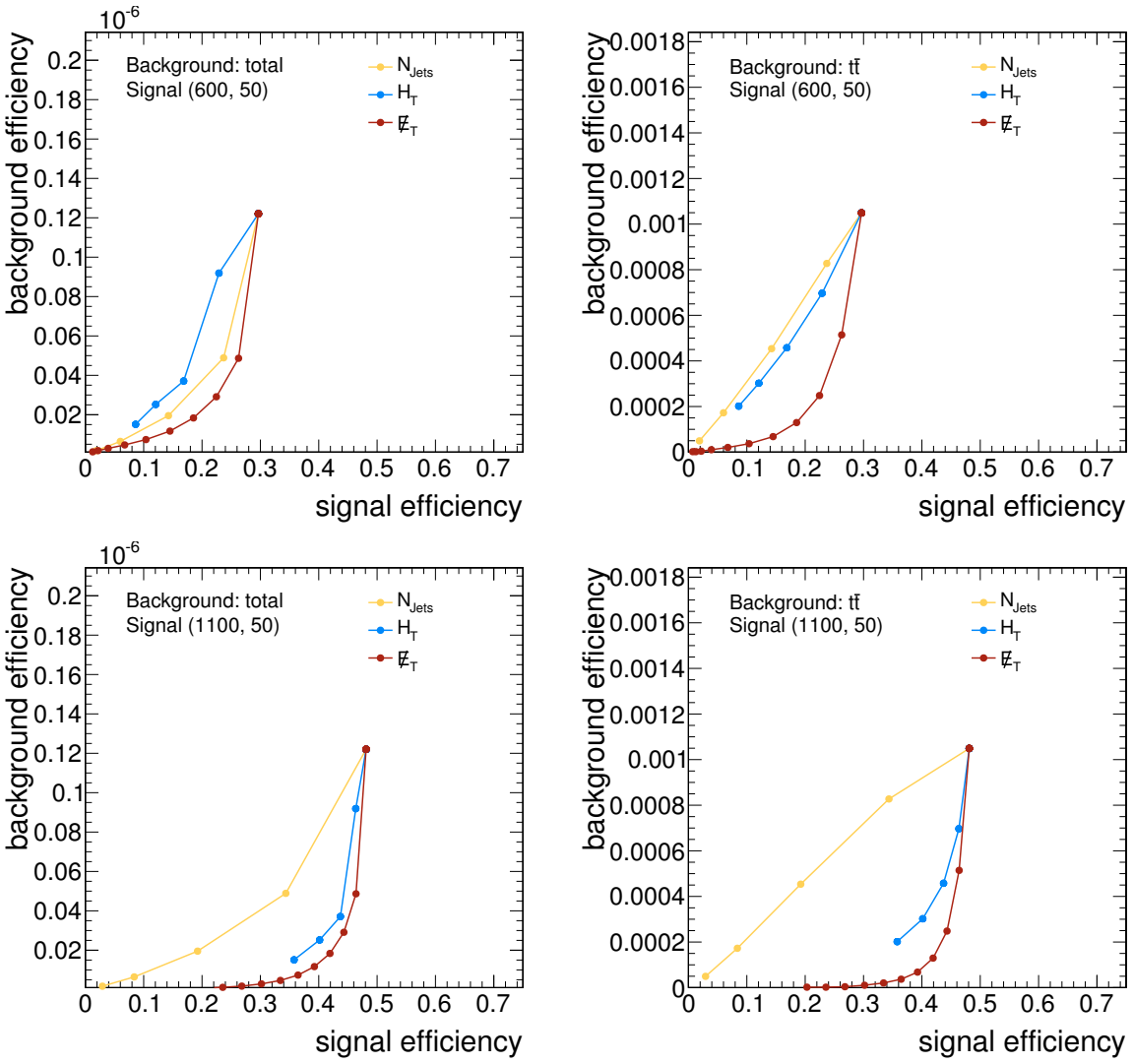

Figure 8.2: Evolution of the signal versus background efficiency for a top squark mass of 600 GeV (*top*) and 1100 GeV (*bottom*) in case the background is the sum of all backgrounds (*left*) or only the $t\bar{t}$ background (*right*) after applying baseline selection criteria.

Table 8.2: Exclusive search regions (SR) used in the analysis binned in H_T and $\not{E}_\mathrm{T}$.

SR1	$500 < H_\mathrm{T} < 1000\,\mathrm{GeV}$	$200 < \not{E}_\mathrm{T} < 400\,\mathrm{GeV}$
SR2	$500 < H_\mathrm{T} < 1000\,\mathrm{GeV}$	$400\,\mathrm{GeV} < \not{E}_\mathrm{T}$
SR3	$1000\,\mathrm{GeV} < H_\mathrm{T}$	$200 < \not{E}_\mathrm{T} < 400\,\mathrm{GeV}$
SR4	$1000\,\mathrm{GeV} < H_\mathrm{T}$	$400\,\mathrm{GeV} < \not{E}_\mathrm{T}$

Table 8.3: Total event yields obtained from simulated samples after the baseline selection described in the text (*first column*) as well as event yields for the various signal regions (*column two to five*). All numbers are scaled to 19.5 fb^{-1}. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV. Furthermore, the signal over background ratios are displayed for the two signal points in squared brackets.

	total	SR1	SR2	SR3	SR4
$t\bar{t}$	16461	13525	754	1868	314
W + jets	12481	9660	1516	985	320
Z + jets	11837	8155	2425	810	447
QCD multijet	20013	19574	0	397	42
Signal (600, 50)	1012 [$16.6{\cdot}10^{-3}$]	416 [$8.2{\cdot}10^{-3}$]	389 [$82.8{\cdot}10^{-3}$]	101 [$24.9{\cdot}10^{-3}$]	106 [$94.6{\cdot}10^{-3}$]
Signal (1100, 50)	29 [$0.5{\cdot}10^{-3}$]	2 [$0.05{\cdot}10^{-3}$]	8 [$1.7{\cdot}10^{-3}$]	3 [$0.7{\cdot}10^{-3}$]	16 [$14.0{\cdot}10^{-3}$]

processes have to be considered. These are not explicitly estimated but chosen in correspondence to the uncertainties obtained for similar kinematic regimes in SUS-13-015. The respective uncertainties taken into account for the different processes are

- QCD multijet events: 100%
- Z + jets: 50%
- W + jets: 20%
- $t\bar{t}$: 20% + additional 20% in the high $\not{E}_{\text{T}}$ search regions (SR2, SR4)

These are treated as total uncertainties which means that the relative rate is assumed to be known for the first three backgrounds and allowed to vary within the additional 20% for $t\bar{t}$ background. 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 [205]. The obtained exclusion curve is shown in Fig. 8.3 for the signal strength μ which is the excluded production cross section divided by the theoretical cross section for direct top squark production as a function of the top squark mass.[1] A particular mass point can be excluded if the expected limit drops below one. As expected from the signal to background ratios, this baseline selection cuts and the subsequent binning in exclusive search 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.

8.3 Improvement using B Tagging

As discussed above, the targeted signal final state involves the presence of bottom quarks emerging from the decay of the top quarks. Thus, an obvious option to enhance the

[1] As stated in the beginning, the main focus of these studies is on scenarios with large mass differences between top squark and LSP. Thus, the LSP mass is fixed to 50 GeV here. The sensitivity of the analysis towards other LSP masses is discussed later in Sec. 8.9.

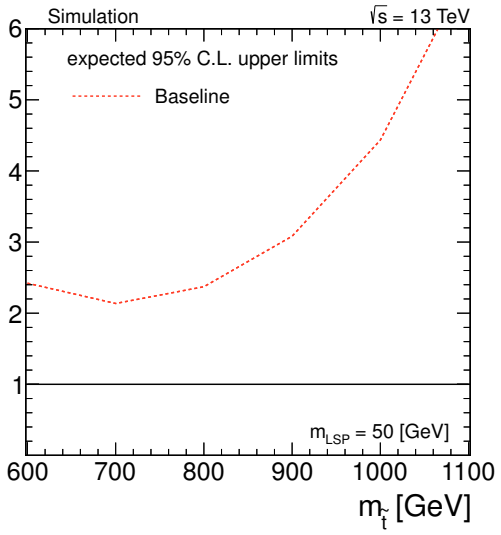

Figure 8.3: Expected 95% C.L. upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV.

sensitivity of such an analysis is to employ b-tagging techniques to identify 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 Sec. 5.3.
In this analysis, b-quark jets are identified based on the CSV algorithm using the medium working point. Furthermore, they are required to have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ in order to be not sensitive to potentially large flavour-dependent JEC uncertainties for jet transverse momenta smaller than 30 GeV. In Fig. 8.4, the b-tag multiplicity, i. e. the number of b-tagged jets in the event, is illustrated. When imposing in addition to the baseline selection also a requirement of at least one b-tagged jet, the total signal efficiencies for the signal samples with top squark masses of 600 GeV and 1100 GeV decrease from around 30% and 48% to 25% and 38%, respectively. However, also the total background efficiency decreases significantly from $1.2{\cdot}10^{-7}$ to $3.6{\cdot}10^{-8}$. Concerning background events, the main reduction occurs for $W+$jets, $Z+$jets and QCD events as these contain in most cases no b-tagged jet. The relative background composition after applying the b-tag requirement is: 77.5% $t\bar{t}$, 8.8% $W+$jets, 8.6% $Z+$jets and 5.1% QCD events.
Because of the improved signal to background ratios, also the expected exclusion limit of the analysis is expected to improve when applying the b-tag requirement in addition to the baseline selection. The same exclusive search regions as defined in Tab. 8.2 are used considering the same total uncertainties as described in Sec. 8.2 for a performance comparison of this improved selection with respect to the baseline requirements. The expected limit for the baseline selection including the additional b-tag requirement is shown in Fig 8.5.
It turns out that the b-tag requirement significantly improves the sensitivity of the analysis for the whole specified top squark mass range and μ drops below one for lower masses so that these can be probed with such an analysis strategy. However, since the focus of this analysis is put on higher top squark masses, it is discussed in the next section how this mass range can be adressed better.

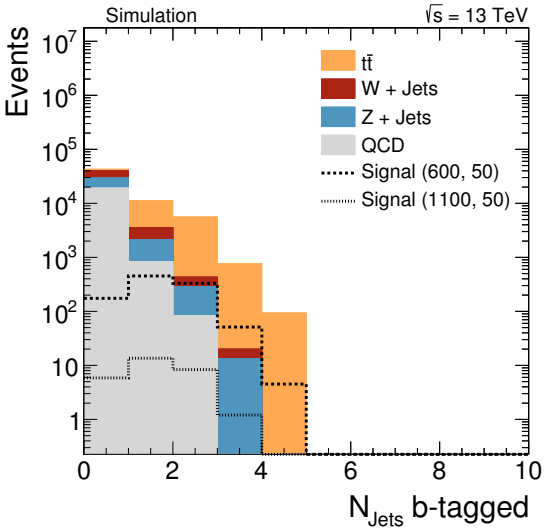

Figure 8.4: B-tag multiplicity after applying the baseline selection in simulated events. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

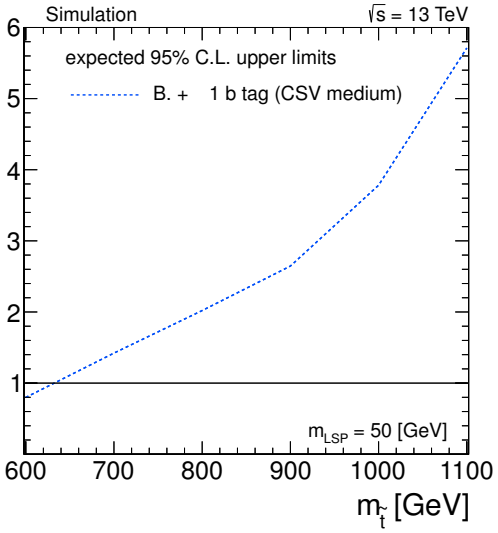

Figure 8.5: Expected 95% C.L. upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV. The label B. denotes the baseline selection.

8.4 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 hadronically decaying top quark is illustrated in Fig. 8.6 for $t\bar{t}$ background and two selected signal points without applying any selection criteria. As expected, the p_{T} spectrum of the signal is significantly harder than that of the $t\bar{t}$ background. For instance for a top squark mass of 1100 GeV the maximum lies at a p_{T} range around 400–500 GeV. As discussed in Sec. 5.4, decay products emerging from the

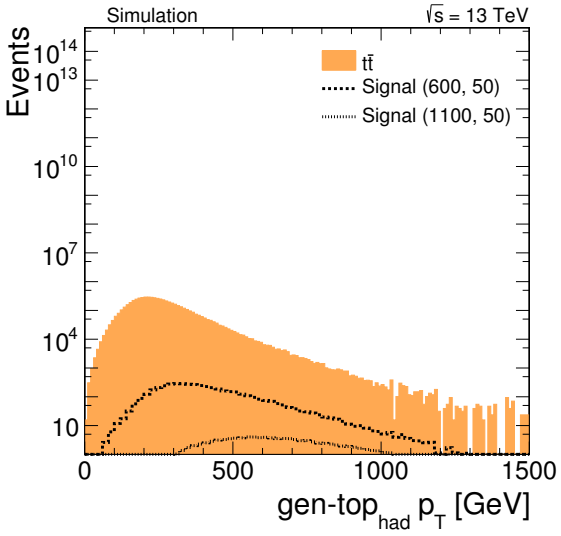

Figure 8.6: Transverse momentum spectrum of the leading generated hadronically decaying top quark without applying any selection criteria. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

decay of a top quark with large transverse momentum can be reconstructed as a single jet with large radius parameter. Following Eq. 5.4, the opening angle of decay products from a top quark with transverse momentum between 400–500 GeV is expected to be $R = 0.8$. Thus, such topologies are well suited to utilize the top tagging techniques described in Sec. 5.4.
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-tagging 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 efficiencies as a function of the transverse momentum of the generated hadronically decaying top quark are shown in the left part of Fig. 8.7 for both the CMS and the HEP Top Tagger. It is visible that the turn-on of the HEP Top Tagger starts already around a p_{T} of 200 GeV while the CMS Top Tagger begins to become efficient not before around a p_{T} of 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

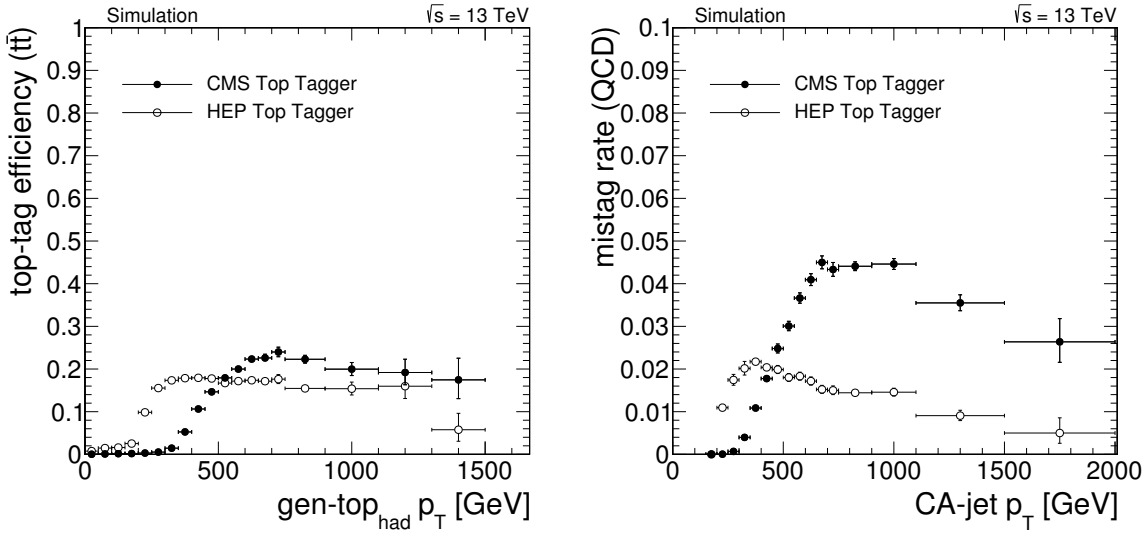

Figure 8.7: Top-tagging efficiency for $t\bar{t}$ events as a function of the transverse momentum of the generated hadronically decaying top quark (*left*) and misidentification rate for QCD multijet events as a function of the CA-jet p_{T} (*right*). In case of the CMS Top Tagger CA-jets with a distance parameter of $R = 0.8$ are used while the HEP Top Tagger is based on CA-jets with $R = 1.5$.

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-tagged CA-jets by the number of all CA-jets. The misidentification rate as a function of the respective CA-jet transverse momentum is shown in the right part of Fig. 8.7. 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.
Often, in analyses based on top quarks with moderate transverse momenta, that 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, if the main background contains actual top quarks, as it is the case for top squark searches, the more important property is a good top-tagging efficiency for high top quark momenta in order to exploit the difference in the p_{T} spectrum between top quarks from SM processes and top quarks stemming from top squark decays. Here, the misidentification rate only plays a less important role.

The impact on the analysis when applying top-tagging requirements can be studied when looking for instance at the top-tag multiplicity. In Fig. 8.8, 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. Although a significant amount of signal events does not have a CA-jet identified as top-jet by the respective algorithm, the distributions exhibit that

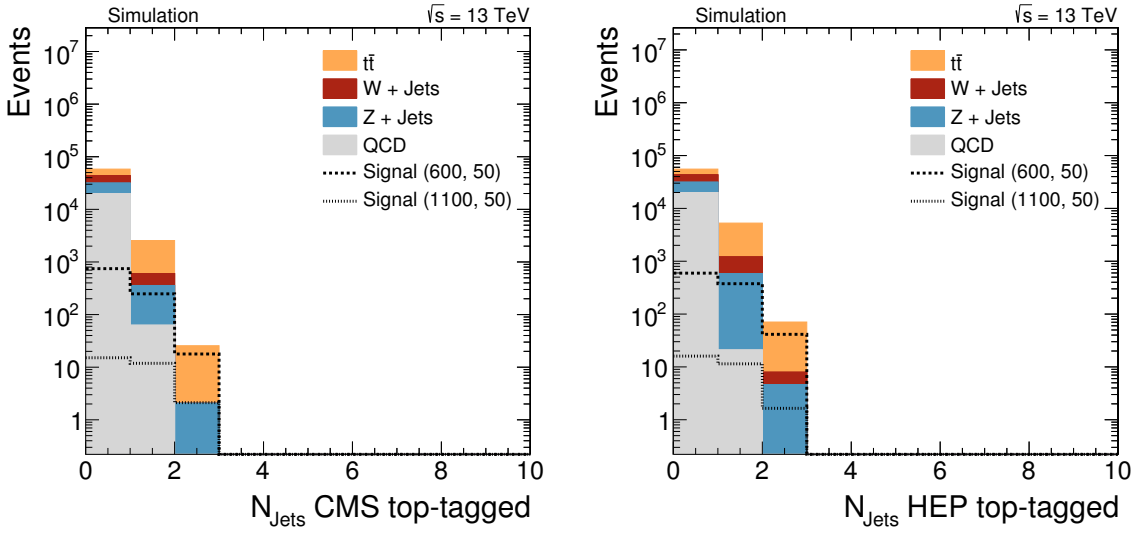

Figure 8.8: Top-tag multiplicity for the CMS Top Tagger (*left*) and the HEP Top Tagger (*right*) after application of the baseline selection. Only CA-jets of the corresponding jet size with a transverse momentum above 150 GeV are considered.

the background can be significantly reduced when requiring at least one top-tagged jet. The respective signal and background efficiencies, when using in addition to the baseline selection a requirement of at least one top-tagged jet, evolve as follows when considering the total efficiencies not separated according to the individual search regions:

- HEP Top Tagger:
  $\epsilon_{\text{sig}} = 12\%$ ($m_{\tilde{t}} = 600\,\text{GeV}$), $\epsilon_{\text{sig}} = 22\%$ ($\text{m}_{\tilde{t}} = 1100\,\text{GeV}$)
  $\epsilon_{\text{bg}} = 1.1 \cdot 10^{-6}$ % (total bg.) , $\epsilon_{\text{bg}} = 0.03\%$ ($t\bar{t}$ bg.)
- CMS Top Tagger:
  $\epsilon_{\text{sig}} = 8\%$ ($m_{\tilde{t}} = 600\,\text{GeV}$), $\epsilon_{\text{sig}} = 23\%$ ($\text{m}_{\tilde{t}} = 1100\,\text{GeV}$)
  $\epsilon_{\text{bg}} = 0.5 \cdot 10^{-6}$ % (total bg.) , $\epsilon_{\text{bg}} = 0.01\%$ ($t\bar{t}$ bg.)

