

# **Probing supersymmetry at the CMS experiment based on precise jet measurements**

**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



## 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 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 lies 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 observed one 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 and missing transverse momentum. 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 production of gluino-mediated 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. A method relying on the jet- $p_T$  resolution to estimate QCD background contributions is presented and necessary modifications to correctly predict 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 study based on simulated events at a centre of mass energy of  $\sqrt{s} = 13 \text{ TeV}$  is shown, investigating different search strategies towards the identification of direct pair production of top squarks. Utilizing algorithms for the identification of boosted hadronically decaying top quarks arising from the decay of heavy top squarks, a search sensitivity of top squark masses up to 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 improves 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 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 Jet-Transversalimpuls-Auflösung 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 Zweijet-Ereignissen 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 und fehlendem Transversalimpuls. Es wird eine Suche vorgestellt, die auf Kollisionsdaten basiert, welche bei einer Schwerpunktsenergie von  $\sqrt{s} = 8 \text{ TeV}$  aufgezeichnet wurden, und auf Signaturen abzielt, welche hauptsächlich sensitiv sind auf die Produktion von leichten-flavour Squarks und Gluinos, sowie die Produktion von gluino-induzierten Teilchen der dritten Generation. Die größte Herausforderung ergibt sich in dieser Analyse durch eine genaue Bestimmung der Untergrundbeiträge aus Standardmodell Prozessen, da die Analyse in einem extremen kinematischen Phasenraum durchgeführt wird. Es wird eine Methode vorgestellt, die basierend auf der jet- $p_T$  Auflösung Untergrundbeiträge aus QCD Ereignissen 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 leichten-flavour Squarks unter 780 GeV und die von Gluinos unter 1,1–1,2 TeV im Kontext von vereinfachten supersymmetrischen Modellen mit 95% confidence level für eine Masse des leichtesten supersymmetrischen Teilchens (LSP) unter 100 GeV ausgeschlossen. Weiterhin wird eine Studie basierend auf simulierten Ereignissen bei einer Schwerpunktsenergie von  $\sqrt{s} = 13 \text{ TeV}$  vorgestellt, welche unterschiedliche Analysestrategien zur Identifikation von direkt produzierten Top-Squarks untersucht. Unter Verwendung von Algorithmen zur Identifikation von geboosteten hadronisch zerfallenden Top-Quarks aus den Zerfällen von Top-Squarks, kann mit denselben integrierten Luminosität wie bei  $\sqrt{s} = 8 \text{ TeV}$  aufgezeichnet werden, eine Sensitivität der Suche für Top Squark Massen bis 1 TeV für LSP Massen unter 300 GeV erreicht werden. Diese Selektion verbessert die Sensitivität der Suche gegenüber bestehenden Analysen. Darüberhinaus ist die identifizierte Selektion auch geeignet um gluino-induzierte Squarks der dritten Generation zu studieren und bietet einen komplementären Ansatz zu existierenden Multijet-Analysen.



# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>1</b>  |
| <b>2</b> | <b>Phenomenological Aspects of the Standard Model and Beyond</b>           | <b>3</b>  |
| 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 . . . . .   | 14        |
| 2.3.2    | Direct Searches at Collider Experiments . . . . .                          | 14        |
| <b>3</b> | <b>Experimental Setup</b>  | <b>23</b> |
| 3.1      | The Large Hadron Collider . . . . .  | 23        |
| 3.2      | The CMS Experiment . . . . .   | 25        |
| 3.2.1    | Coordinate Conventions and Kinematic Variables . . . . .                   | 27        |
| 3.2.2    | Superconducting Magnet . . . . .   | 28        |
| 3.2.3    | Inner Tracking System . . . . .  | 28        |
| 3.2.4    | Electromagnetic Calorimeter . . . . .                                      | 29        |
| 3.2.5    | Hadron Calorimeter . . . . .   | 31        |
| 3.2.6    | Muon System . . . . .  | 33        |
| 3.2.7    | Trigger System . . . . .   | 34        |
| 3.3      | LHC Operation and Data Taking . . . . .                                    | 35        |
| <b>4</b> | <b>Event Simulation</b>  | <b>39</b> |
| 4.1      | Event Modeling . . . . .   | 39        |
| 4.2      | Event Generators . . . . .   | 41        |
| 4.3      | Detector Simulation . . . . .  | 42        |
| <b>5</b> | <b>Object Reconstruction and Particle Identification</b>                   | <b>43</b> |
| 5.1      | Global Event Description with the Particle-Flow Algorithm at CMS . . . . . | 43        |
| 5.2      | Reconstruction of Jets . . . . .   | 44        |
| 5.2.1    | Jet Algorithms . . . . .   | 44        |
| 5.2.2    | Jet Types at CMS . . . . .   | 47        |
| 5.2.3    | Jet Transverse-Momentum Response . . . . .                                 | 48        |
| 5.2.4    | Jet Energy Calibration . . . . .   | 48        |
| 5.3      | Identification of b-Quark Jets . . . . .                                   | 51        |
| 5.4      | Identification of Boosted t-Quark Jets . . . . .                           | 53        |

---

|   |           |
|---|-----------|
| <b>6 Measurement of the Jet Transverse-Momentum Resolution</b>              | <b>59</b> |
| 6.1 Components of the Jet Response . . . . .                                | 59        |
| 6.2 Basic Concept of the Dijet Asymmetry Method . . . . .                   | 61        |
| 6.3 Application to Realistic Collision Events . . . . .                     | 61        |
| 6.4 Samples and Event Selection . . . . .                                   | 62        |
| 6.4.1 Datasets and Triggers . . . . .                                       | 62        |
| 6.4.2 Selection Criteria . . . . .  | 63        |
| 6.5 Definition of the Asymmetry Width . . . . .                             | 67        |
| 6.6 Corrections to the Dijet Asymmetry . . . . .                            | 68        |
| 6.6.1 Correction for Additional Jet Activity . . . . .                      | 70        |
| 6.6.2 Correction for Particle-Level Imbalance . . . . .                     | 71        |
| 6.6.3 Results of the Corrections to the Asymmetry . . . . .                 | 72        |
| 6.7 Determination of the Data-to-Simulation Ratio . . . . .                 | 73        |
| 6.8 Validation of the Method . . . . .                                      | 75        |
| 6.8.1 Validation in Simulated Events . . . . .                              | 75        |
| 6.8.2 Validation of the Measured Data-to-Simulation Ratio . . . . .         | 76        |
| 6.9 Systematic Uncertainties . . . . .                                      | 76        |
| 6.10 Extension of the Method to the Forward Detector Region . . . . .       | 82        |
| 6.11 Measurement for Simulated Events Obtained with the HERWIG++ Generator  | 84        |
| 6.12 Results . . . . .  | 85        |
| 6.12.1 Determination of a combined result . . . . .                         | 85        |
| 6.12.2 Comparison to Other Measurements . . . . .                           | 88        |
| 6.12.3 Impact of the Improved JER Measurement . . . . .                     | 89        |
| 6.13 Adjustment of the MC-Resolution to Data . . . . .                      | 90        |
| <b>7 Search for New Physics with Jets and Missing Transverse Momentum</b>   | <b>93</b> |
| 7.1 Event Selection . . . . .   | 94        |
| 7.1.1 Data Samples . . . . .  | 94        |
| 7.1.2 Trigger . . . . .   | 94        |
| 7.1.3 Event Cleaning . . . . .  | 96        |
| 7.1.4 Baseline Selection . . . . .  | 98        |
| 7.1.5 Exclusive Search Regions . . . . .                                    | 99        |
| 7.2 Estimation of Non-QCD Backgrounds . . . . .                             | 99        |
| 7.2.1 Invisible Z Background . . . . .                                      | 99        |
| 7.2.2 Hadronic-tau Background . . . . .                                     | 103       |
| 7.2.3 Lost-Lepton Background . . . . .                                      | 104       |
| 7.3 QCD Background Estimation with the Rebalance-And-Smear Method . . . . . | 106       |
| 7.3.1 General concept of the Rebalance-and-Smear method . . . . .           | 106       |
| 7.3.2 Application to Collision Events . . . . .                             | 112       |
| 7.3.3 Validation in Simulated Events . . . . .                              | 114       |
| 7.3.4 Application to Data Events . . . . .                                  | 117       |
| 7.3.5 Systematic Uncertainties . . . . .                                    | 119       |
| 7.3.6 QCD Background Prediction . . . . .                                   | 121       |
| 7.4 Results and Interpretation . . . . .                                    | 121       |
| 7.4.1 Comparison to Other Measurements . . . . .                            | 127       |
| 7.5 Status of supersymmetry after LHC Run I . . . . .                       | 127       |

|  |            |
|--|------------|
| <b>8 Prospect Studies for a Search for Top Squarks at <math>\sqrt{s} = 13 \text{ TeV}</math></b> | <b>131</b> |
| 8.1 Data Samples . . . . .   | 132        |
| 8.2 Sensitivity of a Basic Selection using $H_T$ and $\cancel{E}_T$ . . . . .                    | 133        |
| 8.3 Sensitivity Improvement using B Tagging . . . . .  | 138        |
| 8.4 Sensitivity Improvement using Top Tagging . . . . .  | 139        |
| 8.5 Performance Comparison of Various Kinematic Selections . . . . .                             | 144        |
| 8.6 Performance Comparison to Selection Based on Published Stop Analysis . .                     | 150        |
| 8.7 Stability Test . . . . .   | 153        |
| 8.8 Sensitivity to Gluino-Mediated Stop Production . . . . .                                     | 155        |
| 8.9 Results and Discussion . . . . .   | 156        |
| <b>9 Conclusions</b>   | <b>159</b> |
| <b>A Jet Transverse-Momentum Resolution</b>  | <b>161</b> |
| A.1 Primary Vertex Distributions . . . . .   | 161        |
| A.2 Extrapolation Graphs . . . . .   | 163        |
| A.3 Correlation of Asymmetry Widths . . . . .  | 168        |
| A.4 Detailed Results of the Forward Extension . . . . .  | 170        |
| <b>B Search for New Physics with Jets and Missing Transverse Momentum</b>                        | <b>173</b> |
| B.1 Signal Trigger Efficiencies . . . . .  | 173        |
| B.2 Data-to-Simulation Ratios: Core . . . . .  | 175        |
| B.3 Data-to-Simulation Ratios: Tails . . . . .   | 176        |
| B.4 Input Truth Resolutions for Rebalancing in Kinematic Fit . . . . .                           | 177        |
| B.5 Trigger Efficiencies of Control Triggers for QCD Background Prediction . .                   | 179        |



# 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 opportunities 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<sup>1</sup> is currently the most powerful particle accelerator and provides proton-proton collisions at a centre of mass energy of up to  $\sqrt{s} = 8 \text{ TeV}$  to date. In order to search for supersymmetry and to further investigate the SM, the CMS experiment has been built. The CMS experiment is a particle detector designed to analyse particle collisions delivered by the LHC and in this thesis studies are presented that are based on data recorded by this experiment.

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  $\gamma + \text{jet}$ ,  $Z + \text{jet}$  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 and several hard jets. Previous versions of this analysis have been performed at  $\sqrt{s} = 7 \text{ TeV}$  and were especially 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 arise from mismeasured

---

<sup>1</sup>*European Organization for Nuclear Research near Geneva, Suisse*

jets and decays of heavy-flavour quarks. In this thesis, a method relying on the jet- $p_T$  resolution to estimate QCD background contributions is presented and special considerations to correctly predict high jet multiplicity events are discussed.

Since analyses of  $\sqrt{s} = 8$  TeV data exclude gluino and light-flavour squark masses below around 1 TeV, 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  $\sqrt{s} = 13$  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  $\sqrt{s} = 13$  TeV are discussed. Special emphasis is put on the study of several kinematic variables and the application of jet substructure tools.

This 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 up 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 at  $\sqrt{s} = 8$  TeV as well as simulated events.

**Chapter 7:** A search for supersymmetry in final states containing several hard jets and missing transverse momentum at  $\sqrt{s} = 8$  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  $\sqrt{s} = 13$  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 (SM) 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 some open questions which can not be answered within the SM. Thus, several theories have been developed to address problems which go beyond the SM. One of such 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 where the particles are denoted together with their interactions<sup>1</sup>.

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

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

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

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

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

*Leptons:* The SM contains, in total, six leptons, which are three negatively charged leptons ( $e$ ,  $\mu$ ,  $\tau$ ) and three neutral leptons ( $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$ ), the neutrinos. In

---

<sup>1</sup>Gravity is not included in the standard model and thus it is not discussed in this thesis.

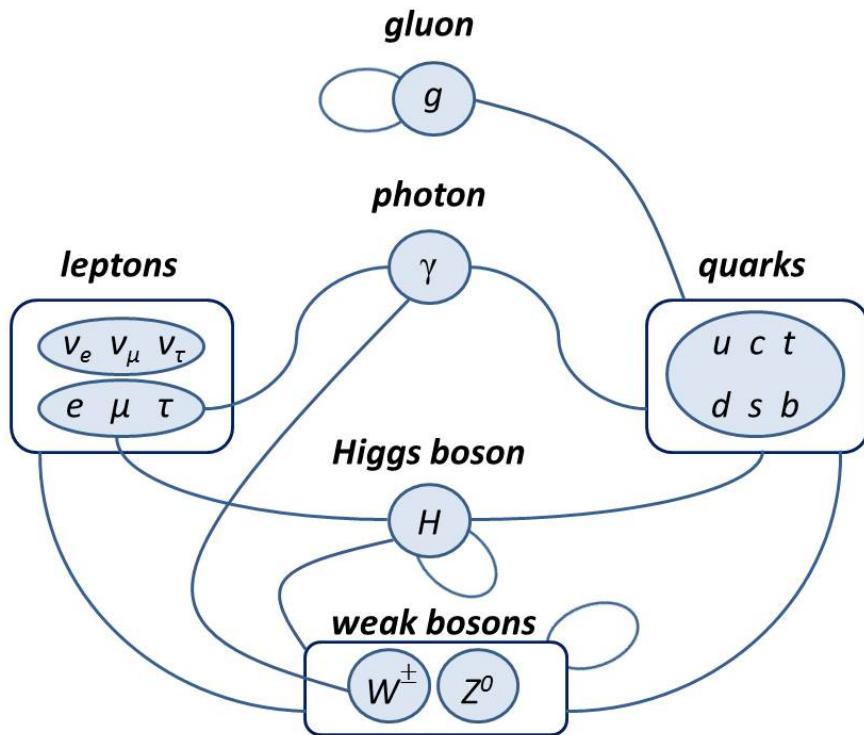


Figure 2.1: Overview of particles contained in the standard model. Blue lines indicate interactions between different particles.

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  $\nu_e$  belong to the first generation,  $\mu$  and  $\nu_\mu$  to the second and  $\tau$  and  $\nu_\tau$  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 where the up-type flavour eigenstates correspond to the mass eigenstates and the down-type quarks mix. This mixing is described by the CKM-matrix [2, 3]. This is 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 phenomenon 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  $B_\mu$  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  $W_\mu^1$  and  $W_\mu^2$ , the fields  $A_\mu$  of the photon and  $Z_\mu$  of the neutral vector boson are obtained by a mixing of the gauge fields  $W_\mu^3$  and  $B_\mu$  according to

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

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

$$e = g \sin \theta_W . \quad (2.2)$$

The fields  $W^\mu$  couple only to left-handed fermions, such that the same holds also for  $W^\pm$ . Since however, the  $B^\mu$  couples to left- and right-handed states, a coupling to left- and right-handed fermions takes place for  $\gamma$  and  $Z^0$  as well. Unlike the photon, the  $W^\pm$  and  $Z$  vector bosons are massive with masses<sup>2</sup> of  $W^\pm = 80.385 \pm 0.015$  GeV and  $Z = 91.1876 \pm 0.0021$  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 mechanism 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  $\Phi$  described by a potential

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

with parameters  $\mu$  and  $\lambda$ . Choosing  $\mu^2$  to be negative and  $\lambda$  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

---

<sup>2</sup>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 antimatter 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  $\lambda_f$  to the Higgs field, the corrections to  $m_H^2$  are

$$\Delta m_H^2 \propto -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{\text{UV}}^2 \propto m_f^2 \quad (2.3)$$

where  $\Lambda_{\text{UV}}$  is an ultraviolet cut-off scale. 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  $\mathcal{O}(100)$  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

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

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 recent lower limit on the proton lifetime is found to be  $5.9 \times 10^{33}$  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.<sup>3</sup> 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 cancellation 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 correspondents 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 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 principle 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 constrained by experimental results (cf. for instance [28] for a comprehensive review).

---

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

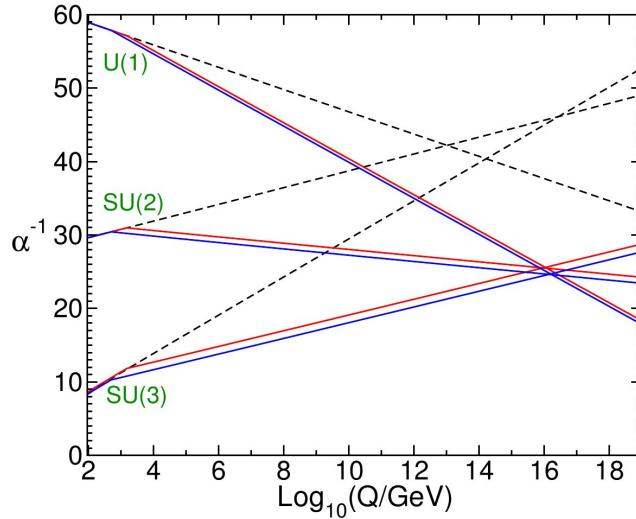


Figure 2.2: Comparison of the renormalization group evolution of the couplings  $\alpha_a^{-1}$  in the SM (dashed lines) and the MSSM (solid lines) including two-loop effects. The masses of the supersymmetric particles in the MSSM are considered as a common threshold changing between 500 GeV and 1.5 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  $\lambda_0$ . This constraint is often imposed by quantifying the amount of necessary fine-tuning  $\Delta$  according to

$$\left| \frac{\lambda_0}{m^2} \frac{\partial m^2}{\partial \lambda_0} \right| < \Delta . \quad (2.4)$$

While the value of  $\Delta$  has been set to  $\sim 10$  in the past, it increased over the years such that also values of  $\sim 100$  or even  $\sim 1000$  are considered as reasonable values [32]. Such boundaries 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

| 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    | $\tilde{u}_L \ \tilde{u}_R \ \tilde{d}_L \ \tilde{d}_R$<br>$\tilde{s}_L \ \tilde{s}_R \ \tilde{c}_L \ \tilde{c}_R$<br>$\tilde{t}_L \ \tilde{t}_R \ \tilde{b}_L \ \tilde{b}_R$ | see left<br>see left<br>$\tilde{t}_1 \ \tilde{t}_2 \ \tilde{b}_1 \ \tilde{b}_2$ |
| Sleptons     | 0    | $\tilde{e}_L \ \tilde{e}_R \ \tilde{\nu}_e$<br>$\tilde{\mu}_L \ \tilde{\mu}_R \ \tilde{\nu}_\mu$<br>$\tilde{\tau}_L \ \tilde{\tau}_R \ \tilde{\nu}_\tau$                      | see left<br>see left<br>$\tilde{\tau}_1 \ \tilde{\tau}_2 \ \tilde{\nu}_\tau$    |
| 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].

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.

### 2.2.2 The MSSM

In general, it is possible to have theories with more than one supersymmetry transformation. 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. 2.2.2. This extension is minimal in that 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 where one doublet  $H_u$  gives mass to the up-type quarks and the other one  $H_d$  to the down-type quarks, respectively. These two doublets have together eight degrees of freedom of which three are needed to give mass to the gauge bosons of the weak interaction as in the SM. This results in five physical Higgs bosons which are the two scalar Higgs particles  $h^0, H^0$ , the pseudoscalar  $A^0$  as well as the charged Higgs bosons  $H^\pm$ . As further consequence, there are two vacuum expectation values  $v_u$  and  $v_d$  present, each assigned to one Higgs doublet, whose ratio  $\tan\beta = v_u/v_d$  is a free parameter of the model. 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

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

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 stop quark 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, 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 stop 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 stop mass, they are expected to not significantly exceed the 1 TeV range, as a second order effect [32].

### 2.2.3 SUSY-Breaking

As discussed above, SUSY particles are expected to have the same mass as their corresponding SM partner particle. However, none of such particles has been observed so far which implies that supersymmetry in fact must be broken and that sparticles are actually heavier than the SM counterpart. If SUSY is expected to provide a solution to the hierarchy problem, the mass difference to the SM particles should consequently not be too large and the characteristic mass scale needs to be around 1 TeV [26].

Typically, SUSY breaking is introduced such that so-called *soft breaking* terms, i.e. of

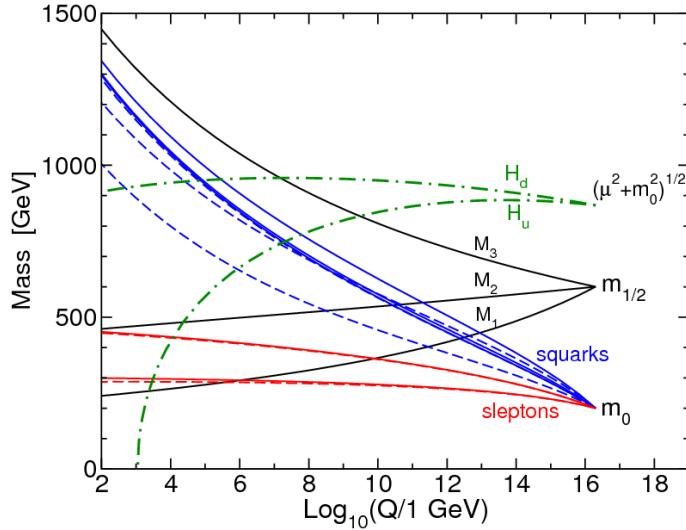


Figure 2.3: Evolution of scalar and gaugino mass parameters in the MSSM with mSUGRA boundary conditions imposed at  $Q_0 = 2 \times 10^{16}$  GeV. The parameter  $\mu^2 + m_{H_u}^2$  runs negative, provoking electroweak symmetry breaking. Taken from [26].

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 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  $\mu$ . In Fig. 2.3 the evolution of the corresponding mass parameters to the electroweak scale are illustrated.

## 2.3 Searches for Supersymmetry and Current Constraints

The appealing attributes of supersymmetry, which have been discussed above, initiated a couple of indirect and direct searches looking for hints of supersymmetric particles. Although no sign for SUSY has been observed in nature so far, several results have been used in order to constrain the allowed parameter space. This section gives a short introduction to general search strategies for supersymmetry with main focus on collider experiments

and provides a short overview of the status of experimental results. Results of direct searches at the LHC discussed in this section are restricted to results from searches performed previous to the analysis presented in this thesis. Further discussion of the status of supersymmetry after LHC Run-I follows in Sec. 7.5.

### 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 contributions might impact for instance electroweak precision data, rare decays of B-mesons or the anomalous magnetic moment of the muon. However, global fits to electroweak precision data using several precision measurements of SM parameters and particle masses, like  $m_W$  and  $m_t$ , together with theoretical calculations have found no evidence for any inconsistency of the SM only hypothesis so far [41–44]. Furthermore, precise measurements of rare processes in B meson decays, like  $B_s^0 \rightarrow \mu^+ \mu^-$ , are in good agreement with expectations from the SM [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  $3.6\sigma$  occur. However, discussions about the accuracy of the SM calculation are ongoing [50].

Further constraints arise from astrophysical and cosmological observations. These occur from direct [51] and indirect [52] dark matter searches. Moreover, several observations suggest that a considerable amount of cold dark matter contributes to the composition of the universe [22, 53]. 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 relic density can put constraints on the MSSM parameter space assuming that it is caused by a neutralino LSP.

### 2.3.2 Direct Searches at Collider Experiments

Although the exploitation of indirect searches for supersymmetry is very useful in order to constrain the allowed SUSY parameter space, the most stringent exclusion limits are derived from direct searches at collider experiments. Typically, searches for SUSY at colliders make use of the specific production and decay properties 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:** At HERA, searches for  $R$ -parity conserving supersymmetric models have mainly targeted processes involving the production of a selectron and a squark. The dominant MSSM process at HERA is the production of a selectron and a squark via the

exchange of a neutralino where  $\tilde{e}$  and  $\tilde{q}$  subsequently decay into any lighter gaugino and their respective SM partner. Thus, a distinct experimental signature is given by an electron, hadrons and missing energy and momentum. Since no excess of such events over the SM expectation has been observed, exclusion limits for the masses of the selectron and squark in the context of the MSSM have been derived. Depending on the specific assumptions made to constrain the MSSM parameter space, the excluded region extends to around 60–70 GeV for squark masses and around 40 GeV for the LSP mass [54, 55].

**LEP:** At LEP, various searches for supersymmetry were performed targeting different species of supersymmetric particles. 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. Typically, 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  $Z/\gamma$   $s$ -channel exchange or  $t$ -channel selectron or sneutrino exchange, respectively. Typically, charginos decay into  $\tilde{\chi}_1^0 l \nu$  or  $\tilde{\chi}_1^0 q q'$  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  $\tilde{\chi}_1^0 q q'$ . Hence, the final state in case of chargino production is characterized by missing energy accompanied by four jets, two jets and one lepton or only leptons, depending on the specific decay mode of the chargino while the most important signature for neutralino production are acoplanar pairs of jets or leptons coming along with large missing momentum. However, the exact decay topologies strongly depend on the particular mass spectrum of the supersymmetric particles so that the above mentioned topologies could also be accompanied by photons or manifest in multiple jets or leptons in cascade decays.

Interpretations of combined results from all four experiments in the mSUGRA model lead to exclusion limits showing that  $m_{1/2}$  has to be greater than 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 [56–60].

**Tevatron:** The Tevatron accelerator made a further SUSY parameter space for searches accessible, 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 below.

Results from SUSY searches at the Tevatron have been interpreted in the context of mSUGRA and extended the LEP results in the parameter region of  $m_0 = 70\text{--}300$  GeV and  $m_{1/2} = 125\text{--}165$  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 [61–63].

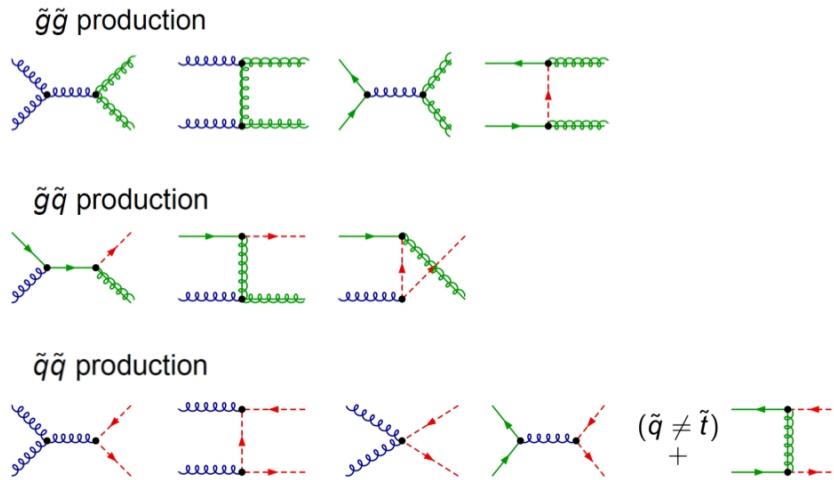


Figure 2.4: Example diagrams for the production of supersymmetric particles in hadron collisions at parton level.

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. It made already with a centre of mass energy of 7 TeV an even larger SUSY parameter space accessible than the Tevatron. At the LHC, a variety of searches targeting the various different production and decay modes of supersymmetric particles are performed. In general, they can be classified into searches for electroweakinos, third generation sfermions and searches for squarks and gluinos. Of particular interest are searches for coloured particles, as the LHC is, as well as the Tevatron, a hadron collider.

At leading order sparticles in  $R$ -parity conserving models are predominantly produced in processes like [64–68]

$$pp \rightarrow \tilde{g}\tilde{g}, \tilde{g}\tilde{q}, \tilde{q}\tilde{q}. \quad (2.5)$$

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, e.g. by quark radiation or gluon loops, typically up to next-to-leading order (NLO). Production cross sections for the processes  $pp \rightarrow \tilde{g}\tilde{g}$ ,  $pp \rightarrow \tilde{q}\tilde{q}$  and  $pp \rightarrow \tilde{t}\tilde{t}$  are illustrated in Fig. 2.5 for different sparticle masses as 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.

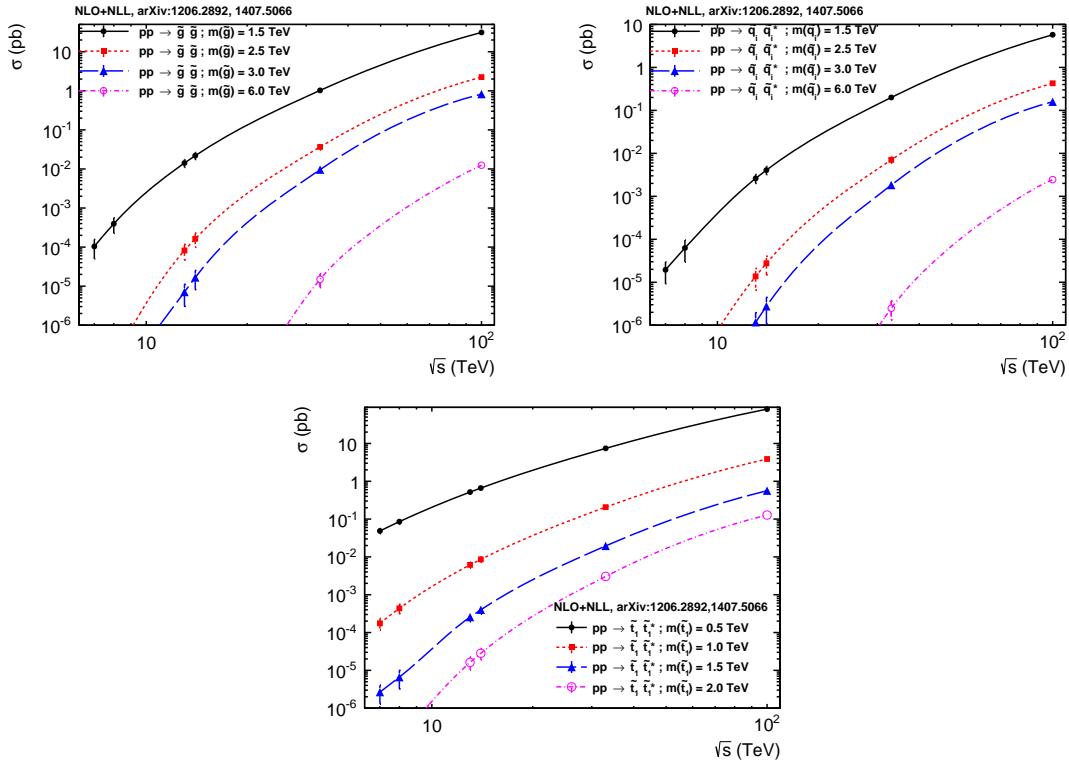


Figure 2.5: SUSY production cross sections of processes  $pp \rightarrow \tilde{g} \tilde{g}$  (top left),  $pp \rightarrow \tilde{q} \tilde{q}$  (top right) and  $pp \rightarrow \tilde{t} \tilde{t}$  (bottom) displayed for different sparticle masses shown as function of the centre of mass energy [69, 70].

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 decay of the gluinos in the upper two diagrams have to be understood as 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. While squarks are expected to decay preferably into a quark and the LSP, gluinos predominantly decay into a quark pair and the LSP. In case of gluino decays, the quark pair can also be, for instance, a pair of top quarks representing the case of gluino-mediated stop production. In all such cases, a multijet final state accompanied by missing transverse momentum with no isolated leptons is expected as experimental signature<sup>4</sup>.

However, final states containing multiple jets accompanied by large values of missing transverse energy do not only arise from SUSY events, but are also realised for several SM processes. For any new physics search, such SM processes have to be considered as background, which is a crucial task in each SUSY analysis. In case of multijet +  $\cancel{E}_T$  searches these are typically  $Z + \text{jets} \rightarrow \nu\bar{\nu} + \text{jets}$  events in which large genuine  $\cancel{E}_T$  is caused by the

<sup>4</sup>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.