In general, the signal efficiency for low masses is larger when using the HEP Top Tagger than for the CMS Top Tagger and comparable for the higher top squark mass. However, also the background efficiency shows a larger value for the HEP Top Tagger case and in particular more $t\bar{t}$ events are selected. Thus, the CMS Top Tagger is expected to provide a better search sensitivity.
In order to test the impact of the top-tagging requirements on the search sensitivity more quantitatively, the expected limits are derived again based on the search regions defined in Tab. 8.2 with the same uncertainties as assumed in Sec. 8.2. The results are illustrated in Fig 8.14. As expected, 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. The improvement amounts to a factor of two to three. Hence, mass points up to $(m_{\tilde{t}}, m_{\text{LSP}}) = (700, 50)\,\text{GeV}$ or even $(800, 50)\,\text{GeV}$ can be probed with such selections. This selection is also more sensitive than the result obtained when adding a b-tag requirement, as discussed in Sec. 8.3, which allows to probe top squark masses around 600 GeV for an LSP mass of 50 GeV. In general, the selection based on the CMS

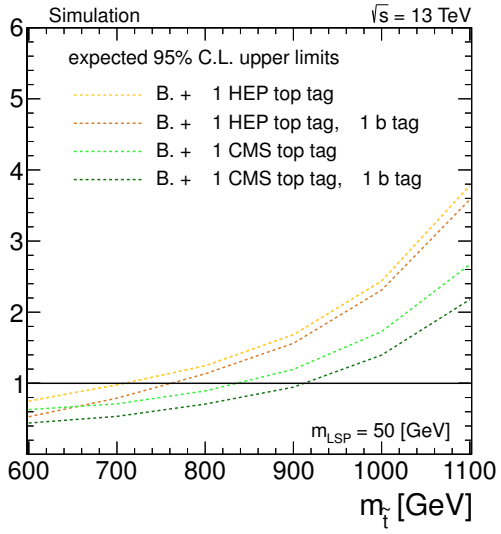

Figure 8.9: Expected 95% C.L. upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV. The label B. denotes the baseline selection.

Top Tagger performs better than the selection involving the HEP Top Tagger since the signal to background ratio is higher as discussed above.
Furthermore, the selections including the top-tagging requirements are also compared when imposing an additional b-tag requirement as defined in Sec. 8.3. In both cases, the b-tag requirement further improves the sensitivity. The selection involving the CMS Top Tagger still performs better. Thus, selections utilizing the HEP Top Tagger are not pursued in the following.

8.5 Performance Comparison of Various Kinematic Selections

The sensitivity of the analysis targeting direct top squark production can be improved already significantly when employing in addition to the baseline criteria also b-tag and top-tag requirements as shown in Sec. 8.3 and 8.4. 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 more detail. In Fig. 8.10, 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 are the semi-leptonic top decays which contain electrons (channel two), muons (channel three) and tau leptons (channel four). Dileptonic $t\bar{t}$ events are not displayed, as these are found to be negligible. The distributions exhibit that after the application of the top-tagging requirement, similarly to the b-tag requirement, the main background contribution arises from $t\bar{t}$ events (around 77%). 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

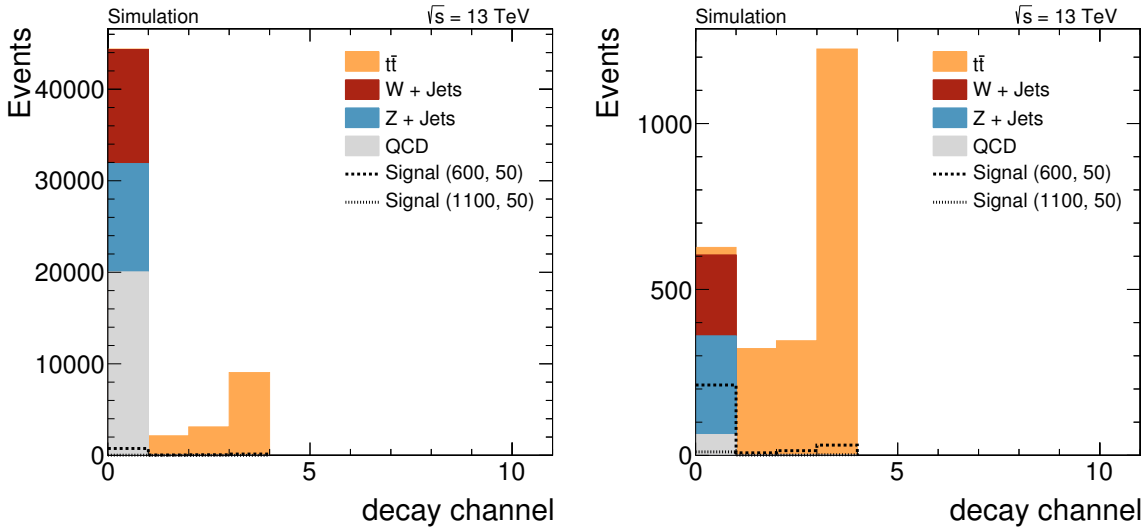

Figure 8.10: Comparison of decay modes of signal and background processes defined according to the top quark properties based on generator information after the baseline selection (*left*) and after the baseline selection and the requirement of ≥ 1 CMS top tag (*right*). For a definition of the decay channels see text.

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:

M_{T2} : The M_{T2} variable represents a generalized version of the transverse mass [180, 206–208]. In events with pair-produced particles which decay further and eventually contain undetectable particles in the decay products, 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

$$M_{T2}(m_{\tilde{\chi}}) = \min_{\vec{p}_T^{\,\tilde{\chi}(1)} + \vec{p}_T^{\,\tilde{\chi}(2)} = \vec{p}_T^{\,\text{miss}}} \left[\max \left(M_T^{(1)}, M_T^{(2)} \right) \right] \tag{8.2}$$

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

$$(M_T^{(i)})^2 = (m^{\text{vis}(i)})^2 + m_{\tilde{\chi}}^2 + 2 \left(E_T^{\text{vis}(i)} E_T^{\tilde{\chi}(i)} - \vec{p}_T^{\,\text{vis}(i)} \cdot \vec{p}_T^{\,\tilde{\chi}(i)} \right) \tag{8.3}$$

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 $\vec{p}_T^{\,\tilde{\chi}(i)}$ with mass $m_{\tilde{\chi}}$. 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 $\vec{p}_T^{\,\text{miss}}$, the missing transverse momentum,[2] is performed. This minimization is carried out

[2] More commonly denoted by $\vec{\not{E}}_T$ in this thesis.

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 the studies shown here, the two visible systems are assumed to be described by the two leading CA8 jets in the event and the mass of the neutralinos is considered to be zero. The calculation of M_{T2} is performed according to [209].
The obtained distributions for M_{T2} in background and signal samples are shown in Fig 8.11 (top right). They exhibit that the maximum of M_{T2} is higher for signal events than for background such that this variable is suitable for distinguishing signal and background events.

Razor variables: The kinematic razor variables are used to describe the generic process of pair production of two heavy particles which subsequently decay into visible products usually represented by (large) jets and undetected particles [181, 210, 211]. 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

$$M_{\mathrm{R}} = \sqrt{\left[(|\vec{p}^{\,j_1}| + |\vec{p}^{\,j_2}|)^2 - (p_z^{j_1} + p_z^{j_2})^2\right]} \tag{8.4}$$

$$M_{\mathrm{T}}^{\mathrm{R}} = \sqrt{\frac{1}{2}\left(\not{E}_{\mathrm{T}}(p_{\mathrm{T}}^{j_1} + p_{\mathrm{T}}^{j_2}) - \vec{\not{E}}_{\mathrm{T}} \cdot (\vec{p}^{\,j_1} + \vec{p}^{\,j_2})\right)} \tag{8.5}$$

$$R = \frac{M_{\mathrm{T}}^{\mathrm{R}}}{M_{\mathrm{R}}} \tag{8.6}$$

with the transverse momenta of the two jets p_{T}^{j1} and p_{T}^{j2}. For signal events, $M_{\mathrm{T}}^{\mathrm{R}}$ has an endpoint and R a maximum of approximately one.
For these studies, the two leading CA8 jets and $\not{E}_{\mathrm{T}}$ are used to compute the razor variables. In Fig. 8.11 the variable R^2 is illustrated (bottom) after the application of baseline selection criteria.

α_{T} : The α_{T} variable is mainly used to reject QCD multijet events which have no intrinsic $\not{E}_{\mathrm{T}}$ [178, 212]. In case of dijet events, it is defined as:

$$\alpha_{\mathrm{T}} = \frac{E_{\mathrm{T}}^{j_2}}{M_{\mathrm{T}}} \text{ and } M_{\mathrm{T}} = \sqrt{\left(\sum_{i=1}^{2} E_{\mathrm{T}}^{j_i}\right)^2 - \left(\sum_{i=1}^{2} p_{\mathrm{x}}^{j_i}\right)^2 - \left(\sum_{i=1}^{2} p_{\mathrm{y}}^{j_i}\right)^2} \tag{8.7}$$

with the transverse energy $E_{\mathrm{T}}^{j_2}$ 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 fat jet, the α_{T} variable is in these 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. 8.11 (top left). Typically, α_{T} is 0.5 for an ideal dijet event with $E_{\mathrm{T}}^{j_1} = E_{\mathrm{T}}^{j_2}$ 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 $\not{E}_{\mathrm{T}}$ as for instance resulting from LSPs.

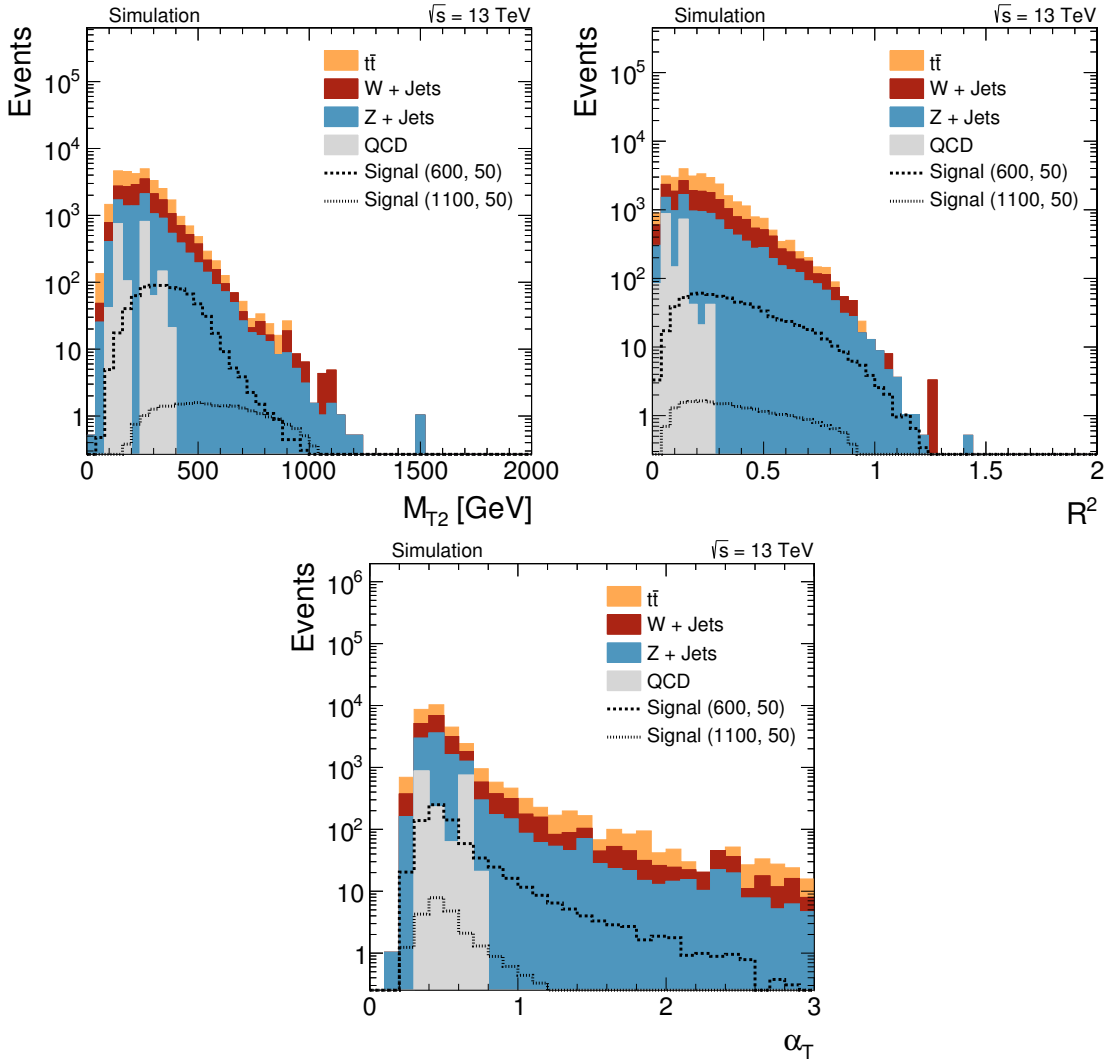

Figure 8.11: Distribution of the M_{T2} (*top left*), the razor variable R^2 (*top right*) and the α_T variable (*bottom*) after application of the baseline selection requirements for background and two selected signal samples. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

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

Transverse mass $\mathbf{m_T}$: As shown in Fig. 8.10, background contributions from $t\bar{t}$ events arise predominantly from semi-leptonic top quark decays. In such decays, the missing transverse energy mainly arises 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 accompanied by a b-quark jet also stemming from the top decay.

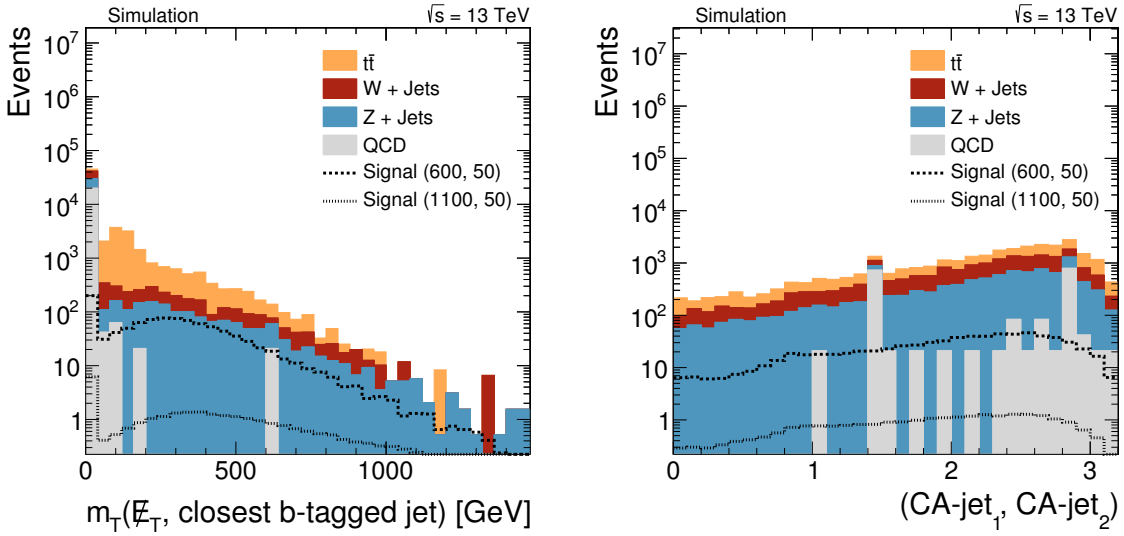

Figure 8.12: Distribution of the m_{T} variable (*left*) and $\Delta\phi(\mathrm{CA\text{–}jet}_1, \mathrm{CA\text{–}jet}_2)$ (*right*) after application of the baseline selection requirements for background and two selected signal samples. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass.

Thus, the missing transverse energy and the closest b-tagged jet in $\Delta\phi$ can be utilized to calculate a transverse mass according to

$$m_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{\mathrm{jet}}\not{E}_{\mathrm{T}} \cdot (1 - \cos[\Delta\phi(\mathrm{jet}, \not{E}_{\mathrm{T}})])}\,. \tag{8.8}$$

The distribution obtained after applying the baseline selection is shown for background and two selected signal samples in Fig. 8.12 (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. The transverse mass is considered 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 shifted to higher values than the top-quark mass for signal events.

$\Delta\phi(\mathrm{CA\text{–}jet_1}, \mathrm{CA\text{–}jet_2})$: The selection of events with a back-to-back topology has been used in Chap. 6 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 fat jets balanced against each other. For signal events however, such a topology is not expected since the event is balanced against genuine $\not{E}_{\mathrm{T}}$ arising from the LSPs. The selected $\Delta\phi$ distributions after applying baseline selection criteria are shown in Fig. 8.12 (right).