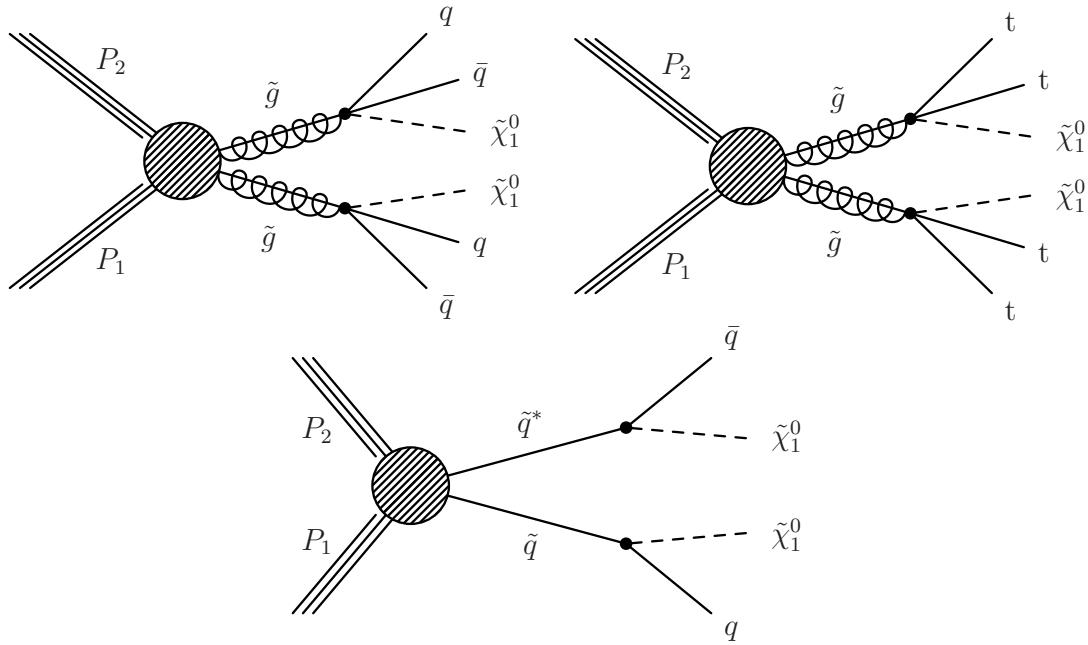


Figure 2.6: Example diagrams of different SUSY decay channels showing  $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_1^0$  (top left),  $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$  (top right) and  $\tilde{g} \rightarrow q\tilde{\chi}_1^0$  (bottom) [71].

neutrinos. This background is denoted *invisible Z background* in this thesis. Furthermore, events with intrinsic missing energy stem from  $W + \text{jets}$  and  $t\bar{t}$  events. The top quark is the heaviest quark in the SM with a mass of  $173.34 \pm 0.76$  GeV [72]. Since its lifetime is smaller than the hadronisation timescale, it decays before it forms colour-neutral hadrons. The decay takes place via the weak interaction and as denoted in Sec. 2.1, the quark mixing is parametrized by the CKM-matrix. Since the corresponding matrix element  $V_{tb} \approx 1$ , the top quark decays almost exclusively into a  $W$  boson and a  $b$  quark. Experimentally, the  $W$  boson can decay into a charged lepton and its corresponding neutrino or a pair of light quarks of the first two generations. Taking into account the three possible colour states for each quark pair, this 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,  $W + \text{jets}$  and  $t\bar{t}$  events featuring a decay containing an electron or muon that is not reconstructed, not isolated or falling out of the detector acceptance have to be considered as possible background. This source of background events is referred to as *lost-lepton background*. Furthermore, also  $W + \text{jets}$  and  $t\bar{t}$  in which the lepton is a hadronically decaying  $\tau$  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, severely mismeasured jets can give rise to large amounts of  $\cancel{E}_T$  for instance because of instrumental effects or semi-leptonically decaying heavy-flavour quarks. This background process is denoted *QCD background* in the following. Contributions from other SM processes are found to be negligible [73, 74].

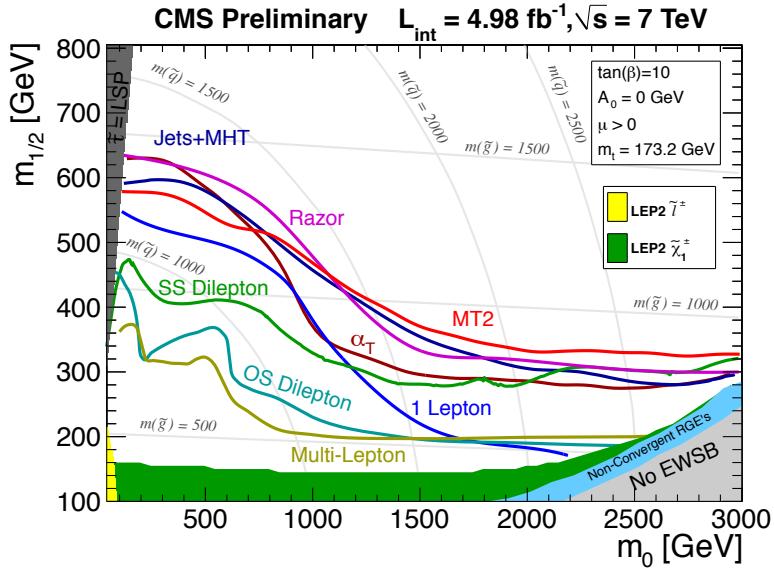


Figure 2.7: Interpretation of searches for supersymmetry at the CMS experiment within the CMSSM. Shown are the 95% CL exclusion limits in the  $m_0$  and  $m_{1/2}$  plane for various searches performed using different final state topologies [71].

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. 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 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 all-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. Thus, 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* [75–78]. 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. Thus, the characterization 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



Figure 2.8: Interpretation of searches for supersymmetry at the CMS experiment with simplified models in  $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}^0$  (left) and  $\tilde{q} \rightarrow q\tilde{\chi}^0$  (right) topologies. Shown are the 95% CL upper limits on the produced particle and LSP masses. The grey area represents the region where the respective decay mode is forbidden [78].

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 illustrated in Fig. 2.8 for  $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}^0$  and  $\tilde{q} \rightarrow q\tilde{\chi}^0$ . These illustrate the 95% confidence level upper limit on the product of the cross section and branching fraction as function of the sparticles masses. Hence, the values of cross sections times branching ratio can be compared to any theoretical prediction in order to determine whether the specific model is compatible with data. The exclusion curves shown in Fig. 2.8 indicate that, in the context of these specific simplified models, gluinos with masses up to around 1 TeV and light-flavour squarks around 800 GeV are excluded in case of LSP masses up to around 100 GeV.

However, interpretations in simplified models typically target only well-defined isolated SUSY topologies and thus do 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 [79]. 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 [80]. 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  $\sqrt{s} = 7$  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  $\sqrt{s} = 8$  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  $\sqrt{s} = 8$  TeV data.



## 3 Experimental Setup

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

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

### 3.1 The Large Hadron Collider

The LHC [81, 82] is a ring accelerator designed to provide particle collisions of hadrons. It is built in the tunnel of the former LEP [83] 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.<sup>1</sup>

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  $1.15 \times 10^{11}$  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 ( $pp$ ) 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  $\sigma$  and the integral  $L = \int \mathcal{L} dt$  of the instantaneous luminosity  $\mathcal{L}$  over time, such that

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

---

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

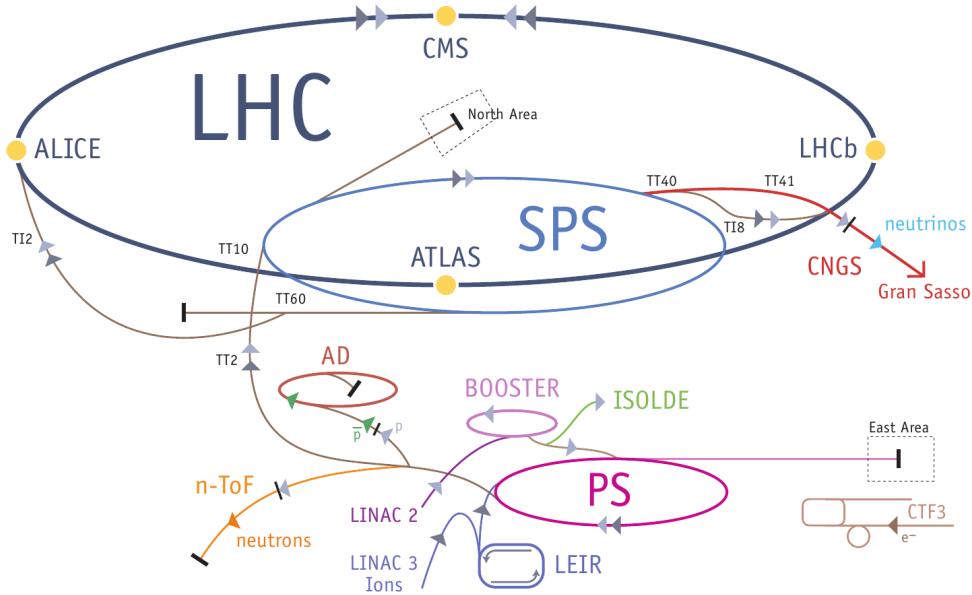


Figure 3.1: Illustration of the CERN accelerator complex including the injector chain of the LHC ring [84].

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

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

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

The four main experiments are located at four of eight locations along the LHC ring where the beams cross and can be brought to collision. The two high luminosity experiments ATLAS [86] and CMS [87, 88] 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 [89], 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 [90] detector with the main emphasis on the physics of strongly interacting matter at extreme energy densities, like quark-gluon plasma.

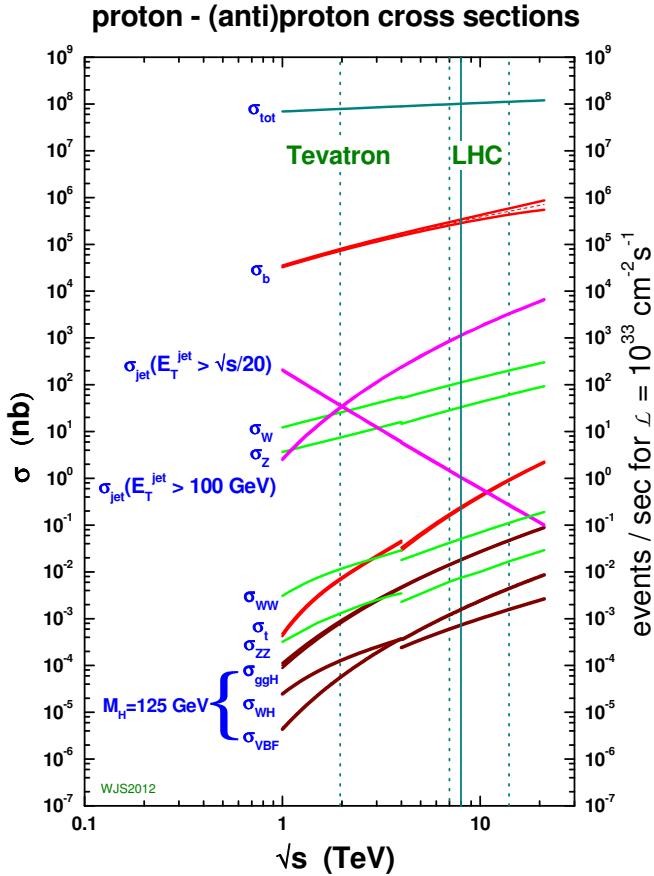


Figure 3.2: Summary of cross sections for various standard model processes in proton-antiproton ( $\sqrt{s} < 4$  TeV) and proton-proton ( $\sqrt{s} > 4$  TeV) collisions as a function of the centre of mass energy. The right axis displays the corresponding event rate at a luminosity of  $10^{33} \text{ cm}^{-2}\text{s}^{-1}$  [85].

## 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 electroweak symmetry breaking and searches for so far unknown effects are the primary targets of these experiments. These ambitious goals can only be achieved by fully exploiting the provided collision energy and luminosity with a suitable detector concept.

The CMS detector with its typical cylindrical design of different sub-detector components around the beam line is designed to meet these requirements. A sketch of the CMS detector and the different sub-detectors is shown in Fig. 3.3. Such as other high-energy particle experiments, the CMS detector makes use of tracking detectors and calorimeters to measure momenta of particles, energy depositions and flight directions in order to identify the objects emerging from the particle collisions.

A typical characteristic of calorimeters is given by the ratio  $e/h$ . This quantifies the

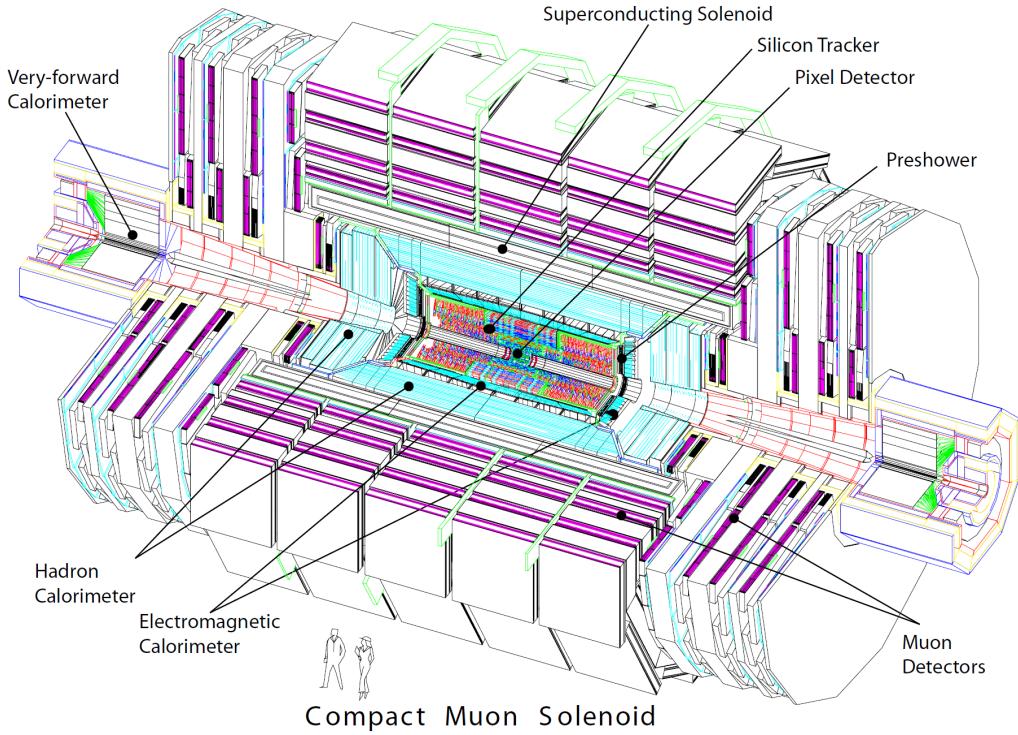


Figure 3.3: A perspective view of the CMS detector [87].

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  $\pi/e$  of the same energy

$$\frac{\pi}{e} = \frac{f_{\text{em}} e + (1 - f_{\text{em}})h}{e} = \frac{1 + (e/h - 1)f_{\text{em}}}{e/h} \quad (3.3)$$

with the *electromagnetic fraction*  $f_{\text{em}}$ , i. e. the fraction of the hadronic shower transferred into an electromagnetic component via the decay of neutral pions into two photons initiating an electromagnetic shower. Since  $f_{\text{em}}$  depends on the energy, the same holds for  $\pi/e$  and the calorimeter response is said to be *non-linear*.

Furthermore, the calorimeter performance can be characterized by the relative calorimeter energy-resolution

$$\frac{\sigma(E)}{E} = \frac{N}{E} \oplus \frac{S}{\sqrt{E}} \oplus C \quad (3.4)$$

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*  $C$ .

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 [87, 88]. After the introduction of the CMS coordinate conventions in Sec. 3.2.1, the superconducting magnet is discussed (Sec. 3.2.2). Afterwards, the tracking system (Sec. 3.2.3), the calorimeters (Sec. 3.2.4 and 3.2.5) and the muon system (Sec. 3.2.6) are described moving from the inside of the detector outwards. The detector description ends with a short review of the trigger system in Sec. 3.2.7.

### 3.2.1 Coordinate Conventions and Kinematic Variables

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

$$\eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right] \quad (3.5)$$

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

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (3.6)$$

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 unknown as well. Thus, conservation of the total momentum can not be utilized directly to describe the momentum balance in the final state. However, it is known that the initial particles have no significant momentum orthogonal to the beam axis which is referred to as *transverse momentum*. Consequently, the transverse momentum of a particle is defined as

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (3.7)$$

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. Hence, the momentum imbalance in the transverse plane ( $\vec{E}_T$ ) is determined as the negative vector sum of the momenta of all  $N$  particles in the event

$$\vec{E}_T = - \sum_{i=1}^N \vec{p}_{T,i}. \quad (3.8)$$

The absolute value of the vector momentum imbalance  $E_T$  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 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<sup>2</sup> 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  $\mu\text{m}^2$  providing similar track resolution quality in  $r-\phi$  and  $z$  direction. This configuration provides for almost the whole range up to  $|\eta| = 2.5$  three precise hits which is especially important for the reconstruction of secondary vertices.

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

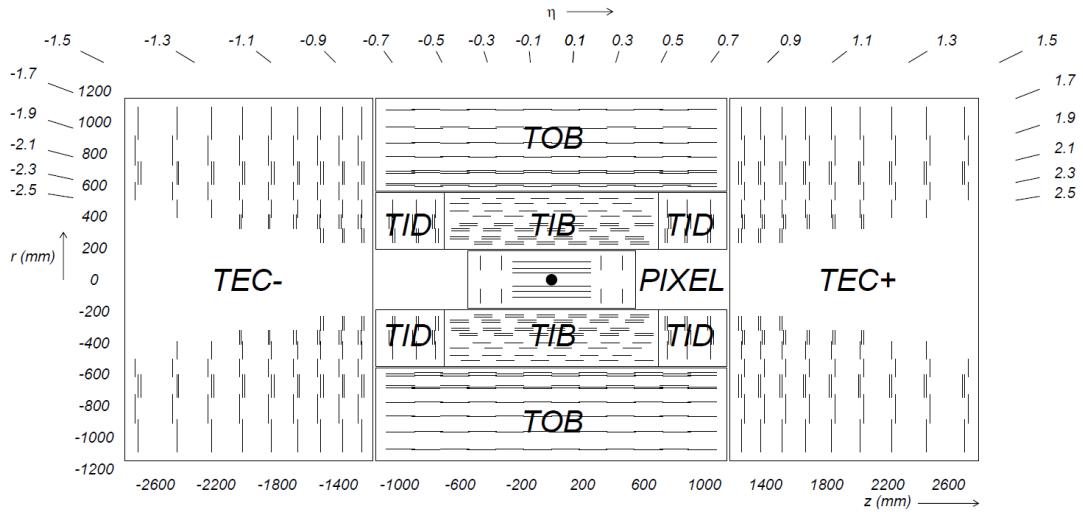


Figure 3.4: Sketch of the CMS tracking system in a  $rz$ -view. Each tracker module is represented by one line [87].

The tracking system with the design described above provides a precise impact parameter resolution and high tracking efficiency [91]. The impact parameter resolution is of the order of  $\lesssim 35 \mu\text{m}$  in the plane perpendicular to the beam (for particles with  $p_T > 10 \text{ GeV}$ ) and reaches  $75 \mu\text{m}$  in the longitudinal direction. Furthermore, the track reconstruction efficiency of high energetic electrons is above 90%, that of charged hadrons up to 95% (for  $p_T > 10 \text{ GeV}$ ) and that for muons even better than 98% in the whole covered region up to  $|\eta| = 2.5$ . This is achieved already for muons with very low transverse momenta around 1 GeV. Altogether, the relative transverse momentum resolution reaches a level of 1–2% for high momentum tracks ( $\approx 100 \text{ GeV}$ ) in the barrel for  $|\eta| < 1.6$ .

### 3.2.4 Electromagnetic Calorimeter

The CMS experiment makes use of a homogeneous electromagnetic calorimeter (ECAL), in order to precisely measure the energy deposits of electrons and photons. It is installed around the inner tracking system covering a range up to  $|\eta| = 3.0$  and consists of lead tungstate ( $\text{PbWO}_4$ ) crystals. These have been chosen as they provide a high density, short radiation length<sup>2</sup>  $X_0$  and a small Molière radius<sup>3</sup>  $R_m$  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:

**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

<sup>2</sup>The radiation length corresponds to the mean distance after which an electron traversing a material has lost all but  $1/e$  of its energy.

<sup>3</sup>The Molière radius corresponds to the radius of a cylinder containing on average 90% of the energy deposition of an electromagnetic shower.

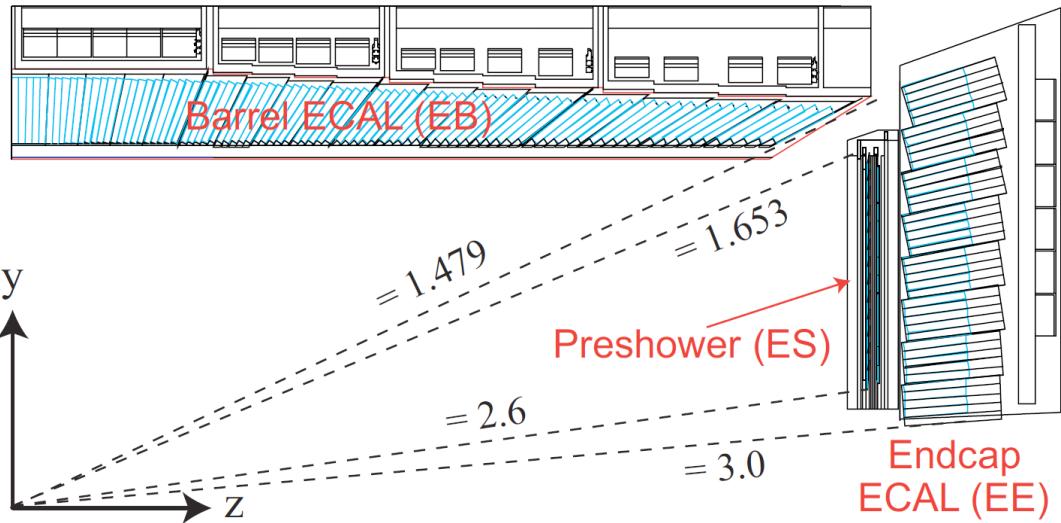


Figure 3.5: View of one quarter section of the CMS electromagnetic calorimeter in a  $y$ - $z$ -view [88].

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  $(\eta, \phi)$  is  $(0.0174, 0.0174)$ . Avalanche photodiodes are used to detect the emitted scintillation light.

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

**Preshower (ES):** In front of the endcap crystals, a preshower detector is placed. It covers the pseudorapidity range of  $1.653 < |\eta| < 2.6$  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  $H \rightarrow \gamma\gamma$  decay channel.

The performance of the ECAL has already been evaluated based on test-beam results [92, 93]. The ratio  $e/h$  has been found to be 1.6 while the relative resolution of electrons with energy  $E$  is determined to be

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

Thus, the typical relative energy resolution for electrons with a transverse momentum of 120 GeV with this calorimeter configuration is of the order of 0.5%.

In addition to the calibration of the absolute energy scale, especially channel-to-channel effects referred to as *intercalibration* have to be accounted for. This intercalibration is

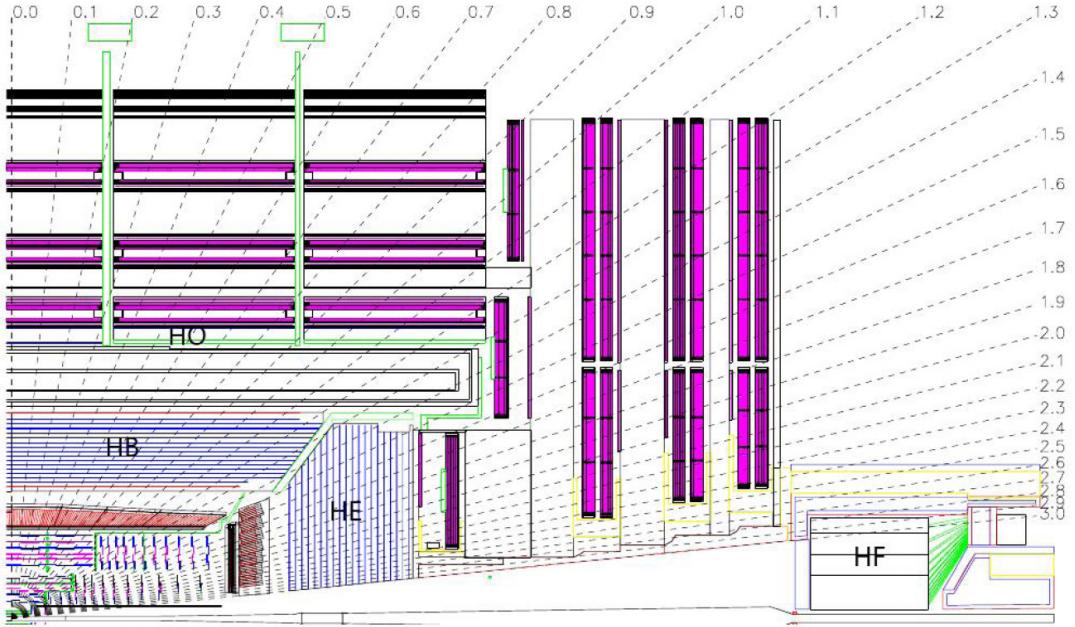


Figure 3.6: Longitudinal view of one quarter of the CMS detector showing the location of the individual HCAL sub-detector parts [87].

performed based on  $\pi^0 \rightarrow \gamma\gamma$ ,  $W \rightarrow e\nu$  and  $Z \rightarrow ee$  events and results in a crystal-intercalibration accuracy of 0.6% [94]. 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.

### 3.2.5 Hadron Calorimeter

The calorimetry of the CMS experiment is completed by the hadron calorimeter (HCAL). It is designed to provide an accurate energy measurement of hadron jets and indirectly also of invisible particles, e. g. neutrinos, by the determination of missing transverse energy. In order to measure the missing transverse energy, it is important that the calorimeter is hermetic, meaning that it provides a large geometric coverage to measure all particles emerging from an interaction. Thus, the HCAL is 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 wedges. These wedges hold the absorber plates which are flat brass plates arranged parallel to the axis of the beam. For reasons of stability, the first and last layers are made of stainless steel. The total thickness of the absorber material ranges from  $5.82 \lambda_I$  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  $\eta$ -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 [92, 93]. 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

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

which corresponds to a relative energy resolution of roughly 14% for pions with an incident energy of  $E = 100 \text{ GeV}$ .

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 continuous update of this calibrations is performed using isolated energetic particles, e. g. from decays of  $W$  or  $Z$  bosons.

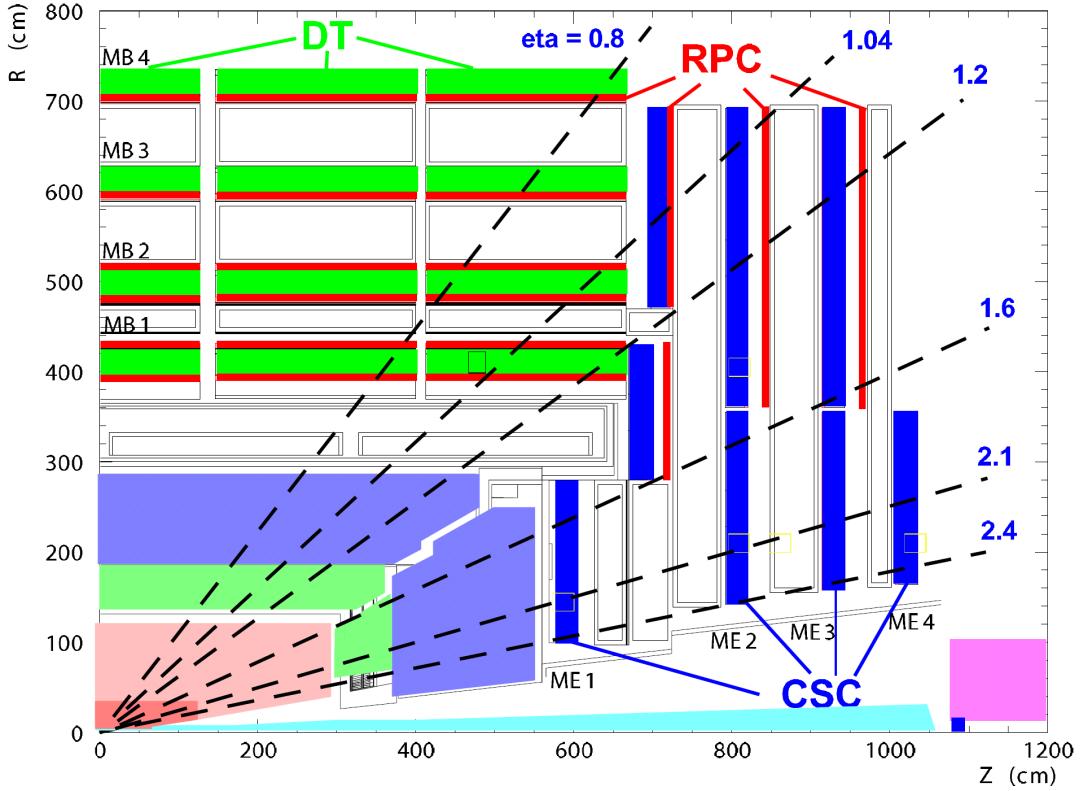


Figure 3.7: View of one quarter section of the muon system in the  $rz$ -plane [88].

### 3.2.6 Muon System

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

Three different types of gaseous detectors are used to achieve a good muon momentum resolution given the different radiation conditions and variations in the homogeneity 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  $100\text{ }\mu\text{m}$ .

**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  $75\text{--}150\text{ }\mu\text{m}$ .

**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  $\approx 200\text{ GeV}$  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 beneficial due to 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%.

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

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

In the two trigger stages, various different conditions can be tested in parallel which define the so-called *trigger paths*. For instance, there exist dedicated trigger paths to select events with single electrons or muons above a certain  $p_T$  threshold or paths to collect events with a specific amount of missing transverse energy. All of these individual trigger paths form together the *trigger menu* and are operated with a dedicated rate such that the total manageable output rate 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  $n^{\text{th}}$  triggered event is kept. The respective value of  $n$  is denoted *prescale factor*.

### 3.3 LHC Operation and Data Taking

The first operation of the LHC took place in September 2008. After a major cooling incident only a few days later requiring a longer technical stop, beams were circulated again in November 2009. The first collisions at a center of mass energy of  $7\,\text{TeV}$  finally happened end of March 2010 [96]. In the following running period in 2010, data corresponding 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  $8\,\text{TeV}$  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

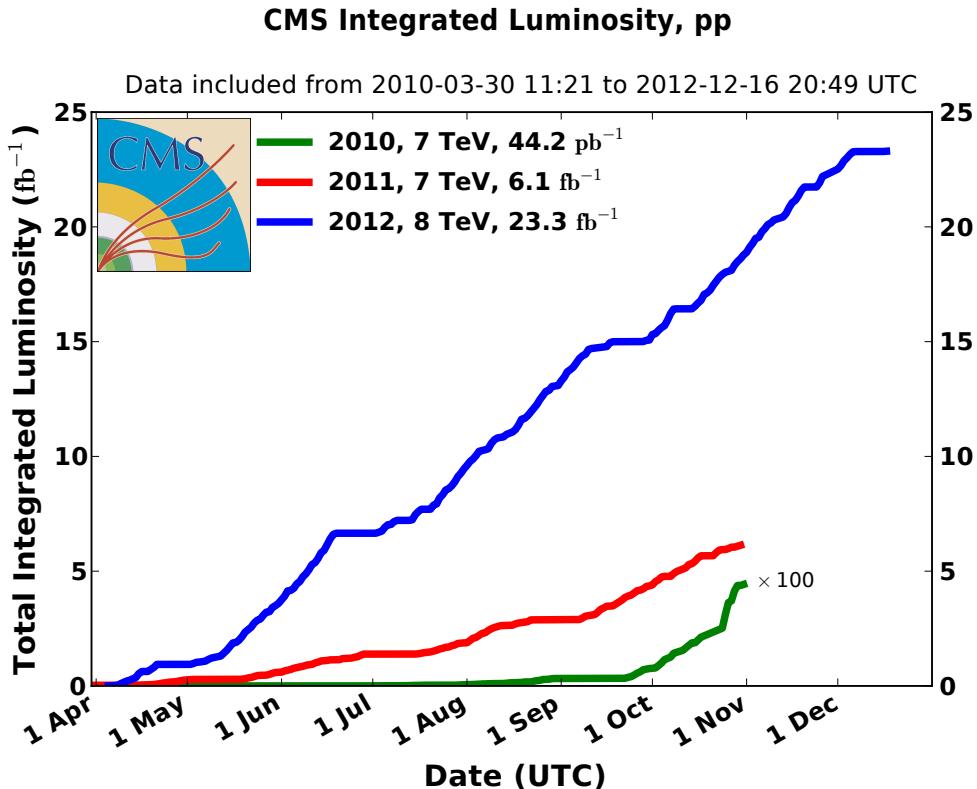


Figure 3.8: Cumulative integrated luminosity versus day delivered to CMS during stable beams for  $pp$  collisions. Different data-taking periods are indicated as follows: green for 2010, red for 2011 and blue for 2012 [95].

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  $1.7 \times 10^{11}$  exceeding even the design value [95, 97].

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 occurring within the same bunch-crossing as the primary hard collision and leads to additional tracks in the tracking system and energy deposits in the calorimeters, OOT PU arises from  $pp$  collisions in previous and following bunch crossings and contributes further energy deposits in the calorimeters to the hard interaction due to the finite signal decay time in the calorimeters. An overview of the pileup profile in collision data taken in 2012 is given in Fig. 3.9. Here, the mean number of interactions per bunch crossing is illustrated with the mean of this distribution located at 21. The maximum number of mean pileup events even reached a value of  $\approx 40$ . Dedicated techniques in order to mitigate effects from pileup have been developed of which some are discussed in Chap. 5.

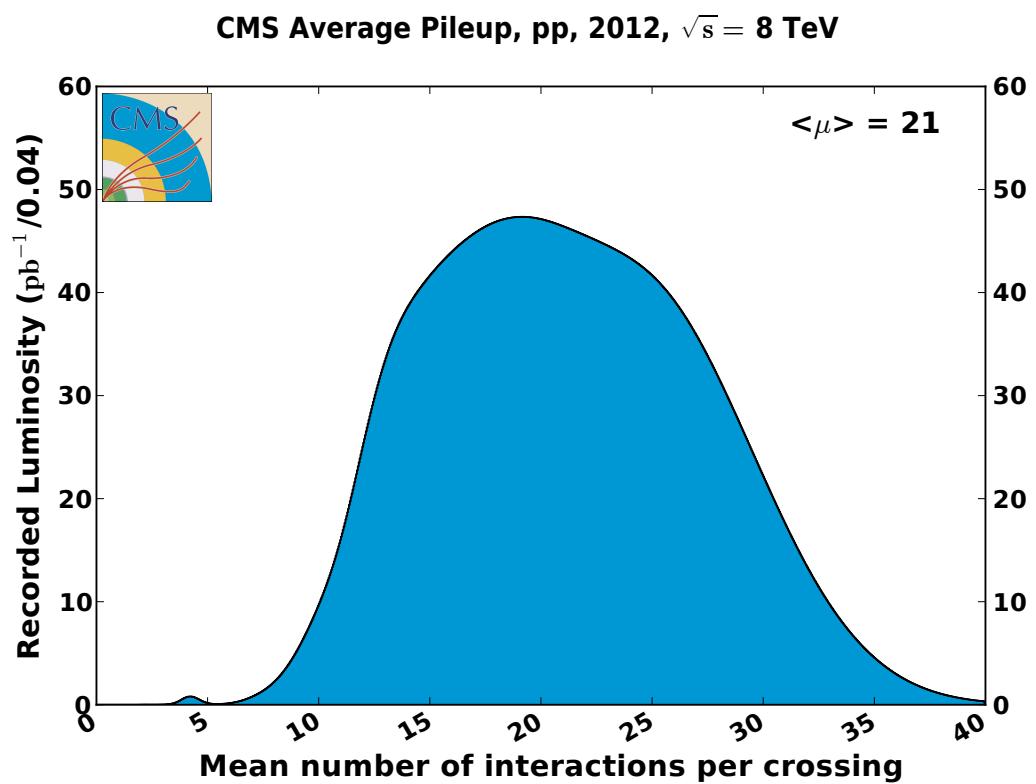


Figure 3.9: Mean number of interactions per bunch crossing in  $pp$  collisions at a centre of mass energy of  $\sqrt{s} = 8 \text{ TeV}$  [95].



## 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 e. g. selection efficiencies.

This section provides a brief introduction to the principles of event simulation in hadron collisions and introduces some event generators including different approaches for the simulation of the 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 [98]. 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 [99,100]. 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-inelastic scattering at the HERA collider [102]. 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* [103,104]. 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

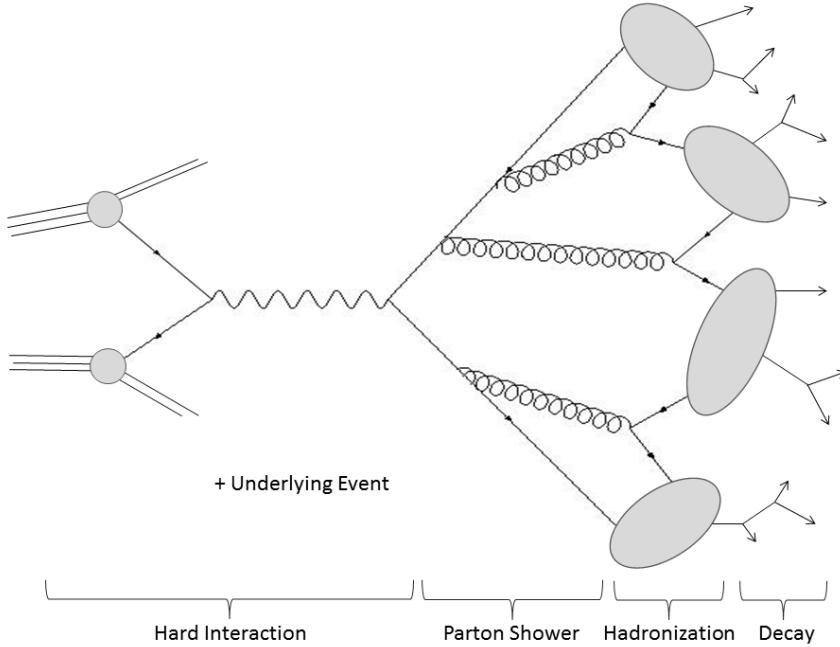


Figure 4.1: Sketch of individual steps in the generation process of simulated events.  
Adapted from [101].

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

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

Here,  $f(x, \mu_F)$  are the PDFs of the interacting protons which depend on the factorization scale  $\mu_F$  while  $\sigma_{ab \rightarrow n}$  indicates the parton-level cross section for the production of a final state  $n$  from partons  $a$  and  $b$ . This parton-level cross section depends on the final-state phase space, the factorization scale and the renormalisation scale  $\mu_R$  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 [105]) and *cluster models* (cf. for instance [106, 107]) in order to characterize the hadronization. While the first model describes directly the transition from the parton to the hadron based on the assumption of linear confinement, i. e. a linearly rising potential, the second model introduces an intermediate step of cluster objects with mass scales around a few GeV. Particles arising from the hadronization are denoted with *generator-level* particles and labelled 'gen' within this thesis.

**Decay:** After the hadronization process, a couple of unstable hadrons are present in the event whose decay into stable particles must be modelled. Stable in this context means that they do not decay further within the typical collider timescales ( $c\tau > 10 \text{ mm}$ ). 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, these choices are based on experimental results in combination with theoretical assumptions. Typical differences occur for instance in the consideration of excited mesons or heavy baryon multiplets as well as the treatment of matrix elements and spin correlations during the decay.

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

## 4.2 Event Generators

The various sub-processes described in Sec. 4.2 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** [108]: PYTHIA is a general purpose event generator which has been used already extensively at previous collider experiments. It is designed to simulate collisions of either hadrons or same-generation leptons. This means that it is well suited to model  $pp$  collisions. The highest multiplicity of particles involved in the hard interaction that can be simulated with PYTHIA are  $2 \rightarrow 3$  processes. Nonetheless, more than three final state particles can arise for instance from the parton shower. This is traditionally  $Q^2$  ordered. The hadronization model in PYTHIA is based on

the aforementioned string model. Besides the simulation of whole physics events including all relevant steps, PYTHIA can also be interfaced to other generator programs in order to carry out the parton shower, hadronization and decay steps only.

**HERWIG++** [109]: 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** [110, 111]: MADGRAPH is a general-purpose matrix-element program and designed to provide accurate descriptions of multiparton processes. It is the only generator that is capable of computing any arbitrary process at tree level or at NLO. In principle, it is possible to generate any  $2 \rightarrow n$  process. However, in order to achieve reasonable computing times, the number of partons is typically limited to values  $< 10$ . MADGRAPH can commonly be interfaced to PYTHIA for the realisation of the showering, hadronization and decay.

**POWHEG** [112]: The POWHEG programme provides accurate QCD computations based on matrix elements up to next-to-leading order following the concept proposed in [113]. 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 [114] 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* [115]. Especially in analyses in which many signal samples are needed to scan a wide parameter range, e.g. in SUSY analyses, the usage of the fast simulation is very beneficial. A comparison of several relevant kinematic distributions in real collision events to simulated events obtained from fast simulation shows a quite good agreement [116]. 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 for the reconstruction of collision events and is designed to identify stable particles in an event. Types of particles which are identified by the PF algorithm are electrons and photons, charged and neutral hadrons as well as muons. In order to reconstruct the four-momenta of these particles all sub-components of the detector are utilized.

The CMS detector is very well suited for this task. The silicon tracker enclosed by the uniform magnetic field enables a very efficient track reconstruction yielding only a small track fake rate even down to small transverse momenta of 150 MeV, 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. A detailed introduction to the PF algorithm can be found in [117].

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 [118]. 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 Section 5.2, missing transverse energy  $\cancel{E}_T$  or decay products of tau leptons. More detailed information on the specific quality criteria required for the identification of certain particles is given in Chap. 6, 7 and 8 for each analysis presented in this thesis.

## 5.2 Reconstruction of Jets

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

Different jet algorithms are introduced in Sec. 5.2.1 followed by a discussion of different jet types used at CMS in Sec. 5.2.2. The jet transverse-momentum response is defined in Sec. 5.2.3 and the calibration of jet energies is reviewed in Sec. 5.2.4.

### 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*)

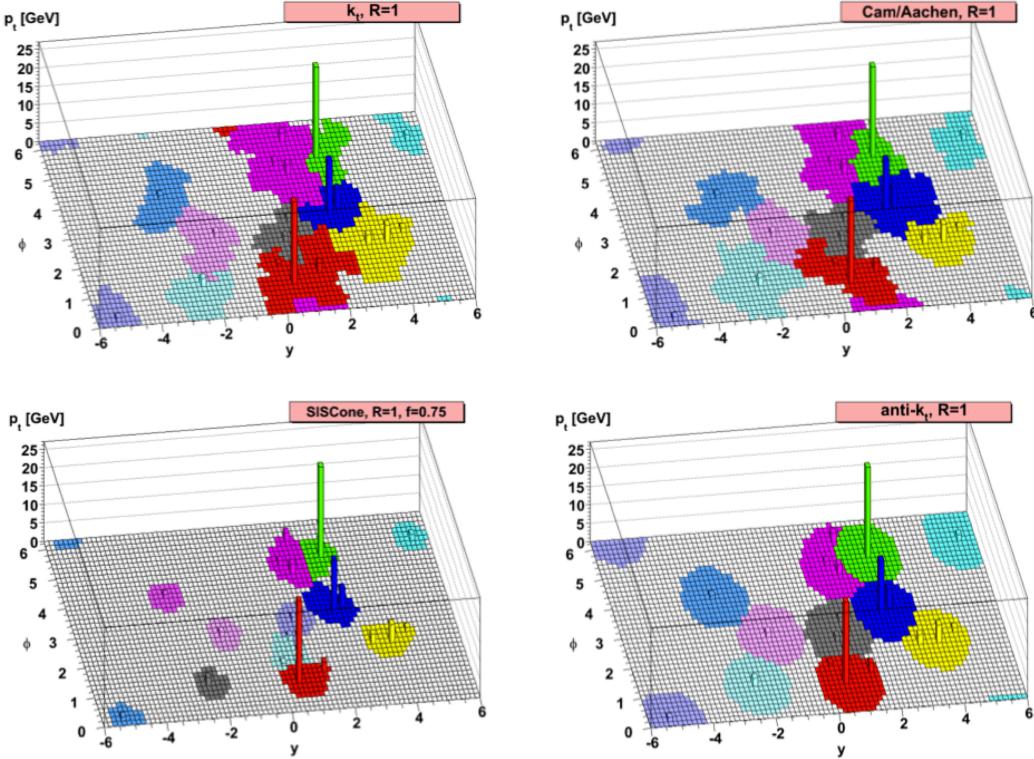


Figure 5.1: Sample parton-level event generated with HERWIG++, adding many random soft particles, that is clustered with the  $k_T$ -algorithm (*top left*), the C/A-algorithm (*top right*), the SIScone-algorithm (*bottom left*) and the anti- $k_T$ -algorithm (*bottom right*). Colours illustrate all particles clustered to the resulting hard jets. Taken from [119].

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.

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 [119]. Technical implementations of various jet algorithms are provided by the *FastJet* package [120, 121].

**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

$$\Delta R_{i,j}^2 < R^2 \quad (5.1)$$

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

However, cone algorithms are IRC safe when instead of iteratively forming stable cones all stable cone solutions are identified at once. This procedure is denoted *seedless cone* (SC) algorithm. The usage of such algorithms though is typically impractical as the computation time increases exponentially with the number of particles to be considered so that even for 100 particles it is not solveable at any reasonable timescale. A feasible implementation of a seedless cone algorithm featuring an  $\mathcal{O}(N \log(N))$  time-dependence is given by the SIScone algorithm [122]. 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

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

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

with the distance  $d_{ij}$  between final state objects  $i$  and  $j$  carrying transverse momentum  $k_T$  and the distance of the object to the beam  $d_{iB}$ . While  $\Delta R_{ij}$  denotes the spatial separation in the  $(\eta, \phi)$ -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  $p = 1$  [123] 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-* or short *CA-algorithm* [124, 125]

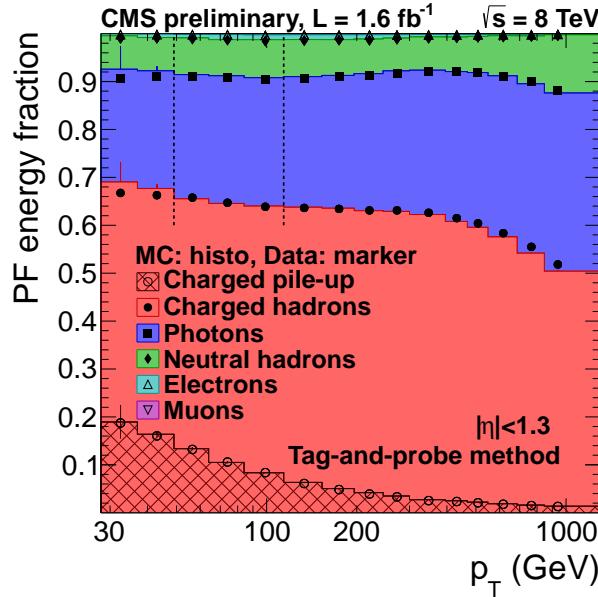


Figure 5.2: Composition of the PF jet energy versus jet  $p_T$  in the barrel detector region  $|\eta| < 1.3$  in simulated events (coloured histograms) and data (solid markers) [128].

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 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  $p = -1$  [126] 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 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 [127]:

**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  $5 \times 5$

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 [129]. 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  $(\eta, \phi)$ -plane. Associated tracks are projected to the jet cone and are exploited to correct the jet energy and direction.

**Particle-Flow (PF) jets:** PF jets are clustered from the four-momentum vectors of particle-flow candidates as identified by the PF algorithm described in Section 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 pile-up, charged hadrons unambiguously associated to vertices other than the primary vertex can be removed from the jet algorithm input list before the actual jet clustering is performed. This technique is referred to as *charged hadron subtraction* (CHS) and used as default throughout this thesis.

### 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

$$\mathcal{R} = \frac{p_T}{p_T^{\text{particle}}} \quad (5.2)$$

where  $p_T$  denotes the transverse momentum of the jet measured at detector level and  $p_T^{\text{particle}}$  is the transverse momentum of the original particle-level jet. Thus, the jet response provides a measure of the jet momentum visible in the detector compared to the actual momentum of the particle after hadronisation and decay. The jet response usually depends on the jet momentum as well as on the pseudorapidity. This is expected since the precision of the jet measurement is directly related to the energy of the particles and the resolution of the detector sub-components. This is for instance caused by the fluctuating electromagnetic fraction  $f_{\text{em}}$  from shower to shower, 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).

### 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

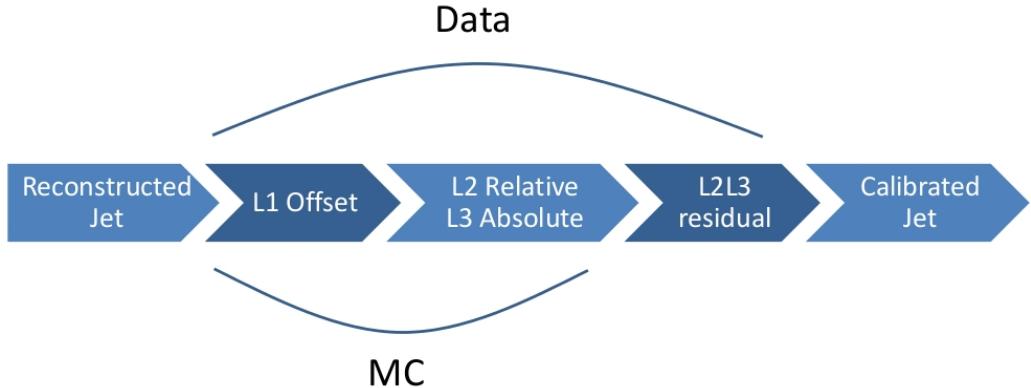


Figure 5.3: Sketch of the factorized approach used for jet energy corrections at the CMS experiment.

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 [127] and illustrated in Fig. 5.3. The actual set of correction factors used in this thesis, if not stated otherwise, is documented in [130].

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

$$p^{\text{cor}} = C \cdot p^{\text{raw}} = C_{\text{offset}}(p_T^{\text{raw}}, \eta) \cdot C_{\text{rel}}(\eta) \cdot C_{\text{abs}}(p'_T) \cdot C_{\text{res}}(p''_T, \eta) \cdot p^{\text{raw}} \quad (5.3)$$

where  $C$  is composed of the offset correction  $C_{\text{offset}}$ , the calibration factors  $C_{\text{rel}}$  and  $C_{\text{abs}}$  as well as residual correction factors  $C_{\text{res}}$ . While  $C_{\text{offset}}$ ,  $C_{\text{rel}}$  and  $C_{\text{abs}}$  are applied to both data and simulation, the residual correction factors  $C_{\text{res}}$  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 pile-up events. The  $p_T$  offset is estimated in dependence of  $\eta$ , the effective jet area  $A_j$  and the  $p_T$ -density  $\rho$  (*hybrid jet area method* [131]).

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  $\rho$  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.

**L2 Relative + L3 Absolute Correction:** The L2 relative correction is designed to make the jet energy scale uniform with respect to  $\eta$  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

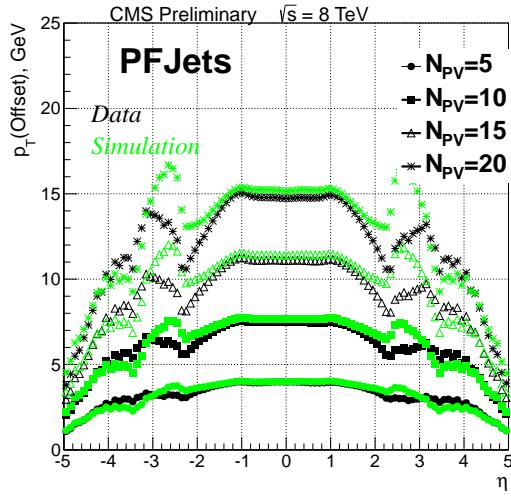


Figure 5.4: L1 offset transverse momentum correction for AK5 PF jets as a function of jet pseudorapidity in data (black) and simulation (green). Different intervals of reconstructed primary vertices ( $N_{PV}$ ) are shown with different markers [130].

average response  $1/\langle R \rangle$  at fixed  $p_T^{\text{gen}}$  and illustrated in Fig. 5.5.

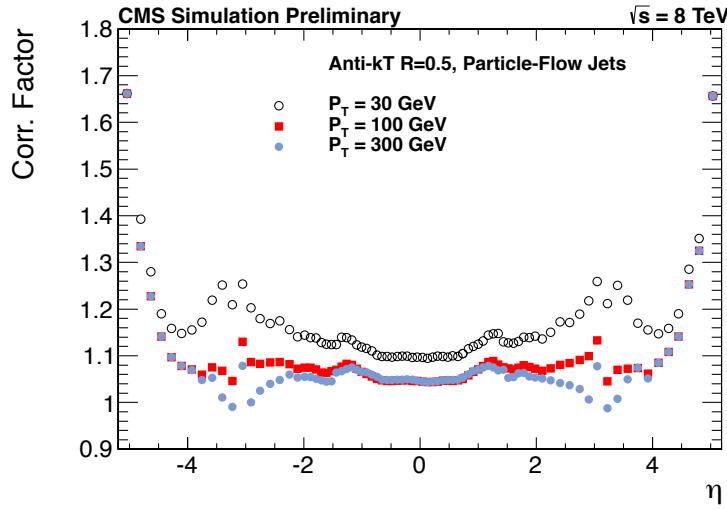


Figure 5.5: MC-Truth corrections for AK5 PF jets as a function of jet pseudorapidity for three reference transverse momentum values: 30 GeV (white hollow circles), 100 GeV (red squares) and 300 GeV (blue circles) [130].

**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

L2 residual determination) or  $Z/\gamma + \text{jet}$  events (used for the measurement of the L3 residual correction). The L2 residual correction is illustrated in Fig. 5.6.

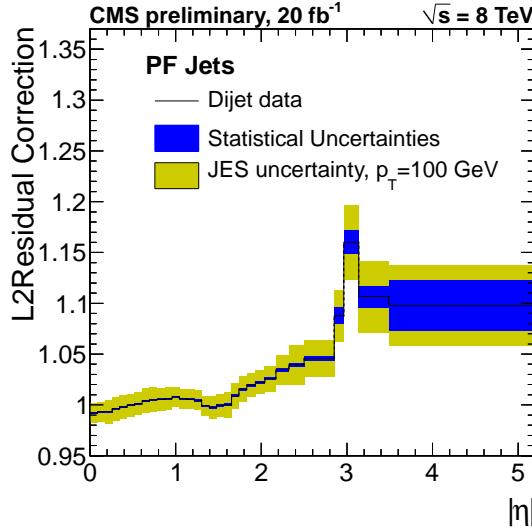


Figure 5.6: L2 residual corrections for AK5 PF jets as a function of jet pseudorapidity, obtained from dijet events, shown with JES uncertainty (yellow band) and the statistical uncertainty (blue band) [130].

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

### 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 these specific  $b$ -jet properties in dedicated  $b$ -tagging algorithms for an efficient  $b$ -jet identification [132].

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:

**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 significances for the tracks in the jet. These variables are used to compute two likelihood ratios which can be used to distinguish either c and b jets or light-parton and b jets.

In order to determine the quality of a particular b-tagging algorithm, typically the misidentification probability as a function of the b-jet efficiency is compared for various taggers. Such a performance comparison is illustrated in Fig 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  $p_T > 60$  GeV. For loose working points, the b-jet efficiency is around

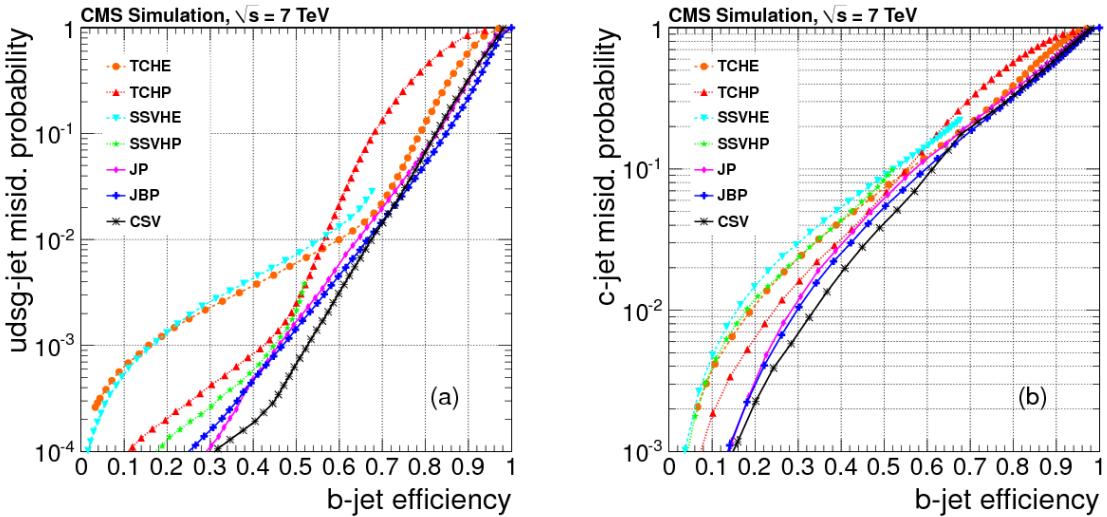


Figure 5.7: Performance curves obtained from simulation for the algorithms described in the text. (a) light-parton- and (b) c-jet misidentification probabilities as a function of the b-jet efficiency. Taken from [132].

$\approx 80 - 85\%$ , 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  $\sqrt{s} = 8$  TeV in 2012 were the TCHP, JP and CSV algorithms [133].

## 5.4 Identification of Boosted $t$ -Quark Jets

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

As discussed in Sec. 2.3.2, the decay of the top-quark is experimentally characterized by the decay of the W boson which makes two thirds of top quark decays resulting exclusively in hadrons. This decay mode is here referred to as *hadronic top*. If top quarks are produced with  $p_T \ll m_t$ , the top quark decay products show up as 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 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  $\Delta R$  is expected to scale as

$$\Delta R \approx 2 m_t / p_T \quad (5.4)$$

where  $m_t$  and  $p_T$  are the mass and the transverse momentum of the decaying particle, respectively. Schematic diagrams of resolved and boosted top quark decays are illustrated

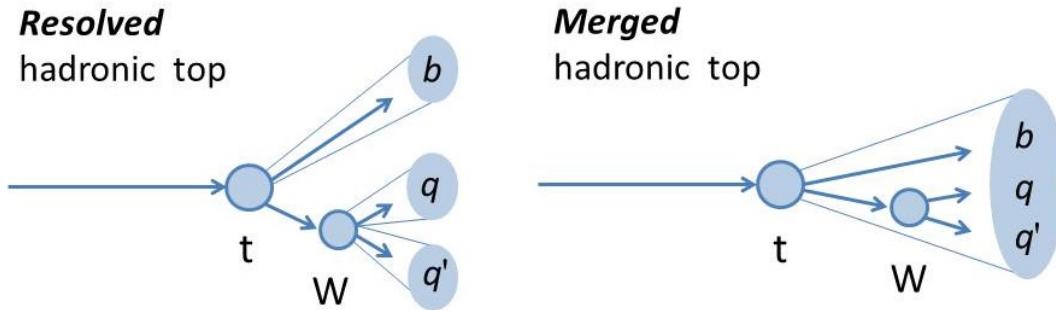


Figure 5.8: Schematic diagrams of the decay of a top-quark in the resolved case (*left*) and the boosted scenario (*right*).

in Fig. 5.8.

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

$$t \rightarrow W + b \rightarrow qq' + b \quad (5.5)$$

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

**Filtering [134]:** Jet constituents are recombined by rerunning the Cambridge-Aachen algorithm with a smaller distance parameter, e.g.  $R = 0.3$ . With this procedure, subjets within the fat jet are identified: The filtering is performed by keeping only  $n$  of the identified subjets. Thus, the dominant contribution from the hard process is captured while contamination from the underlying event or pileup is removed.

**Pruning [135]:** In contrast to filtering, pruning does not attempt to find a particular number of subjets. Rather, the idea is to systematically remove soft and large angle recombinations in the jet clustering process. For the recombination of objects  $i$  and  $j$ , two conditions are defined:

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

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

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

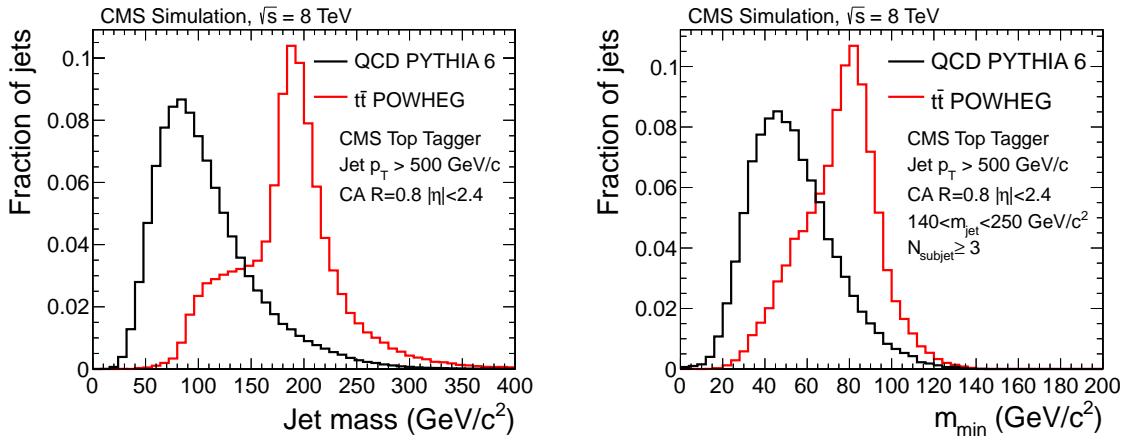


Figure 5.9: Jet mass (*left*) and minimum pairwise subjet mass (*right*) for CA8 jets with  $p_T > 500 \text{ GeV}$  from a simulated  $t\bar{t}$  Powheg sample (red) and from a simulated QCD Pythia sample (black). Taken from [137].

**Trimming [136]:** Similarly to the filtering procedure, the jet trimming is based on the concept of identifying subjets within an initially clustered large jet, here called *seed jet*. The constituents of this seed jet are reclustered using a jet clustering algorithm with a smaller distance parameter than for the seed jet. If

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

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

In order to identify decays of boosted hadronic top quarks, as discussed above, several top tagging algorithms (or short *taggers*) are commissioned within the CMS experiment [137, 138]. Two of them are used within this thesis:

**CMS Top Tagger:** The CMS Top Tagger [139] is based on the top tagger developed by Kaplan et al. [140] and acts on jets clustered by the Cambridge-Aachen algorithm with distance parameter  $R = 0.8$ . These jets, used as input for the algorithm, are denoted as *hard jets*. Since the decay products of the hadronic top are not expected to be all contained within one jet with  $R = 0.8$  for low transverse momenta of the top quark, only jets with  $p_T^{\text{jet}} > 350 \text{ GeV}$  are considered. In order to identify subjets within the hard jets, a two-stage decomposition procedure is applied which performs the pairwise clustering sequence used to form the hard jet in reverse order. First, the algorithm aims at splitting the hard jet into two subclusters (*primary decomposition*). Second, it is attempted to further split the clusters emerging from the first step (*secondary decomposition*). Typically, subjets are found when they are spatially well separated and carry a significant fraction of the momentum of the hard jet. Details on the actual splitting criteria can be found in [137]. 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  $\geq 3$
- The jet mass  $m_{\text{jet}}$ , i. e. the mass of the four-vector sum of the constituents of the hard jet, has to be close to the top-mass with  $140 < m_{\text{jet}} < 250 \text{ GeV}$ .
- The invariant mass of each pair of the three subjets highest in  $p_T$  is calculated. The minimum of the pairwise masses  $m_{\min}$  has to be  $> 50 \text{ GeV}$ .

In Fig. 5.9, the distributions for jet mass and minimum pairwise mass are illustrated for a simulated  $t\bar{t}$  (red) and a QCD multijet (black) sample at  $\sqrt{s} = 8 \text{ TeV}$ .