In order to quantify the quality of these various kinematic selections, the separation power of the different variables is tested by studying again the background versus signal efficiency curves as introduced in Sec. 8.2. The resulting curves after applying baseline selection

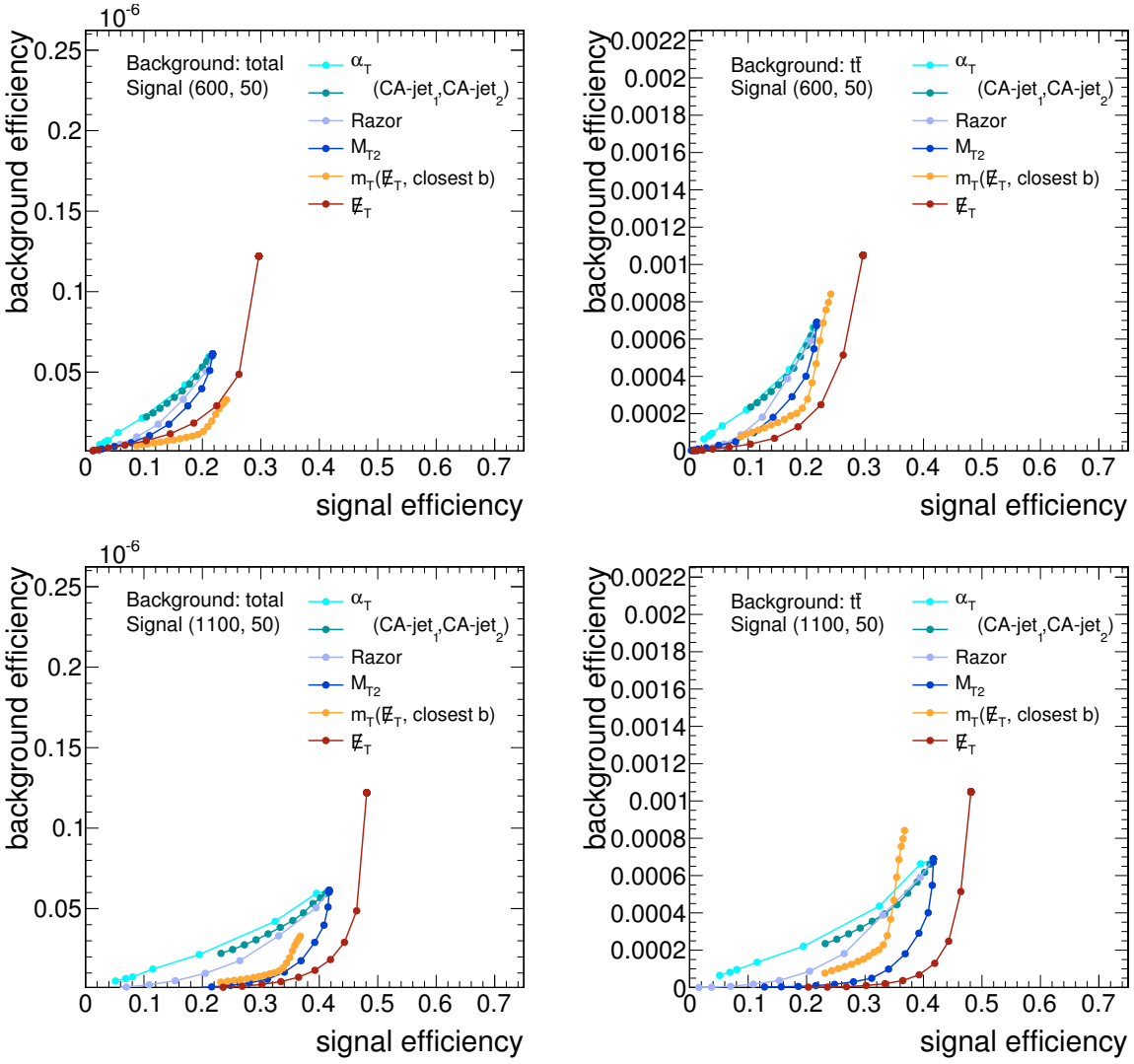

Figure 8.13: Evolution of the signal versus background efficiency for a top squark mass of 600 GeV (*top*) and 1100 GeV (*bottom*) in case the background is the sum of all backgrounds (*left*) or the $t\bar{t}$ background (*right*) after applying baseline selection criteria. The definitions of the variables α_{T}, $\Delta\phi(\mathrm{CA\text{–}jet}_1, \mathrm{CA\text{–}jet}_2)$, razor and M_{T2} imply a requirement of at least two CA-jets and the definition of m_{T} implies a requirement of at least one b-tagged jet.

criteria are shown in Fig. 8.13. For comparison, also the signal versus background efficiency curve for $\not\!\!E_{\mathrm{T}}$ is illustrated.
The calculation of α_{T}, $\Delta\phi(\mathrm{CA\text{–}jet}_1, \mathrm{CA\text{–}jet}_2)$, M_{T2} and the razor variable R^2 imply that there exist two CA8 jets with $p_{\mathrm{T}} > 150$ GeV in the event. Consequently, the starting point of that scan curves is different than that of the $\not\!\!E_{\mathrm{T}}$ curve which represents the signal and background efficiencies after the baseline selection. Similarly the m_{T} curve implies that each event has at least one b-tagged jet. The point of highest efficiency for the m_{T} curve corresponds to a requirement of $m_{\mathrm{T}} > 20$ GeV.

Table 8.4: Signal over background ratios are displayed for two signal points labelled as (X, Y), where X is the top squark mass and Y is the LSP mass in GeV, in the four exclusive search regions. For definitions of selections see text.

		S/B			
		top tag	top tag + M_{T2}	top tag + b tag	top tag + m_T
SR1	Signal (600, 50)	0.04	0.2	0.05	0.1
	Signal (1100, 50)	0.0004	0.002	0.0004	0.001
SR2	Signal (600, 50)	0.3	0.3	0.5	0.7
	Signal (1100, 50)	0.008	0.01	0.01	0.02
SR3	Signal (600, 50)	0.08	0.2	0.1	0.2
	Signal (1100, 50)	0.003	0.01	0.003	0.007
SR4	Signal (600, 50)	0.6	0.8	1.1	2.7
	Signal (1100, 50)	0.1	0.2	0.2	0.5

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 and outperforms M_{T2} especially for smaller top squark 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.

In order to compare the quality of these different criteria not only for the total signal and background efficiencies, the signal over background ratios are summarized in Tab. 8.4 for the four different search regions. Here, the signal over background ratios are shown for the baseline selection including at least one CMS top-tagged jet (denoted *top tag*) and when adding in addition to these requirement $M_{T2} > 400\,\text{GeV}$ (*top tag* + M_{T2}), at least one b-tagged jet (*top tag* + *b tag*) or $m_T > 180\,\text{GeV}$ (*top tag* + m_T). Although the signal versus background efficiency curves have indicated that the selection including M_{T2} provides the best signal to background ratios when considering the whole sample, the signal to background ratios for distinct search regions are highest in case a selection with m_T is used. For a further quantification, expected exclusion curves are derived again based on the same exclusive search regions, as defined in Tab. 8.2, considering the uncertainties for each background source, as described in Sec. 8.2. In Fig. 8.14, the derived expected exclusion limits are compared for various selections as a function of the top squark mass for a LSP mass of 50 GeV. The selection discussed in Sec. 8.3, in which at least one CMS top tag is required in addition to the baseline selection, is illustrated in light green. This selection is combined with either a selection of $M_{T2} > 400\,\text{GeV}$ (pink), a b-tag requirement as discussed already in Sec. 8.3 (dark green) or a requirement of $m_T > 180\,\text{GeV}$ (light blue). It turns out that the latter performs best among those described selections over the whole top squark mass range as indicated already by the achieved signal to background ratios.

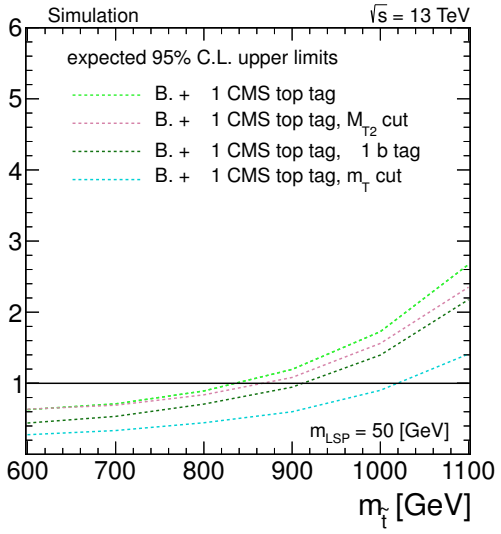

Figure 8.14: Expected 95% C.L. upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV. The label B. denotes the baseline selection.

8.6 Performance Comparison to Selection Based on Published Top Squark 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 SUS-13-015 is carried out in this section.
The comparison is done by performing the same selections as done in SUS-13-015 based on the simulated samples discussed in Sec. 8.1. Jets in this analysis are clustered with the anti-k_T algorithm with a distance parameter of $R = 0.5$ and corrected for pileup effects by applying charged-hadron subtraction. The preselection criteria used here are:

- Events with isolated electrons and muons with $p_{\mathrm{T}} > 10$ GeV, as described in Sec. 8.2, are vetoed.
- Events have to have at least five jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$. The two highest p_{T} jets further have to have $p_{\mathrm{T}} > 70$ GeV while the next two highest jets in p_{T} must fulfill $p_{\mathrm{T}} > 50$ 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 $\Delta\phi(\mathrm{jet}_n, \not{E}_{\mathrm{T}}) > 0.5$, $n = 1, 2$ and $\Delta\phi(\mathrm{jet}_3, \not{E}_{\mathrm{T}}) > 0.3$.[3]

The only difference between these selection criteria employed here and SUS-13-015 is related to the lepton veto. While in SUS-13-015 it is required to have no events with identified and isolated electrons and muons with $p_{\mathrm{T}} > 5$ GeV, here only events with electrons and muons with $p_{\mathrm{T}} > 10$ 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

[3]The missing transverse energy in the published analysis is denoted with $p_{\mathrm{T}}^{\mathrm{miss}}$.

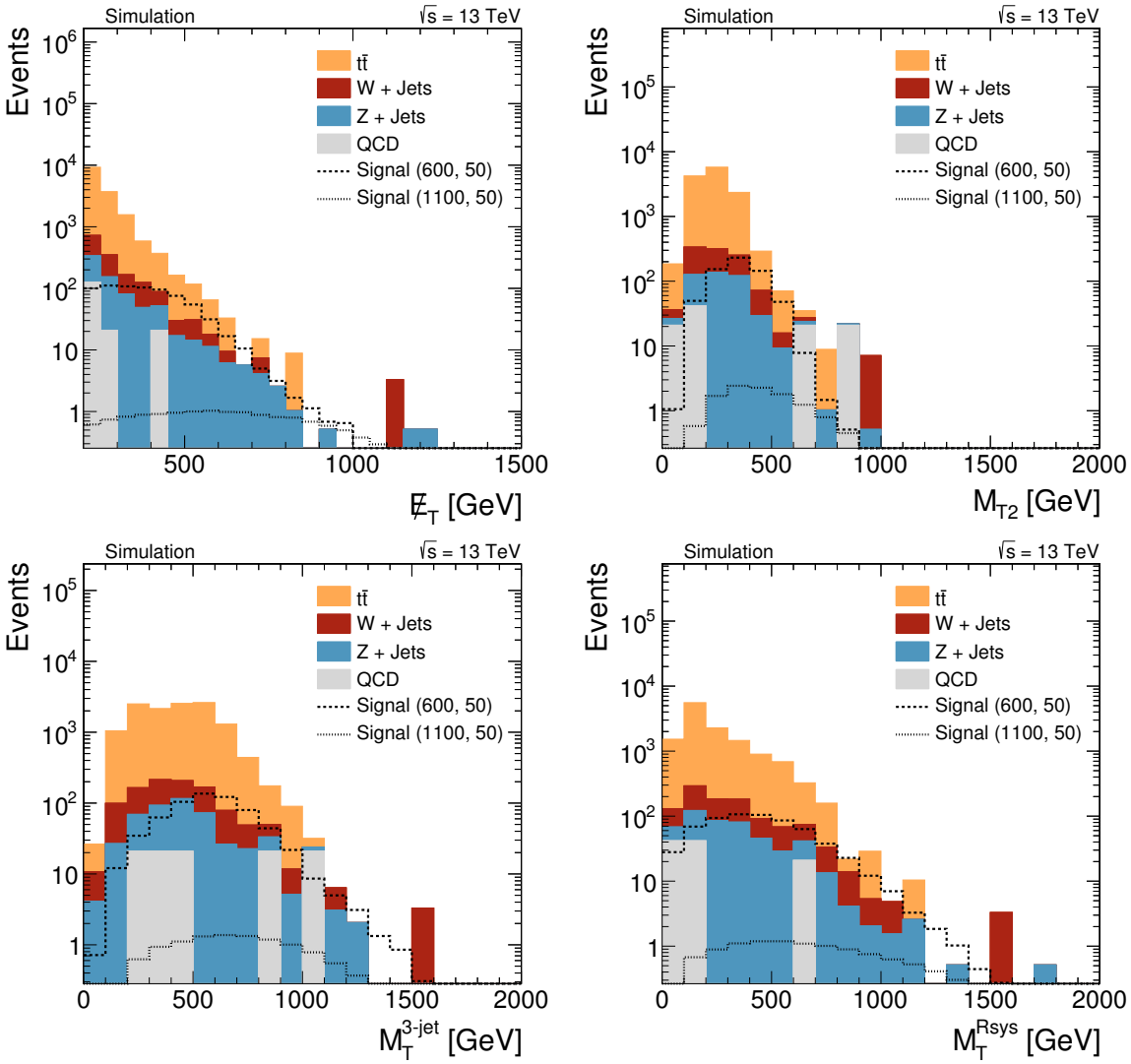

Figure 8.15: Comparison of $\not{E}_T$, M_{T2}, M_T^{3-jet} and M_T^{Rsys} distributions for signal and background events after applying preselection criteria and $\not{E}_T > 200\,GeV$. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

to compare the quality of the different selections applied to the sample after selecting the all-jet final state. In addition to those preselection 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 containing 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 (denoted *3-jet*) while the other is based on the remnant system and referred to as *partially-reconstructed* top (denoted *Rsys*). Details on the reconstruction process of the two top quark systems can be found in [203]. In particular, this process involves that the fully-reconstructed top quark has to satisfy the criteria de-

Table 8.5: Signal over background ratios are displayed for two signal points labelled as (X, Y), where X is the top squark mass and Y is the LSP mass in GeV, in the four search regions defined in SUS-13-015.

		SUS-13-015: S/B
$SR1_{ref}$	Signal (600, 50)	0.2
	Signal (1100, 50)	0.005
$SR2_{ref}$	Signal (600, 50)	0.8
	Signal (1100, 50)	0.02
$SR3_{ref}$	Signal (600, 50)	0.5
	Signal (1100, 50)	0.01
$SR4_{ref}$	Signal (600, 50)	1.2
	Signal (1100, 50)	0.03

scribed in Eq. 5.5, 5.6 and 5.7 using $R_{\text{min}} = 0.85 \cdot (m_W/m_{\text{top}})$, $R_{\text{max}} = 1.25 \cdot (m_W/m_{\text{top}})$, $m_W = 80.4\,\text{GeV}$ and $m_{\text{top}} = 173.1\,\text{GeV}$. If there is more than one trijet system satisfying these criteria, the combination with $m^{3-\text{jet}}$ closest to m_{top} is selected.
After the successful identification of the two top quark systems according to the above mentioned criteria, further topological requirements are used:

- $\not{E}_{\text{T}} > 200\,\text{GeV}$
- The variable M_{T2}, as defined in Eq. 8.2, is required to be $\geq 300\,\text{GeV}$. It is calculated from the four momenta of the fully- and the partially-reconstructed top quark as well as $\not{E}_{\text{T}}$ assuming the invisible particles to be massless.
- $(0.5 \cdot M_{\text{T}}^{3-\text{jet}} + M_{\text{T}}^{\text{Rsys}}) \geq 500\,\text{GeV}$. $M_{\text{T}}^{3-\text{jet}}$ and $M_{\text{T}}^{\text{Rsys}}$ denote the transverse mass of the fully-reconstructed and the remnant system which are calculated with the angle $\Delta\phi$ between the missing energy and the momentum vector of the three-jet or remnant system, respectively, according to

$$(M_{\text{T}}^{3-\text{jet}})^2 = (m^{3-\text{jet}})^2 + 2 \cdot (E_{\text{T}}^{3-\text{jet}} \not{E}_{\text{T}} - p_{\text{T}}^{3-\text{jet}} \not{E}_{\text{T}} \cos\Delta\phi)$$

and

$$(M_{\text{T}}^{\text{Rsys}})^2 = (m^{\text{Rsys}})^2 + 2 \cdot (E_{\text{T}}^{\text{Rsys}} \not{E}_{\text{T}} - p_{\text{T}}^{\text{Rsys}} \not{E}_{\text{T}} \cos\Delta\phi)\,.$$

A comparison of the $\not{E}_{\text{T}}$, M_{T2}, $M_{\text{T}}^{3-\text{jet}}$ and $M_{\text{T}}^{\text{Rsys}}$ distributions for signal and background events after the preselection and a requirement of $\not{E}_{\text{T}} > 200\,\text{GeV}$ is shown in Fig. 8.15. These distributions illustrate 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 overlapping search regions:

- $SR1_{ref}$: $\not{E}_{\text{T}} > 200\,\text{GeV}$, $N_{\text{b-jets}} \geq 1$
- $SR2_{ref}$: $\not{E}_{\text{T}} > 350\,\text{GeV}$, $N_{\text{b-jets}} \geq 1$

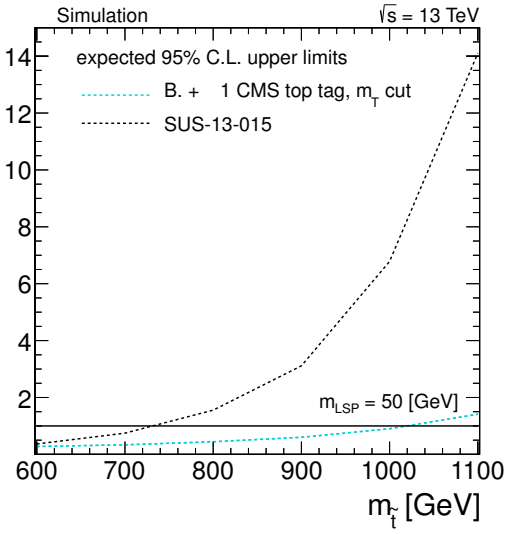

Figure 8.16: Expected 95% C.L. upper limit for signal strength versus $m_{\tilde{t}}$. The LSP mass is chosen to be 50 GeV. The label B. denotes the baseline selection.

- SR3_{ref}: $\not{E}_{\text{T}} > 200\,\text{GeV}$, $N_{\text{b-jets}} \geq 2$
- SR4_{ref}: $\not{E}_{\text{T}} > 350\,\text{GeV}$, $N_{\text{b-jets}} \geq 2$

The respective signal over background ratios for these regions are summarized in Tab. 8.5 for two signal points. The best sensitivity is provided by the signal region defined by $\not{E}_{\text{T}} > 350\,\text{GeV}$ and $N_{\text{b-jets}} \geq 2$ for these signal mass scenarios.
In order to compare the sensitivity of this selection to the selection with best sensitivity studied in this chapter, expected exclusion limits are calculated. Background uncertainties assumed here are considered to be the same as those assumed in Sec. 8.2. This allows an easier comparison of the performance of the different 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 top squark masses ranging from 600–1100 GeV and a LSP mass of 50 GeV, the selection requiring $\not{E}_{\text{T}} > 350\,\text{GeV}$ and $N_{\text{b-jets}} \geq 2$ as seen already from the signal over background ratios.
The comparison of the expected limits is illustrated in Fig 8.16. The limit curves exhibit that the selection based on the requirements used in SUS-13-015 (black curve) shows a quite good sensitivity for low top squark masses, while the sensitivity drops rapidly towards higher top squark masses. However, the selection proposed in Sec. 8.5 (blue curve) performs better for all mass scenarios.

8.7 Stability Test

In order to study the stability of the sensitivity of the identified selection towards the assumed background uncertainties, the expected limit for the selection using the baseline requirements, at least one CMS top tag and $m_{\text{T}} > 180\,\text{GeV}$ is evaluated when varying the assumed $t\bar{t}$ uncertainty. Since for this selection $t\bar{t}$ is the main background, the variation of the $t\bar{t}$ uncertainty is expected to cause the largest deviation.

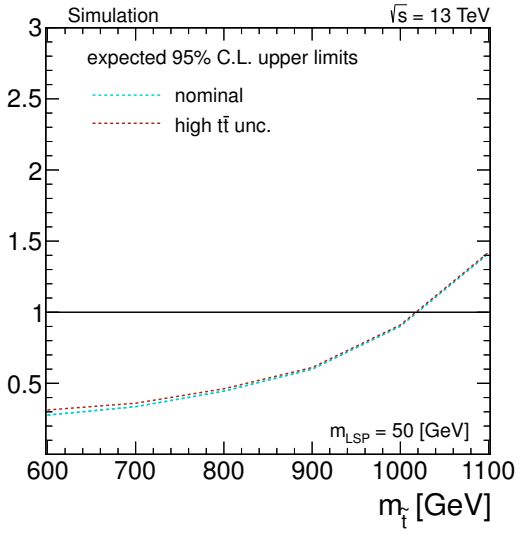

Figure 8.17: Expected 95% C.L. upper limit for signal strength versus $m_{\tilde{g}}$. The LSP mass is chosen to be 50 GeV.

The dependence on the assumed uncertainties is studied by increasing the uncertainties considered for $t\bar{t}$ background from 20% to 50% while others stay unchanged. The expected limits derived using the nominal and the increased $t\bar{t}$ uncertainty are illustrated in Fig. 8.17. The resulting change for the sensitivity caused by this variation is small and only impacts small top squark masses. Consequently, the derived selection is able to probe the top squark mass region up to 1 TeV even for a much higher uncertainty on the $t\bar{t}$ background. Accordingly, if the $t\bar{t}$ background can not be determined in data to a precision of 20%, the sensitivity of the analysis is expected to be not much degraded for the studied top squark masses.