**HEP Top Tagger:** The HEP Top Tagger [141] uses jets with a larger distance parameter than the CMS Top Tagger of  $R = 1.5$ . This makes the HEP top-tagging algorithm especially suited for top quark decays with moderate boost and thus, uses fat jets with a transverse momentum greater than  $200 \text{ GeV}$  as input. Since a larger jet size is in principle more prone to disturbing effects from the underlying event or pile-up, a sophisticated decomposition procedure is applied to distinguish hard subjets from soft components. Similar to the decomposition done for the CMS tagger, the identification of subjets is based on going through the cluster history of the jet in reversed order. First, the jet is decomposed into subclusters applying a *mass drop condition* discarding too soft components. Afterwards, subclusters resulting from the mass drop decomposition are reclustered into subjets and filtering criteria are applied. Details on the mass drop decomposition and reclustering criteria can be found in [137]. Finally, the combination with the mass determined from the filtered subjets closest to the top mass is kept and reclustered to force three subjets. Kinematic selections are applied to these three final subjets in order to identify 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 mass, the jet has to satisfy at least one of the following conditions based on the subjet pairwise masses

$$0.2 < \text{atan} \frac{m_{13}}{m_{12}} < 1.3 \text{ and } R_{\min} < \frac{m_{23}}{m_{123}} < R_{\max} \quad (5.6)$$

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

$$R_{\min}^2 \left( 1 + \left( \frac{m_{12}}{m_{13}} \right)^2 \right) < 1 - \left( \frac{m_{23}}{m_{123}} \right)^2 < R_{\max}^2 \left( 1 + \left( \frac{m_{12}}{m_{13}} \right)^2 \right) \text{ and } \frac{m_{23}}{m_{123}} > 0.35 \quad (5.8)$$

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

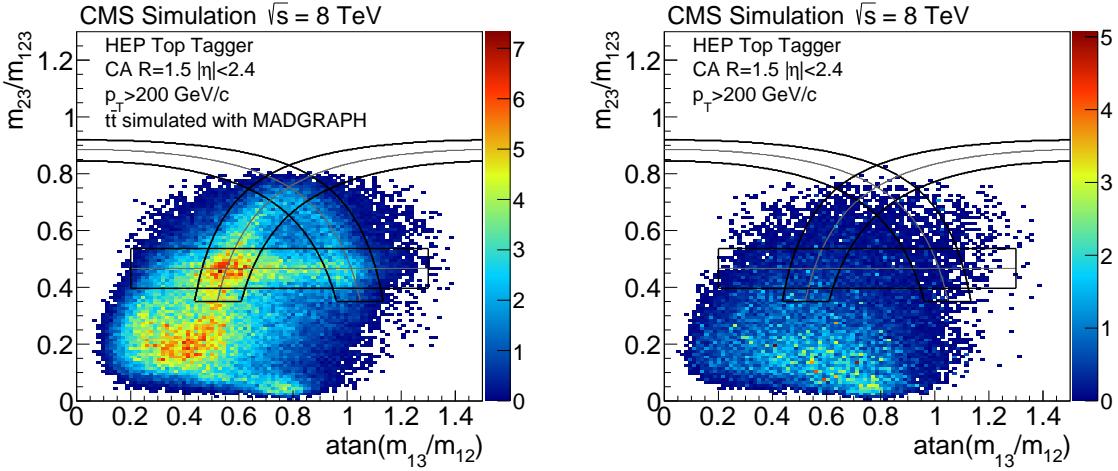


Figure 5.10: Two-dimensional distributions of  $m_{23}/m_{123}$  versus  $\text{atan}(m_{13}/m_{12})$  for HEP Top Tagger subjets from CA15 jets with  $p_T > 200 \text{ GeV}$  for a simulated  $t\bar{t}$  MADGRAPH sample (*left*) and for a simulated background sample composed of cross-section weighted boson + jets, diboson, single-top,  $t\bar{t}$  all-hadronic and  $t\bar{t}$  leptonic events (*right*). The A-shaped region indicates the selected region by the HEP Top Tagger. Taken from [137].

The criteria related to the selection of the W mass are illustrated in Fig. 5.10 for signal events as well as background events and indicated by the black solid lines resulting in an A-shaped region.

In analogy to the b-tagging algorithms, different working points for each top tagger can be defined characterized by a specific top-tag efficiency and misidentification rate. Typically, the working points are chosen such that they have a minimum mistag rate for a given signal efficiency. In addition, further selection criteria can be added to the taggers as described above to achieve a potential performance improvement. These criteria are subject b-tagging, as introduced in [133], and selections on the  $n$ -*subjettiness* [142,143] ratio  $\tau_3/\tau_2$  that denotes how likely a fat jet is composed of three subjets rather than two.

The performance of various taggers is compared by determining the top-tag efficiency versus the misidentification rate. Commonly, the top-tag efficiency for simulated events is defined as the number of jets passing the top-tagging selection divided by the number of jets associated to a simulated hadronic top or anti-top passing a certain  $p_T$  selection. Similarly, the mistag rate is defined as the number of jets passing the top-tagging selection divided by the number of jets associated to a simulated quark or gluon from the hard process passing the  $p_T$  selection. In Fig. 5.11 the mistag rate versus the top-tagging efficiency is illustrated for various taggers and parton momenta. The top-tag efficiency determined from simulated  $t\bar{t}$  events for the selection criteria described above amounts to 35.3% for a matched parton- $p_T$  of  $> 200 \text{ GeV}$  for the 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 pile-up such that in high pile-up environments a performance degradation is expected. For instance, in the case of the CMS Top Tagger the efficiency stays rather stable as a function of the number

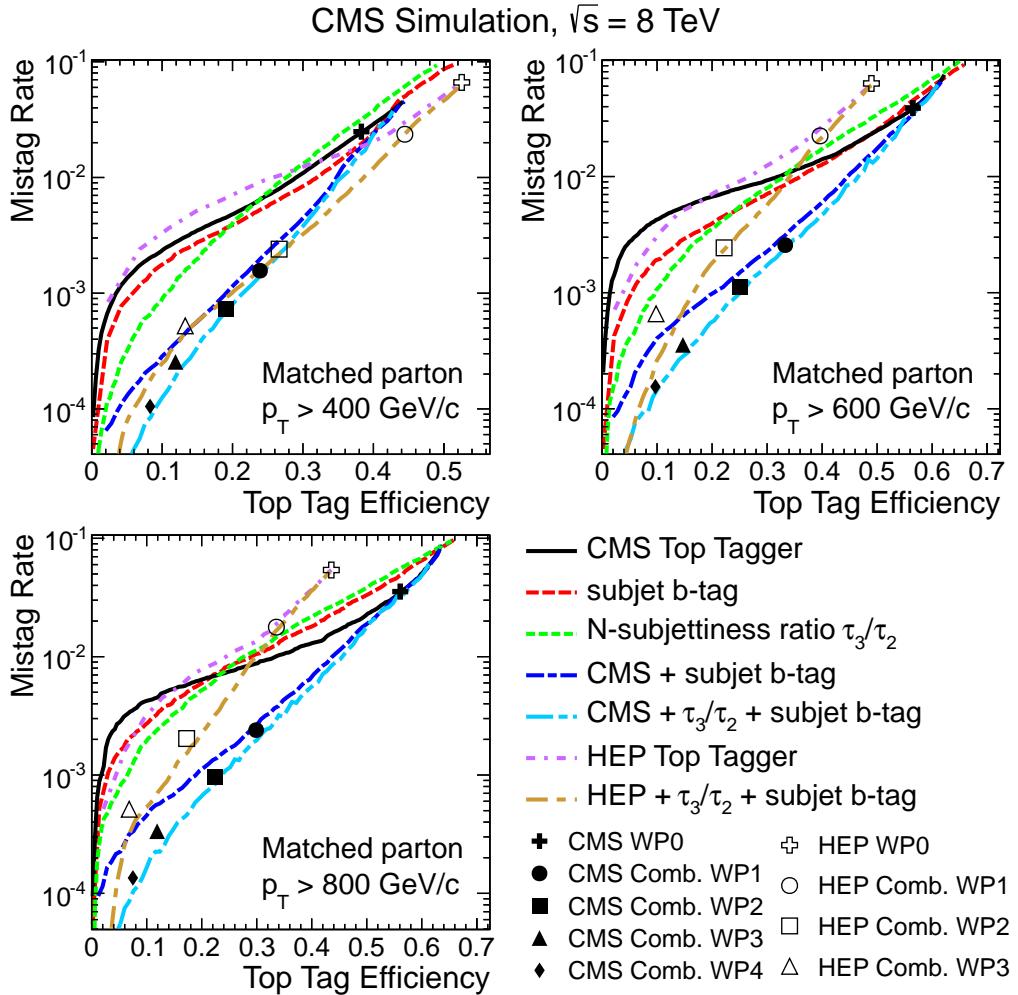


Figure 5.11: Mistag rate versus top-tagging efficiency as measured from QCD PYTHIA6 and POWHEG  $t\bar{t}$  simulated events, respectively. In the cases where a jet mass cut is applied, the cut is not varied but fixed at  $140 < m_{\text{jet}} < 250 \text{ GeV}$ . N-subjettiness is calculated using  $R = 0.8$  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 all-hadronic generated top quarks, while background jets are matched to simulated partons from the hard scatter. Distributions are shown for three  $p_T$  selections, where the  $p_T$  cut is applied to the matched generated parton [137].

of primary vertices with a slope of  $0.031\% \pm 0.034\%$  while 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 and several jets. 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  $\eta$ . A similar approach was already used in previous studies at  $\sqrt{s} = 7\text{ TeV}$  [127, 144] while a complementary approach utilizes  $\gamma + \text{jet}$  events [145–147]. The measurement shown here is based on collision data corresponding to an integrated luminosity of  $19.7\text{ fb}^{-1}$  recorded at  $\sqrt{s} = 8\text{ TeV}$  in 2012. Parts of this chapter are taken from [148], 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

$$\mathcal{R} = \frac{p_T}{p_T^{\text{gen}}} . \quad (6.1)$$

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 [149] using the CTEQ6L1 PDF set [150] 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  $p_T^{\text{gen}}$  and  $|\eta^{\text{gen}}|$ . This is done since a dependence on the momentum and the respective detector region is expected, as discussed in Sec. 5.2.3.

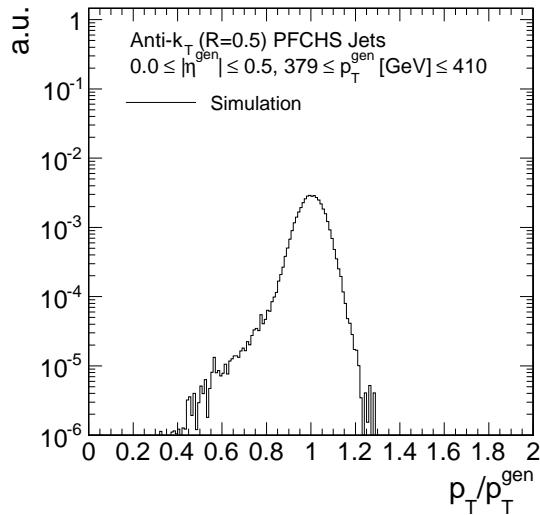


Figure 6.1: Jet response distribution for one example  $|\eta^{\text{gen}}|$  and  $p_T^{\text{gen}}$  interval derived from simulation.

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 precision 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 for instance be due to detector effects. Shower leakage can occur, if not the whole shower is deposited in the calorimeters. Due to the limited length of the calorimeters, some particles might even cross the whole apparatus (*punch-through*). Also malfunctioning detector components can lead to high or low response tails by creating artificial signals. Beyond that, also physics processes lead to response tails. For instance, the lower response tails

get populated by semi-leptonic decays of heavy-flavour quarks. These contain neutrinos that carry a certain amount of the momentum and leave the detector unnoticed resulting in an overall reduced response.

## 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. In simulated events, the particle-level jet momenta are given by the generator-level jet momenta. In data events, however, no such equivalent is present so that the jet resolution is not accessible directly. One possibility to measure the resolution of the jet transverse momenta in data as well as in simulated events is to utilize the dijet asymmetry  $A$ . For events with at least two jets it is defined as

$$A = \frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}}. \quad (6.2)$$

In this equation  $p_{T,1}$  and  $p_{T,2}$  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

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

In an ideal dijet topology, the two jets are exactly balanced at particle level. If they belong in addition to the same  $\eta$  region, then  $\langle p_{T,1} \rangle = \langle p_{T,2} \rangle = \langle p_T \rangle$  and  $\sigma(p_{T,1}) = \sigma(p_{T,2}) = \sigma(p_T)$ . This allows the simplification of Eq. 6.3 and provides the following important relation between the width of the asymmetry  $\sigma_A$  and the jet- $p_T$  resolution  $\sigma(p_T)$

$$\frac{\sigma(p_T)}{\langle p_T \rangle} = \sqrt{2} \cdot \sigma_A. \quad (6.4)$$

This relationship was already used at the Tevatron experiments [151, 152], the ATLAS experiment [153] or in previous CMS analyses [127, 144] 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  $\eta$ . 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, as e.g. the  $p_T$  of the leading jet, as well. However, the jet- $p_T$  spectrum is affected by migration effects such that momentum measures based on single jets are not preferable. Due to 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  $p_T^{\text{gen}}$  migrate into a specific interval than jets with high  $p_T^{\text{gen}}$ . 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

$$p_T^{\text{ave}} = \frac{1}{2}(p_{T,1} + p_{T,2}) . \quad (6.5)$$

Beyond that, 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  $\alpha$  which is defined as the ratio of the transverse momentum of the third jet to the average momentum according to

$$\alpha = \frac{p_{T,3}}{p_T^{\text{ave}}} . \quad (6.6)$$

The presence of jets beyond the third is neglected in the parametrization of the additional activity in the event, as these have consecutively declining momentum due to the strongly decreasing jet production cross section versus jet- $p_T$  [154]. The presence of additional jets and the thereby introduced imbalance leads to a broadening of the observed asymmetry distribution. This effect is also illustrated later in Sec. 6.4.2. In order to determine the intrinsic resolution from such events, 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 associated 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  $\sqrt{s} = 8$  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  $19.7 \text{ fb}^{-1}$  with an uncertainty of  $2.5\% \text{ (syst.)} \pm 0.5\% \text{ (stat.)}$  [155].

Multijet events are pre-selected by a set of triggers based on the average transverse momentum of the two leading jets in the event. In order to obtain a good coverage of the  $p_T^{\text{ave}}$  spectrum, different trigger paths are combined. Since they have different minimum  $p_T^{\text{ave}}$  thresholds, a broad range in  $p_T^{\text{ave}}$  is considered. In Tab. 6.1 the different trigger paths used for this analysis are listed and the particular offline  $p_T^{\text{ave}}$  value for which the respective trigger reaches 99% of the efficiency plateau is given [156]. However, the low  $p_T^{\text{ave}}$  triggers have been operated with prescale factors applied in order to keep the data rate sufficiently

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

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

low. In fact, only the trigger with the highest  $p_T^{\text{ave}}$  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 Section 6.1. Since the cross-section of the process has been scaled by  $\hat{p}_T^{4.5}$ , with the scale parameter  $\hat{p}_T$  describing the momentum transfer in the hard process, the sample is reweighted with the inverse in order to regain the physical spectrum.

#### 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 in  $\phi$  direction. 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 [157, 158]. In order to mitigate effects from pileup, only jets with  $p_T > 10 \text{ GeV}$  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 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

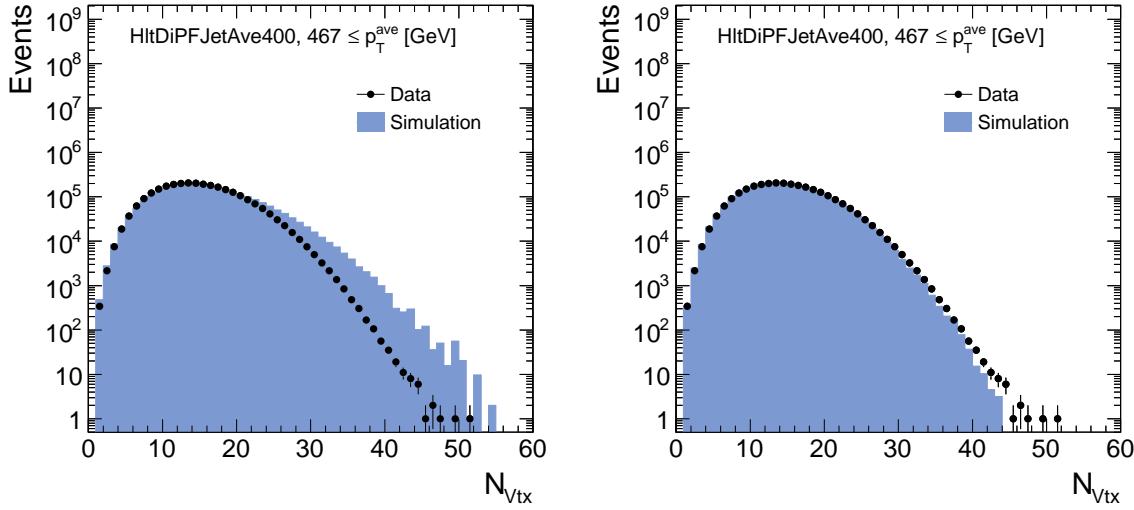


Figure 6.2: Distribution of number of primary vertices in data (black dots) and simulation (blue histogram) before (*left*) and after (*right*) reweighting of the pileup scenario in simulation for trigger path HltDiPFJetAve400.

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

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

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

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

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

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

$$\Delta|\eta|_{\text{jet},1} = \Delta|\eta|_{\text{jet},2} \quad (6.8)$$

To account for the  $\alpha$ -dependence of the measured asymmetry distributions, the  $p_T^{\text{ave}}$  and  $|\eta|$  intervals are further subdivided in various  $\alpha$  intervals. These are chosen such that each  $\alpha$  interval ranges from  $\alpha = 0.0$  to a particular upper boundary  $\alpha_{\text{max}}$ . The respective upper boundaries of the  $\alpha_{\text{max}}$  intervals are 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25. This inclusive definition of the  $\alpha$  intervals implicates that one specific event might be assigned to more than one  $\alpha$  bin.

The resulting inclusive  $p_T^{\text{ave}}$  spectrum after applying the 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  $p_T^{\text{ave}}$  interval is normalized to the respective integral in data. It is visible that the shapes of the spectra in data and simulation agree well. Furthermore, the effect of the pre-scales applied to the low- $p_T^{\text{ave}}$ -threshold triggers is nicely visible resulting in the sawtooth-shaped  $p_T^{\text{ave}}$ -spectrum.

Finally, 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

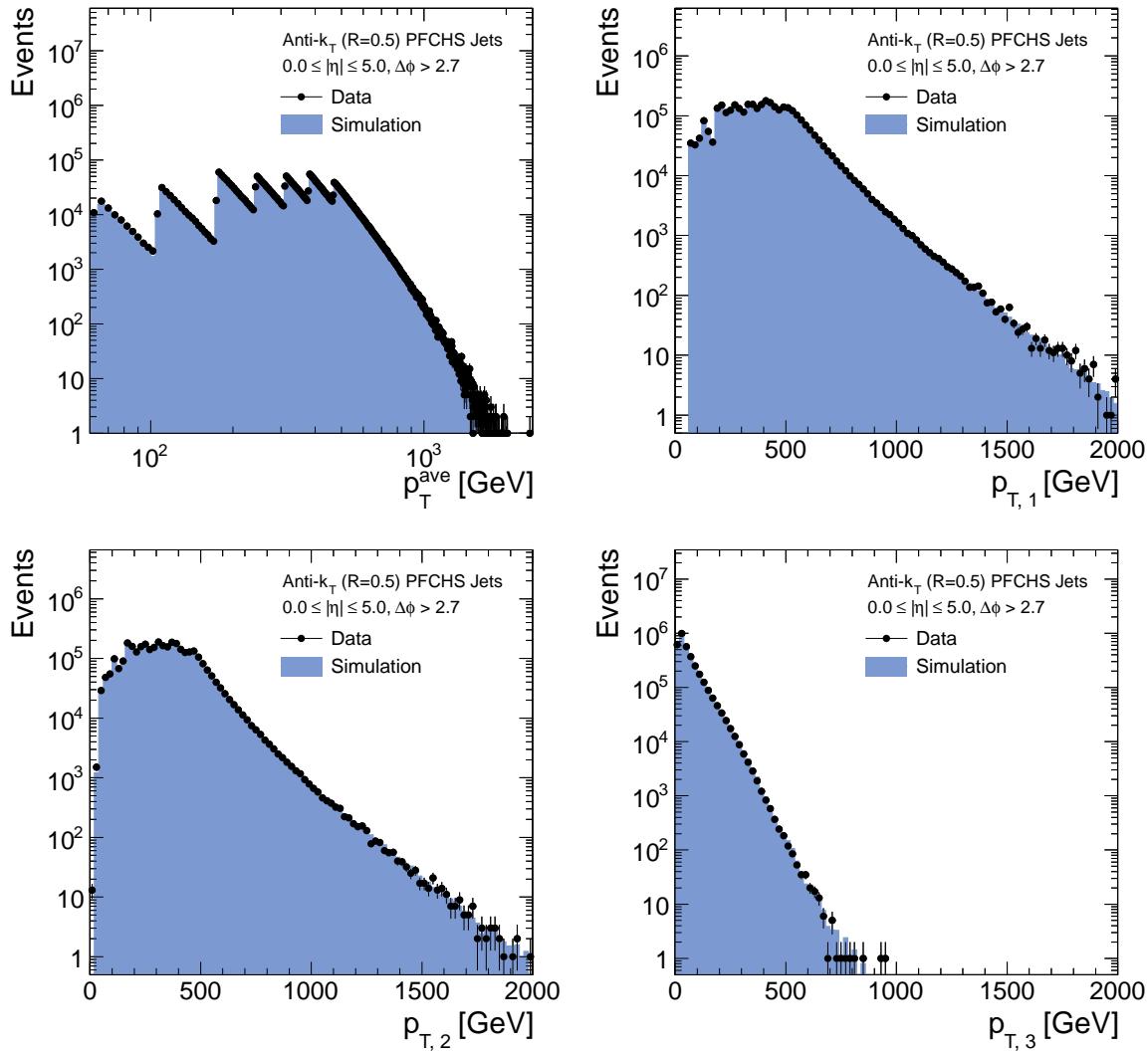


Figure 6.3: Inclusive  $p_T^{\text{ave}}$  spectrum of events after the total selection in data and in simulation (top left), leading jet  $p_T$  (top right), subleading jet  $p_T$  (bottom left) and third jet  $p_T$  (bottom right).

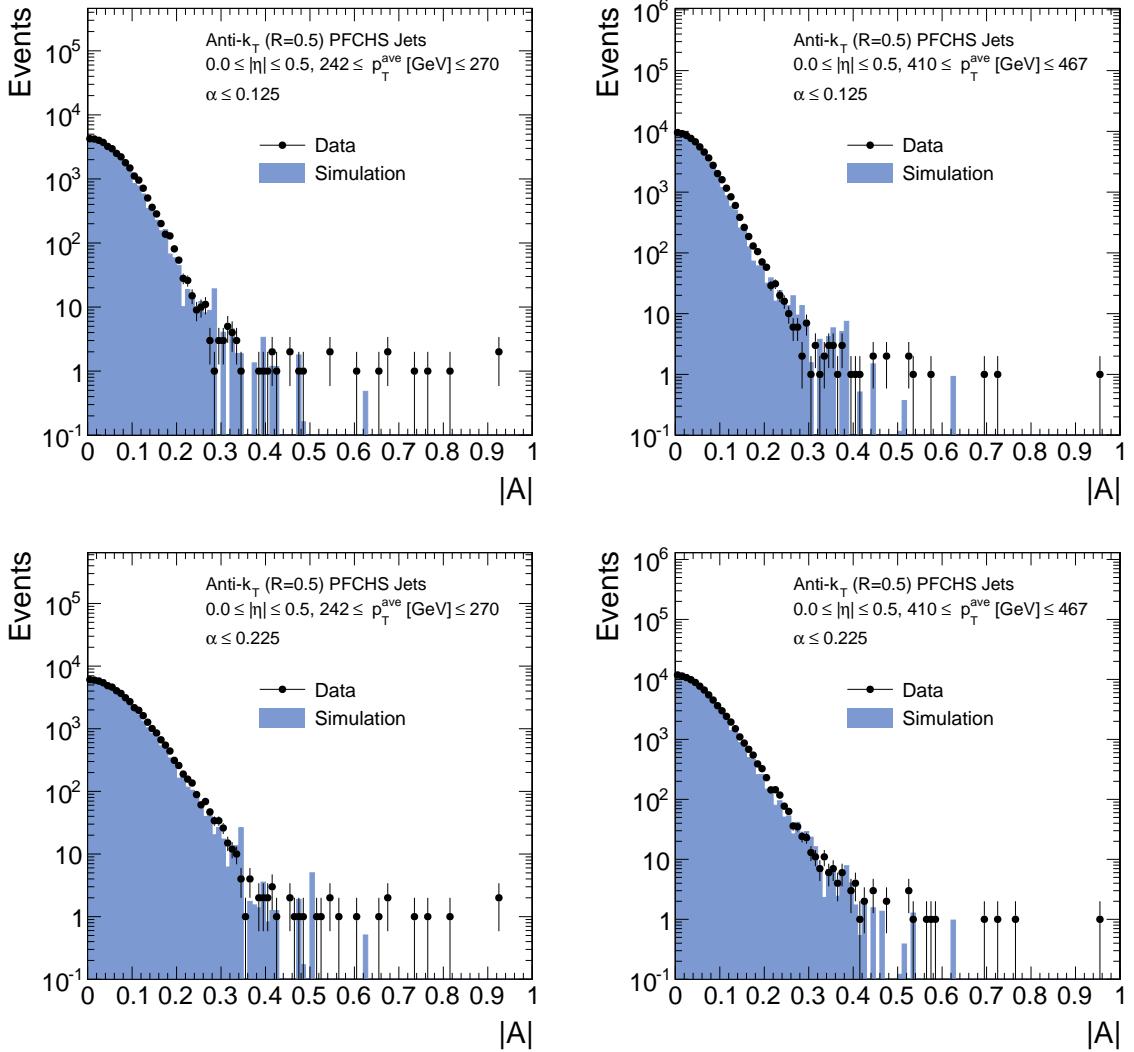


Figure 6.4: Some example asymmetry distributions for the lowest  $|\eta|$  region, two medium  $p_T^{\text{ave}}$  intervals and for a low (*top*) and a high (*bottom*)  $\alpha$  interval.

is used. The effect that the asymmetry distribution gets broadened for larger values of  $\alpha$  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  $p_T^{\text{ave}}$ ,  $\eta$  and  $\alpha$  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. The standard deviation of the Gaussian function can be chosen according to different definitions of the width of the asymmetry distribution and so the definition of the asymmetry width can be identified which features a good description of the asymmetry distribution by a corresponding Gaussian function.

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

$$\sigma_A = \text{RMS}_{t\%} = \sqrt{\frac{1}{\sum_i y_i} \sum_i y_i \cdot A_i^2} \quad (6.9)$$

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

$$\Delta\sigma_A = \Delta\text{RMS}_{t\%} = \frac{\text{RMS}_{t\%}}{\sqrt{2 \cdot n_{\text{eff}}}} \quad (6.10)$$

with the number of effective entries  $n_{\text{eff}}^1$  in the specified  $(p_T^{\text{ave}}, |\eta|, \alpha)$ -interval. The truncation is chosen such that the whole distribution, 98.5% or only 95% are considered. A comparison of the asymmetry distributions and Gaussian functions where the standard deviation has been set to the value of the determined asymmetry width is illustrated in Fig. 6.5 for a certain  $|\eta|$ ,  $p_T^{\text{ave}}$  and  $\alpha$  interval. It is visible that the core part of the asymmetry distribution can in good approximation be described by a Gaussian function when choosing  $\text{RMS}_{98.5\%}$  as the asymmetry width (cf. middle row in Fig. 6.5). Hence, this is the default definition of the asymmetry width used for this measurement. It is chosen to be the same for data events as well as for simulated events.

## 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 Section 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. These corrections are explained in the following Sections 6.6.1 and 6.6.2.

---

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

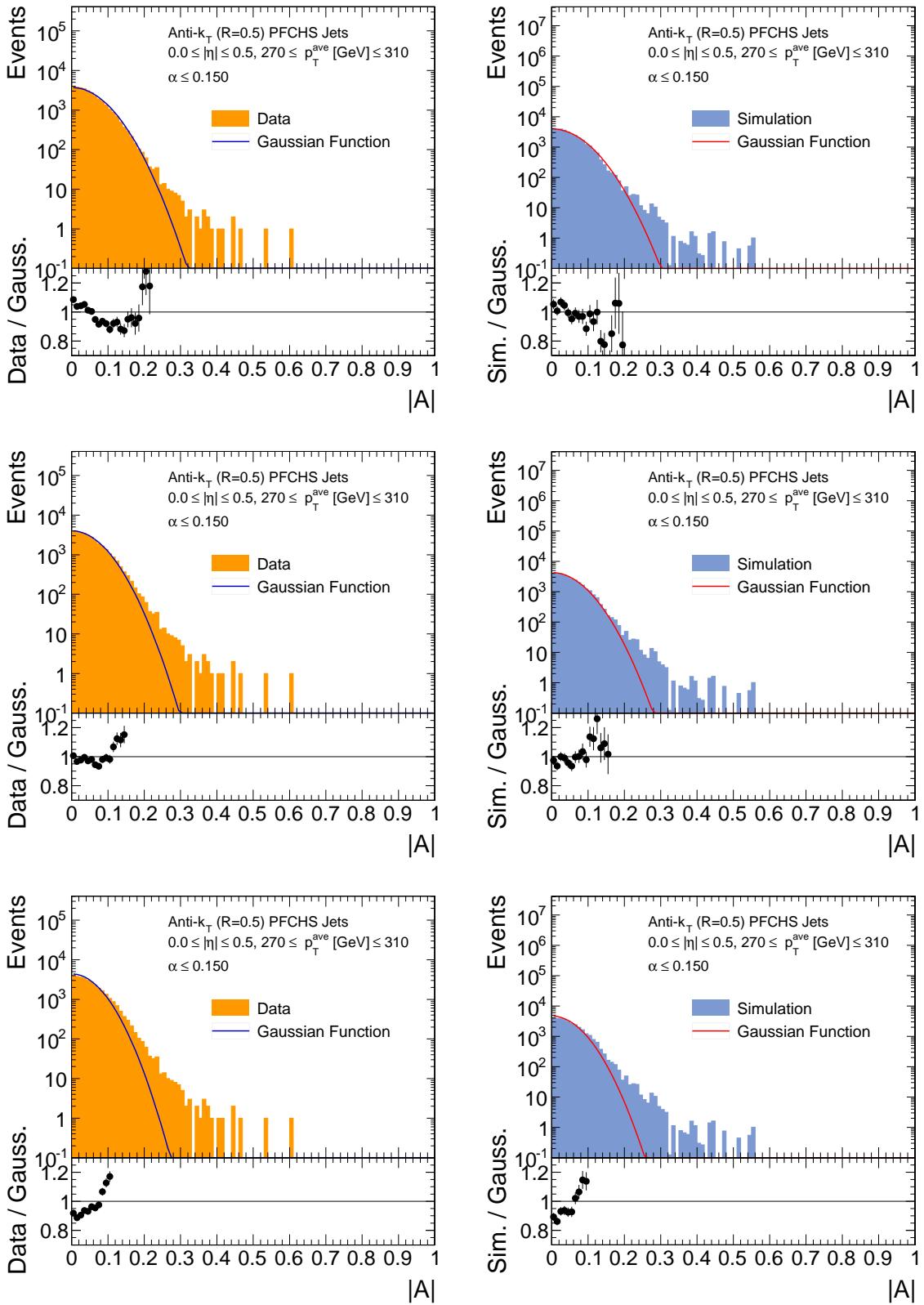


Figure 6.5: Asymmetry distributions for data (left) and simulation (right) compared to Gaussian distributions obtained with standard deviations corresponding to 0% truncation (top), 1.5% truncation (middle) and 5% truncation (bottom) of the respective upper tail.

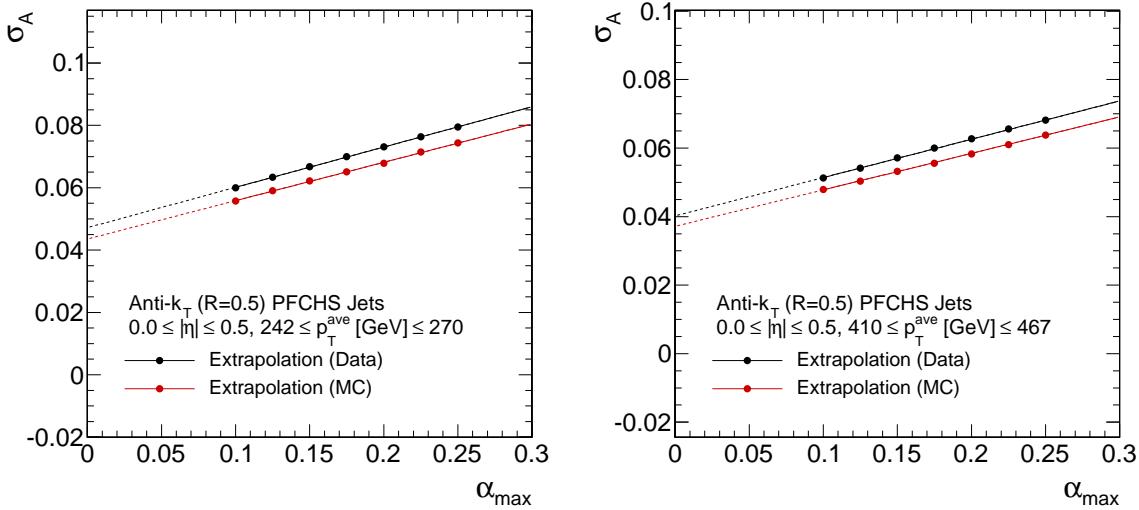


Figure 6.6: Two example extrapolations of measured values for  $\sigma_A$  in data (black) and simulation (red) to obtain the asymmetry width for zero additional jet activity.