8.8 Sensitivity to Gluino-Mediated Stop Production

The selections studied in this chapter are developed for selecting events from direct pair production of top squarks when each subsequently decays into a top quark and a LSP. Special emphasis is put on studying scenarios with large mass splittings between top squark and LSP which is suitable for employing top tagging techniques. These studied selections offer an interesting alternative to classical selections targeting gluino-mediated production of third generation squarks. In those scenarios, pair produced gluinos are considered in which each subsequently decays into a pair of top quarks and a LSP as studied in Chap. 7. Thus, the final state contains four top quarks. For gluino masses exceeding 1 TeV, this gives rise to boosted top quark decays in the final state. In order to study this model, 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 $m_{\mathrm{T}} > 180$ GeV is evaluated with respect to the gluino-mediated top squark production. A comparison of H_{T} and $\not{E}_{\mathrm{T}}$ for one selected signal point and respective background events after applying these selection criteria is shown in Fig. 8.18.
The evaluation of the sensitivity towards gluino-mediated top squark production is per-

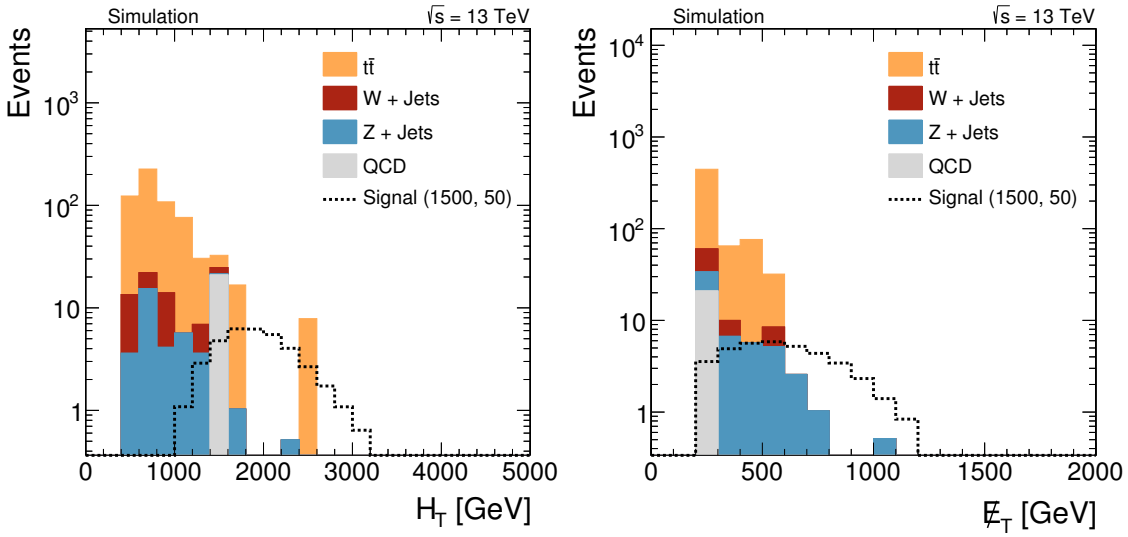

Figure 8.18: Comparison of selected H_{T} (*left*) and $\not\!\!E_{\mathrm{T}}$ (*right*) distributions in simulated events found from applying the baseline selection criteria, at least one CMS top tag and $m_{\mathrm{T}} > 180\,\mathrm{GeV}$. The signal points for gluino-mediated top squark production are labelled as (X, Y) where X is the gluino mass and Y is the LSP mass in GeV.

formed for three different mass scenarios, as specified in Sec. 8.1, with gluino masses of 1.3 TeV, 1.5 TeV and 1.7 TeV for a LSP mass of 50 GeV. Based again on the search strategy to use exclusive H_{T} and $\not\!\!E_{\mathrm{T}}$ search regions, the derived expected exclusion curve for gluino-mediated top squark production is illustrated in Fig. 8.19 and shows that the derived selection described in Sec. 8.5 is also sensitive to the tested mass scenarios of

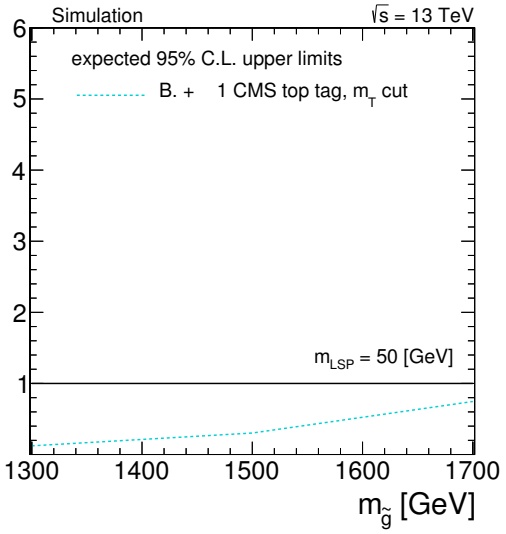

Figure 8.19: Expected 95% C.L. upper limit for signal strength versus $m_{\tilde{g}}$. The LSP mass is chosen to be 50 GeV. The label B. denotes the baseline selection.

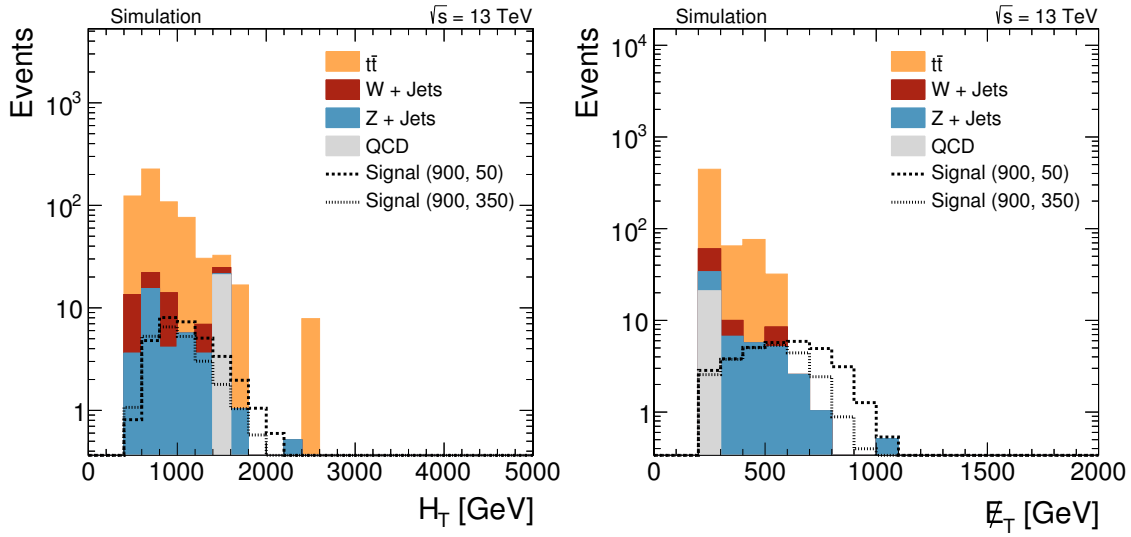

Figure 8.20: Comparison of selected H_{T} (*left*) and $\not{E}_{\mathrm{T}}$ (*right*) distributions in simulated events found from applying the baseline selection criteria, at least one CMS top tag and $m_{\mathrm{T}} > 180\,\mathrm{GeV}$. The signal points for direct top squark production are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

gluino-mediated top squark production.
Thus, selections involving top tagging also offer the opportunity to probe SUSY scenarios other than direct top squark 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 as employed in Chap. 7 for studies at $\sqrt{s} = 13\,\mathrm{TeV}$.

8.9 Results and Discussion

Selections based on different kinematic properties of direct top squark pair production have been studied and analysis strategies were identified that allow to probe direct top squark production up to 1 TeV in case of an LSP mass of 50 GeV (cf. Sec. 8.5). However, the sensitivity of this selection targeting large mass splittings between top squark and LSP can be evaluated also for other mass scenarios. For illustration, the selected H_{T} and $\not{E}_{\mathrm{T}}$ distributions for the baseline selection, at least one CMS top tag and $m_{\mathrm{T}} > 180\,\mathrm{GeV}$ are shown in Fig. 8.20 for a top squark mass of 900 GeV and a LSP mass of 50 GeV and 350 GeV, respectively.
As expected, the spectra for the larger mass difference of top squark and LSP are harder than in the case that the mass difference is smaller. However, both variables still provide separation power such that the same selection is expected to provide sensitivity also to various scenarios with smaller mass splittings.
Thus, the same search strategy and assigned uncertainties have been used to probe the expected exclusion reach for the same top squark masses as before, but when considering the LSP mass to be larger. Here, LSP masses up to 350 GeV are considered. The derived exclusion curves are summarized in Fig. 8.21 in the $m_{\tilde{t}}$ versus m_{LSP} plane. The illustrated

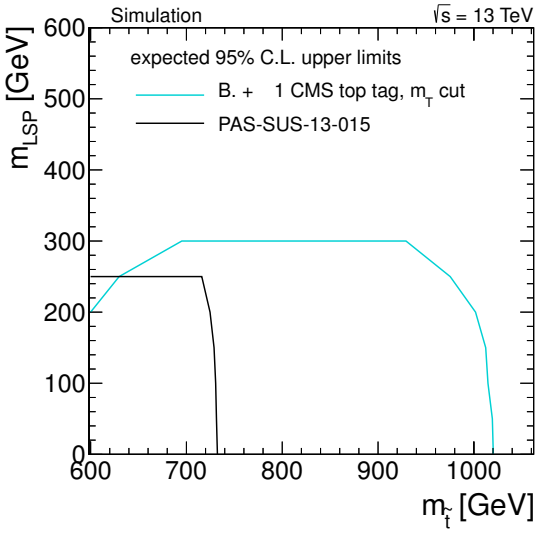

Figure 8.21: Expected 95% C.L. upper limit for $m_{\tilde{t}}$ versus m_{LSP}. The label B. denotes the baseline selection.

selections are on the one hand baseline requirements adding at least one CMS top tag and $m_{\mathrm{T}} > 180\,\mathrm{GeV}$ (light blue) discussed in Sec. 8.5 and on the other hand the selection based on SUS-13-015 discussed in Sec. 8.6 (black). The areas below the lines can be excluded. In particular for the selection indicated by the light blue line, top squark masses up to 1 TeV can be tested for LSP masses not exceeding 300 GeV.
In general, the conclusions found in previous sections also hold 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. 8.5. Although the selection following SUS-13-015 is able to probe top squark masses up to 700–800 GeV also for other LSP masses than 50 GeV, the sensitivity towards larger top squark masses is not achieved. These studies suggest that in order to test the whole top squark mass range up to 1 TeV a search strategy as proposed in this chapter can be employed.
Nonetheless, it has to be kept in mind that the studies shown here, have been made with some simplified assumptions, as introduced in Sec. 8.1. The pileup scenario considered in the simulation is not the actual expected situation during the next running period, but uses the pileup conditions form $\sqrt{s} = 8\,\mathrm{TeV}$. As discussed already in Sec. 8.1, 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. 4.3, no dedicated studies have been performed yet if those holds also for all jet substructure variables used for the top tagging algorithms. Nevertheless, the derived efficiency and misidentification rates illustrated in Fig. 8.7 show a similar performance to those quantities at $\sqrt{s} = 8\,\mathrm{TeV}$ (cf. Sec. 5.4). Consequently, the performance improvement of the analysis due to an employment of top tagging is expected to be estimated at the correct order of magnitude. In addition to those simplifications, 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 [203], 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 data-based background estimation procedure which still has to be defined. Thus, the absolute expected analysis reach might be better or worse by some 10 GeV than found in these studies. However, as stated in Sec. 8.1, the main goal of the studies was to identify general concepts and suitable selection criteria. Since it is expected that employed simplifications affect the various selections studied in this chapter in a similar manner, the sensitivity of the selections relative to each other is reflected correctly. Hence, the general conclusions stay valid and data at $\sqrt{s} = 13\,\text{TeV}$ can be utilized to shed light on the question if there is a realization of supersymmetry manifesting in top squarks below the TeV mass range when employing selections introduced in this thesis.

9 Conclusions

Supersymmetry is among the favoured extensions of the standard model of particle physics and one of the main targets of searches for new physics at the CMS experiment. Since in natural supersymmetric models sparticle masses are expected to be around $\mathcal{O}(1\,\text{TeV})$, the respective phase space can be well explored with *pp* collision data obtained during LHC Run I. Here, especially coloured SUSY particles are expected at a high rate which predominantly manifest in final states containing jets and missing transverse momentum. In order to fully exploit all-jet final states in such searches, a precise knowledge of jet-related quantities, like the jet transverse-momentum resolution, is of crucial importance.

In this thesis, a measurement of the jet transverse-momentum resolution in dijet events corresponding to data with an integrated luminosity of $19.7\,\text{fb}^{-1}$ recorded at $\sqrt{s} = 8\,\text{TeV}$ in 2012 by the CMS experiment has been presented. Here, systematic limitations of previous analyses have been overcome. This was achieved by an improved treatment of statistical uncertainties by considering correlations among inclusive distributions which then also allowed a revised treatment of systematic uncertainties. Since the systematic uncertainties have been conservatively overestimated in the past, the total precision of the measurement could be significantly improved. Furthermore, the method has been extended to be able to measure the resolution in the forward part of the detector with higher precision. Since no significant trend of the data-to-simulation ratio of the jet transverse-momentum resolution as a function of $p_{\text{T}}^{\text{ave}}$ was apparent, the ratio is parametrized as a function of $|\eta|$ only and has been determined for $0.0 \leq |\eta| \leq 5.0$. The ratios obtained for the various $|\eta|$ regions increase from 1.08 ± 0.03 in the central region up to 1.40 ± 0.06 for $2.8 \leq |\eta| \leq 3.2$ and drop again for the outermost region $3.2 \leq |\eta| \leq 5.0$ down to 1.06 ± 0.19. A distinct reason for that particular difference between data and simulation could not yet be identified, but in general noise effects, inhomogenities in the detector, miscalibration or inaccurate modelling of the hadronization process are expected to contribute. The determined data-to-simulation ratios can be utilized to adjust the resolution in simulation to match the one observed in data. A publication of the results is currently in preparation [213].

In the second part of this thesis, a search for supersymmetry in proton-proton collisions at $\sqrt{s} = 8$ TeV in data corresponding to an integrated luminosity of 19.5 fb^{-1} has been presented. This is based on events with high jet multiplicities, large values of hadronic energy, missing transverse momentum and no isolated leptons. The main goal of this analysis was to study scenarios arising from supersymmetric models. These involve especially the production and decay of gluinos, light-flavour squarks and gluino-mediated production of third generation squarks. A crucial requirement is a precise estimate of background contributions arising from SM processes. The main focus of the work presented here is the determination of the QCD multijet background which is the most challenging to model for such searches as it requires a precise knowledge of the particle-level jet spectrum. 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 necessary. 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 for instance the challenging conditions due to pileup appropriately. In total, the QCD multijet background could be estimated with a precision of approximately 50% 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 these 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% C.L. for LSP masses not exceeding 100 GeV. The respective analysis is published in [168].
In addition to inclusive searches targeting gluinos and squarks, CMS has also performed searches for direct production of supersymmetric top quark partners with data obtained at $\sqrt{s} = 8\,\text{TeV}$. With this data, direct poduction of top squarks decaying into top and LSP could be excluded up to top squark masses of approximately 750 GeV for LSP masses below around 100–200 GeV. The next run period of the LHC is going to start in 2015 with a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. These data will open a to the present date unexplored parameter space also for the direct top squark production. In order to extend the mass reach of respective analyses to top squark masses up to the 1 TeV scale, suitable selection criteria for such searches have been investigated in the third part of this thesis. Studies presented there are based on events with several jets, large momentum imbalance and no isolated leptons. One key aspect is the application of dedicated algorithms for the identification of decay products emerging from boosted hadronic top quark decays in order to separate possible SUSY signal events from standard model background. Moreover, various different discriminating kinematic variables have been investigated. Finally, the sensitivity of several selections has been compared by deriving the expected exclusion reach for data corresponding to an integrated luminosity of $19.5\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$. Selection criteria could be identified extending the mass reach of direct top squark searches up to roughly 1 TeV for LSP masses below around 300 GeV. It has been shown that the proposed selections are also more sensitive than a selection following closely the hadronic search for direct top squark production at $\sqrt{s} = 8\,\text{TeV}$ [203]. In addition, it has been demonstrated that such analysis strategies are in general also suitable to study gluino-mediated production of third generation squarks. Thus, this thesis provides a rich variety of strategies to further investigate the question if supersymmetry is realised in nature.

A Jet Transverse-Momentum Resolution

A.1 Primary Vertex Distributions

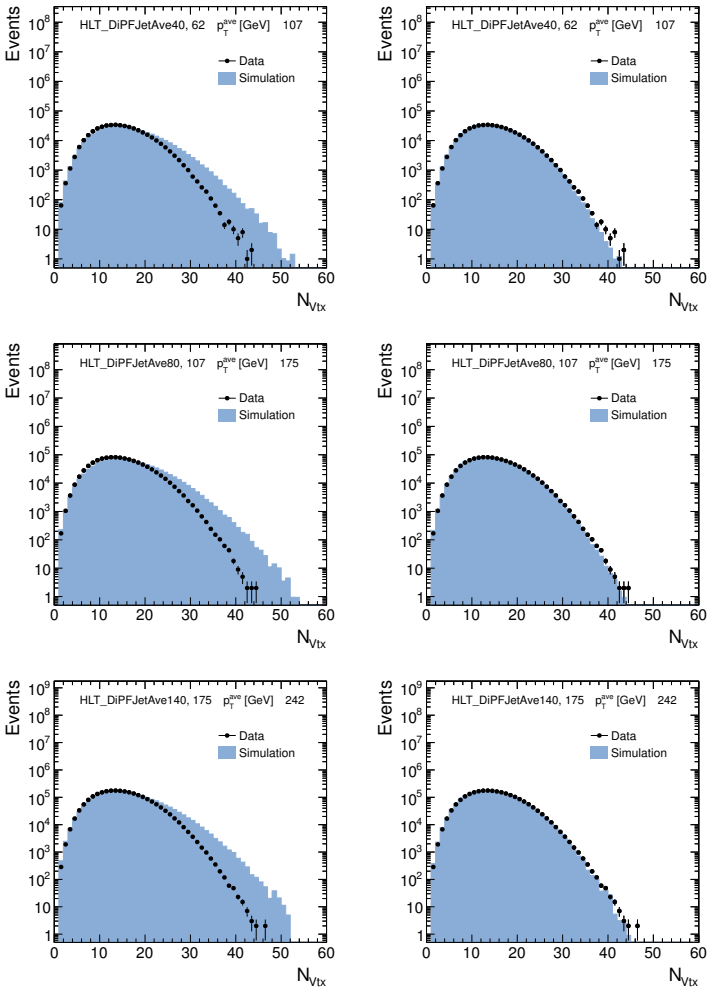

Figure A.1: Distribution of number of primary vertices in data (black dots) and simulation (blue histogram) before (*left*) and after (*right*) reweighting of the pileup scenario in simulation for trigger paths HLT_DiPFJetAve40 (*top*), HLT_DiPFJetAve80 (*middle*) and HLT_DiPFJetAve140 (*bottom*).