### 6.6.1 Correction for Additional Jet Activity

The imbalance contribution arising from additional jet activity is taken care of by an extrapolation procedure. As described in Section 6.4.2, the asymmetry distribution is calculated for several intervals in  $|\eta|$  and  $p_T^{\text{ave}}$  with different selections on the maximum value of  $\alpha$  ( $= \alpha_{\max}$ ). For each of these individual selections, the width of the asymmetry is determined as stated in Section 6.5. The measured values of  $\sigma_A(\alpha_{\max})$  are extrapolated to  $\alpha_{\max} \rightarrow 0$  assuming a linear behaviour. Thus, the measured asymmetry widths in one particular ( $p_T^{\text{ave}}$ ,  $|\eta|$ ,  $\alpha$ )-interval are fitted with a linear function. The  $y$ -intercept of the fitted linear function represents the resolution without further jet activity in the event. The statistical uncertainty is given by the respective fit uncertainty for the intercept.

The extrapolation procedure for two exemplary  $|\eta|$  and  $p_T^{\text{ave}}$  intervals is illustrated in Fig. 6.6. The performed extrapolation fits for all other non-empty intervals are shown in App. A.2.

As stated in Section 6.4.2, the selection is performed in inclusive intervals of  $\alpha < \alpha_{\max}$ . This results in a correlation of the measured values of  $\sigma_A$  for a particular ( $p_T^{\text{ave}}$ ,  $|\eta|$ )-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 [127, 144, 153].

More specifically this means that the measured data points are described by a linear function

$$a \cdot \alpha_{\max} + b = \sigma_A(\alpha_{\max}) \quad (6.11)$$

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

$$\chi^2 = dy^T \cdot C^{-1} \cdot dy \quad (6.12)$$

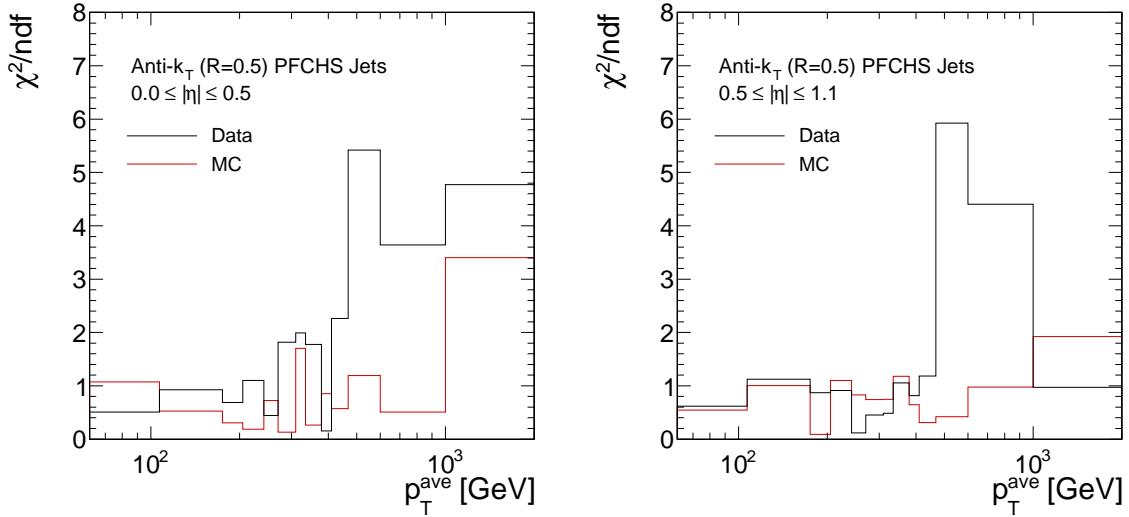


Figure 6.7: Goodness-of-fit test for the extrapolation fits in data and simulation in the two central  $|\eta|$  regions as function of  $p_T^{\text{ave}}$ .

with  $dy = y_{\text{measured}} - y_{\text{predicted}}$  and the covariance matrix  $C$ . Consequently, the asymmetry width for vanishing additional jet activity is given by  $b = \sigma_A(\alpha_{\max} \rightarrow 0)$ . With the assumption that all events that belong to  $\alpha$  interval  $i$  are also completely included in the next higher  $\alpha$  interval  $j$ <sup>2</sup> the covariance matrix is given by

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

where  $n$  is the number of events in that particular  $\alpha$  interval. A discussion regarding the derivation of that expression is summarized in App. A.3. The function minimization itself is performed with the Minuit package [159]. 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  $\chi^2$  over the number of degrees of freedom are shown as function of  $p_T^{\text{ave}}$  exemplary for the two central  $|\eta|$  intervals. This is expected to be distributed around 1. Consequently, the fit quality is in general quite good but worsens for larger  $p_T^{\text{ave}}$  values. Since especially for higher  $p_T^{\text{ave}}$  intervals in the central  $|\eta|$  regions, the statistical uncertainty is rather low, the fit is very sensitive to even small deviations from a linear behaviour which results in a minor fit quality. However, a possible non-linearity is considered in the systematic uncertainties of the measurement, as discussed in Section 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.

<sup>2</sup>This assumption is only almost true since the asymmetry distributions are truncated to reject non-Gaussian components and some events might not fulfill this criterion.

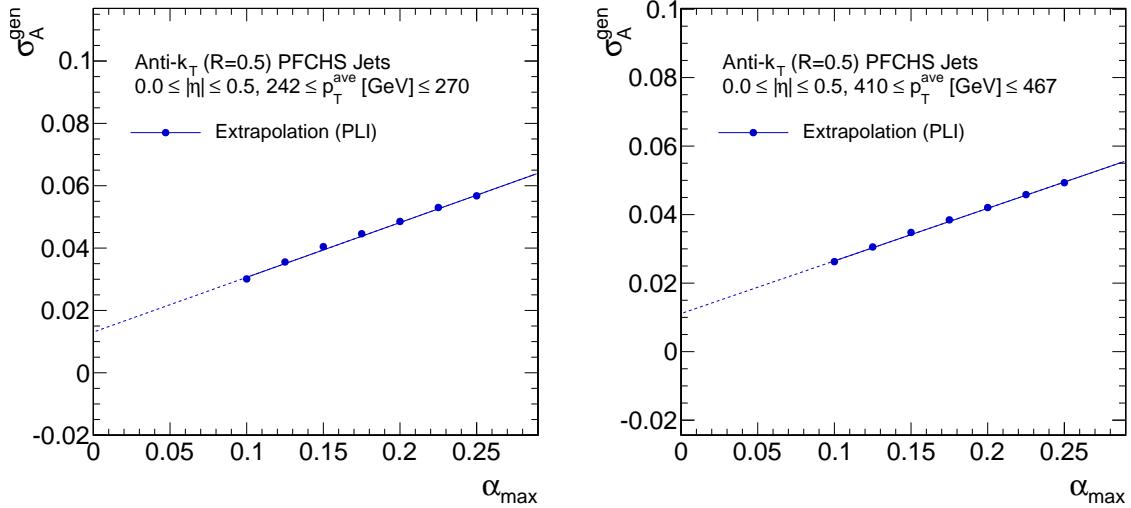


Figure 6.8: Two example extrapolations of measured values for  $\sigma_A^{\text{gen}}$  in simulation to determine the contribution from PLI.

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

$$A^{\text{gen}} = \frac{p_{T,1}^{\text{gen}} - p_{T,2}^{\text{gen}}}{p_{T,1}^{\text{gen}} + p_{T,2}^{\text{gen}}} \quad (6.14)$$

with  $p_{T,1}^{\text{gen}}$  and  $p_{T,2}^{\text{gen}}$  referring to the momenta of the two leading generated jets. This distribution is affected by additional parton radiation as well. Thus, the procedure to obtain  $\sigma_A^{\text{gen}}(\alpha_{\max} \rightarrow 0)$  is the same as for the asymmetry at detector level. The generator asymmetry is calculated in the same ( $p_T^{\text{ave}}$ ,  $|\eta|$ ,  $\alpha$ )-intervals as the detector-level asymmetry in order to derive the size of the particle-level imbalance for exactly the same events. The asymmetry width of the generator asymmetry is also calculated as RMS<sub>98.5%</sub>. In order to obtain the values of  $\sigma_A^{\text{gen}}$  for zero additional jet activity, the analogue extrapolation procedure is performed, as described in Sec. 6.6.1. Some example extrapolations are illustrated in Fig. 6.8.

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

$$\sigma_{\text{PLI}} = \sqrt{2} \cdot \sigma_A^{\text{gen}}(\alpha_{\max} \rightarrow 0). \quad (6.15)$$

### 6.6.3 Results of the Corrections to the Asymmetry

The results of the various extrapolation fits to determine  $\sigma_A(\alpha_{\max} \rightarrow 0)$  in data and simulation as well as for the particle-level imbalance are shown together in Fig. 6.9 as a function of  $p_T^{\text{ave}}$  for the 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

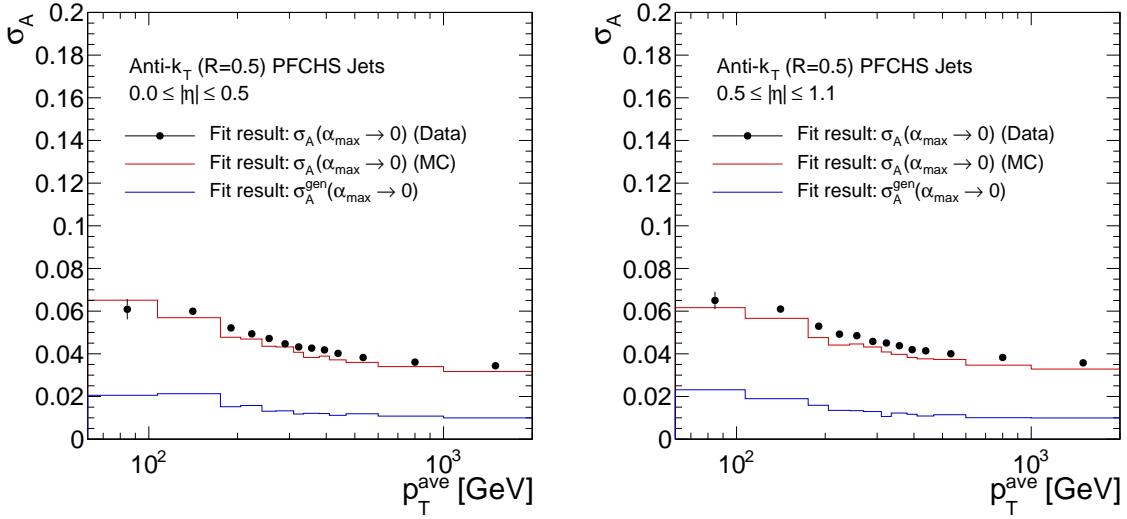


Figure 6.9: Results of extrapolation fits in the two central  $|\eta|$  regions as a function of  $p_T^{\text{ave}}$  for data (black), simulation (red) and particle-level imbalance (blue).

by subtracting the PLI correction in quadrature from the measured  $\sigma_A$  after the correction for additional jet activity

$$\sigma_{\text{JER}} = \sqrt{2} \cdot \sigma_A(\alpha_{\max} \rightarrow 0) \ominus \sigma_{\text{PLI}}. \quad (6.16)$$

The same PLI correction is subtracted both from data as well as from simulation results. As can be seen in Fig. 6.9, the correction due to particle-level imbalance is small compared to the measured asymmetry widths. This justifies to take this correction from simulation, also for the data resolution. However, in order to account for a possible imprecise modelling of the particle-level imbalance contribution, a systematic uncertainty will be considered, as discussed in 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

$$c(\text{Data}/\text{MC}) = \frac{\sigma_{\text{JER, Data}}}{\sigma_{\text{JER, MC}}}. \quad (6.17)$$

The resulting distributions as function of  $p_T^{\text{ave}}$  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 also visualized in Fig. 6.10 (red line) together with the statistical uncertainty resulting from the fit (grey shaded area). The description with a constant seems justified, as the values of  $\chi^2/\text{ndf}$ , also displayed in Fig. 6.10, resulting from the fit are in agreement with

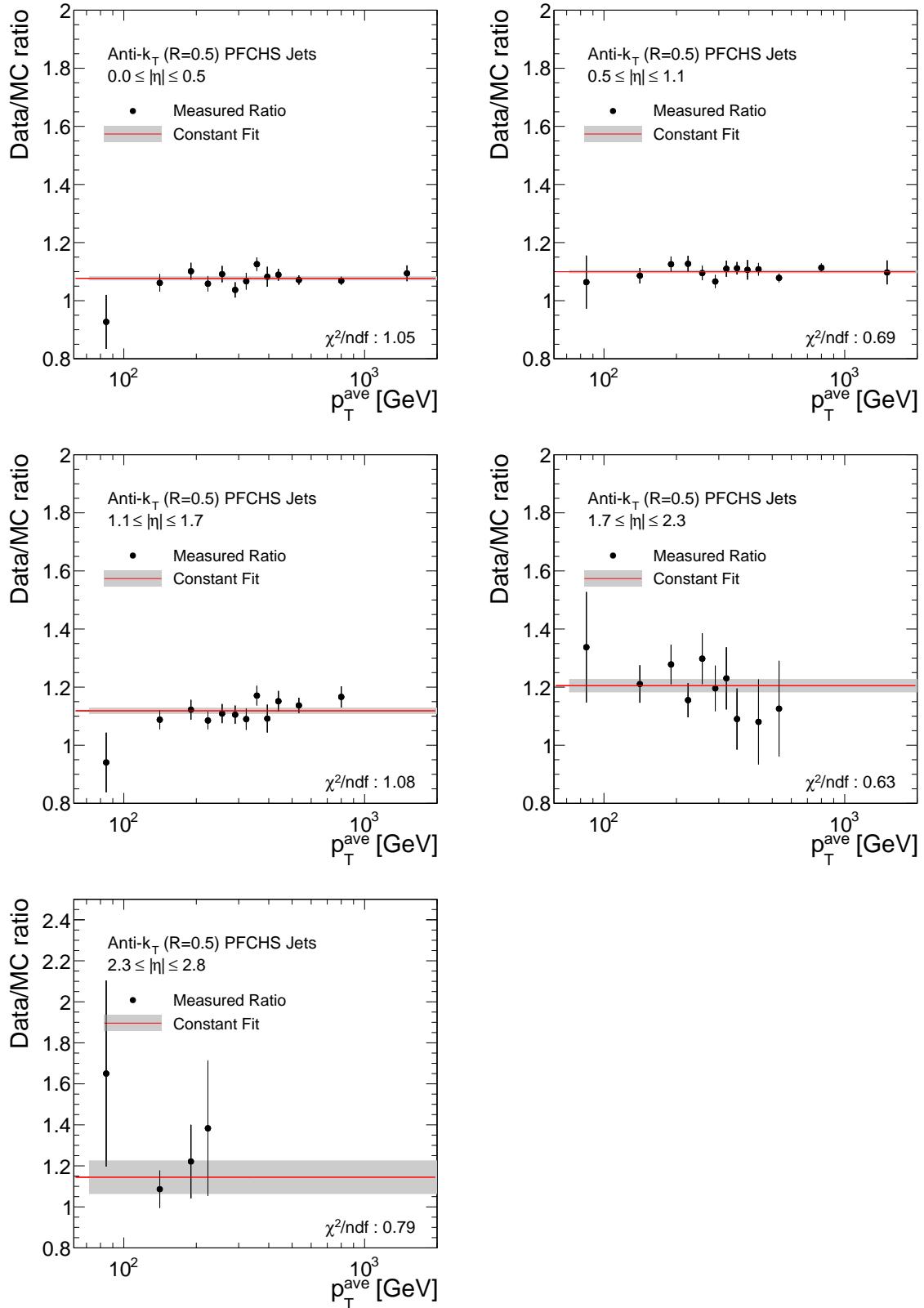


Figure 6.10: Ratio of the resolution in data to the resolution in the PYTHIA QCD sample in non-empty  $|\eta|$  regions as a function of  $p_T^{\text{ave}}$  shown together with a constant fit (red line) and corresponding statistical uncertainty (grey shaded area).

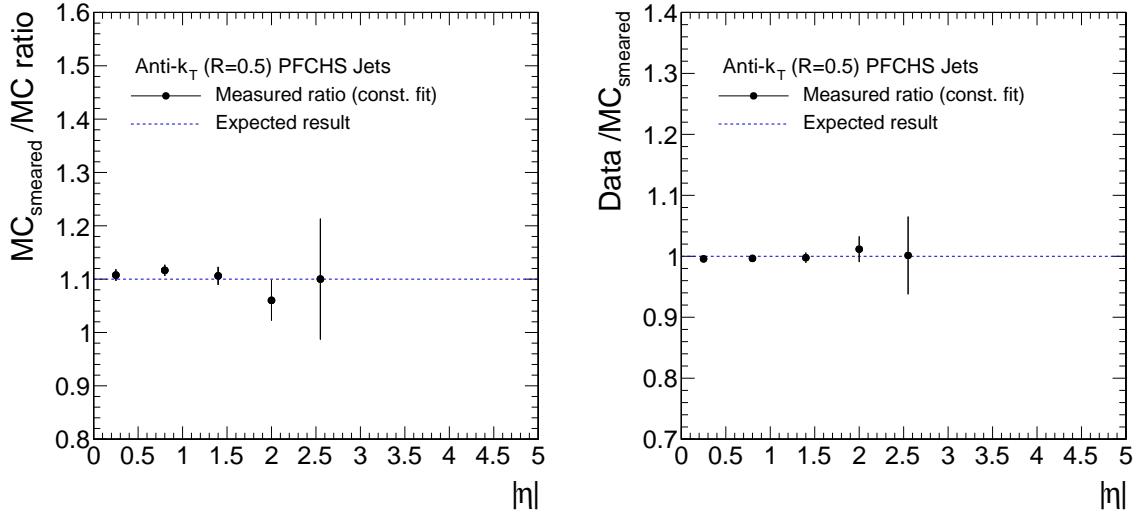


Figure 6.11: Ratio  $c(MC_{\text{smeared}}/MC)$  of the resolution in the smeared QCD sample to the resolution of the unsmeared QCD sample as a function of  $|\eta|$  with statistical uncertainty (left) and ratio  $c(\text{Data}/MC_{\text{smeared}})$  of the resolution in data to the resolution of the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

one. However, a potential hidden trend as function of  $p_T^{\text{ave}}$  is considered as systematic uncertainty of the measurement.

The determined data-to-simulation ratios are summarized later together with statistical and systematic uncertainties in Tab. 6.3. Due to statistical limitations, no data-to-simulation ratio could be determined for the two highest  $|\eta|$  intervals. As jets in the forward region of the detector have essentially low transverse momentum, these  $|\eta|$  intervals are mainly affected by the high pre-scales of the triggers with low momentum thresholds.

## 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 corrected jets<sup>3</sup> in  $p_T$  in each event are smeared with a smearing factor  $c$  which increases the  $p_T$  resolution. This is done for jets which pass a minimum  $p_T$  threshold of 10 GeV. More precisely this means that for each reconstructed jet that has a corresponding generated jet within a  $\Delta R < 0.25$  cone,

<sup>3</sup>It is necessary to correct 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 corrected.

the transverse momentum is scaled according to

$$p'_T = p_T^{\text{gen}} + c \cdot (p_T - p_T^{\text{gen}}). \quad (6.18)$$

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

For this test, the smearing factor has been chosen to be  $c = 1.1$  for each  $|\eta|$  interval. This is expected to be the resulting ratio when determining the ratio of the resolution in the smeared QCD sub-sample to the resolution in the unsmeared sub-sample derived with the asymmetry method including all corrections as described in Sec. 6.6. Consequently, the test is passed when the measured ratio recovers the input value according to  $c(\text{MC}_{\text{smeared}}/\text{MC}) = 1.1$ . The resulting values obtained from a constant fit of the ratios as a function of  $p_T^{\text{ave}}$  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 well recovered with this procedure within the statistical uncertainties. Thus, a potential resolution difference when comparing data and simulation can be determined with this approach as well.

### 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 Section 6.8.1.

For the test of the measured data-to-simulation ratio, the individual jet momenta in the simulated PYTHIA QCD sample are scaled with the measured resolution ratio values  $c(\text{Data}/\text{MC})$  according to Eq. 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 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 re-determination of the data-to-simulation ratio, it is expected to obtain ratios of  $c(\text{Data}/\text{MC}_{\text{smeared}}) = 1$ . The result of this test is illustrated in Fig. 6.11 (right) with statistical uncertainties. A good agreement with the expected value of 1 is visible, even without the consideration of systematic uncertainties in this consistency test. This again proves that the dijet asymmetry method is well suited to determine resolution differences between data and simulation, expressed in the data-to-simulation ratio.

## 6.9 Systematic Uncertainties

In addition to the uncertainty of the resolution measurement due to statictical 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  $\delta c$  is calculated by the determination of the ratio for a certain shift  $\Delta$  and comparing it to the nominal ratio

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

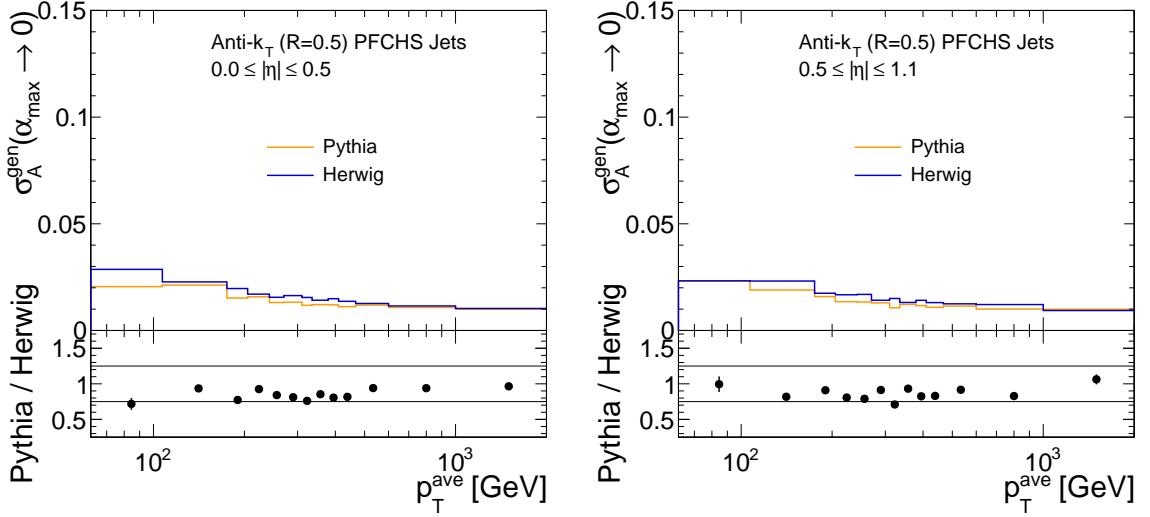


Figure 6.12: Comparison of the contributions to the asymmetry width due to particle-level imbalance in QCD multijet events generated with PYTHIA (yellow) to the same quantity derived with the HERWIG++ generator (blue) in the two central  $|\eta|$  regions as a function of  $p_T^{\text{ave}}$ .

Thus, all systematic uncertainties are determined as absolute shift of the nominal ratio. Typically, an upward and downward shift is evaluated. Resulting uncertainties are afterwards symmetrized by taking the average shift. If, however, the up- and downward variation both result in either an upward or downward shift, the absolute shift is determined and the maximum of both is taken and quoted as symmetric uncertainty. In case only an up- or downward variation is performed, the systematic uncertainty 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  $\sigma_{\text{PLI}}$ , the PLI-correction factor for each measured resolution value is shifted by  $\pm 25\%$ . Consequently, the changed ratio is calculated as

$$c(\text{Data/MC})_{\text{PLI}} = \frac{\sqrt{2} \cdot \sigma_A^{\text{Data}}(\alpha_{\max} \rightarrow 0) \ominus f \cdot \sigma_{\text{PLI}}}{\sqrt{2} \cdot \sigma_A^{\text{MC}}(\alpha_{\max} \rightarrow 0) \ominus f \cdot \sigma_{\text{PLI}}} \quad (6.20)$$

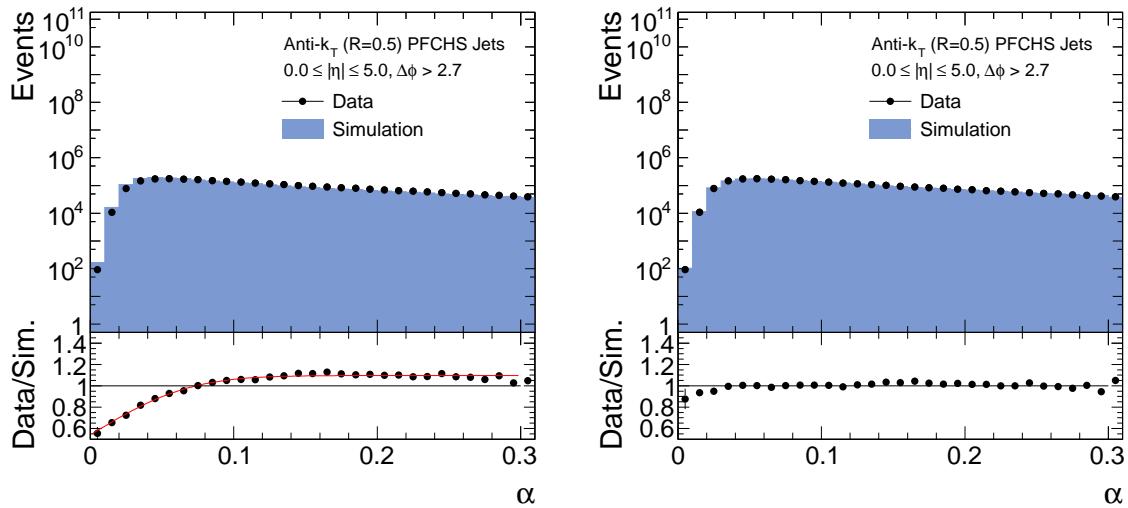


Figure 6.13: Inclusive  $\alpha$ -spectrum before (*left*) and after (*right*) reweighting the  $\alpha$ -spectrum of the simulated sample. The red curve in the bottom of the left plot illustrates the function used for the  $\alpha$ -spectrum reweighting.

with  $f = 0.75, 1.25$  respectively. The size of this variation is chosen by comparing the size of the PLI correction in simulated events from the nominal PYTHIA sample to the PLI correction estimated from simulated events obtained by the HERWIG++ generator with tune EE3C. This comparison is illustrated in Fig. 6.12 for the two central  $|\eta|$  regions. Since the size of the PLI correction agrees within 25% between both generators, the size of this uncertainty is justified.

**Jet energy scale:** The jet energy scale has been corrected to particle level by the application of dedicated calibration factors. 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 determined again based on the varied jet momenta.

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

The observed inclusive  $\alpha$ -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/MC. Since the ratio shows that both spectra do not agree, the influence of

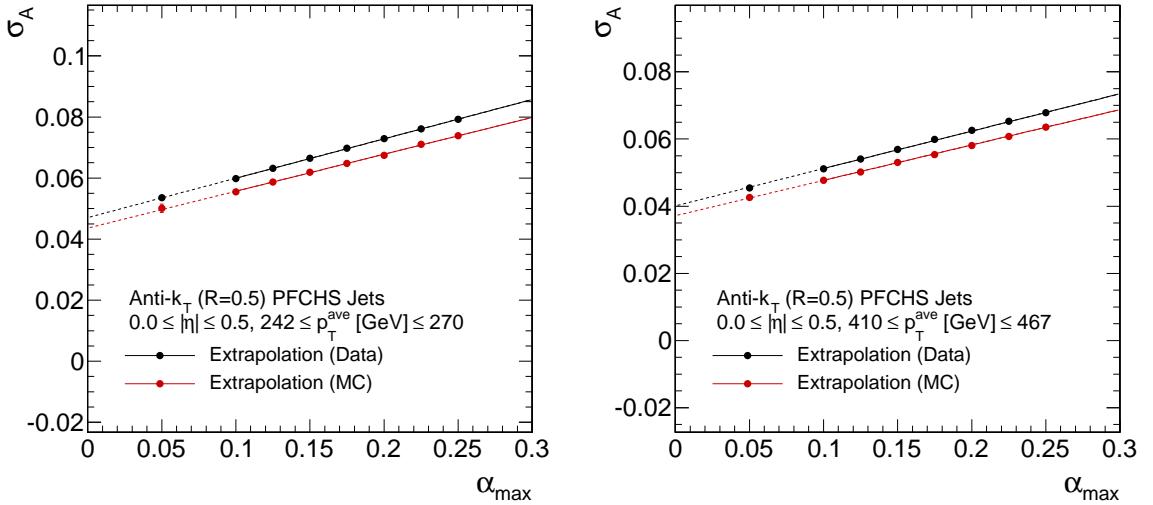


Figure 6.14: Two example extrapolations of measured values for  $\sigma_A$  in data and simulation to obtain the result at zero additional jet activity when adding one additional  $\alpha$  interval 0.0–0.05.

the  $\alpha$ -spectrum on the data-to-simulation ratio is evaluated by reweighting the  $\alpha$ -spectrum in simulated events to roughly match the one observed in data. The red curve overlaid in the ratio of the left distribution is used to reweight the events in the simulation. For each event a weight  $w(\alpha)$  is calculated according to

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

with the error function  $\text{erf}^4$ . This weight is considered as multiplicative factor onto the usual event weight in simulation. For comparison, the reweighted  $\alpha$ -spectrum is shown in the right part of Fig. 6.13.

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

In order to study the linear behaviour of the extrapolation also towards smaller  $\alpha$ -values, the minimum  $p_T$  cut of 10 GeV for the third jet is dropped and an additional  $\alpha$ -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  $\alpha$ -interval are shown. The resulting difference to the nominal data-to-simulation ratio is considered as systematic uncertainty.

<sup>4</sup> $\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt$

**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 ( $u, d, s$ ) and heavy-flavour quarks ( $c, b$ ). If the flavour composition is the same in data and simulation, this flavour difference cancels out in the data-to-simulation ratio. Since it is known that especially the rate of gluon splitting events with  $g \rightarrow b\bar{b}$  is modelled wrong by about a factor of two [160], 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. All events identified as gluon splitting into heavy-flavour quarks get an event weight of 1.5. Afterwards, the data-to-simulation ratio is derived as the resolution from data to the resolution in the reweighted simulated sample.

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

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

$$f(p_T) = \sqrt{\frac{N^2}{p_T^2} + \frac{S^2}{p_T} + C^2} \quad (6.22)$$

with  $p_T = p_T^{\text{ave}}$  and the free parameters  $N, S$  and  $C$ . This function is chosen in accordance to the parametrization typically used for the description of the relative energy-resolution in calorimeters (cf. Sec. 3.2). If the resolution in simulation is fitted with this NSC-function, it is expected to find one common scale factor  $k_{\text{NSC}}$  for the  $N, S$  and  $C$  parameters to describe the data, in case the data-to-simulation ratio is flat as a function of  $p_T^{\text{ave}}$ . Significant differences between data and the actual modelling of the detector behaviour consequently can result in the necessity to use individual scale factors for the  $k_N, k_S$  and  $k_C$  parameters in order to describe the data from the fit results of the function to the simulation.

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

$$f(p_T) = \sqrt{\frac{(k_{\text{NS}} \cdot N)^2}{p_T^2} + \frac{(k_{\text{NS}} \cdot S)^2}{p_T} + (k_C \cdot C)^2} \quad (6.23)$$

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

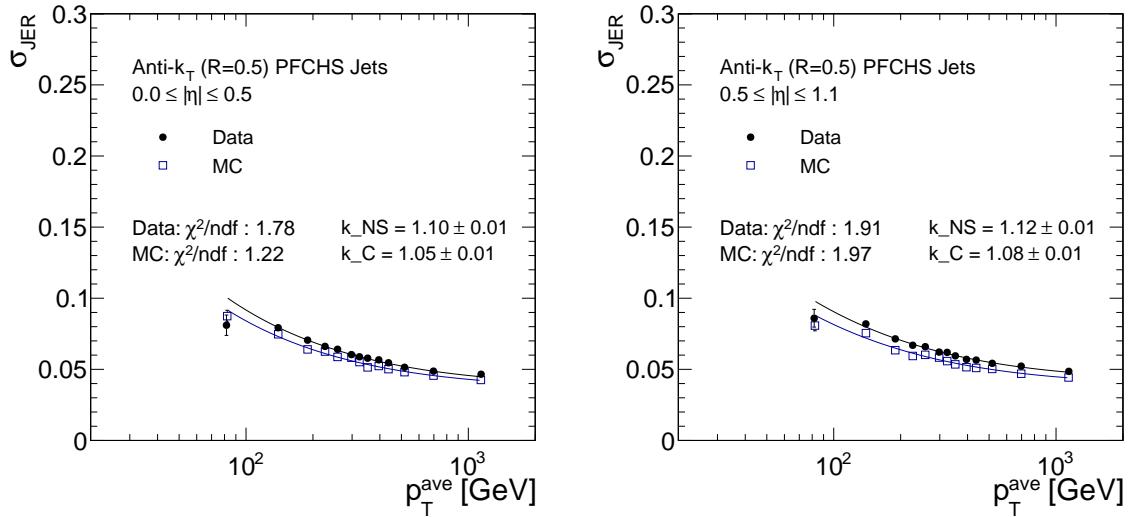


Figure 6.15: Results of fitting the resolution in data and simulation with the respective NSC-functions in the two central  $|\eta|$  intervals.