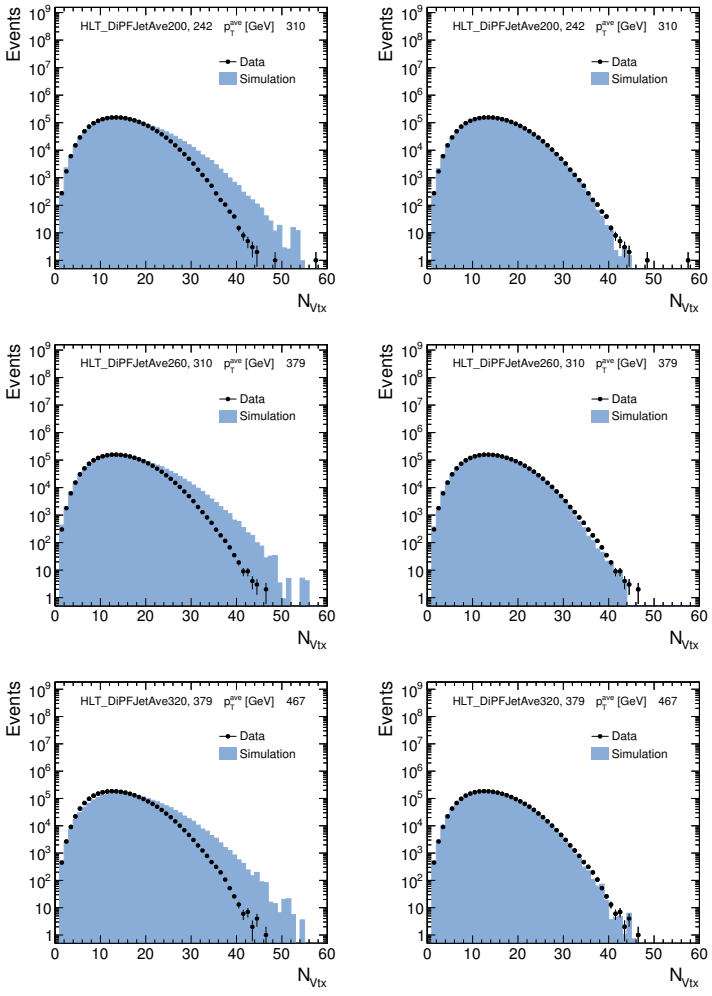

Figure A.2: Distribution of number of primary vertices in data (black dots) and simulation (blue histogram) before (*left*) and after (*right*) reweighting of the pileup scenario in simulation for trigger paths HLT_DiPFJetAve200 (*top*), HLT_DiPFJetAve260 (*middle*) and HLT_DiPFJetAve320 (*bottom*).

A.2 Extrapolation Graphs

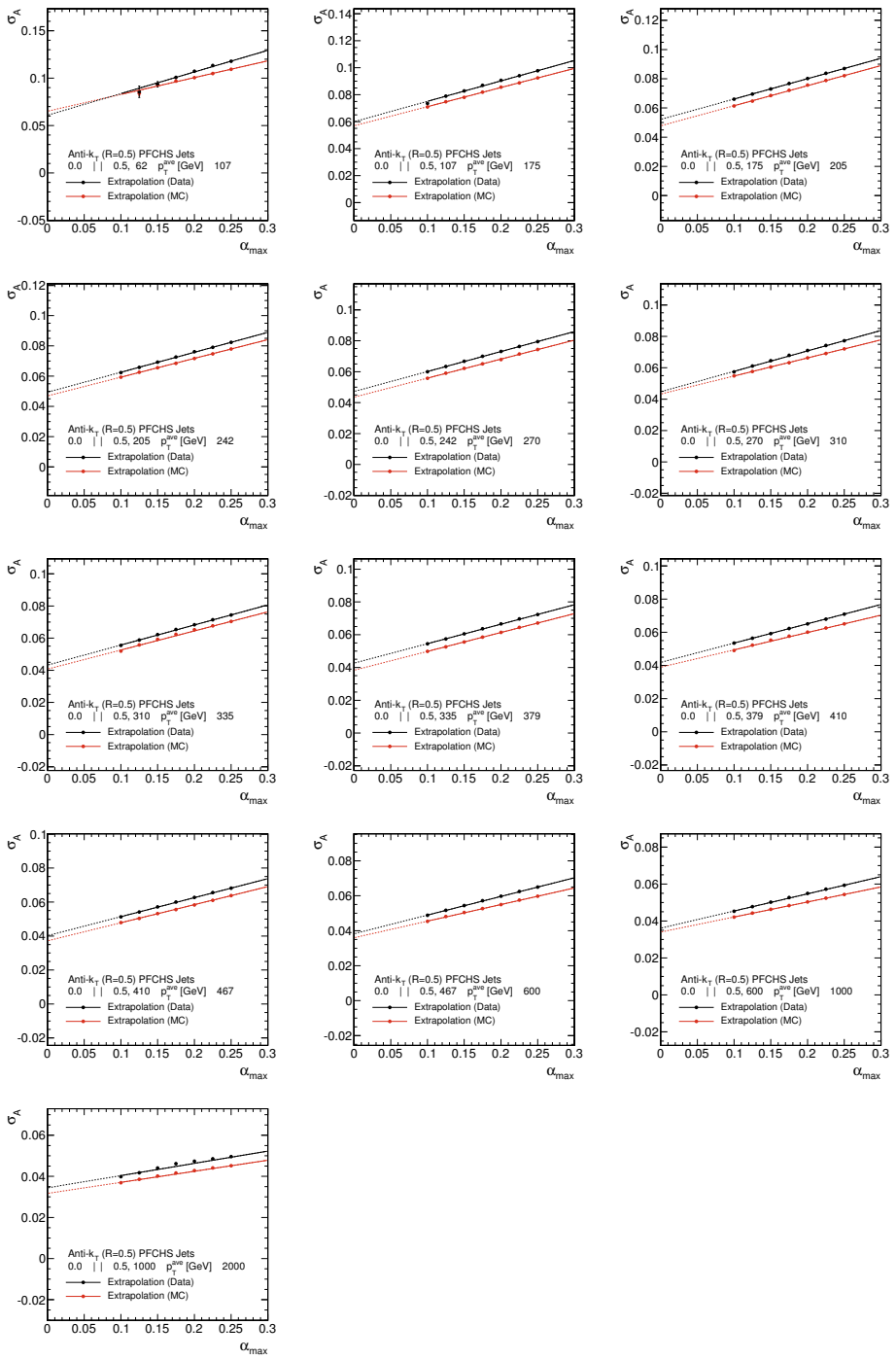

Figure A.3: Extrapolations of measured values for σ_{A} in data and simulation to obtain the result for zero additional jet activity for $0.0 \leq |\eta| \leq 0.5$.

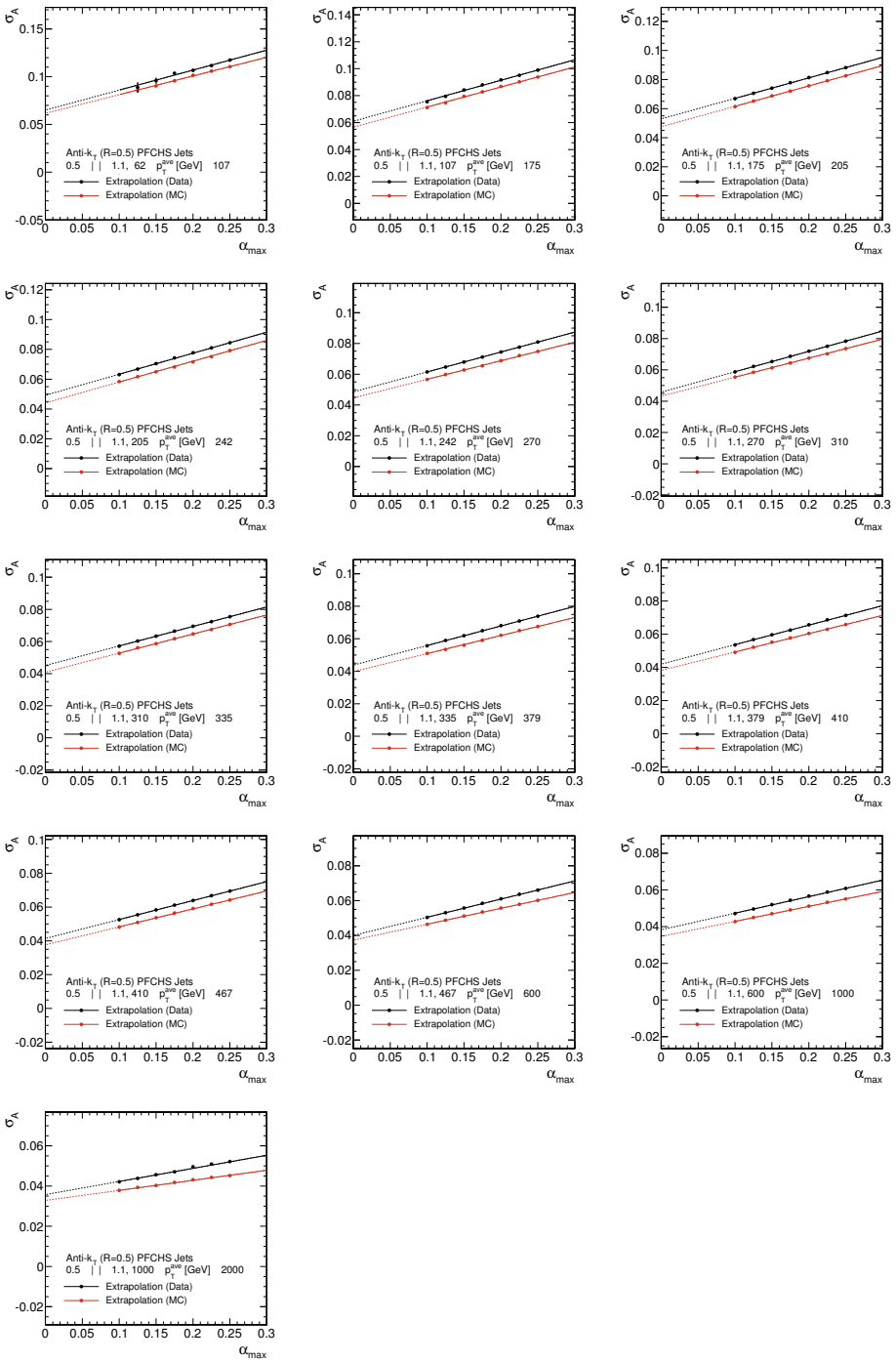

Figure A.4: Extrapolations of measured values for σ_{A} in data and simulation to obtain the result for zero additional jet activity for $0.5 \leq |\eta| \leq 1.1$.

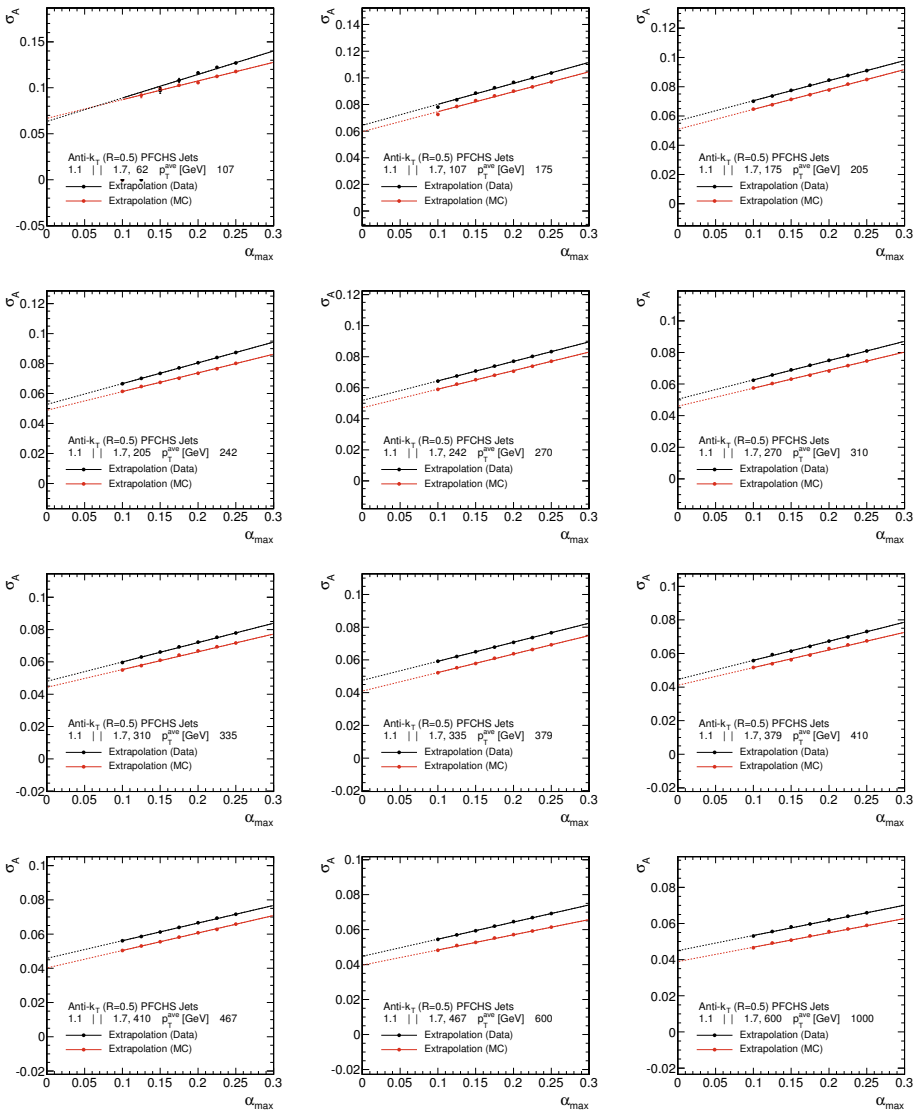

Figure A.5: Extrapolations of measured values for σ_{A} in data and simulation to obtain the result for zero additional jet activity for $1.1 \leq |\eta| \leq 1.7$.

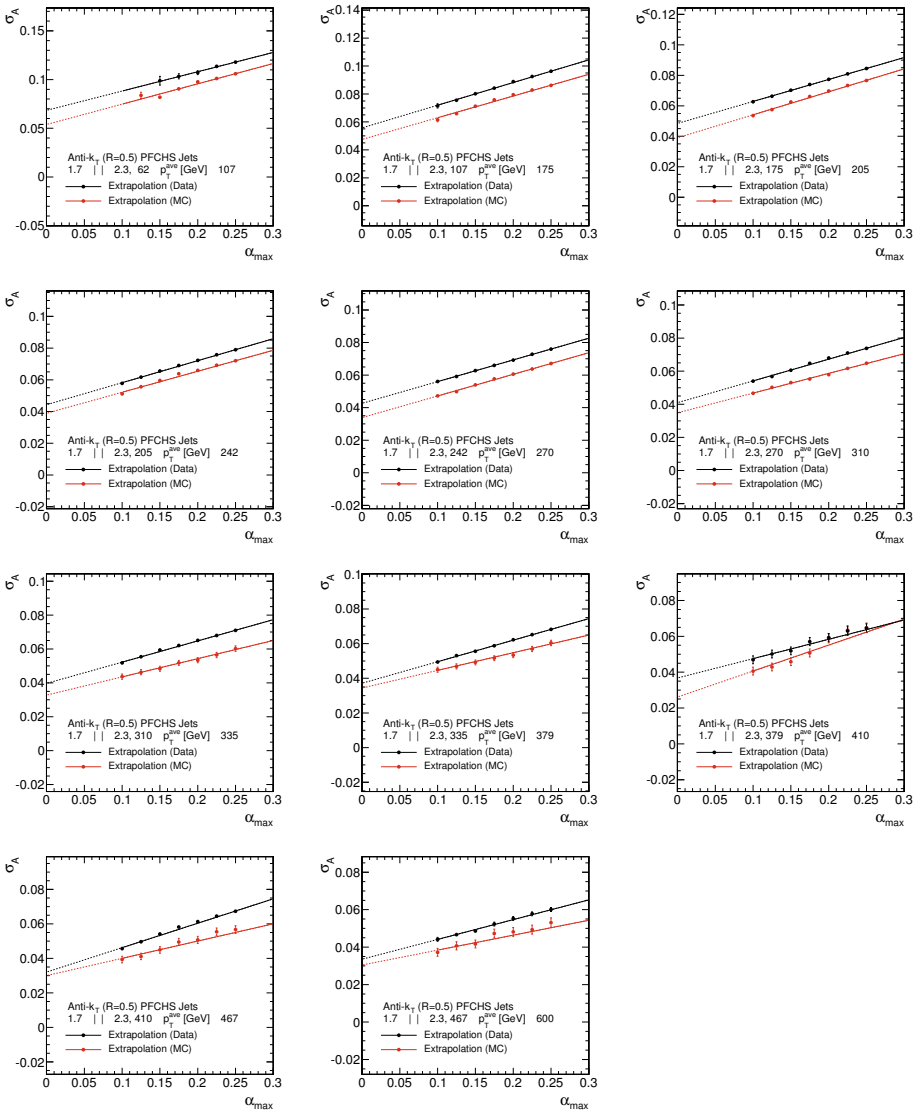

Figure A.6: Extrapolations of measured values for σ_{A} in data and simulation to obtain the result for zero additional jet activity for $1.7 \leq |\eta| \leq 2.3$.

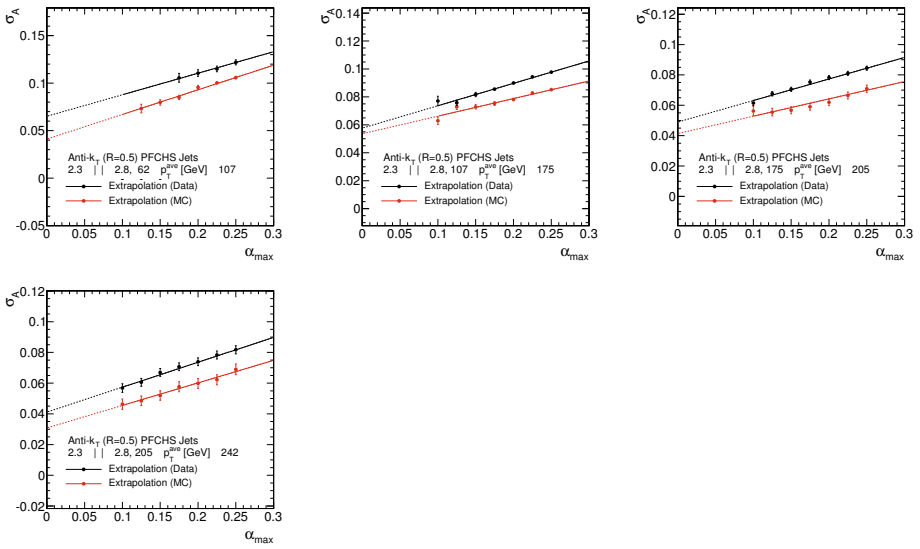

Figure A.7: Extrapolations of measured values for σ_{A} in data and simulation to obtain the result for zero additional jet activity for $2.3 \leq |\eta| \leq 2.8$.

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 ($p_{\mathrm{T}}^{\mathrm{ave}}$, $|\eta|$)-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 [161].
Given a distribution of random variables x_i, with $i = 1, ..., N$ 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 $\alpha_{\mathrm{max,i}}$ selection. Given another distribution of random variables y_j, with $j = 1, ..., M$ 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 $\alpha_{\mathrm{max,j}}$ ($\alpha_{\mathrm{max,j}} > \alpha_{\mathrm{max,i}}$).
In general, the covariance for two random variables X and Y is defined as

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

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

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

since the values x_i and y_i are independent resulting in $\mathrm{cov}(\hat{\sigma}_x^2, \hat{\sigma}_y^2) = 0$. The expression $\mathrm{cov}(\hat{\sigma}_x^2, \hat{\sigma}_x^2)$ is the variance of the estimator $\hat{\sigma}_x^2$.
In order to convert $\mathrm{cov}(\hat{\sigma}_x^2, \hat{\sigma}_{xy}^2)$ into $\mathrm{cov}(\hat{\sigma}_x, \hat{\sigma}_{xy})$, 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 exploited 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

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

Consequently,

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

Using a result for the *variance of the variance* in case of a normal distribution,

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

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

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

the result is determined to be

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

which is the same as expressed in Eq. 6.13.