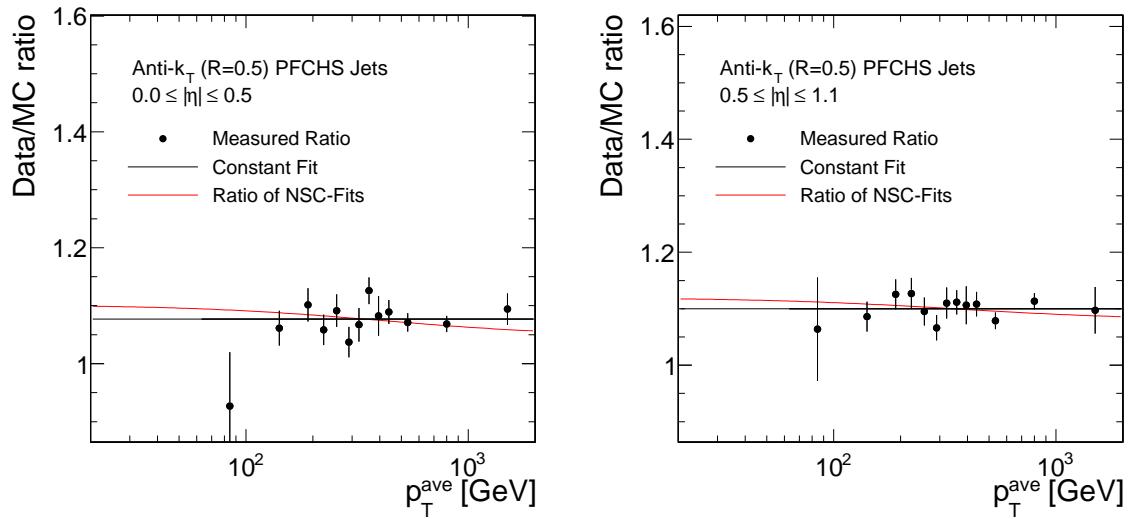


Figure 6.16: Ratio of the results fitting the resolution in data and simulation with the NSC-functions in the two central  $|\eta|$  intervals together with the measured ratio and a constant fit.

$C$ . Since there are no events with low transverse momenta available, no sensitivity to the special characteristics of  $N$  is expected and hence no separate scale factor for  $N$  is defined. Furthermore, this means that this procedure can not be used to study influences from pileup.

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  $p_T^{\text{ave}}$  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. It is visible that the ratio derived from the NSC-fits is also compatible with the measured values for the ratio. Consequently, a trend versus  $p_T^{\text{ave}}$  could be hidden in the measured points. The difference of the measured scale factors  $k_{NS}$  and  $k_C$  to the central value of the measurement from the constant fit amounts to about 2% in the two central  $|\eta|$  intervals. This is considered as systematic uncertainty for all  $|\eta|$  intervals in order to compensate for this potentially hidden shape in the data-to-simulation ratio. Furthermore, this also covers possible differences of the scaling factors for jets with very low or very high transverse momenta which can not be directly measured in this analysis. Overall, this systematic uncertainty constitutes the major contribution to the total systematic uncertainty.

A summary of the data-to-simulation ratios measured with the dijet asymmetry method requiring that the two leading jets belong to the same  $|\eta|$  interval together with all systematic uncertainties considered for the ratio is given in Tab. 6.3. Due to limited event numbers, no results could be determined for the two highest  $|\eta|$  regions with this approach. In general, the measurement is limited by systematic effects which amount to uncertainties of about 2–2.5% in the central detector part. However, in the  $|\eta| \in [2.3, 2.8]$  region the measurement is limited by the statistical uncertainty.

## 6.10 Extension of the Method to the Forward Detector Region

The presented measurement with results summarized in Tab. 6.3 is utilizing the dijet asymmetry method based on the requirement that the two leading jets in an event belong both to the same  $|\eta|$  region. Consequently, the resolution of these two jets is equal which results in the simple relation expressed in Eq. 6.2 between asymmetry width and jet resolution. However, the requirement that both leading jets have to belong to the same  $|\eta|$  interval reduces significantly the available number of events. Especially in the forward region of the detector this is a problem, as the jets in these events have low transverse momenta and consequently have to be triggered by the lowest  $p_T^{\text{ave}}$  triggers which are highly prescaled. 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  $i_{\text{jet},1}^{|\eta|} = i_{\text{jet},2}^{|\eta|}$  has to be dropped. Instead, events are selected in which the two leading jets belong to different pseudorapidity regions. Nonetheless, the jet resolution  $\sigma(p_T^{\text{probe}})$  in a probe interval  $|\eta_{\text{probe}}|$  can be determined, if the resolution  $\sigma(p_T^{\text{ref}})$  in a reference interval  $|\eta_{\text{ref}}|$  is known.

Table 6.3: Summary of the measured data-to-simulation ratios  $c(\text{Data}/\text{MC})$  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     |
| $c(\text{Data}/\text{MC})$ | 1.077       | 1.100       | 1.119       | 1.205       | 1.145       |
| Stat. uncertainty          | $\pm 0.007$ | $\pm 0.006$ | $\pm 0.010$ | $\pm 0.027$ | $\pm 0.078$ |
| 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       |
| $\alpha$ -spectrum         | 0.004       | 0.007       | 0.004       | 0.006       | 0.003       |
| $\alpha$ -range            | 0.005       | 0.009       | 0.004       | 0.011       | 0.012       |
| Non-Gaussian tails         | 0.004       | 0.003       | 0.003       | 0.013       | 0.007       |
| Jet Flavour                | 0.007       | 0.004       | 0.005       | 0.006       | 0.003       |
| Ratio shape                | 0.022       | 0.022       | 0.022       | 0.024       | 0.023       |
| Total syst. uncertainty    | $\pm 0.025$ | $\pm 0.027$ | $\pm 0.026$ | $\pm 0.038$ | $\pm 0.038$ |

Based on Eq. 6.3, it can be shown that for  $\langle p_T^{\text{probe}} \rangle = \langle p_T^{\text{ref}} \rangle = \langle p_T \rangle$  and  $\sigma(p_T^{\text{probe}}) \neq \sigma(p_T^{\text{ref}})$  the relation

$$\frac{\sigma(p_T^{\text{probe}})}{\langle p_T \rangle} = \sqrt{4 \cdot \sigma_A^{|\eta_{\text{probe}}| \neq |\eta_{\text{ref}}|} - \left( \frac{\sigma(p_T^{\text{ref}})}{\langle p_T \rangle} \right)^2} \quad (6.24)$$

is obtained. Here,  $\sigma_A^{|\eta_{\text{probe}}| \neq |\eta_{\text{ref}}|}$  is the width of the asymmetry calculated from the two leading jets in an event of which one jet belongs to a reference  $|\eta|$  interval and the probe jet 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  $\sigma_{A,\text{gen}}^{|\eta_{\text{probe}}| \neq |\eta_{\text{ref}}|}$ , i.e. 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 preferred to be in the central detector region where the statistical precision from the same- $|\eta|$  measurement has already been sufficient. The statistical uncertainty of  $\frac{\sigma(p_T^{\text{ref}})}{\langle p_T \rangle}$  is propagated to the statistical uncertainty of  $\frac{\sigma(p_T^{\text{probe}})}{\langle p_T \rangle}$ .

As long as both jets belong to the same  $|\eta|$  region, residual effects from jet energy scale differences in data and simulation affect both jets the same way and should not have an impact on the resolution measurement. Since residual effects from the jet energy scale,

where the mean of the asymmetry is shifted and not exactly located at zero, become more important, if both jets belong to different probe and reference intervals, a slightly modified definition of the asymmetry of

$$A = \frac{p_T^{\text{probe}} - p_T^{\text{ref}}}{p_T^{\text{probe}} + p_T^{\text{ref}}} \quad (6.25)$$

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

$$\sigma_A = \text{RMS}_{98.5\%} = \sqrt{\frac{1}{y_i} \cdot \sum_i (A_i - A_{\text{mean}})^2} \quad (6.26)$$

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 symmetric around the mean.

The measurement with the reference-and-probe-interval selection is performed taking the three innermost  $|\eta|$  regions 0.0–0.5, 0.5–1.1 and 1.1–1.7 as reference intervals each. This allows a measurement of the data-to-simulation ratio also in the  $|\eta|$  intervals 2.8–3.2 and 3.2–5.0. In addition, it provides for the innermost  $|\eta|$  intervals an enhancement of the available number of events and thus a further reduction of the statistical uncertainties. The systematic uncertainties of the two new measurements are evaluated according to same- $|\eta|$  measurement by varying a certain aspect in the determination of the data-to-simulation ratio and taking the absolute deviation from the nominal ratio as uncertainty. The variation is done simultaneously for the  $\sigma_A^{||\eta_{\text{probe}}| \neq |\eta_{\text{ref}}|}$  value as well as for  $\frac{\sigma(p_T^{\text{ref}})}{\langle p_T \rangle}$  for each uncertainty source. 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. Thus, 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 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 same selection criteria, as described in Sec. 6.4.2. Also the pileup reweighting procedure is performed as described there. In Fig. 6.17, the inclusive  $p_T^{\text{ave}}$  spectrum is compared in data and simulation for simulated events taken from

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

| $ \eta $  | $ \eta_{\text{ref}}  \in [0.0, 0.5]$<br>$c(\text{Data}/\text{MC}) \pm \text{stat.} \pm \text{syst.}$ | $ \eta_{\text{ref}}  \in [0.5, 1.1]$<br>$c(\text{Data}/\text{MC}) \pm \text{stat.} \pm \text{syst.}$ | $ \eta_{\text{ref}}  \in [1.1, 1.7]$<br>$c(\text{Data}/\text{MC}) \pm \text{stat.} \pm \text{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 | $1.171 \pm 0.116 \pm 0.079$  | $0.829 \pm 0.082 \pm 0.149$  | $1.137 \pm 0.105 \pm 0.096$  |

HERWIG++. The agreement of the  $p_T^{\text{ave}}$  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 have been studied. The resulting data-to-simulation ratios for simulated events taken from HERWIG++ together with statistical uncertainties for the different measurements are summarized in Tab. 6.5.

In general, the obtained data-to-simulation ratios for HERWIG++ agree among each other and are quite similar to the values obtained from PYTHIA for most  $|\eta|$  regions. Some further discussion follows in Sec. 6.12.1.

## 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 for all  $|\eta|$  regions – except for the very last  $|\eta|$  region ranging from 3.2–5.0 which is discussed separately – derived from the measurements done with simulated events from PYTHIA while the HERWIG++ results are only used as cross check.

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

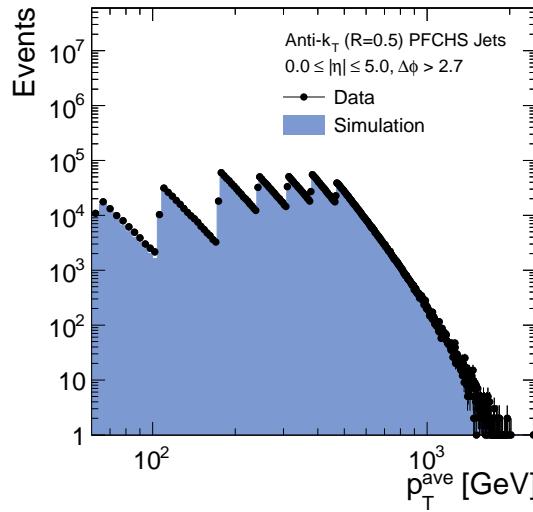


Figure 6.17: Inclusive  $p_T^{\text{ave}}$  spectrum of events after the event selection in data and in simulated events generated with HERWIG++.

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

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

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

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

the systematic uncertainties for the combined result the systematic shift of the weighted mean caused by the systematic shifts of the single measurements is determined. This means that for each uncertainty source the individual up and down shifts to the nominal ratio of each measurement are taken and combined as the weighted mean using the statistical uncertainty just as for the nominal mean. Consequently, the difference of each systematically shifted weighted mean to the nominal weighted mean is the respective systematic uncertainty for the combination.

The  $|\eta|$  region covering the most forward region of the detector from 3.2–5.0, i. e. the hadronic forward, behaves a bit special compared to the others. In contrast to the other  $|\eta|$  regions, the individual measurements from PYTHIA and HERWIG++ have a very large spread of ratios from 0.83 up to 1.32 in this particular interval. If those are combined for this interval, it is in this case reasonable to use all six individual measurements from PYTHIA as well as from HERWIG++ to find a result which covers all of these partly very different ratios. If the combination is done following the same procedure as described for the other intervals considering all nominal values obtained with PYTHIA and HERWIG++ while taking the systematic uncertainty from the combined PYTHIA results, the total scale factor would be  $1.056 \pm 0.048$  (stat. unc.)  $\pm 0.079$  (syst. unc.) ( $= 0.092$  total unc.). In order to test the consistency of this combined value with the single measurements the  $\chi^2/\text{ndf}$  is calculated. This results in  $\chi^2/\text{ndf} = 4.3$ . This shows that the obtained total uncertainty is too small to reasonably cover all individual results. Hence, the systematic uncertainty is increased to a value of  $\pm 0.185$  which gives a total uncertainty of  $\pm 0.191$  and results in a  $\chi^2/\text{ndf} = 1.0$ , such that the single measurements are well represented by this combined result and the individual shifts are covered by the assigned uncertainty.

The obtained data-to-simulation ratios for the various  $|\eta|$  regions are summarized in Tab. 6.6, together with the statistical and systematic uncertainties. In addition, Fig. 6.18 shows the measured ratios together with the total uncertainty. This is given by the

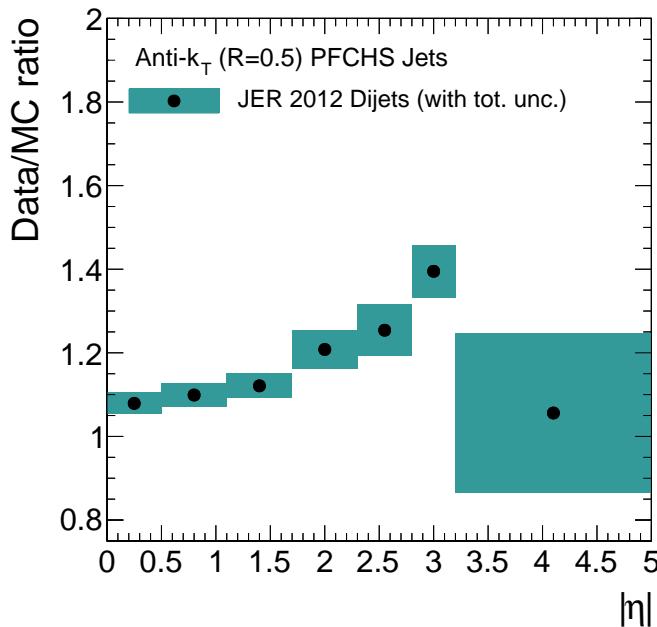


Figure 6.18: Measured data-to-simulation ratio in various  $|\eta|$  regions displayed with total uncertainty.

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 inhomogeneities 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  $\sqrt{s} = 7$  TeV have obtained similar results showing data-to-simulation ratios greater than one. A complementary approach is followed using  $\gamma +$  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  $\sqrt{s} = 7$  TeV [144] and results for  $\gamma +$  jet measurements at  $\sqrt{s} = 8$  TeV [147] are shown together in Fig. 6.19.

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

In comparison to the  $\sqrt{s} = 7$  TeV results from dijet events, the total uncertainty could be significantly reduced by incorporating the correlation among the different inclusive  $\alpha$  regions in the extrapolation procedure. In the previous analyses, the statistical uncertainties were underestimated by not considering the above mentioned correlation. This

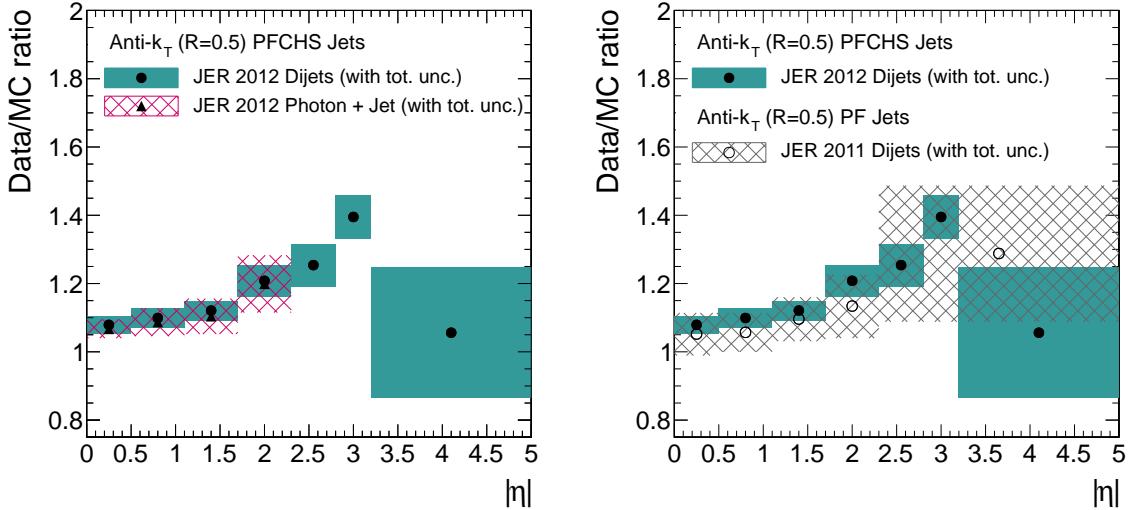


Figure 6.19: Measured data-to-simulation ratio in various  $|\eta|$  regions displayed with total uncertainty. Comparison to results obtained for  $\sqrt{s} = 8$  TeV data from  $\gamma + \text{jet}$  events (left) and to  $\sqrt{s} = 7$  TeV data from dijet events (right).

effect was compensated by introducing a rather conservative systematic uncertainty on the extrapolation to zero additional jet activity. Since the treatment of the statistical uncertainty has been changed in this analysis, the estimation of systematic effects could be adjusted accordingly and lead to the overall reduced uncertainty values summarized in Table 6.6.

### 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  $|\eta|$  and denote difference data and simulation, 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. Closed and open symbols represent two different methods used to derive the L2 residual correction. Especially in the forward region of the detector, the uncertainty is reduced by about 50%.

Furthermore, the improved jet resolution measurement has 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  $m_t = 172.04 \pm 0.19$  (stat. + JSF)  $\pm 0.75$  (syst.) GeV [161]. The propagated

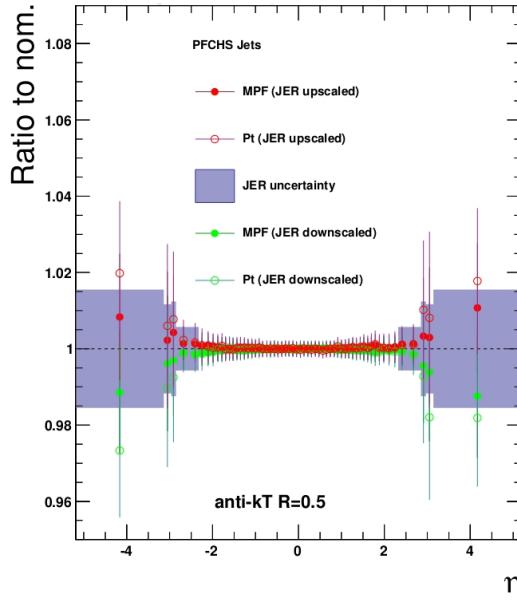


Figure 6.20: Impact of the JER uncertainty on the L2 residual correction. The variation of the correction with respect to the nominal correction factor is shown as a function of the pseudorapidity [156]. For further explanation see text.

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 to 0.71 GeV. 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 [162].

## 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 already in Eq. 6.18.

Alternatively, the resolution in simulation can be adjusted by convolution with a Gaussian. Taking a Gaussian function with appropriate width  $\sigma_c$ , the width of the response function

after convolution corresponds to the resolution in data according to

$$\sigma_c = \sqrt{c^2 - 1} \cdot \left( \frac{\sigma_{MC}}{p_T} \right) . \quad (6.27)$$

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

Further discussion about the adjustment of the resolution in simulation to data can be found in Section 7.3 where this is actually applied in the estimation of background contributions arising from QCD multijet events to a search for new physics in an all-jet final state.



## 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. 2.3.2, searches performed at  $\sqrt{s} = 7$  TeV have constrained the allowed parameter space for light-flavour squarks and gluinos already up to around 800 GeV and 1 TeV for light LSP masses, respectively. However, the increased centre of mass energy from 7 TeV to 8 TeV and the recorded dataset, which is around four times larger than at 7 TeV, provide the opportunity to extend the reach of such searches into entirely unexplored parameter regions. In Fig. 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  $20 \text{ fb}^{-1}$  of  $pp$  collision data at the LHC at  $\sqrt{s} = 8$  TeV.

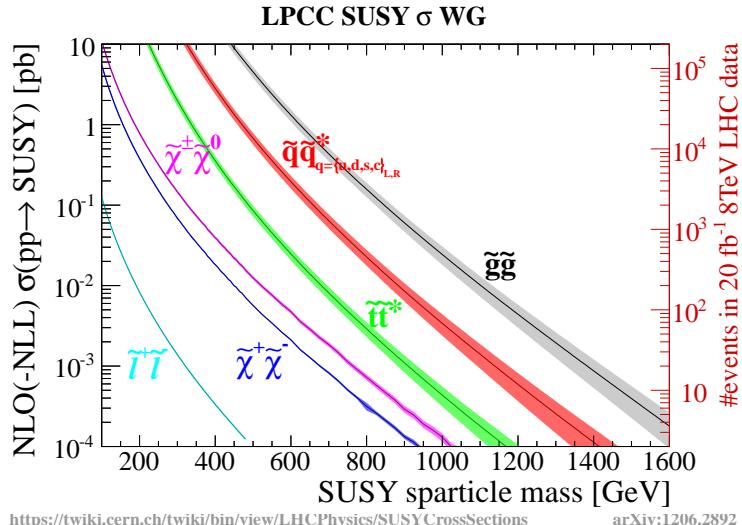


Figure 7.1: Theory cross sections for selected SUSY processes as a function of the sparticle mass. The  $y$ -axis on the right indicates the expected number of events in  $20 \text{ fb}^{-1}$  of  $pp$  collision data at the LHC at  $\sqrt{s} = 8$  TeV [69].

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. 2.3.2, the expected experimental signature contains several hard jets and a certain amount of missing transverse energy. Thus, events are selected based on the number of jets ( $N_{\text{Jets}}$ ), the scalar sum of the jet transverse momenta ( $H_T$ ) and the missing transverse momentum calculated from the jet momenta ( $\cancel{H}_T$ ). 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). Finally, results are presented and interpreted in various simplified supersymmetric models (Sec. 7.4). Parts of this chapter are taken from [163], written by the author. This analysis follows previous inclusive searches [73, 74] and is published in [164].

## 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 [110] generator at leading order and are interfaced with the parton-shower model in PYTHIA 6.4.24 [108]. They are scaled to cross-section predictions at next-to-leading order or next-to-next-to-leading order, when available [165, 166]. The events are processed with the full detector simulation.

Furthermore, SUSY signal samples are obtained from simulation. They are generated with MADGRAPH [110] (with up to two additional partons), the CTEQ6L parton distribution functions [150] and processed with the fast detector simulation. Cross sections are determined at NLO with a resummation of soft gluon emission at the accuracy of next-to-leading-log [69, 167–171]. 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  $\cancel{E}_T$ , the missing transverse energy. An overview of the considered runs and the integrated luminosity is shown together with the respective HLT trigger paths in Tab. 7.1.  $H_T$  and  $\cancel{E}_T$  are calculated from particle-flow objects at trigger level with nominal thresholds of  $350\text{ GeV}$  and  $100\text{ GeV}$ , 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  $\cancel{H}_T$  (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. 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

Table 7.1: Signal trigger paths used in different run ranges listed together with the integrated luminosity.

| Trigger path             | Run range     | Luminosity [fb <sup>-1</sup> ] |
|--------------------------|---------------|--------------------------------|
| 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                            |

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 paths providing enough statistical precision are 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 ( $3 \leq N_{\text{Jets}} \leq 5$ ,  $6 \leq N_{\text{Jets}} \leq 7$  and  $N_{\text{Jets}} \geq 8$ ). The obtained trigger turn-on curves for the two different HLT paths are shown as a function of  $H_T$  and  $\cancel{H}_T$ , as used in the analysis, for jet multiplicity  $3 \leq N_{\text{Jets}} \leq 5$  in Fig. 7.2. The respective turn-on curves for jet multiplicities  $6 \leq N_{\text{Jets}} \leq 7$  and  $N_{\text{Jets}} \geq 8$  are shown in App. B.1 and App. B.2. For both trigger paths, the efficiency plateau is reached around values of  $H_T = 500$  GeV and  $\cancel{H}_T = 200$  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  $N_{\text{Jets}} \geq 8$  only few events were selected such that statistical uncertainties are few 10% large. Though, no hints for a systematic inefficiency have been observed and the signal triggers are considered as fully efficient with an uncertainty of 2% for values of  $H_T > 500$  GeV and  $\cancel{H}_T > 200$  GeV independent of the jet multiplicity.

Table 7.2: Summary of total trigger efficiencies of the signal triggers for selections of  $H_T > 500$  GeV and  $\cancel{H}_T > 200$  GeV in different jet multiplicity intervals determined with respect to the reference trigger path.

| $N_{\text{Jets}}$ | HLT_PFHT350_PFMET100   | HLT_PFNopUHT350_PFMET100 |
|-------------------|------------------------|--------------------------|
| 3–5               | $99.4^{+0.2}_{-0.3}$   | $99.8^{+0.1}_{-0.1}$     |
| 6–7               | $99.1^{+0.7}_{-2.0}$   | $100.0^{+0.0}_{-0.6}$    |
| $\geq 8$          | $100.0^{+0.0}_{-36.9}$ | $100.0^{+0.0}_{-10.9}$   |

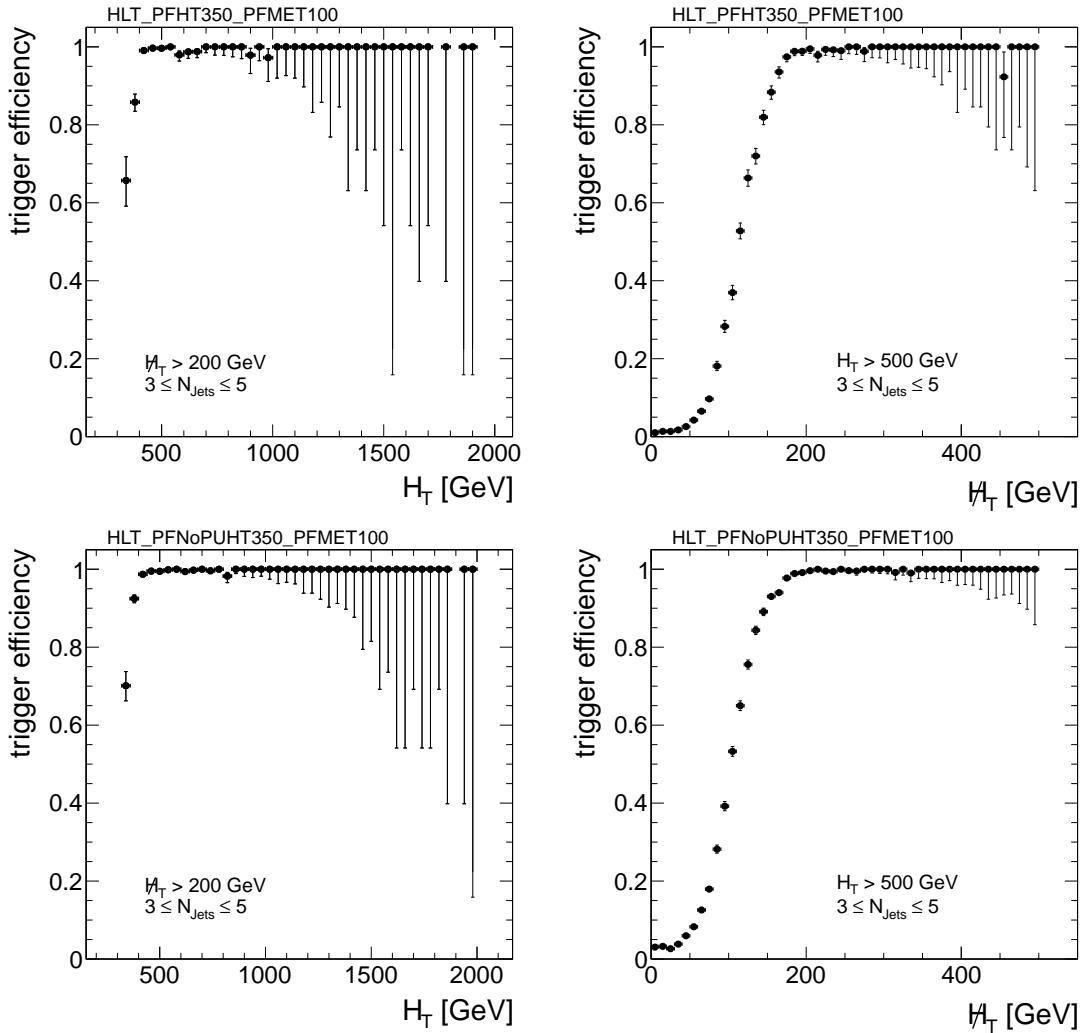


Figure 7.2: Measured relative trigger efficiency for paths HLT\_PFHT350\_PFMET100 (*top*) and HLT\_PFNNoPUHT350\_PFMET100 (*bottom*) as a function of  $H_T$  (*left*) and  $\cancel{H}_T$  (*right*) shown for  $3 \leq N_{\text{jets}} \leq 5$ .

### 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 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 artificial energy deposits at random times. Such events are identified based on timing and pulse-shape information (*HBHE noise filter*). Furthermore, two  $5 \times 5$  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  $\cancel{H}_T$  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- $\cancel{H}_T$ . 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 from detector noise, events are discarded, if the energy of a jet with  $p_T > 30 \text{ GeV}$  is composed of more than

95% from PF photon candidates or more than 90% from PF neutral hadron energy (*PBNR filter*).

#### 7.1.4 Baseline Selection

The physics objects used in the analysis are reconstructed with the PF algorithm. Jets are clustered from the particle-flow objects with the anti- $k_T$  algorithm using  $R = 0.5$ . Furthermore, they are calibrated, as discussed in Sec. 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 ( $N_{\text{Jets}}$ ) is required to be  $\geq 3$ .  $N_{\text{Jets}}$  is defined as the number of jets with  $p_T > 50 \text{ GeV}$  and  $|\eta| < 2.5$ . This requirement is imposed in order to select multijet events.
- The scalar sum of jet momenta ( $H_T$ ) is required to be  $\geq 500 \text{ GeV}$  with

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

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

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

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

for all jets with  $p_T > 30 \text{ GeV}$  and  $|\eta| < 5.0$ . This selection reduces contributions from standard model processes where missing transverse momentum is expected to be small. 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  $\cancel{H}_T$  is not aligned with any of the leading three jets and events with

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

are selected. The value of 0.5 is chosen according to the jet 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  $W + \text{jets}$  events are reduced by rejecting events containing isolated electrons or muons with  $p_T > 10 \text{ GeV}$ . These are required to have a good quality track that can be associated with the primary interaction vertex [172, 173]. 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  $H_T$ ,  $\cancel{H}_T$  and  $N_{\text{jets}}$  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:  $3 \leq N_{\text{jets}} \leq 5$ ,  $3 \leq N_{\text{jets}} \leq 5$ ,  $N_{\text{jets}} \geq 8$ . 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 [174–177].

In order to gain sensitivity to a variety of models, the jet categories are further classified according to  $H_T$  and  $\cancel{H}_T$ . With this approach, various exclusive search regions are defined. An overview of the exact definition of all 36 exclusive regions in  $N_{\text{jets}}$ ,  $H_T$  and  $\cancel{H}_T$  is given in Tab. 7.3.

## 7.2 Estimation of Non-QCD Backgrounds

As discussed in Sec. 2.3.2, events from the SM processes  $Z(\nu\bar{\nu}) + \text{jets}$ ,  $W + \text{jets}$  or semi-leptonic  $t\bar{t}$  events (in which either the lepton is lost or a hadronically decaying  $\tau$  lepton) and mismeasured QCD multijet events constitute important background contributions to all-hadronic final states. In this analysis, dedicated data-based methods are employed to estimate their size in the selected data.

In this section, those data-based background estimation methods are described. First, the estimation of non-QCD backgrounds (Sec. 7.2.1, Sec. 7.2.2 and Sec. 7.2.3) is reviewed. This is followed by a detailed discussion of the QCD background estimation (Sec. 7.3).

### 7.2.1 Invisible Z Background

The irreducible background contributions arising from  $Z(\nu\bar{\nu}) + \text{jets}$  events are estimated using  $\gamma + \text{jets}$  events. This is a well suited method as for high transverse momenta of the

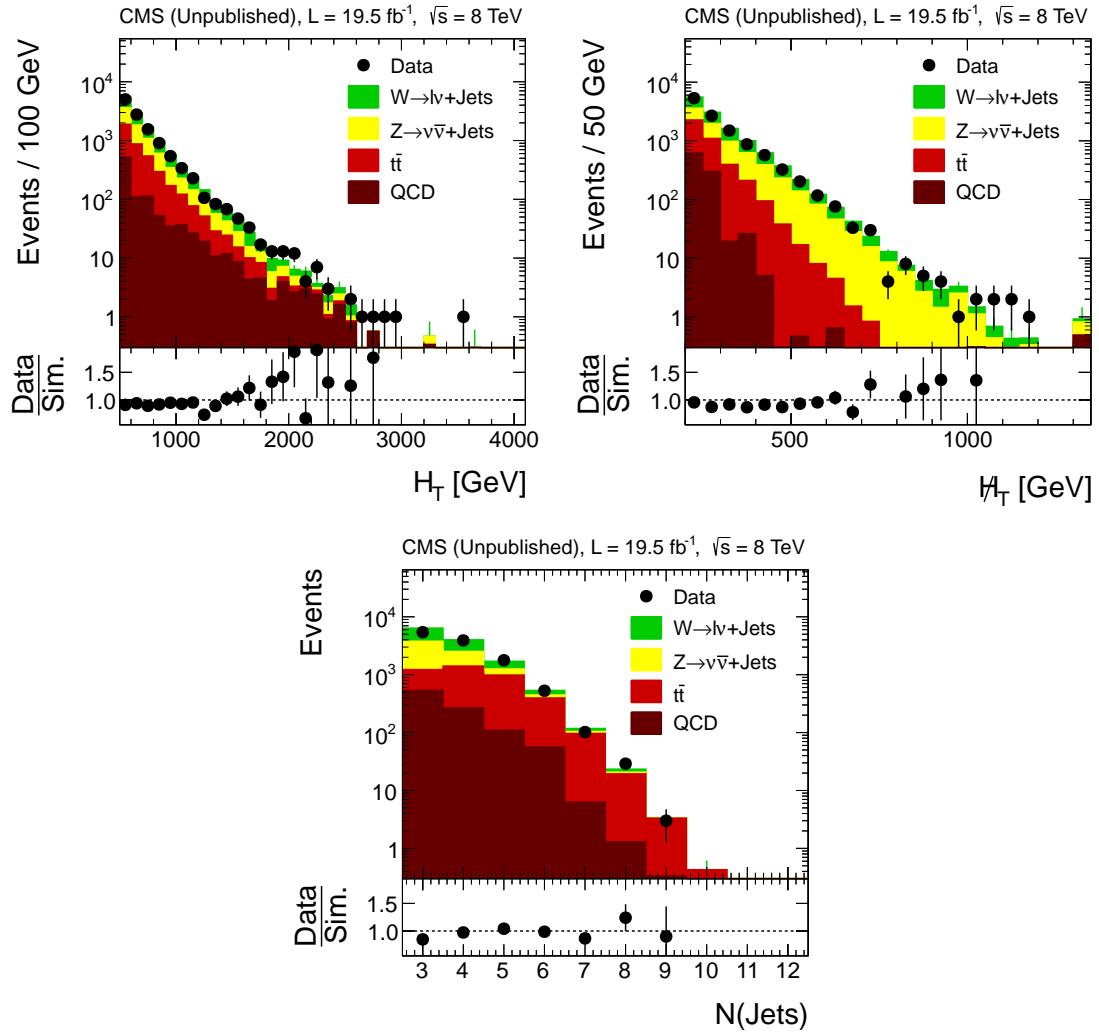


Figure 7.3: Comparison of selected  $H_T$  (top left),  $\H_T$  (top right) and  $N_{\text{Jets}}$  (bottom) distributions in data (black dots) and simulated events (shaded curve) found from applying the event cleaning and baseline selection criteria described in Sec. 7.1.3 and 7.1.4. Only statistical uncertainties are shown. Taken from [164].

vector boson the event kinematics are basically the same and the cross sections differ only according to the different couplings [178, 179].

The  $\gamma +$  jets sample is collected by triggering on a  $\gamma$  candidate and large values of  $H_T$ . Photon candidates are selected, if they satisfy  $p_T > 100 \text{ GeV}$  and  $|\eta| < 1.44$  or  $1.566 < |\eta| < 2.5$ . Furthermore, they have to have a shower profile consistent with that of a prompt photon produced directly in the hard interaction. In order to distinguish photons from misidentified electrons they must not have an associated track in the pixel detector. In addition, photon candidates are required to be isolated meaning that in a cone of 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

Table 7.3: Exclusive search regions used in the analysis binned in  $H_T$ ,  $\cancel{H}_T$  and  $N_{\text{Jets}}$ .

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

efficiency. Furthermore, the purity of the  $\gamma + \text{jets}$  sample, which is the fraction of selected photon candidates emerging from direct production, has to be taken into account. The number of background photons caused e. g. by misidentified jet fragments, is estimated by exploiting the difference between the shower profile of prompt and background photons. The average purity of the  $\gamma + \text{jets}$  sample is measured to be 93%.

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

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

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

The behaviour of  $R_{Z/\gamma}$  is examined in simulated events as function of  $\cancel{H}_T$ ,  $H_T$  and  $N_{\text{Jets}}$ . The obtained distributions are shown in Fig. 7.4(a) – (c). While a strong dependence on  $\cancel{H}_T$  for small values ( $\lesssim 500$  GeV) is observed, the variation as function of  $H_T$  amounts

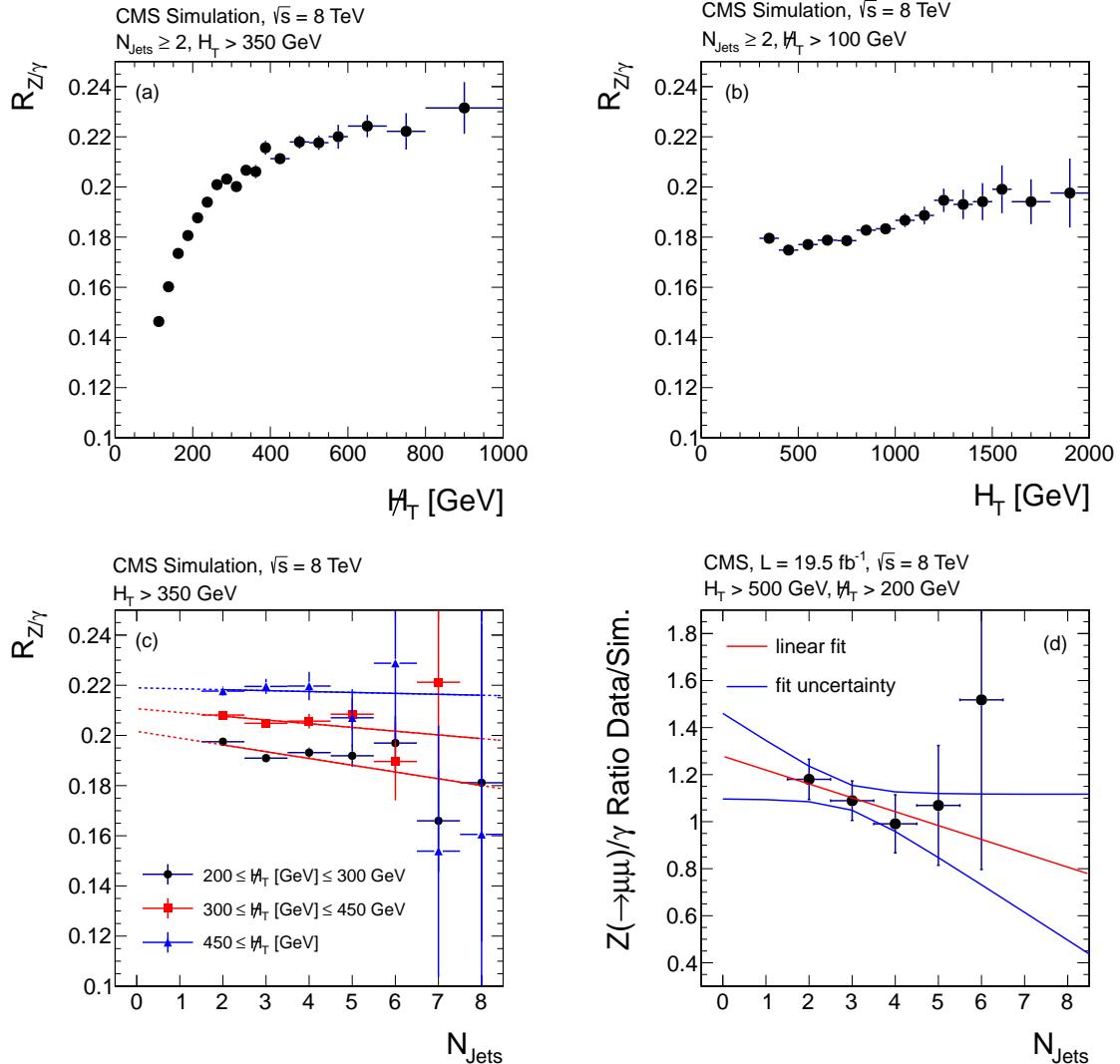


Figure 7.4: The simulated ratio  $R_{Z/\gamma}$  as a function of (a)  $\cancel{H}_T$ , (b)  $H_T$ , (c)  $N_{\text{jets}}$ , where the values for three  $\cancel{H}_T$  selections are shown with linear fits, and (d) the double ratio of  $R_{Z(\mu\mu)/\gamma}$ , using events from data to those from simulation; the linear fit and its uncertainty band are overlaid [164].

to only  $(12 \pm 5)\%$  in the relevant range for this analysis of  $500 < H_T < 1500$  GeV. The ratio as function of the jet multiplicity is determined for different  $\cancel{H}_T$  ranges of  $200 < / H_T < 300$  GeV,  $300 < \cancel{H}_T < 450$  GeV and  $\cancel{H}_T > 450$  GeV. The behaviour in each of these  $\cancel{H}_T$  ranges is described by a linear function also displayed in Fig. 7.4 (c). It is found that the ratio decreases slightly with increasing jet multiplicity which is consistent with findings from theory [179, 180].

In order to take the theoretical uncertainty on  $R_{Z/\gamma}$  into account this phenomenological ratio is also determined for  $Z(\mu\mu)$  events in data and simulation, respectively. The double ratio of  $R_{Z(\mu\mu)/\gamma}^{\text{data}}$  to  $R_{Z(\mu\mu)/\gamma}^{\text{MC}}$  is derived as function of the jet multiplicity and shown in Fig. 7.2.1 (d). It is fitted with a linear function, and the deviation from unity is considered

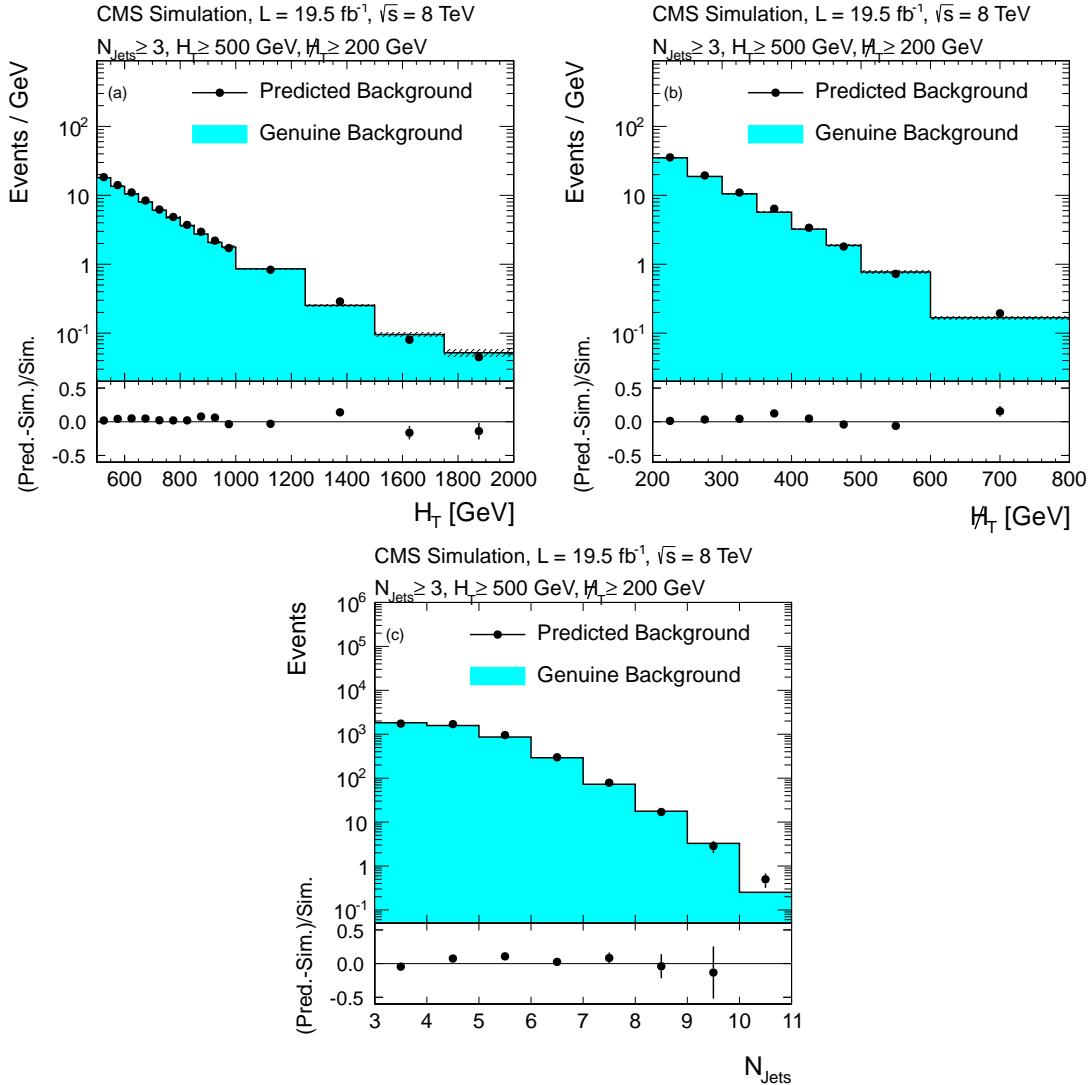


Figure 7.5: Predicted (a)  $H_T$ , (b)  $\cancel{H}_T$ , and (c)  $N_{\text{Jets}}$  distributions found from applying the hadronic-tau background evaluation method to simulated  $t\bar{t}$  and  $W + \text{jets}$  events (solid points) in comparison to the genuine  $t\bar{t}$  and  $W + \text{jets}$  background from simulation (shaded curve). Only statistical uncertainties are shown [164].

as correction for the  $R_{Z/\gamma}$  ratio in each jet multiplicity selection.

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

### 7.2.2 Hadronic-tau Background

Background contributions arising from  $W + \text{jets}$  and  $t\bar{t}$  events with a hadronically decaying  $\tau$  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  $\mu$  and a  $\tau_h$ .

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  $\mu$  with  $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.1$  is required. In order to prevent the control sample from signal contamination, a selection on the transverse mass  $m_T = \sqrt{2p_T^\mu E_T [1 - \cos(\Delta\phi)]}$  of  $m_T \leq 100 \text{ GeV}$  is imposed with the azimuthal angle  $\Delta\phi$  between the direction of the muon four-momentum and the  $E_T$  vector.

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

In the following,  $H_T$ ,  $\cancel{H}_T$  and  $N_{\text{Jets}}$  are recalculated for each event, including the transverse momentum of the  $\tau_h$  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  $W \rightarrow \tau_h\nu$  to  $W \rightarrow \mu\nu$  events. The statistical uncertainty of the background prediction is estimated with pseudo-experiments using a bootstrap method [181].

The validity of this background estimation procedure is tested by comparing the event yields obtained from applying the prediction method to simulated events from  $t\bar{t}$  and  $W + \text{jets}$  events to the respective genuine background obtained from simulation. This comparison is shown as a function of  $H_T$ ,  $\cancel{H}_T$  and  $N_{\text{Jets}}$  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  $3 \leq N_{\text{Jets}} \leq 5$  and 20% for jet multiplicities  $6 \leq N_{\text{Jets}} \leq 7$  and  $N_{\text{Jets}} \geq 8$ . These uncertainties mainly reflect the statistical precision of the validation test.

Further systematic uncertainties taken into account for the hadronic-tau background prediction cover differences between data and MC for the muon isolation and reconstruction efficiencies as well as uncertainties on the kinematic and geometric acceptance, the  $\tau_h$  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  $\mu + \text{jets}$  control sample. This is selected with the same trigger as used as signal trigger for the search. The sample is

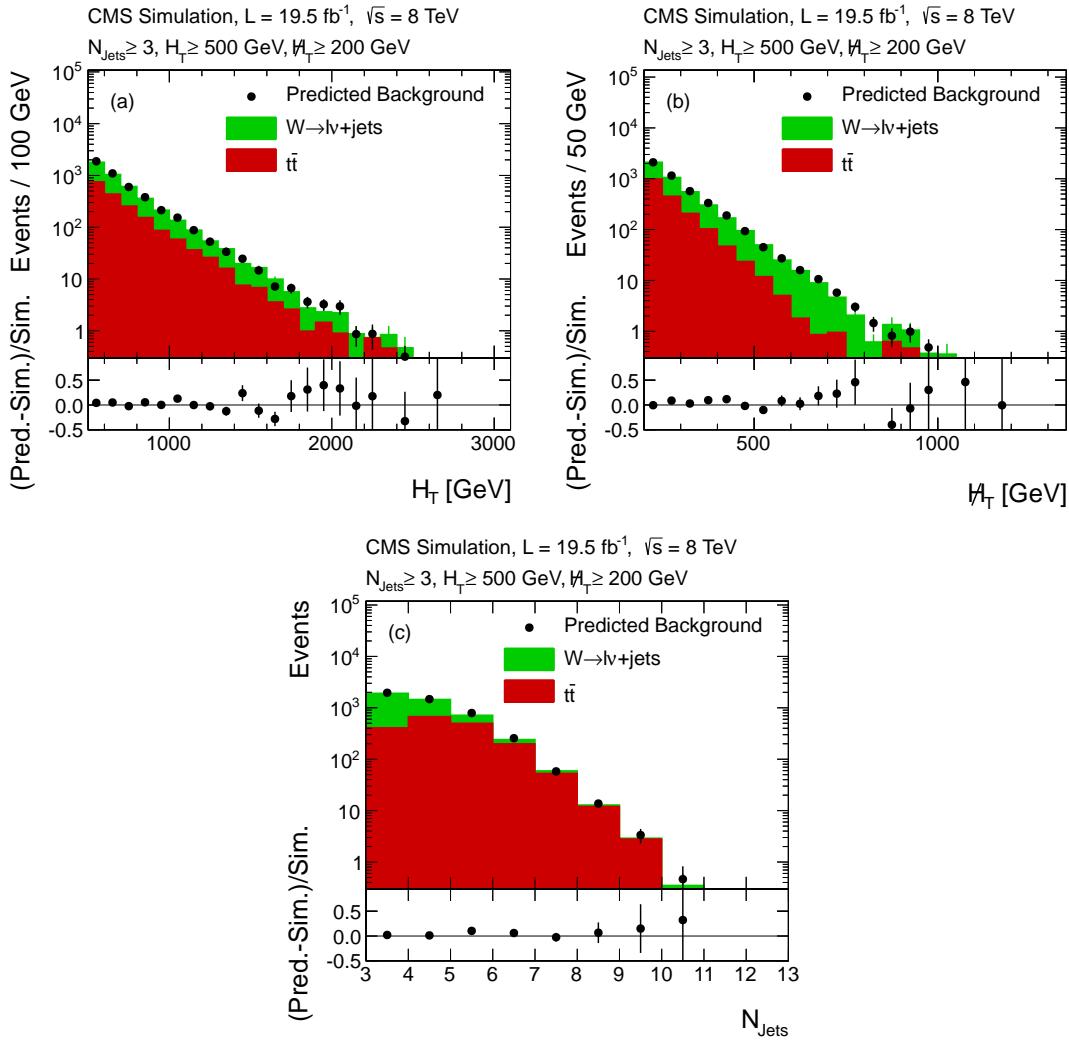


Figure 7.6: Predicted (a)  $H_T$ , (b)  $\cancel{H}_T$ , and (c)  $N_{\text{jets}}$  distributions found from applying the lost lepton background evaluation method to simulated  $t\bar{t}$  and  $W + \text{jets}$  events (solid points) in comparison to the genuine  $t\bar{t}$  and  $W + \text{jets}$  background from simulation (shaded curves). Only statistical uncertainties are shown [164].

selected by requiring exactly one well-reconstructed and isolated muon with  $p_T > 10 \text{ GeV}$ . Furthermore, the same transverse mass requirement of  $m_T < 100 \text{ GeV}$ , as for the hadronic tau background, is applied.

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

The number of events due to unidentified leptons is determined by weighting the events

in the control sample according to

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

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

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

The method is validated in simulation by comparing the predicted event yields for lost-lepton events in  $t\bar{t}$  and  $W + \text{jets}$  from a single-muon control sample after the baseline selection to the simulated genuine background. This comparison is illustrated in Fig. 7.6 as a function of  $H_T$ ,  $\cancel{H}_T$  and  $N_{\text{Jets}}$  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 all-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 [73, 74]. 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 multiplicity intervals at  $\sqrt{s} = 8 \text{ TeV}$ . After the discussion of the general concept of the R+S method in Sec. 7.3.1, the adjustment of the procedure to the actual conditions for 8 TeV data are introduced in Sec. 7.3.2. This is followed by a description of the validation procedure of the method in Sec. 7.3.3. Furthermore, the application of the R+S method to data is introduced in Sec. 7.3.4 and systematic uncertainties are discussed in Sec. 7.3.5. This section concludes with a presentation of the results of the QCD background prediction in Sec. 7.3.6.

#### 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

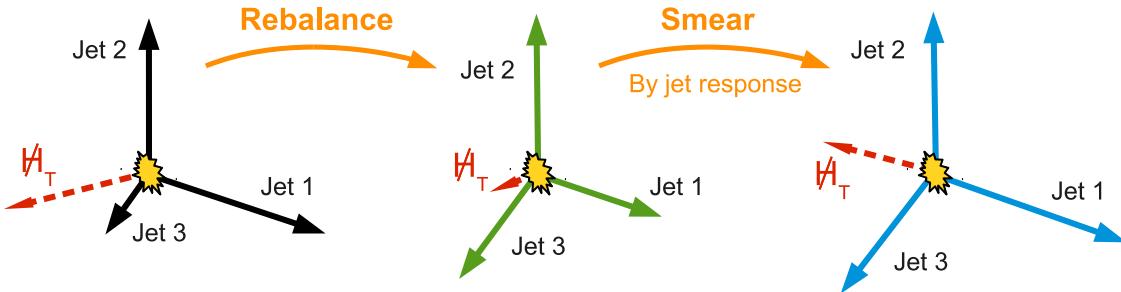


Figure 7.7: Outline of the two steps performed in the R+S method for estimation of QCD background events. Sketch from [144].

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. The resulting seed events reflect the event kinematics before the actual detector measurement is performed and are thus estimators of the true particle-level jet momenta. In a second step, all jets in the event are *smearied* 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  $\cancel{H}_T$ . 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  $p_T^{\text{gen}}$  and  $|\eta^{\text{gen}}|$  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  $p_T^{\text{gen}}$  and  $|\eta^{\text{gen}}|$  binning is chosen, the response is averaged over a certain part of the  $p_T^{\text{gen}}$  spectrum in each interval. Thus, the response distribution tends to overestimate the width of the resolution for high- $p_T$  jets while it behaves oppositely for low- $p_T$  jets.

The truth jet response is derived by performing an unambiguous 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 re-

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

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

constructed or generated jet  $j \neq i$  around a matched jet pair is vetoed in a cone of size  $R < 0.7$  by requiring

$$p_T^{\text{GenJet}_j} / p_T^{\text{GenJet}_i} < 0.05 \quad (7.1)$$

and

$$p_T^{\text{Jet}_j} < 30 \text{ GeV} \text{ and } p_T^{\text{Jet}_j} / p_T^{\text{Jet}_i} < 0.05. \quad (7.2)$$

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.

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  $\sigma_c$  according to Eq. 6.27. The considered correction factors are listed in Tab. B.1 in the appendix and correspond to the data-to-simulation ratios obtained from dijet data at  $\sqrt{s} = 7 \text{ TeV}$ , as illustrated in Fig. 6.19 (right).<sup>1</sup> The residual tail contributions are scaled according to the correction factors  $\rho_{\text{tail}}$  listed in Tab. B.2 in the appendix derived from dijet asymmetry parts fulfilling ( $A > 2\sigma_c$ ) [144].

### 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* [182].

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.

<sup>1</sup> 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.

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

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

with  $d\vec{y}^i = y_{\text{true}}^i - y_{\text{measured}}^i$  and covariance matrix  $C$  is minimised. In this particular case, the measured jet four-momenta correspond to the values of  $y_{\text{measured}}^i$  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  $p_T^{\text{gen}}$  for two example  $|\eta^{\text{gen}}|$  intervals in Fig. 7.8. Subsequently, the obtained resolution values are fitted with the function

$$\frac{\sigma_{\text{MC}}(p_T)}{p_T} = \sqrt{\text{sgn}(N) \cdot \left(\frac{N}{p_T}\right)^2 + S^2 \cdot p_T^{m-1} + C^2}, \quad (7.4)$$

with 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  $\text{sgn}(N)$  considers the improved momentum resolution at low  $p_T$  due to the employed tracking information. Since also at medium  $p_T$  the tracking information still compensates for non-linearities of the calorimeters, the parameter  $m$  is introduced. The fitted functions are illustrated in Fig. 7.8 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  $|\eta^{\text{gen}}|$  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  $(|\cancel{H}_T^x| + |\cancel{H}_T^y| < 0.02 \text{ GeV})$  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  $\cancel{H}_T$ , since selections suppressing high tails of missing energy often tend to bias the QCD kinematics. Typically,  $H_T$  and  $\cancel{H}_T$  are correlated quantities in QCD events, as high values of  $\cancel{H}_T$  caused by severe jet mismeasurements can only occur, if there is a certain amount of energy in the event. Consequently, selection cuts removing high- $\cancel{H}_T$  tails also remove parts of the  $H_T$  spectrum which results in an overall underestimation of

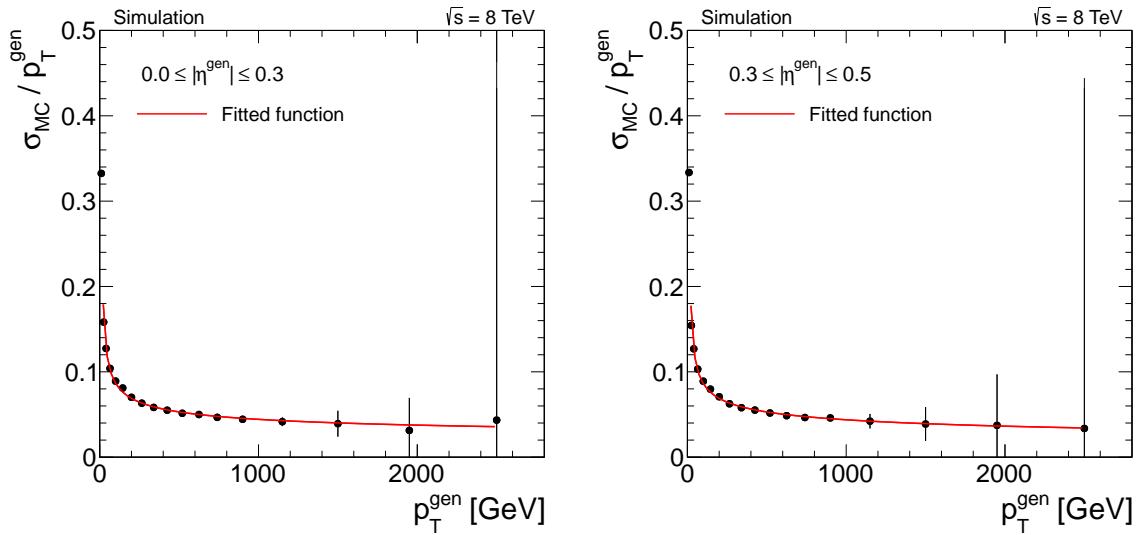


Figure 7.8: Relative truth- $p_T$  resolution derived from simulated events shown as a function of  $p_T^{\text{gen}}$ . The distribution is fitted with a function as described in the text used as input for the kinematic fit employed to gain a balanced seed sample.

QCD contributions to the high  $H_T$  and  $\cancel{H}_T$  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 pre-selection at

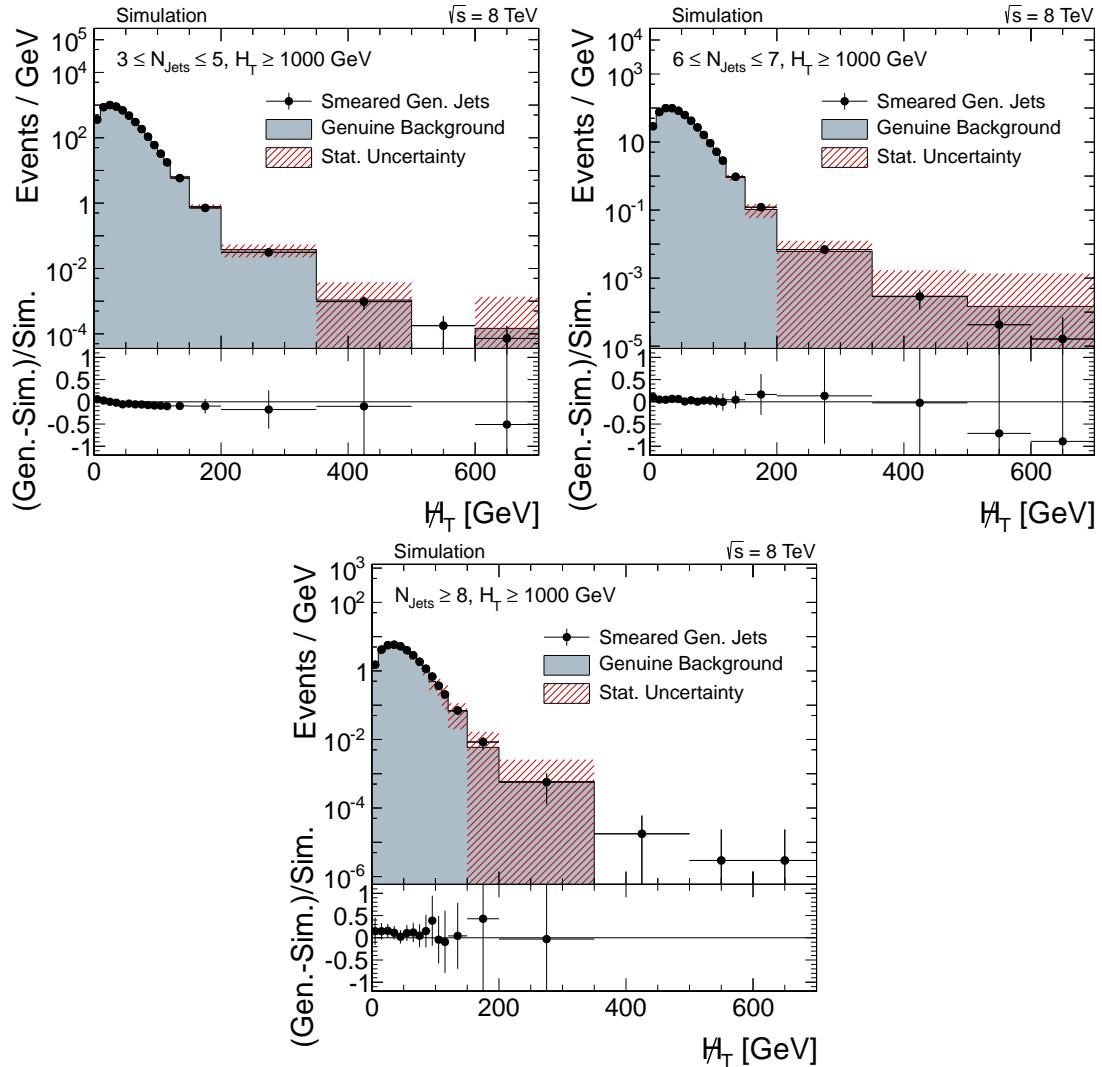


Figure 7.9: Particle-level jets obtained from a QCD multijet sample generated with MADGRAPH smeared with the truth response templates are compared to the expectation from full simulation. This comparison is shown for search regions with non-negligible QCD multijet background contributions as a function of  $H_T$  defined by  $H_T \geq 1000 \text{ GeV}$  for  $3 \leq N_{\text{jets}} \leq 5$  (top left),  $6 \leq N_{\text{jets}} \leq 7$  (top right) and  $8 \leq N_{\text{jets}}$  (bottom) after the application of the minimum  $\Delta\phi$  selection.

detector-level of  $N_{\text{jets}} \geq 2$  and  $H_T > 350 \text{ GeV}$ .<sup>2</sup> The result of this generator-jet smearing is shown in Fig. 7.9. 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 *non-closure* of the method and the respective tests of the agree-

<sup>2</sup>The same pre-selection 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.