A.4 Detailed Results of the Forward Extension

Table A.1: Summary of the measurement with reference bin $|\eta| \in [0.0, 0.5]$ showing the nominal data-to-simulation ratio c(Data/MC) with absolute statistical uncertainty and systematic uncertainty for each uncertainty source in different $|\eta|$ regions.

	$\|\eta_{\text{ref}}\| \in [0.0, 0.5]$ $\|\eta_{\text{probe}}\|$						
	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(Data/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	—	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-extrapolation	—	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

Table A.2: Summary of the measurement with reference bin $|\eta| \in [0.5, 1.1]$ showing the nominal data-to-simulation ratio c(Data/MC) with absolute statistical uncertainty and systematic uncertainty for each uncertainty source in different $|\eta|$ regions.

	$\|\eta_{\text{ref}}\| \in [0.5, 1.1]$ $\|\eta_{\text{probe}}\|$						
	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(Data/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	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-extrapolation	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

Table A.3: Summary of the measurement with reference bin $|\eta| \in [1.1, 1.7]$ showing the nominal data-to-simulation ratio c(Data/MC) with absolute statistical uncertainty and systematic uncertainty for each uncertainty source in different $|\eta|$ regions.

	$\|\eta_{\text{ref}}\| \in [1.1, 1.7]$ $\|\eta_{\text{probe}}\|$						
	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(Data/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	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-extrapolation	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

B Search for New Physics with Jets and Missing Transverse Momentum

B.1 Signal Trigger Efficiencies

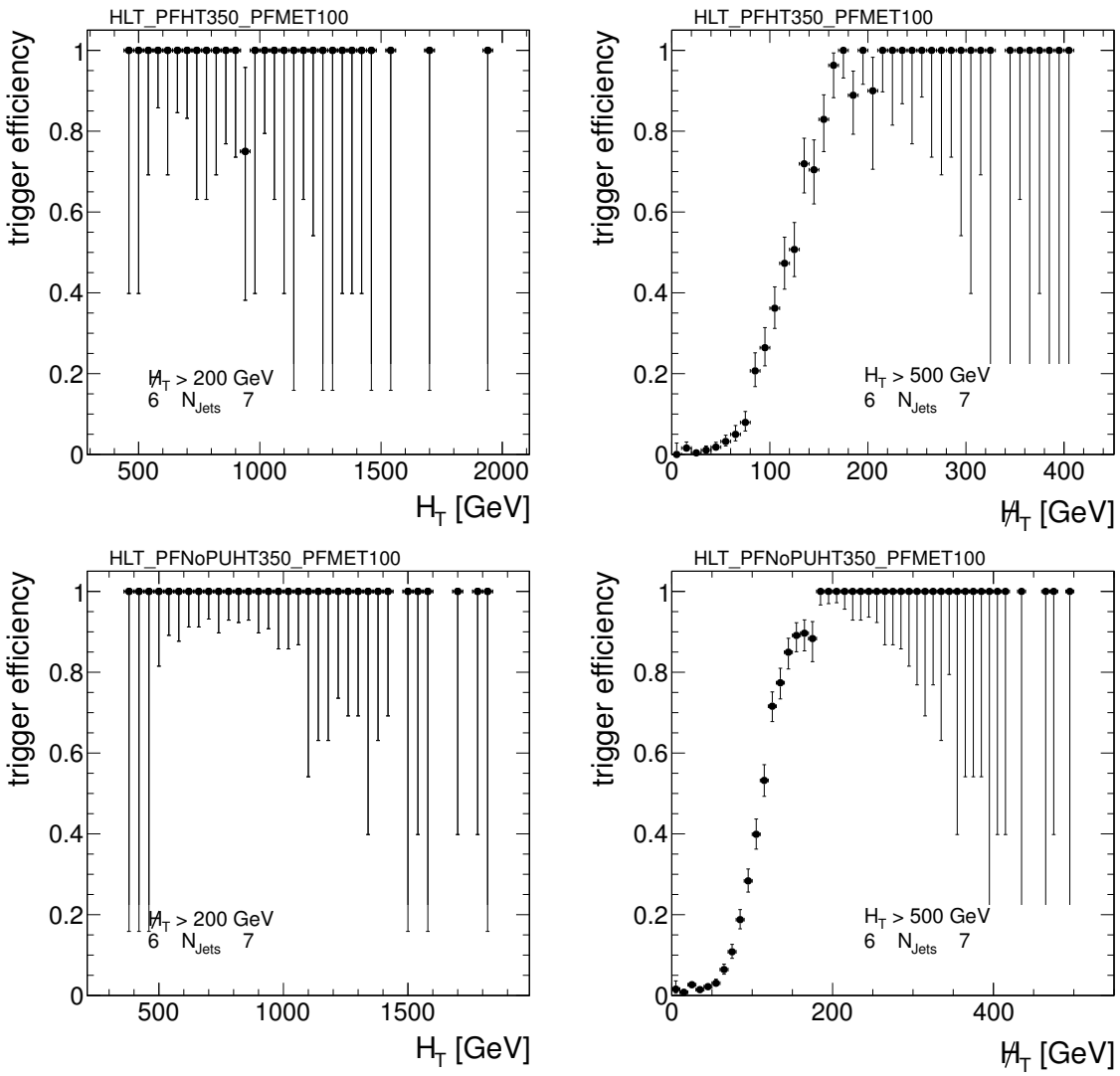

Figure B.1: Measured relative trigger efficiency for paths HLT_PFHT350_PFMET100 (*top*) and HLT_PFNoPUHT350_PFMET100 (*bottom*) as a function of H_{T} (*left*) and $\not{H}_{\mathrm{T}}$ (*right*) shown for $6 \leq N_{\mathrm{Jets}} \leq 7$.

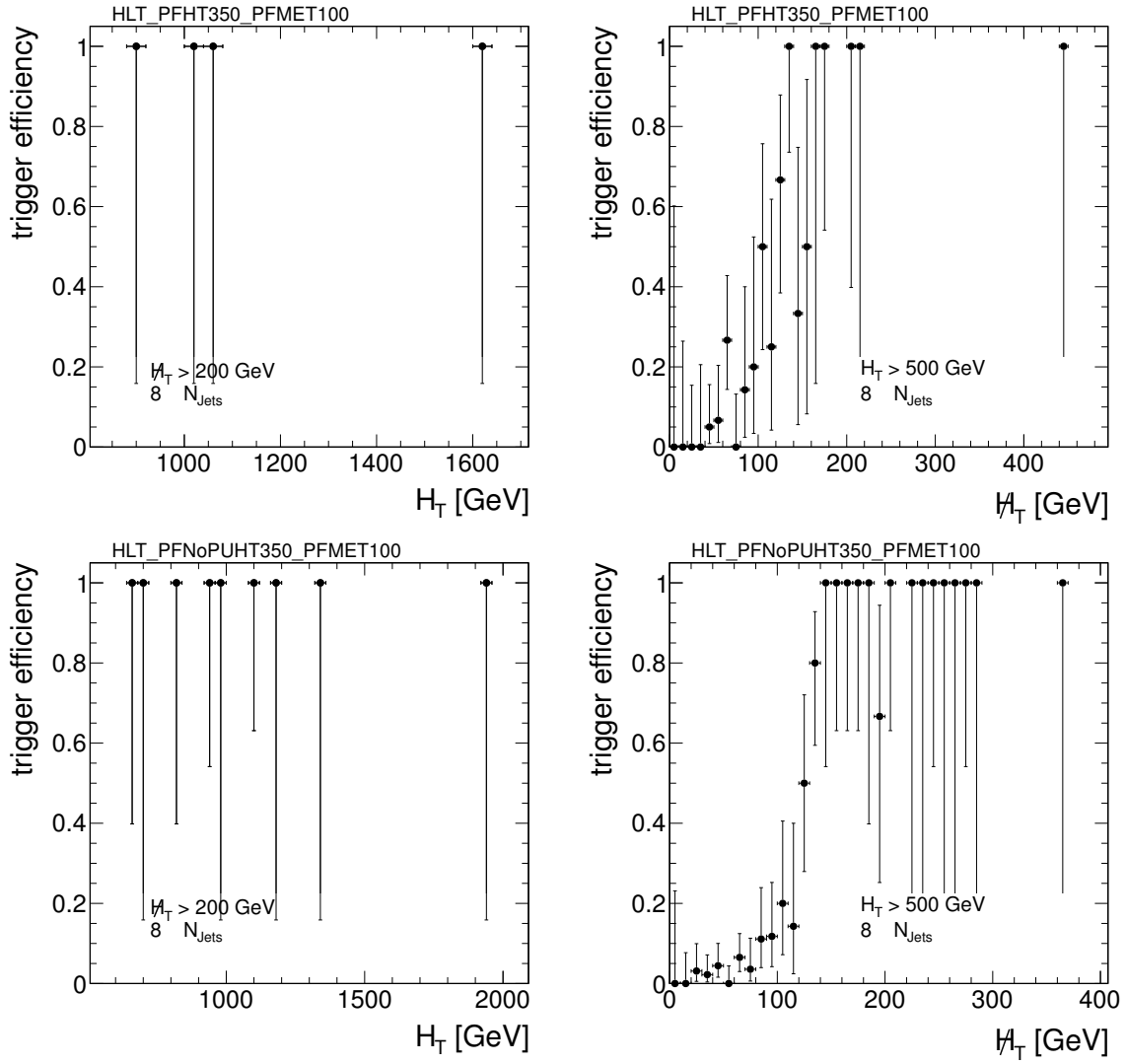

Figure B.2: Measured relative trigger efficiency for paths HLT_PFHT350_PFMET100 (*top*) and HLT_PFNoPUHT350_PFMET100 (*bottom*) as a function of H_T (*left*) and $\not{H}_T$ (*right*) shown for $8 \leq N_{\text{Jets}}$.

B.2 Data-to-Simulation Ratio: Core Resolution

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.

$\lvert\eta\rvert$	c(Data/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.3 Data-to-Simulation Ratio: Response Tails

Table B.2: Data-to-simulation ratio ρ_{tail} of the fractional tail-size together with the total uncertainty in different $|\eta| \times p_{\text{T}}^{\text{ave}}$ intervals for the tail regions derived from the dijet asymmetry parts with($\text{A} > 2\sigma_c$) used in the QCD background prediction with the R+S method for the correction of the tail parts of the response [147].

$\lvert\eta\rvert$	$p_{\text{T}}^{\text{ave}}$ [GeV]	ρ_{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

B.4 Input Truth Resolutions for Rebalancing in Kinematic Fit

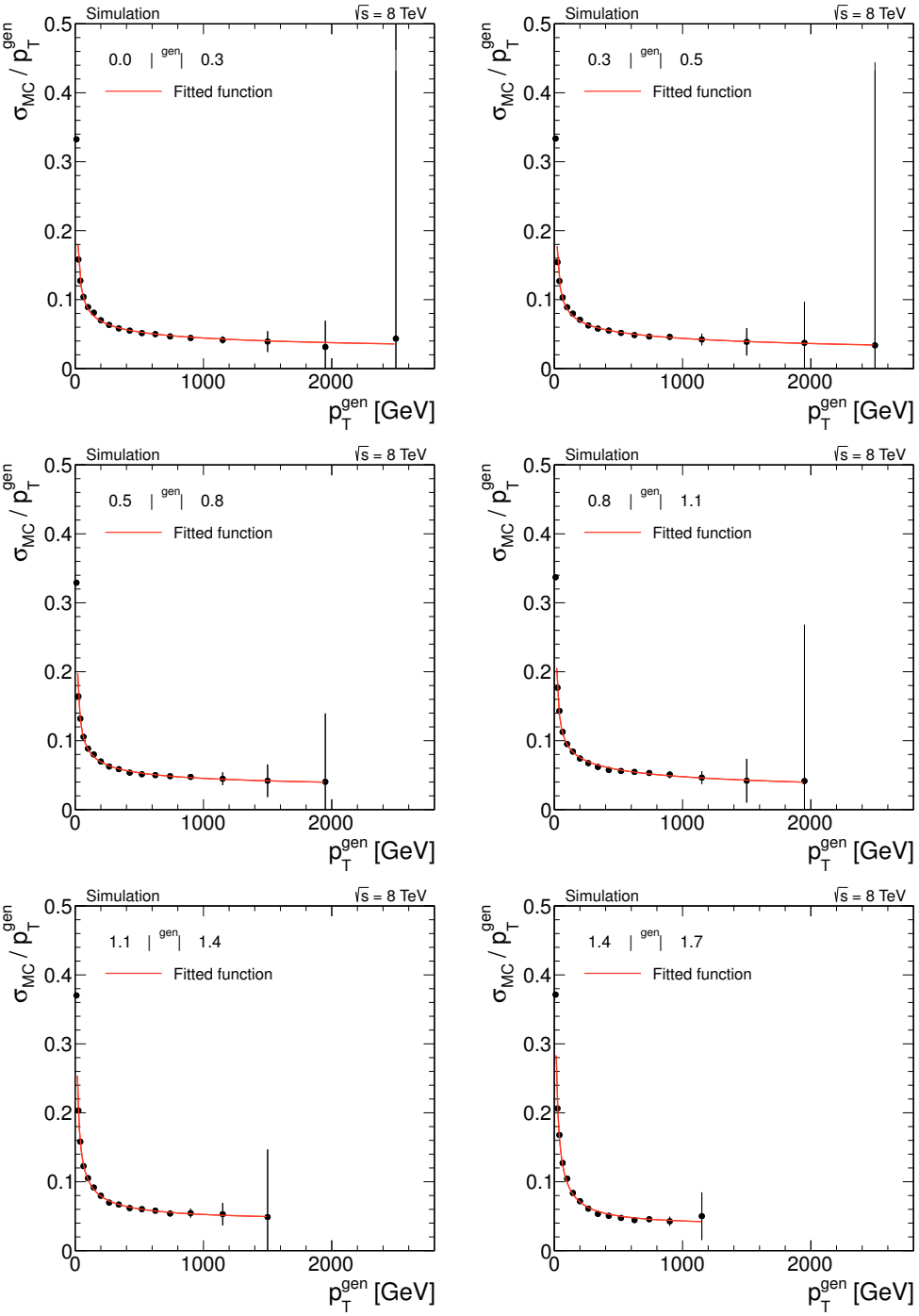

Figure B.3: Relative truth-p_T resolution derived from simulated events shown as a function of $p_\mathrm{T}^\mathrm{gen}$. The distribution is fitted with a function as described in the next used as input for the kinematic fit employed to gain a balanced seed sample.

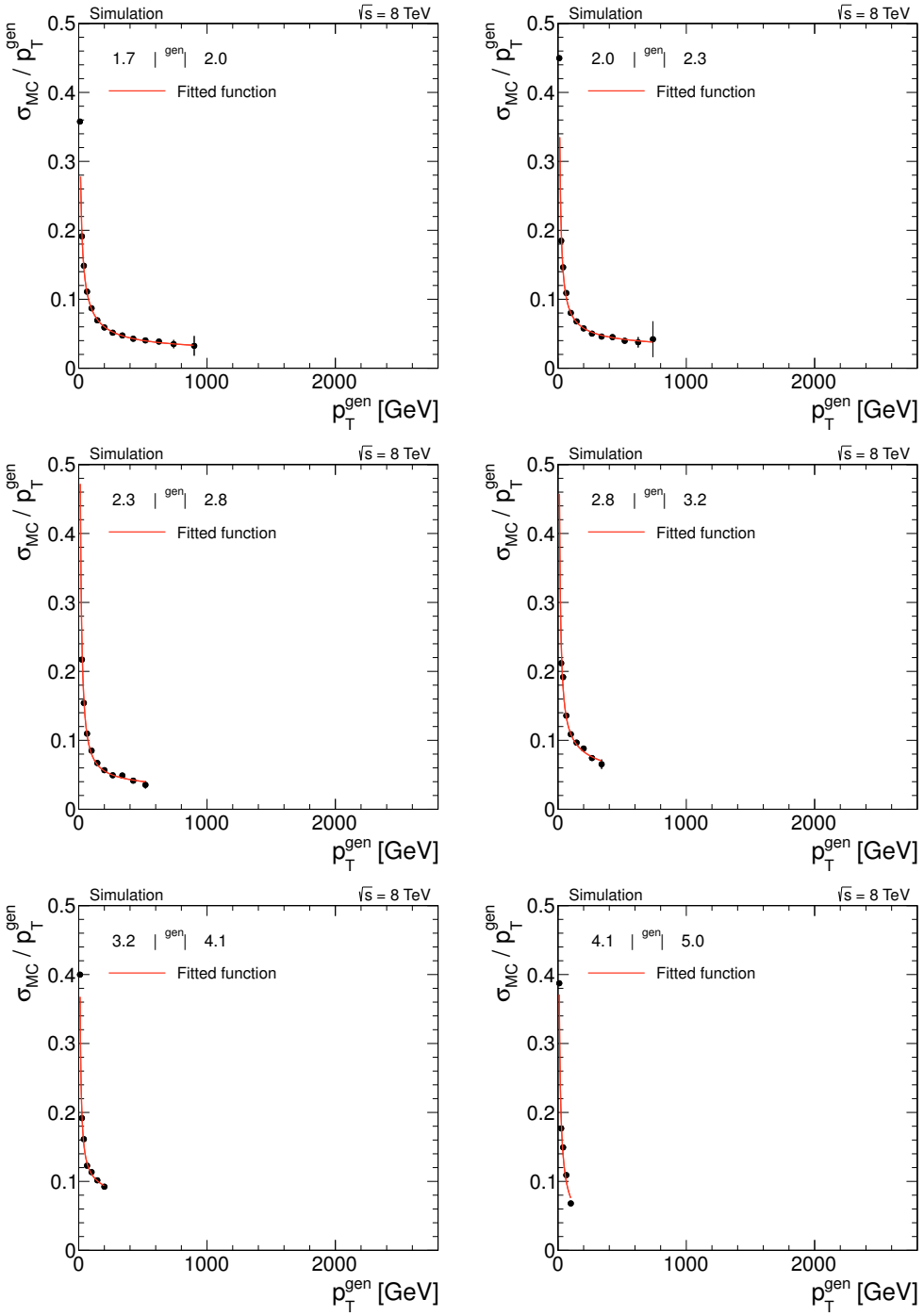

Figure B.4: Continued from Fig. B.3.

B.5 Trigger Efficiencies of Control Triggers for QCD Background Prediction

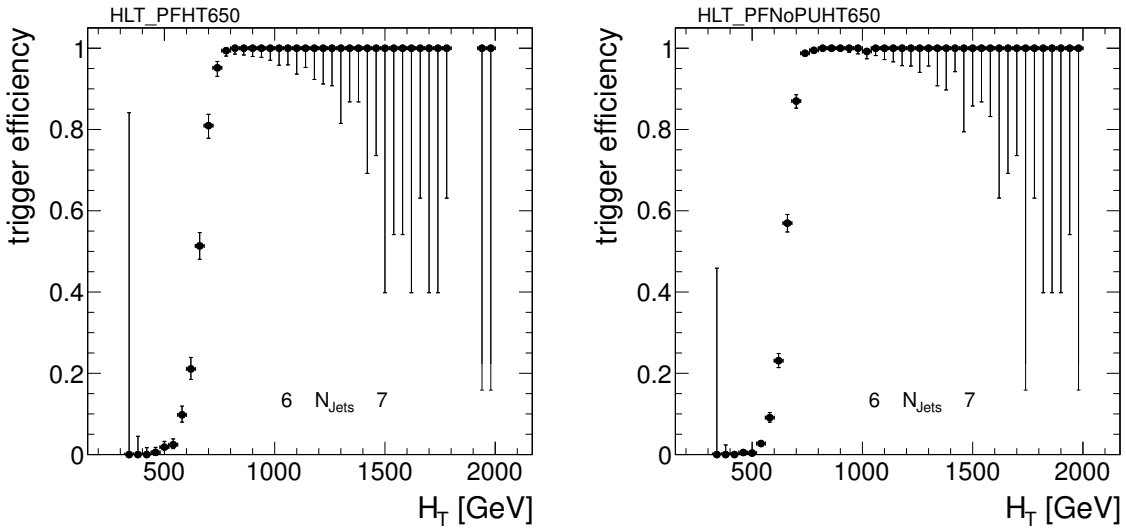

Figure B.5: Measured trigger efficiency for paths HLT_PFHT650 (*left*) and HLT_PFNoPUHT650 (*right*) as a function of H_{T} illustrated for $6 \leq N_{\mathrm{Jets}} \leq 7$.

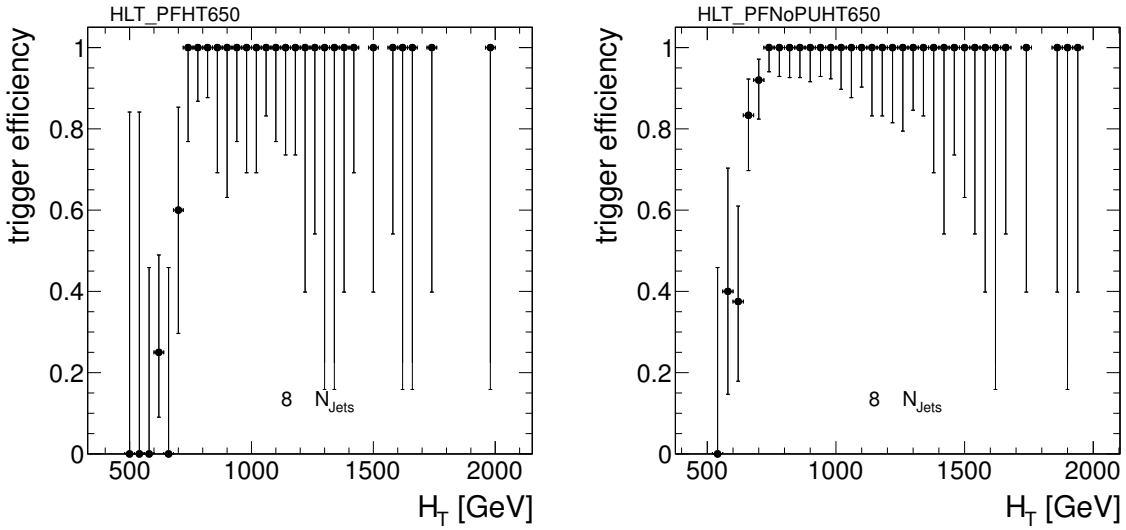

Figure B.6: Measured trigger efficiency for paths HLT_PFHT650 (*left*) and HLT_PFNoPUHT650 (*right*) as a function of H_{T} illustrated for $N_{\mathrm{Jets}} \geq 8$.

Bibliography

[1] 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.

[2] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (1963) 531–533, doi:10.1103/PhysRevLett.10.531.

[3] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652–657, doi:10.1143/PTP.49.652.

[4] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles", *Prog. Theor. Phys.* **28** (1962) 870–880, doi:10.1143/PTP.28.870.

[5] B. Pontecorvo, "Neutrino Experiments and the Problem of Conservation of Leptonic Charge", *JETP* **26** (1968) 984–988.

[6] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (1998) 1562–1567, arXiv:hep-ex/9807003, doi:10.1103/PhysRevLett.81.1562.

[7] R. Alkofer and J. Greensite, "Quark Confinement: The Hard Problem of Hadron Physics", *J. Phys. G* **34** (2007) S21, arXiv:hep-ph/0610365, doi:10.1088/0954-3899/34/7/S02.

[8] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?", *Phys. Rev. Lett.* **30** (1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

[9] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[10] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[11] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[12] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[13] 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.

[14] 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.

[15] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, arXiv:1207.7214, doi:10.1016/j.physletb.2012.08.020.

[16] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, arXiv:1207.7235, doi:10.1016/j.physletb.2012.08.021.

[17] ATLAS Collaboration, "Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC", *Phys. Lett. B* **726** (2013) 88–119, arXiv:1307.1427, doi:10.1016/j.physletb.2013.08.010.

[18] ATLAS 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", (2014) arXiv:1408.5191, Accepted by PRD.

[19] ATLAS 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", (2014) arXiv:1408.7084, Accepted by PRD.

[20] CMS Collaboration, "Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model", *CMS Physics Analysis Summary* **CMS-PAS-HIG-14-009** (2014).

[21] M. Beyer, "CP Violation in Particle, Nuclear, and Astrophysics", (2010), Lecture Notes in Physics, Springer, ISBN 3-642-07830-3.

[22] Planck Collaboration, "Planck 2013 results. XVI. Cosmological parameters", *Astron. Astrophys.* (2014) arXiv:1303.5076, doi:10.1051/0004-6361/201321591.

[23] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints", *Phys. Rept.* **405** (2005) 279–390, arXiv:hep-ph/0404175, doi:10.1016/j.physrep.2004.08.031.

[24] J. Wess and B. Zumino, "Supergauge transformations in four dimensions", *Nucl. Phys. B* **70** (1974) 39–50, doi:10.1016/0550-3213(74)90355-1.

[25] I. J. Aitchison, "Supersymmetry and the MSSM: An Elementary Introduction", (2005) arXiv:hep-ph/0505105.

[26] S. P. Martin, "A Supersymmetry primer", *Adv. Ser. Direct. High Energy Phys.* **21** (2010) 1–153, arXiv:hep-ph/9709356.

[27] Super-Kamiokande Collaboration, "Search for proton decay via $p \to \nu K^+$ using 260 kiloton $\cdot$ year data of Super-Kamiokande", *Phys. Rev. D* **90** (2014) 072005, doi:10.1103/PhysRevD.90.072005.

[28] T. Ibrahim and P. Nath, "CP Violation from the Standard Model to Strings", *Rev. Mod. Phys.* **80** (2008) 577–631, arXiv:0705.2008, doi:10.1103/RevModPhys.80.577.

[29] J. Scherk, "An introduction to the theory of dual models and strings", *Rev. Mod. Phys.* **47** (1975) 123–164, doi:10.1103/RevModPhys.47.123.

[30] M. J. Veltman, "The infrared-ultraviolet connection", *Acta Phys. Polon. B* **12** (1981) 437–457.

[31] R. Barbieri and G. F. Giudice, "Upper bounds on supersymmetric particle masses", *Nuclear Physics B* **306** (1988) 63–76, doi:10.1016/0550-3213(88)90171-X.

[32] N. Craig, "The State of Supersymmetry after Run I of the LHC", (2013) arXiv:1309.0528.

[33] A. Djouadi, "The Anatomy of electro-weak symmetry breaking Tome. II. The Higgs bosons in the minimal supersymmetric model", *Phys. Rept.* **459** (2008) 1–241, arXiv:hep-ph/0503173, doi:10.1016/j.physrep.2007.10.005.

[34] S. Dimopoulos and D. W. Sutter, "The Supersymmetric flavor problem", *Nucl. Phys. B* **452** (1995) 496–512, arXiv:hep-ph/9504415, doi:10.1016/0550-3213(95)00421-N.

[35] A. H. Chamseddine, R. L. Arnowitt, and P. Nath, "Locally Supersymmetric Grand Unification", *Phys. Rev. Lett.* **49** (1982) 970–974, doi:10.1103/PhysRevLett.49.970.

[36] L. Alvarez-Gaume, J. Polchinski, and M. B. Wise, "Minimal Low-Energy Supergravity", *Nucl. Phys. B* **221** (1983) 495–523, doi:10.1016/0550-3213(83)90591-6.

[37] G. L. Kane, C. F. Kolda, L. Roszkowski et al., "Study of constrained minimal supersymmetry", *Phys. Rev. D* **49** (1994) 6173–6210, arXiv:hep-ph/9312272, doi:10.1103/PhysRevD.49.6173.

[38] H. Baer, C. Balazs, A. Belyaev et al., "Updated constraints on the minimal supergravity model", *JHEP* **207** (2002) 050, arXiv:hep-ph/0205325.

[39] M. Dine and W. Fischler, "A Phenomenological Model of Particle Physics Based on Supersymmetry", *Phys. Lett. B* **110** (1982) 227, doi:10.1016/0370-2693(82)91241-2.

[40] L. Alvarez-Gaume, M. Claudson, and M. B. Wise, "Low-Energy Supersymmetry", *Nucl. Phys. B* **207** (1982) 96–110, doi:10.1016/0550-3213(82)90138-9.

[41] 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) 119–244, arXiv:1302.3415.

[42] J. Erler, "Tests of the Electroweak Standard Model", *J. Phys. Conf. Ser.* **485** (2014) 012010, arXiv:1209.3324, doi:10.1088/1742-6596/485/1/012010.

[43] M. Ciuchini, E. Franco, S. Mishima et al., "Electroweak Precision Observables, New Physics and the Nature of a 126 GeV Higgs Boson", *JHEP* **8** (2013) 106, arXiv:1306.4644, doi:10.1007/JHEP08(2013)106.

[44] M. Baak, J. Cuth, J. Haller et al., "The global electroweak fit at NNLO and prospects for the LHC and ILC", *Eur. Phys. J. C* **74** (2014) 3046, arXiv:1407.3792, doi:10.1140/epjc/s10052-014-3046-5.

[45] CMS Collaboration, "Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ with the CMS Experiment", *Phys. Rev. Lett.* **111** (2013) 101804, arXiv:1307.5025, doi:10.1103/PhysRevLett.111.101804.

[46] LHCb Collaboration, "Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment", *Phys. Rev. Lett.* **111** (2013) 101805, arXiv:1307.5024, doi:10.1103/PhysRevLett.111.101805.

[47] CMS and LHCb Collaboration, "Combination of results on the rare decays $B_{(s)}^0 \to \mu^+\mu^-$ from the CMS and LHCb experiments", **CMS-PAS-BPH-13-007, CERN-LHCb-CONF-2013-012** (2014), Paper in preparation.

[48] Muon (g-2) Collaboration, "Final report of the muon E821 anomalous magnetic moment measurement at BNL", *Phys. Rev. D* **73** (2006) 072003, arXiv:hep-ex/0602035, doi:10.1103/PhysRevD.73.072003.

[49] 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, doi:10.1088/0954-3899/38/8/085003.

[50] M. Davier, A. Hoecker, B. Malaescu et al., "Reevaluation of the Hadronic Contributions to the Muon $g-2$ and to $\alpha(M_Z^2)$", *Eur. Phys. J. C* **71** (2011) 1515, arXiv:1010.4180, doi:10.1140/epjc/s10052-010-1515-z.

[51] C. L. Bennett et al., "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results", *ApJS* **208** (2013) 19, arXiv:1212.5226, doi:10.1088/0067-0049/208/2/19.

[52] D. G. Cerdeno and A. M. Green, "Direct detection of WIMPs", (2010) arXiv:1002.1912.

[53] 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:10.1088/1475-7516/2011/03/051.

[54] XENON100 Collaboration, "Dark Matter Results from 225 Live Days of XENON100 Data", *Phys. Rev. Lett.* **109** (2012) 181301, doi:10.1103/PhysRevLett.109.181301.

[55] Fermi-LAT Collaboration Collaboration, "Search for Gamma-ray Spectral Lines with the Fermi Large Area Telescope and Dark Matter Implications", *Phys. Rev. D* **88** (2013) 082002, arXiv:1305.5597, doi:10.1103/PhysRevD.88.082002.

[56] SuperCDMS Collaboration, "Search for Low-Mass Weakly Interacting Massive Particles Using Voltage-Assisted Calorimetric Ionization Detection in the SuperCDMS Experiment", *Phys. Rev. Lett.* **112** (2014) 041302, arXiv:1309.3259, doi:10.1103/PhysRevLett.112.041302.

[57] CRESST-II Collaboration, "Results on low mass WIMPs using an upgraded CRESST-II detector", (2014) arXiv:1407.3146.

[58] AMS Collaboration, "High Statistics Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station", *Phys. Rev. Lett.* **113** (2014) 121101, doi:10.1103/PhysRevLett.113.121101.

[59] HESS Collaboration, "Search for dark matter annihilation signatures in H.E.S.S. observations of Dwarf Spheroidal Galaxies", (2014) arXiv:1410.2589.

[60] H1 Collaboration, "Search for squark production in R-parity-violating supersymmetry at HERA", *Eur. Phys. J. C* **36** (2004) 425–440, arXiv:hep-ex/0403027.

[61] H1 Collaboration, "Search for bosonic stop decays in R-parity-violating supersymmetry in e^+p collisions at HERA", *Phys. Lett. B* **599** (2004) 159–172, arXiv:hep-ex/0405070, doi:10.1016/j.physletb.2004.08.032.

[62] ZEUS Collaboration, "Search for stop production in R-parity-violating supersymmetry at HERA", *Eur. Phys. J. C* **50** (2007) 269–281, arXiv:hep-ex/0611018, doi:10.1140/epjc/s10052-007-0240-8.

[63] H1 Collaboration, "A search for selectrons and squarks at HERA", *Phys. Lett. B* **380** (1996) 461–470, arXiv:hep-ex/9605002, doi:10.1016/0370-2693(96)00640-5.

[64] ZEUS 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.

[65] ALEPH, DELPHI, L3 and OPAL Collaborations, "Joint SUSY Working Group", *LEPSUSYWG/02-06-2* (2004), http://lepsusy.web.cern.ch/lepsusy.

[66] CDF 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.

[67] DØ Collaboration, "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 $\sqrt{s} = 1.96$ TeV", *Phys. Lett. B* **660** (2008) 449–457, arXiv:0712.3805, doi:10.1016/j.physletb.2008.01.042.

[68] 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) 34–43, doi:10.1016/j.physletb.2009.08.011.

[69] G. L. Kane and J. Leveille, "Experimental Constraints on Gluino Masses and Supersymmetric Theories", *Phys. Lett. B* **112** (1982) 227–232, doi:10.1016/0370-2693(82)90968-6.

[70] P. R. Harrison and C. H. Smith, "Hadroproduction of Supersymmetric Particles", *Nucl. Phys. B* **213** (1983) 223–240, doi:10.1016/0550-3213(83)90510-2.

[71] E. Reya and D. P. Roy, "Supersymmetric-particle production at $p\bar{p}$ collider energies", *Phys. Rev. D* **32** (1985) 645–657, doi:10.1103/PhysRevD.32.645.

[72] S. Dawson, E. Eichten, and C. Quigg, "Search for Supersymmetric Particles in Hadron-Hadron Collisions", *Phys. Rev. D* **31** (1985) 1581–1637, doi:10.1103/PhysRevD.31.1581.

[73] H. Baer and X. Tata, "Component Formulae for Hadroproduction of Left-handed and Right-handed Squarks", *Phys. Lett. B* **160** (1985) 159–162, doi:10.1016/0370-2693(85)91484-4.

[74] M. Krämer, A. Kulesza, R. van der Leeuw et al., "Supersymmetry production cross sections in pp collisions at $\sqrt{s} = 7$ TeV", (2012) arXiv:1206.2892.

[75] C. Borschensky, M. Krämer, A. Kulesza et al., "Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13$, 14, 33 and 100 TeV", (2014) arXiv:1407.5066.

[76] CMS Collaboration, "CMS Supersymmetry Physics Results", (2014), https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS, Topic revision: r279.

[77] CMS Collaboration, "Search for new physics with jets and missing transverse momentum in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **8** (2011) 1–46, arXiv:1106.4503, doi:10.1007/JHEP08(2011)155.

[78] CMS Collaboration, "Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **109** (2012) 171803, arXiv:1207.1898, doi:10.1103/PhysRevLett.109.171803.

[79] ATLAS Collaboration, "ATLAS Supersymmetry (SUSY) searches", (2014), https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults, Topic revision: r437.

[80] N. Arkani-Hamed, P. Schuster, N. Toro et al., "MARMOSET: The Path from LHC Data to the New Standard Model via On-Shell Effective Theories", (2007) arXiv:hep-ph/0703088.

[81] J. Alwall, P. Schuster, and N. Toro, "Simplified Models for a First Characterization of New Physics at the LHC", *Phys. Rev. D* **79** (2009) 075020, arXiv:0810.3921, doi:10.1103/PhysRevD.79.075020.

[82] J. Alwall, M.-P. Le, M. Lisanti et al., "Model-Independent Jets plus Missing Energy Searches", *Phys. Rev. D* **79** (2009) 015005, arXiv:0809.3264, doi:10.1103/PhysRevD.79.015005.

[83] CMS Collaboration, "Interpretation of Searches for Supersymmetry with Simplified Models", *Phys. Rev. D* **88** (2013) 052017, arXiv:1301.2175, doi:10.1103/PhysRevD.88.052017.

[84] A. Djouadi, S. Rosier-Lees, M. Bezouh et al., "The Minimal supersymmetric standard model: Group summary report", (1999) arXiv:hep-ph/9901246.

[85] CMS Collaboration, "Phenomenological MSSM interpretation of the CMS 2011 5fb^{-1} results", *CMS Physics Analysis Summary* **CMS-PAS-SUS-12-030** (2013).

[86] O. S. Brüning, P. Collier, P. Lebrun et al., "LHC Design Report: The LHC Main Ring", (2004), CERN-2004-003-V-1.

[87] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[88] CERN, "LEP Design Report: The LEP Main Ring", (1984), CERN-LEP-84-01.

[89] CERN, "LHC: the guide", (2009), CERN-Brochure-2009-003-Eng.

[90] W. J. Stirling, "Parton luminosity and cross section plots", Retrieved Nov. 2014, http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html.

[91] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[92] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[93] CMS Collaboration, "CMS Detector Performance and Software", *Technical Design Report* (2006), CMS-TDR-008-1. CERN-LHCC-2006-001.

[94] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[95] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[96] CMS Collaboration, "The CMS tracker system project", *Technical Design Report* (1997), CMS-TDR-005. CERN-LHCC-98-006.

[97] S. Abdullin, V. Abramov, B. Acharya et al., "Studies of the response of the prototype CMS hadron calorimeter, including magnetic field effects, to pion, electron, and muon beams", *Nucl. Instrum. Meth. A* **457** (2001) 75–100, arXiv:hep-ex/0007045, doi:10.1016/S0168-9002(00)00711-7.

[98] S. Abdullin, V. Abramov, B. Acharya et al., "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c", *Europ. Phys. J. C* **60** (2009) 359–373, doi:10.1140/epjc/s10052-009-0959-5.

[99] CMS Collaboration, "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV", *JINST* **8** (2013) P09009, arXiv:1306.2016, doi:10.1088/1748-0221/8/09/P09009.

[100] CMS Collaboration, "Luminosity Information for the 2012 Proton Run", (2012), https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, Topic revision: r101.

[101] M. Lamont, "Status of the LHC", *J. Phys. Conf. Ser.* **455** (2013) 012001, doi:10.1088/1742-6596/455/1/012001.

[102] N. Metropolis and S. Ulam, "The Monte Carlo Method", *J. Americ. Stat. Assoc.* **44** (1949) 335–341, doi:10.2307/2280232.

[103] M. H. Seymour and M. Marx, "Monte Carlo Event Generators", (2013) arXiv:1304.6677.

[104] A. Buckley, J. Butterworth, S. Gieseke et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, arXiv:1101.2599, doi:10.1016/j.physrep.2011.03.005.

[105] M. Dobbs and J. B. Hansen, "The HepMC C++ Monte Carlo Event Record for High Energy Physics", *Comput. Phys. Commun.* **134** (2001) 41, http://lcgapp.cern.ch/project/simu/HepMC, doi:10.1016/S0010-4655(00)00189-2.

[106] H1 and ZEUS Collaborations, "Combined Measurement and QCD Analysis of the Inclusive $e^{\pm}p$ Scattering Cross Sections at HERA", *JHEP* **01** (2010) 109, arXiv:0911.0884, doi:10.1007/JHEP01(2010)109.

[107] 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.

[108] 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.

[109] B. Andersson, G. Gustafson, G. Ingelman et al., "Parton fragmentation and string dynamics", *Phys. Rep.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.

[110] D. Amati and G. Veneziano, "Preconfinement as a property of perturbative QCD", *Phys. Let. B* **83** (1979) 87–92, doi:10.1016/0370-2693(79)90896-7.

[111] A. Bassetto, M. Ciafaloni, and G. Marchesini, "Color singlet distributions and mass damping in perturbative QCD", *Physics Letters B* **83** (1979) 207–212, doi:10.1016/0370-2693(79)90687-7.

[112] T. Sjöstrand and M. van Zijl, "A multiple-interaction model for the event structure in hadron collisions", *Phys. Rev. D* **36** (1987) 2019–2041, doi:10.1103/PhysRevD.36.2019.

[113] J. Butterworth, J. R. Forshaw, and M. Seymour, "Multiparton interactions in photoproduction at HERA", *Z. Phys. C* **72** (1996) 637–646, arXiv:hep-ph/9601371, doi:10.1007/s002880050286.

[114] CMS Collaboration, "Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7\,\text{TeV}$ and Comparison with $\sqrt{s} = 0.9\,\text{TeV}$", *JHEP* **09** (2011) 109, arXiv:1107.0330, doi:10.1007/JHEP09(2011)109.

[115] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175, doi:10.1088/1126-6708/2006/05/026.

[116] M. Bähr, S. Gieseke, M. Gigg et al., "Herwig++ Physics and Manual", *Eur. Phys. J. C* **58** (2008) 639–707, arXiv:0803.0883, doi:10.1140/epjc/s10052-008-0798-9.

[117] 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* **07** (2014) 079, arXiv:1405.0301, doi:10.1007/JHEP07(2014)079.

[118] C. Oleari, "The POWHEG BOX", *Nucl. Phys. B Proc. Suppl.* **205-206** (2010) 36–41, arXiv:1007.3893, doi:10.1016/j.nuclphysbps.2010.08.016.

[119] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, arXiv:hep-ph/0409146, doi:10.1088/1126-6708/2004/11/040.

[120] S. Agostinelli, J. Allison, K. Amako et al., "G4 – a simulation toolkit", *NIM A* **506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[121] CMS 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.

[122] CMS Collaboration, "Comparison of the Fast Simulation of CMS with the first LHC data", *CMS Detector Performance Summaries* **CMS-DP-2010-039** (2010).

[123] 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).

[124] CMS Collaboration, "Track Reconstruction in the CMS tracker", *CMS Note* **CMS-NOTE-2006-041** (2006).

[125] G. P. Salam, "Towards Jetography", *Eur. Phys. J. C* **67** (2010) 637–686, arXiv:0906.1833, doi:10.1140/epjc/s10052-010-1314-6.

[126] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", (2011) arXiv:1111.6097.

[127] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder", *Phys. Lett. B* **641** (2006) 57–61, arXiv:hep-ph/0512210, doi:10.1016/j.physletb.2006.08.037.

[128] G. P. Salam and G. Soyez, "A Practical Seedless Infrared-Safe Cone jet algorithm", *JHEP* **05** (2007) 086, arXiv:0704.0292, doi:10.1088/1126-6708/2007/05/086.

[129] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D* **48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160.

[130] Y. L. Dokshitzer, G. Leder, S. Moretti et al., "Better jet clustering algorithms", *JHEP* **08** (1997) 001, arXiv:hep-ph/9707323, doi:10.1088/1126-6708/1997/08/001.

[131] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering", (1999) arXiv:hep-ph/9907280, DESY-THESIS-2000-049.

[132] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{T} jet clustering algorithm", *JHEP* **04** (2008) 063, arXiv:0802.1189, doi:10.1088/1126-6708/2008/04/063.

[133] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, arXiv:1107.4277, doi:10.1088/1748-0221/6/11/P11002.

[134] CMS Collaboration, "Jet Energy Corrections and Uncertainties. Detector Performance Plots for 2012", *CMS Detector Performance Summaries* **CMS-DP-2012-012** (2012).

[135] CMS Collaboration, "The Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS", *CMS Physics Analysis Summary* **CMS-PAS-JME-09-002** (2009).

[136] CMS Collaboration, "8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb^{-1} of data in CMS", *CMS Detector Performance Summaries* **CMS-DP-2013-033** (2013).

[137] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077.

[138] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, arXiv:1211.4462, doi:10.1088/1748-0221/8/04/P04013.

[139] CMS Collaboration, "Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events", *CMS Physics Analysis Summary* **CMS-PAS-BTV-13-001** (2013).

[140] CMS Collaboration, "Boosted Top Jet Tagging at CMS", *CMS Physics Analysis Summary* **CMS-PAS-JME-13-007** (2014).

[141] CMS Collaboration, "Boosted Top Jet Tagging at CMS", *CMS Detector Performance Summary* **CMS-DP-2014-036** (2014).

[142] CMS Collaboration, "A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging", *CMS Physics Analysis Summary* **CMS-PAS-JME-09-001** (2009).

[143] 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, doi:10.1103/PhysRevLett.101.142001.

[144] T. Plehn, M. Spannowsky, M. Takeuchi et al., "Stop Reconstruction with Tagged Tops", *JHEP* **10** (2010) 078, arXiv:1006.2833, doi:10.1007/JHEP10(2010)078.

[145] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *JHEP* **03** (2011) 015, arXiv:1011.2268, doi:10.1007/JHEP03(2011)015.

[146] J. Thaler and K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", *JHEP* **02** (2012) 093, arXiv:1108.2701, doi:10.1007/JHEP02(2012)093.

[147] M. Schröder, "Quality of Jet Measurements and Impact on a Search for New Physics at CMS", (2012), PhD thesis, Universität Hamburg, Germany, DESY-THESIS-2012-042.

[148] F. Pandolfi, D. del Re, M. Voutilainen, "Jet Response and Resolution Measurement with Photon+Jet Events at $\sqrt{s} = 7$ TeV", *CMS Analysis Note* **CMS-AN-10-141** (2010), Internal documentation.

[149] F. Pandolfi, D. del Re, M. Voutilainen, "Update on Jet Response and Resolution Measurements with Photon+Jet Events at $\sqrt{s} = 7$ TeV", *CMS Analysis Note* **CMS-AN-11-004** (2011), Internal documentation.

[150] T. Lenz, C. Sander, P. Schleper et al., "Jet Energy Resolution Measurement using Photon+Jet Events at $\sqrt{s} = 8$ TeV", *CMS Analysis Note* **CMS-AN-13-179** (2013), Internal documentation.

[151] K. Goebel, J. Haller, J. Ott et al., "Jet Transverse Momentum Resolution Measurement using Dijet Events at $\sqrt{s} = 8\,\text{TeV}$", *CMS Analysis Note* **CMS-AN-13-416** (2013), Internal documentation.

[152] J. Pumplin, D. R. Stump, J. Huston et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **07** (2002) 012, arXiv:hep-ph/0201195, doi:10.1088/1126-6708/2002/07/012.

[153] DØ Collaboration, "High-p_T jets in $\bar{p}p$ collisions at $\sqrt{s} = 630$ GeV and 1800 GeV", *Phys. Rev. D* **64** (2001) 032003, arXiv:hep-ex/0012046, doi:10.1103/PhysRevD.64.032003.

[154] DØ Collaboration, "Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. D* **85** (2012) 052006, arXiv:1110.3771, doi:10.1103/PhysRevD.85.052006.

[155] ATLAS Collaboration, "Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recoreded in 2010 with the ATLAS detector", *Eur. Phys. J. C* **73** (2013) 2306, arXiv:1210.6210, doi:10.1140/epjc/s10052-013-2306-0.

[156] CMS Collaboration, "Measurement of Differential Jet Cross Sections at $\sqrt{s} = 7\,\text{TeV}$ with the CMS Detector", *CMS Physics Analysis Summary* **CMS-PAS-QCD-11-004** (2012).

[157] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting – Summer 2013 Update", *CMS Physics Analysis Summary* **CMS-PAS-LUM-13-001** (2013).

[158] D. Rathjens (2014), PhD thesis in preparation, University of Hamburg.

[159] CMS Collaboration, "Calorimeter Jet Quality Criteria for the First CMS Collision Data", *CMS Physics Analysis Summary* **CMS-PAS-JME-09-008** (2008).

[160] CMS Collaboration, "Jet Performance in *pp* Collisions at 7 TeV", *CMS Physics Analysis Summary* **CMS-PAS-JME-10-003** (2010).

[161] J. Ott, private communication (2014).

[162] 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.

[163] CMS Collaboration, "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV", *JHEP* **03** (2011) 136, arXiv:1102.3194, doi:10.1007/JHEP03(2011)136.

[164] CMS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in *pp* collisions at $\sqrt{s} = 8$ TeV", **CMS-PAS-TOP-14-001** (2014).

[165] M. Seidel (2014), PhD thesis in preparation, University of Hamburg.

[166] DØ Collaboration, "Precision measurement of the top-quark mass in lepton+jets final states", *Phys. Rev. Lett.* **113** (2014) 032002, arXiv:1405.1756, doi:10.1103/PhysRevLett.113.032002.

[167] C. Autermann et al., "Search for new physics in multijets and missing momentum final state at 8 TeV", *CMS Analysis Note* **CMS-AN-12-350** (2012), Internal documentation.

[168] CMS Collaboration, "Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s}$= 8 TeV", *JHEP* **06** (2014) 055, arXiv:1402.4770, doi:10.1007/JHEP06(2014)055.

[169] N. Kidonakis, "Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution", *Phys. Rev. D* **82** (2010) 114030, arXiv:1009.4935, doi:10.1103/PhysRevD.82.114030.

[170] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$", *Phys. Rev. D* **74** (2006) 114017, arXiv:hep-ph/0609070, doi:10.1103/PhysRevD.74.114017.

[171] W. Beenakker, R. Höpker, M. Spira et al., "Squark and gluino production at hadron colliders", *Nucl. Phys. B* **492** (1997) 51–103, arXiv:hep-ph/9610490, doi:10.1016/S0550-3213(97)80027-2.

[172] A. Kulesza and L. Motyka, "Threshold Resummation for Squark-Antisquark and Gluino-Pair Production at the LHC", *Phys. Rev. Lett.* **102** (2009) 111802, doi:10.1103/PhysRevLett.102.111802.

[173] 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** (2009) 095004, doi:10.1103/PhysRevD.80.095004.

[174] W. Beenakker, S. Brensing, M. Krämer et al., "Soft-gluon resummation for squark and gluino hadroproduction", *JHEP* **12** (2009) 041, arXiv:0909.4418, doi:10.1088/1126-6708/2009/12/041.

[175] W. Beenakker, S. Brensing, M. Krämer et al., "Squark and Gluino Hadroproduction", *Int. J. Mod. Phys. A* **26** (2011) 2637–2664, arXiv:1105.1110, doi:10.1142/S0217751X11053560.

[176] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV", *CMS Physics Analysis Summary* **CMS-PAS-EGM-10-004** (2010).

[177] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, arXiv:1206.4071, doi:10.1088/1748-0221/7/10/P10002.

[178] CMS 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 $\sqrt{s} = 8$ TeV", *Eur. Phys. J. C* **73** (2013) 2568, arXiv:1303.2985, doi:10.1140/epjc/s10052-013-2568-6.

[179] CMS Collaboration, "Search for gluino mediated bottom- and top-squark production in multijet final states in *pp* collisions at 8 TeV", *Phys. Lett. B* **725** (2013) 243–270, arXiv:1305.2390, doi:10.1016/j.physletb.2013.06.058.

[180] CMS Collaboration, "Search for supersymmetry in hadronic final states using M_{T2} with the CMS detector at $\sqrt{s} = 8$ TeV", *CMS Physics Analysis Summary* **CMS-PAS-SUS-13-019** (2014).

[181] CMS 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", *CMS Physics Analysis Summary* **CMS-PAS-SUS-14-011** (2014).

[182] S. Ask, M. Parker, T. Sandoval et al., "Using γ + jets Production to Calibrate the Standard Model $Z(\to \nu\bar{\nu})$ + jets Background to New Physics Processes at the LHC", *JHEP* **10** (2011) 058, arXiv:1107.2803, doi:10.1007/JHEP10(2011)058.

[183] Z. Bern, G. Diana, L. Dixon et al., "Driving Missing Data at Next-to-Leading Order", *Phys. Rev. D* **84** (2011) 114002, arXiv:1106.1423, doi:10.1103/PhysRevD.84.114002.

[184] Z. Bern, G. Diana, L. Dixon et al., "Missing Energy and Jets for Supersymmetry Searches", *Phys. Rev. D* **87** (2013) 034026, arXiv:1206.6064, doi:10.1103/PhysRevD.87.034026.

[185] B. Efron, "The Jackknife, the bootstrap and other resampling plans", (1982), Number 38 in Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia.

[186] J. D'Hondt, S. Lowette, O. L. Buchmüller et al., "Fitting of Event Topologies with External Kinematic Constraints in CMS", *CMS Note* **CMS-NOTE-2006-023** (2006).

[187] CMS Collaboration, "Study of the underlying event at forward rapidity in pp collisions at $\sqrt{s} = 0.9$, 2.76, and 7 TeV", *JHEP* **1304** (2013) 072, arXiv:1302.2394, doi:10.1007/JHEP04(2013)072.

[188] A. L. Read, "Presentation of search results: the CL_{s} technique", *J. Phys. G: Nucl. Part. Phys.* **28** (2002) 2693–2704, doi:10.1088/0954-3899/28/10/313.

[189] T. Junk, "Confidence level computation for combining searches with small statistics", *NIM A* **434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2.

[190] ATLAS and CMS Collaborations, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", **ATL-PHYS-PUB-2011-011. CMS-NOTE-2011-005** (2011).

[191] M. Botje, J. Butterworth, A. Cooper-Sarkar et al., "The PDF4LHC Working Group Interim Recommendations", (2011) arXiv:1101.0538.

[192] 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, doi:10.1140/epjc/s10052-013-2677-2.

[193] 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, doi:10.1103/PhysRevD.78.013004.

[194] A. Martin, W. Stirling, R. Thorne et al., "Parton distributions for the LHC", *Eur. Phys. J. C* **63** (2009) 189–285, arXiv:0901.0002, doi:10.1140/epjc/s10052-009-1072-5.

[195] ATLAS Collaboration, "Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS experiment", *JHEP* **10** (2013) 130, arXiv:1308.1841, doi:10.1007/JHEP10(2013)130.

[196] ATLAS Collaboration, "Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum using $\sqrt{s} = 8$ TeV proton–proton collision data", *JHEP* **09** (2014) 176, arXiv:1405.7875, doi:10.1007/JHEP09(2014)176.

[197] ATLAS Collaboration, "Search for strong production of supersymmetric particles in final states with missing transverse momentum and at least three b-jets at $\sqrt{s} =$ 8 TeV proton-proton collisions with the ATLAS detector", *JHEP* **10** (2014) 24, arXiv:1407.0600, doi:10.1007/JHEP10(2014)024.

[198] CMS Collaboration, "Phenomenological MSSM interpretation of the CMS 7 and 8 TeV results", *CMS Physics Analysis Summary* **CMS-PAS-SUS-13-020** (2014).

[199] P. Bechtle, T. Bringmann, K. Desch et al., "Constrained Supersymmetry after two years of LHC data: a global view with Fittino", (2012) arXiv:1204.4199.

[200] O. Buchmueller, R. Cavanaugh, A. De Roeck et al., "The CMSSM and NUHM1 after LHC Run 1", *Eur. Phys. J. C* **74** (2014) 2922, arXiv:1312.5250, doi:10.1140/epjc/s10052-014-2922-3.

[201] C. P. Robert, "The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation (Springer Texts in Statistics)", (2001), Springer-Verlag New York.

[202] A. O'Hagan, "Bayesian inference", (1994), Number Vol. 2B in Kendall's Advanced Theory of Statistics, Edward Arnold, London.

[203] CMS Collaboration, "Search for top squarks in multijet events with large missing momentum in proton-proton collisions at 8 TeV", *CMS Physics Analysis Summary* **CMS-PAS-SUS-13-015** (2013).

[204] M. Aliev, H. Lacker, U. Langenfeld et al., "HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034–1046, arXiv:1007.1327, doi:10.1016/j.cpc.2010.12.040.

[205] J. Ott, "Theta framework", Retrieved Nov. 2014, www.theta-framework.org.

[206] 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, doi:10.1016/S0370-2693(99)00945-4.

[207] 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, doi:10.1088/0954-3899/29/10/304.

[208] CMS Collaboration, "Search for supersymmetry in hadronic final states using M_{T2} in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **10** (2012) 018, arXiv:1207.1798, doi:10.1007/JHEP10(2012)018.

[209] H.-C. Cheng and Z. Han, "Minimal Kinematic Constraints and M_{T2}", *JHEP* **12** (2008) 063, arXiv:0810.5178, doi:10.1088/1126-6708/2008/12/063.

[210] CMS Collaboration, "Inclusive search for supersymmetry using the razor variables in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **111** (2013) 081802, arXiv:1212.6961, doi:10.1103/PhysRevLett.111.081802.

[211] CMS Collaboration, "Search for supersymmetry with razor variables in pp collisions at $\sqrt{s}$=7 TeV", (2014) arXiv:1405.3961, submitted to PRD.

[212] CMS Collaboration, "Search for supersymmetry in final states with missing transverse energy and 0, 1, 2, or $\geq$ 3 b-quark jets in 7 TeV pp collisions using the variable α_T", *JHEP* **01** (2013) 077, arXiv:1210.8115, doi:10.1007/JHEP01(2013)077.

[213] CMS Collaboration, "Jet energy calibration in the 8 TeV pp data", (2014), Paper in preparation, JME-13-004.

List of Publications

My work during the PhD resulted in this publication:

CMS Collaboration, "Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 8$ TeV", JHEP 06 (2014) 055, arXiv:1402.4770, doi:10.1007/JHEP06(2014)055

The following publication is in preparation:

CMS Collaboration, "Jet energy calibration in the 8 TeV *pp* data", (2014), JME-13-004, paper in preparation

Danksagung

Ich möchte mich an dieser Stelle bei allen bedanken, die mich während der Zeit meiner Doktorarbeit unterstützt haben. Besonderer Dank gilt dabei:

- Prof. Dr. Johannes Haller für die Betreuung meiner Doktorarbeit, die wöchentlichen Diskussionen und insbesondere die Möglichkeit an zahlreichen Workshops, Konferenzen und Sommerschulen teilnehmen zu dürfen.
- Jun.-Prof. Dr. Christian Sander für die Erstellung des Zweitgutachtens und die umfangreiche Betreuung insbesondere in der ersten Hälfte meiner Doktorandenzeit. Danke für die jederzeit geduldige Beantwortung aller meiner Fragen rund um Susy, Jets und R+S.
- Prof. Dr. Dieter Horns, Prof. Dr. Jan Louis und Prof. Dr. Peter Schleper für die Bereitschaft Mitglied meiner Prüfungskommission zu sein, sowie die interessanten und lehrreichen Fragen in meiner Disputation.
- Jochen Ott für die engagierte Unterstützung meiner Arbeit zur Messung der Jetimpulsauflösung und den Studien zur Suche nach Top-Squarks. Deine unerschöpfliche Menge an Ideen und Vorschlägen ist wirklich beeindruckend und ich danke dir besonders für das akribische Lesen meiner Arbeit.
- Roman Kogler und Thomas Peiffer für das Korrekturlesen meiner Arbeit und viele nützliche Hinweise.
- der Arbeitsgruppe am Institut für eine angenehme und entspannte Arbeitsatmosphäre. Besonderer Dank gilt hierbei meinen Kolleginnen Rebekka, Teresa und Mareike für die vielen schönen Stunden gemeinsam im Büro.
- meinen Freunden für die Ablenkung vom Arbeitsalltag.
- meinem Mann Martin für stets motivierende und aufmunternde Worte während der letzten drei Jahre.
- meiner Familie, weil sie immer an mich glaubt und jederzeit bereit ist mich tatkräftig zu unterstützen.

Vielen Dank!