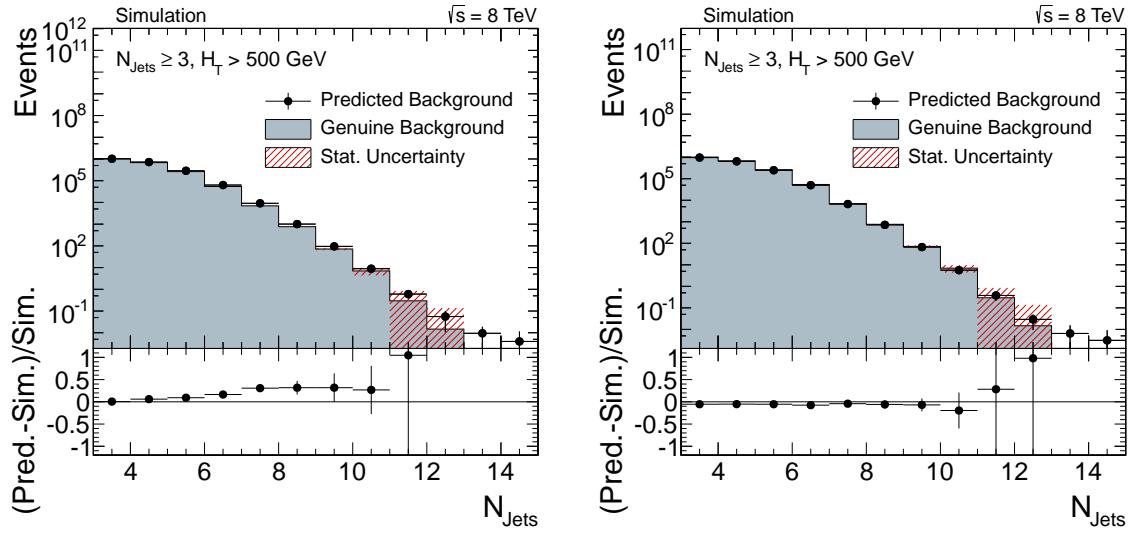


Figure 7.10: Prediction of QCD background on a QCD multijet sample generated with MADGRAPH compared to the expectation from full simulation with a  $p_T$  cut of 10 GeV imposed on jets considered in the rebalancing. The predicted  $N_{\text{jets}}$  distribution is obtained using the R+S method as in 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*).

ment *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. 7.3.2) and is assigned as systematic uncertainty to the QCD background prediction (see Sec. 7.3.5).

### 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 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  $p_T > 10$  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  $\cancel{H}_T$  distributions for an inclusive jet multiplicity selection of  $N_{\text{jets}} \geq 3$ . Since in the analysis presented in this thesis, the search regions have been extended to several jet multiplicity intervals, it is of particular interest to study, if the R+S method is able to correctly predict the jet multiplicity as well.

A comparison of the predicted  $N_{\text{jets}}$  distribution with the R+S method, using a  $p_T$  cut of

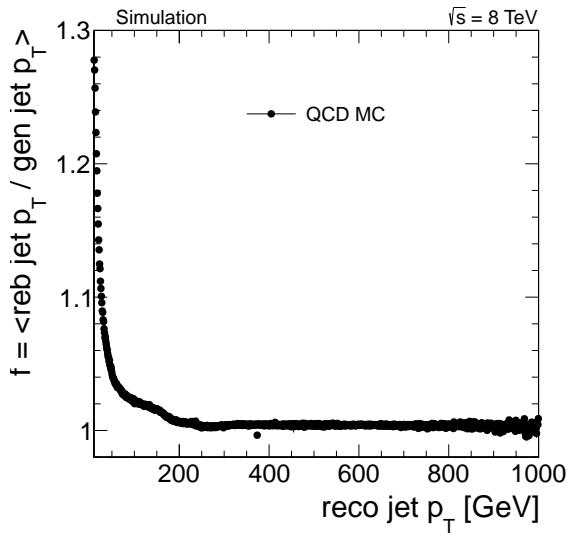


Figure 7.11: Mean of transverse momenta of rebalanced jets divided by transverse momenta of matched generated jets as a function of transverse momentum of matched reconstructed jets obtained from simulated events generated with MADGRAPH.

10 GeV for jets considered in the rebalancing, to the expected genuine background from simulation is shown in Fig. 7.10 (left) after imposing baseline requirements without the  $\cancel{H}_T$  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%. Since the smearing procedure has been shown in Sec. 7.3.1 to provide a reliable performance, 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. 7.11. It is observed that especially jets with small transverse momentum ( $p_T \leq 100$  GeV) are, on average, rebalanced to too high momenta. The observed too high momenta of rebalanced jets, thus result in more jets passing the  $N_{\text{Jets}}$  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 artificial additional imbalance in the event which has to be compensated for in the kinematic fit.

In order to account for this threshold effect, 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 overpredic-

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

tion. The prediction of the  $N_{\text{jets}}$  distribution with the adjusted R+S method compared to the expected background is shown in Fig. 7.10 (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 multi-jet 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 the results from full simulation, as explained in Sec. 7.3.2. The different closure tests as a function of  $\cancel{H}_T$  for various jet multiplicity bins for a low  $H_T = [500, 1000]$  GeV and a high  $H_T \geq 1000$  GeV selection are shown in Fig. 7.12 for simulated events obtained from MADGRAPH. In general, the prediction shows a good agreement with the expected QCD background contributions. Especially for the high  $H_T$  regions, in which QCD background is even more important than for lower  $H_T$  values, the agreement is reasonable. 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. 7.12.

The first choice for the determination of remaining biases is to calculate for the signal region, defined by  $\cancel{H}_T > 200$  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

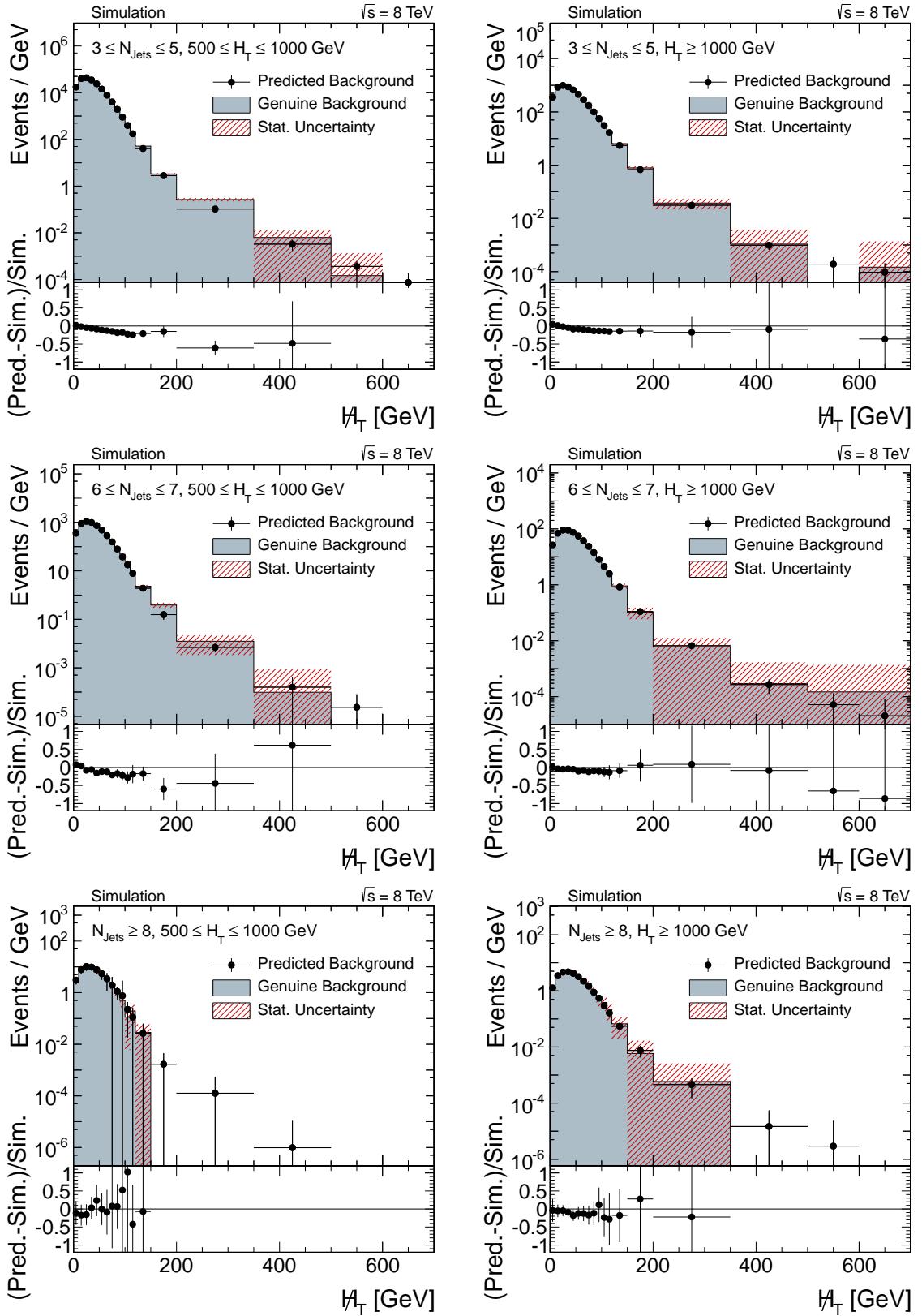


Figure 7.12: Prediction of QCD background on a QCD multijet sample generated with MADGRAPH compared to the expectation from full simulation. The closure test is shown for various jet multiplicity bins and low (*left*) or high (*right*)  $H_T$  selections.

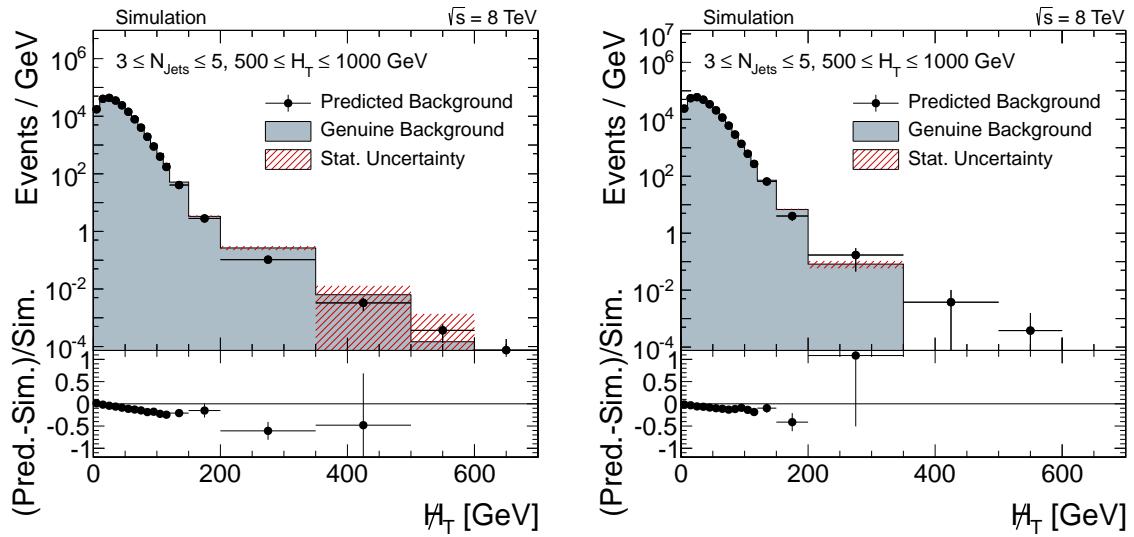


Figure 7.13: Prediction of QCD background on a simulated QCD multijet sample compared to the expectation from full simulation. The closure test is shown for  $3 \leq N_{\text{Jets}} \leq 5$  and  $500 \leq H_{\text{T}} \leq 1000$  GeV obtained from the MADGRAPH QCD sample (*left*) and alternatively for a QCD sample generated with PYTHIA QCD sample (*right*).

(first column) for the signal region. The table shows that there is one bin ( $3 \leq N_{\text{Jets}} \leq 5$  and  $500 \leq H_{\text{T}} \leq 1000$  GeV) in which the signal region shows a statistically significant non-closure. In Fig. 7.13, 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  $200 < H_{\text{T}} < 350$  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  $500 \leq H_{\text{T}} \leq 1000$  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  $H_{\text{T}} \geq 200$  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  $100 \leq H_{\text{T}} \leq 200$  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 is considered as systematic uncertainty.

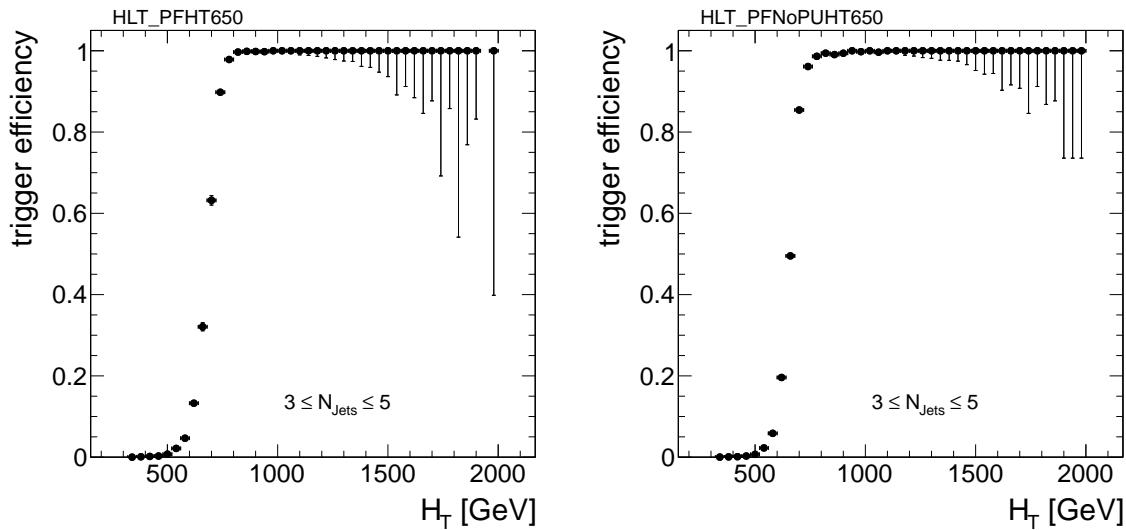


Figure 7.14: Measured trigger efficiency for paths HLT\_PFHT650 (*left*) and HLT\_PFNNoPUHT650 (*right*) as a function of  $H_T$ , shown for  $3 \leq N_{\text{jets}} \leq 5$ .

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

The uncertainty which is 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.

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

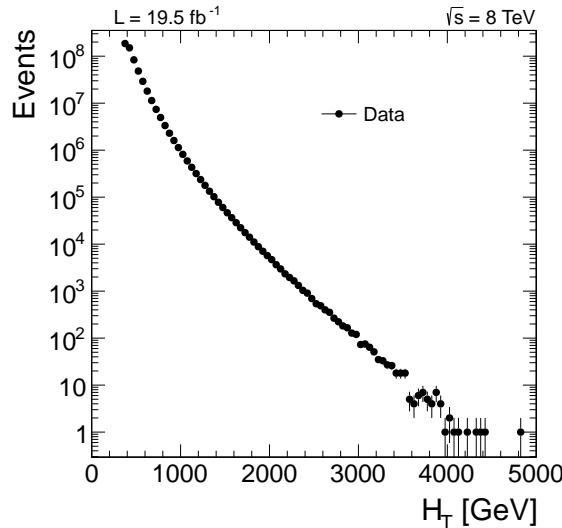


Figure 7.15: Seed  $H_T$  spectrum of full 2012 dataset used as input for the R+S method after correcting for trigger prescales.

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  $H_T = 650$  GeV are shown in Fig. 7.14 for a jet multiplicity of 3–5 and for jet multiplicities  $6 \leq N_{\text{Jets}} \leq 7$  and  $N_{\text{Jets}} \geq 8$  in App. B.5 and App. B.6, respectively. For these jet multiplicity selections, these two trigger paths are fully efficient for a  $H_T$  selection of 800 GeV.

Since two different triggers are used, each event of the multijet control sample has to be unambiguously 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 unambiguous 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. 7.15. The usage of prescaled triggers is important in order to collect also events with  $H_T$  values smaller than 500 GeV which enter the signal region through a fluctuation to large response values. Nevertheless, 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. This problem is solved by smearing the events obtained from the prescaled trigger not only  $N = 100$  times, but  $N = \text{prescale factor} \times 100$  times and weighting them accordingly with one.

The successful 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. This control sample is defined by at least three jets,  $H_T > 1000$  GeV, an inverted  $\Delta\phi$  criterion and  $100 \leq \cancel{H}_T \leq 200$  GeV. The resulting com-

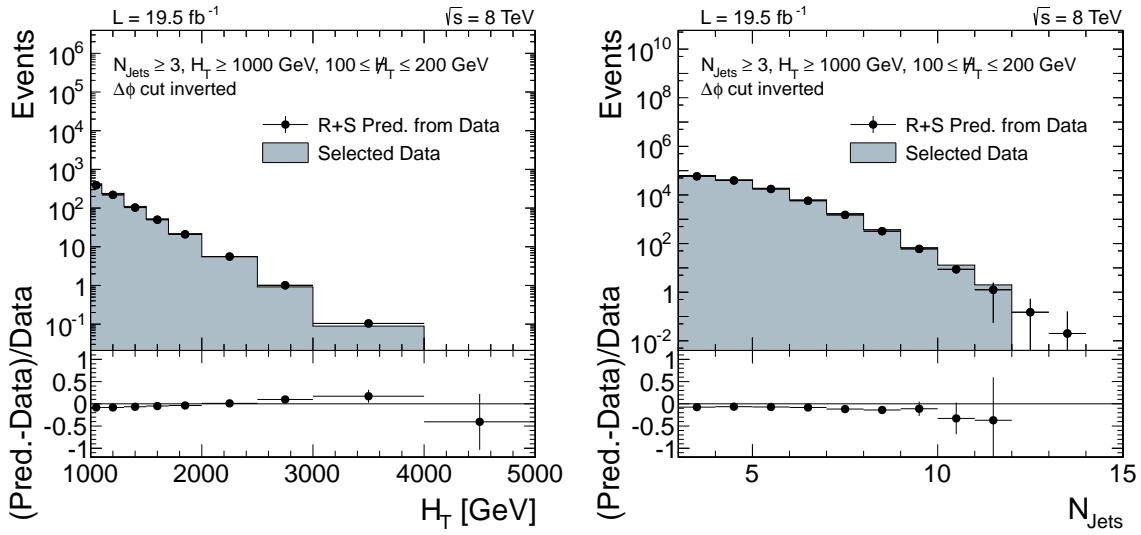


Figure 7.16: Prediction of QCD background on data compared to the expectation from data for a QCD enriched control region with  $H_T > 1000 \text{ GeV}$ , inverted  $\Delta\phi$  cut,  $N_{\text{Jets}} \geq 3$  and  $100 \leq H_T \leq 200 \text{ GeV}$ .

parison between predicted and selected data events in that region is shown in Fig. 7.16 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.

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. However, systematic uncertainties that have to be considered for the prediction of QCD background are discussed in the next section.

### 7.3.5 Systematic Uncertainties

#### Core of response functions

The uncertainties on the factors of the Gaussian core resolution accounting for the differences in data and Monte Carlo 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 due to 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

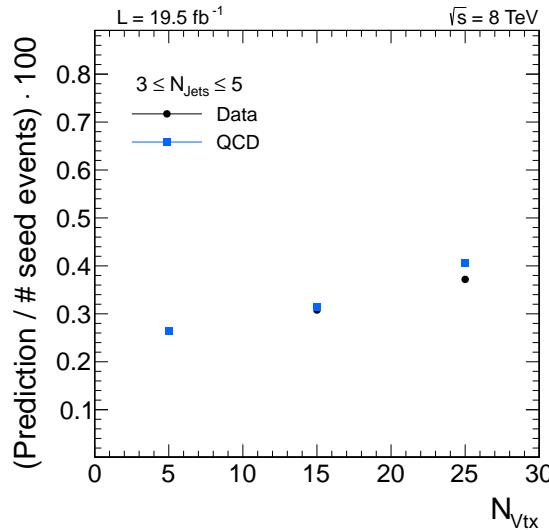


Figure 7.17: Prediction of QCD background in data and Monte Carlo as a function of primary vertices ( $N_{\text{Vtx}}$ ) normalized to the number of contributing seed events used for the determination of the pile-up uncertainty as described in the text.

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  $500 \leq H_{\text{T}} \leq 1000$  GeV and a high  $H_{\text{T}} > 1000$  GeV region.

### Pileup

In general, the R+S method is taking care of 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  $H_{\text{T}}$  search bins, as these are mainly populated due to heavily mismeasured hard jets.

The residual pileup influence is evaluated in a low  $H_{\text{T}}$  control region with  $H_{\text{T}} > 100$  GeV, since residual pileup effects are expected to contribute mostly in such low  $H_{\text{T}}$  regions. Pileup contributions to the high  $H_{\text{T}}$  signal regions might be even less, so that this approach provides a conservative estimate of the pileup uncertainty. Furthermore, the usage of a low  $H_{\text{T}}$  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  $N_{\text{Vtx}} = [0, 10]$ ,  $N_{\text{Vtx}} = [11, 20]$  and  $N_{\text{Vtx}} > 20$ . Then, the QCD prediction is calculated for each vertex bin when applying baseline selection criteria with the relaxed  $H_{\text{T}}$  requirement mentioned above. In order to compensate for the different number of seed events contributing to each  $N_{\text{Vtx}}$  interval, the prediction in each  $N_{\text{Vtx}}$  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 distribution obtained with this procedure is illustrated for jet multiplicity 3–5 in Fig. 7.17. From this distribution, 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, such that the pileup dependent part of the QCD prediction in data is obtained. 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.

### 7.3.6 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  $H_T$  and the highest jet multiplicity search bins.

However, search regions with non-negligible contributions from QCD background, like high  $H_T$  ( $\geq 1000 \text{ GeV}$ ) and low  $H_T$  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.

## 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. These have been obtained by adding statistical and systematic uncertainties in quadrature. Furthermore, the obtained yields in data and the predicted background are visualized in Fig. 7.18. 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  $6 \leq N_{\text{Jets}} \leq 7$ ,  $500 < H_T < 800 \text{ GeV}$

Table 7.6: Predicted event yields for the QCD background in the search regions defined by  $H_T$ ,  $\cancel{H}_T$  and  $N_{\text{Jets}}$  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.

| $N_{\text{Jets}}$ | $H_T$ [GeV] | $\cancel{H}_T$ [GeV] | Pred. | stat. unc. | Core [%]                | Tail [%]                | Bias [%]   | PU [%]     | syst. unc.            |
|-------------------|-------------|----------------------|-------|------------|-------------------------|-------------------------|------------|------------|-----------------------|
| 3–5               | 500–800     | 200–300              | 307.4 | $\pm 18.5$ | +13.0<br>−12.2<br>+7.3  | 36.0<br>−34.4<br>22.9   | $\pm 60.4$ | $\pm 2.9$  | +220<br>−217<br>+22.4 |
| 3–5               | 500–800     | 300–450              | 34.5  | $\pm 5.8$  | −10.5<br>−24.2          | −31.6<br>+37.9          | $\pm 60.4$ | $\pm 2.9$  | +23.8<br>−23.8        |
| 3–5               | 500–800     | 450–600              | 1.3   | $\pm 1.2$  | +24.2<br>−16.7          | −26.5<br>+55.6          | $\pm 60.4$ | $\pm 2.9$  | +1.0<br>−0.9          |
| 3–5               | 500–800     | > 600                | 0.1   | $\pm 0.3$  | +55.6<br>−55.6          | +55.6<br>−55.6          | $\pm 60.4$ | $\pm 2.9$  | +0.09<br>−0.09        |
| 3–5               | 800–1000    | 200–300              | 91.7  | $\pm 10.2$ | +14.7<br>−13.8<br>+5.2  | +33.2<br>−33.5<br>+29.8 | $\pm 60.4$ | $\pm 2.9$  | +64.7<br>64.7         |
| 3–5               | 800–1000    | 300–450              | 9.9   | $\pm 3.2$  | −5.1<br>+65.5           | −27.1<br>+65.5          | $\pm 60.4$ | $\pm 2.9$  | +6.7<br>−6.6          |
| 3–5               | 800–1000    | 450–600              | 0.8   | $\pm 0.9$  | +65.5<br>−65.5          | −65.5<br>+75.0          | $\pm 60.4$ | $\pm 2.9$  | +0.9<br>−0.8          |
| 3–5               | 800–1000    | > 600                | 0.1   | $\pm 0.4$  | +75.0<br>−41.7          | +8.3<br>−41.7           | $\pm 60.4$ | $\pm 2.9$  | +0.1<br>−0.1          |
| 3–5               | 1000–1250   | 200–300              | 59.0  | $\pm 7.2$  | +19.0<br>−14.6<br>+12.2 | +34.7<br>−31.7<br>+32.0 | $\pm 14.5$ | $\pm 2.9$  | +24.9<br>−22.4        |
| 3–5               | 1000–1250   | 300–450              | 5.1   | $\pm 2.2$  | −8.8<br>+35.3           | −16.9<br>+23.5          | $\pm 14.5$ | $\pm 2.9$  | +1.9<br>−1.2          |
| 3–5               | 1000–1250   | 450–600              | 0.5   | $\pm 0.7$  | +35.3<br>−3.9           | −5.9<br>+41.7           | $\pm 14.5$ | $\pm 2.9$  | +0.2<br>−0.1          |
| 3–5               | 1000–1250   | > 600                | 0.1   | $\pm 0.3$  | +41.7<br>−41.7          | +41.7<br>−41.7          | $\pm 14.5$ | $\pm 2.9$  | +0.1<br>−0.1          |
| 3–5               | 1250–1500   | 200–300              | 31.2  | $\pm 5.3$  | +18.3<br>−19.1<br>+16.3 | +30.3<br>−29.7<br>+38.3 | $\pm 14.5$ | $\pm 2.9$  | +12.0<br>11.9         |
| 3–5               | 1250–1500   | 300–450              | 2.3   | $\pm 1.3$  | −5.3<br>+0.0            | −30.4<br>+54.2          | $\pm 14.5$ | $\pm 2.9$  | +1.0<br>−0.8          |
| 3–5               | 1250–1500   | > 450                | 0.2   | $\pm 0.5$  | +0.0<br>−8.3            | +54.2<br>−8.3           | $\pm 14.5$ | $\pm 2.9$  | +0.1<br>−0.1          |
| 3–5               | > 1500      | 200–300              | 35.1  | $\pm 6.1$  | +19.6<br>−20.0<br>+39.9 | +23.3<br>−29.4<br>+39.9 | $\pm 14.5$ | $\pm 2.9$  | +11.9<br>13.5         |
| 3–5               | > 1500      | > 300                | 2.4   | $\pm 1.4$  | +39.9<br>−39.9          | +39.9<br>−39.9          | $\pm 14.5$ | $\pm 2.9$  | +1.4<br>−1.4          |
| 6–7               | 500–800     | 200–300              | 18.2  | $\pm 3.9$  | +8.9<br>−12.5<br>+31.9  | +37.4<br>−33.5<br>+31.9 | $\pm 25.4$ | $\pm 8.0$  | +8.5<br>−8.1          |
| 6–7               | 500–800     | 300–450              | 1.9   | $\pm 1.4$  | −31.9<br>+400.0         | −31.9<br>+400.0         | $\pm 25.4$ | $\pm 8.0$  | +1.0<br>−1.0          |
| 6–7               | 500–800     | > 450                | 0.01  | $\pm 0.1$  | +400.0<br>−100.0        | +400.0<br>−100.0        | $\pm 25.4$ | $\pm 8.0$  | +0.1<br>−0.01         |
| 6–7               | 800–1000    | 200–300              | 13.13 | $\pm 3.4$  | +15.0<br>−8.2<br>+5.1   | +33.7<br>−30.0<br>+30.8 | $\pm 25.4$ | $\pm 8.0$  | +6.0<br>−5.3          |
| 6–7               | 800–1000    | 300–450              | 2.0   | $\pm 1.1$  | −20.0<br>+46.7          | −28.2<br>+46.7          | $\pm 25.4$ | $\pm 8.0$  | +0.8<br>−0.9          |
| 6–7               | 800–1000    | > 450                | 0.2   | $\pm 0.4$  | +46.7<br>−46.7          | +46.7<br>−46.7          | $\pm 25.4$ | $\pm 8.0$  | +0.1<br>−0.1          |
| 6–7               | 1000–1250   | 200–300              | 11.9  | $\pm 3.8$  | +5.9<br>−12.7<br>+31.8  | +33.5<br>−35.8<br>+31.8 | $\pm 10.9$ | $\pm 8.0$  | +4.4<br>−4.8          |
| 6–7               | 1000–1250   | 300–450              | 1.5   | $\pm 1.3$  | −31.8<br>+100.0         | −31.8<br>+100.0         | $\pm 10.9$ | $\pm 8.0$  | +0.7<br>−0.7          |
| 6–7               | 1000–1250   | > 450                | 0.1   | $\pm 0.3$  | +100.0<br>−100.0        | +100.0<br>−100.0        | $\pm 10.9$ | $\pm 8.0$  | +0.2<br>−0.1          |
| 6–7               | 1250–1500   | 200–300              | 6.8   | $\pm 3.0$  | +12.0<br>−11.9<br>+54.4 | +32.6<br>−32.4<br>+54.4 | $\pm 10.9$ | $\pm 8.0$  | +2.5<br>−2.5          |
| 6–7               | 1250–1500   | 300–450              | 0.9   | $\pm 1.0$  | −54.4<br>+44.4          | −54.4<br>+44.4          | $\pm 10.9$ | $\pm 8.0$  | +0.7<br>−0.7          |
| 6–7               | 1250–1500   | > 450                | 0.09  | $\pm 0.3$  | +44.4<br>−44.4          | +44.4<br>−44.4          | $\pm 10.9$ | $\pm 8.0$  | +0.06<br>−0.06        |
| 6–7               | > 1500      | 200–300              | 8.0   | $\pm 2.8$  | +20.9<br>−15.4<br>+47.0 | +31.5<br>−25.8<br>+47.0 | $\pm 10.9$ | $\pm 8.0$  | +3.1<br>−2.6          |
| 6–7               | > 1500      | > 300                | 0.8   | $\pm 0.9$  | +47.0<br>−47.0          | +47.0<br>−47.0          | $\pm 10.9$ | $\pm 8.0$  | +0.6<br>−0.6          |
| $\geq 8$          | 500–800     | > 200                | 0.14  | $\pm 0.38$ | +71.4<br>−71.4<br>+33.3 | +71.4<br>−71.4<br>+33.3 | $\pm 86.0$ | $\pm 33.4$ | +0.19<br>−0.14        |
| $\geq 8$          | 800–1000    | > 200                | 0.54  | $\pm 0.69$ | −33.3<br>+19.2          | −33.3<br>+56.2          | $\pm 86.0$ | $\pm 33.4$ | +0.56<br>−0.54        |
| $\geq 8$          | 1000–1250   | > 200                | 0.73  | $\pm 0.78$ | −1.4<br>+55.6           | −27.4<br>+55.6          | $\pm 86.0$ | $\pm 33.4$ | +0.59<br>−0.44        |
| $\geq 8$          | 1250–1500   | > 200                | 0.54  | $\pm 0.75$ | −55.6<br>+65.2          | −55.6<br>+65.2          | $\pm 86.0$ | $\pm 33.4$ | +0.52<br>−0.52        |
| $\geq 8$          | > 1500      | > 200                | 0.89  | $\pm 0.94$ | +65.2<br>−65.2          | +65.2<br>−65.2          | $\pm 86.0$ | $\pm 33.4$ | +0.95<br>−0.89        |

Table 7.7: Predicted event yields for the different background components in the search regions defined by  $H_T$ ,  $\cancel{H}_T$  and  $N_{\text{Jets}}$ . The uncertainties of the different background sources are added in quadrature to obtain the total uncertainties. Taken from [164].

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

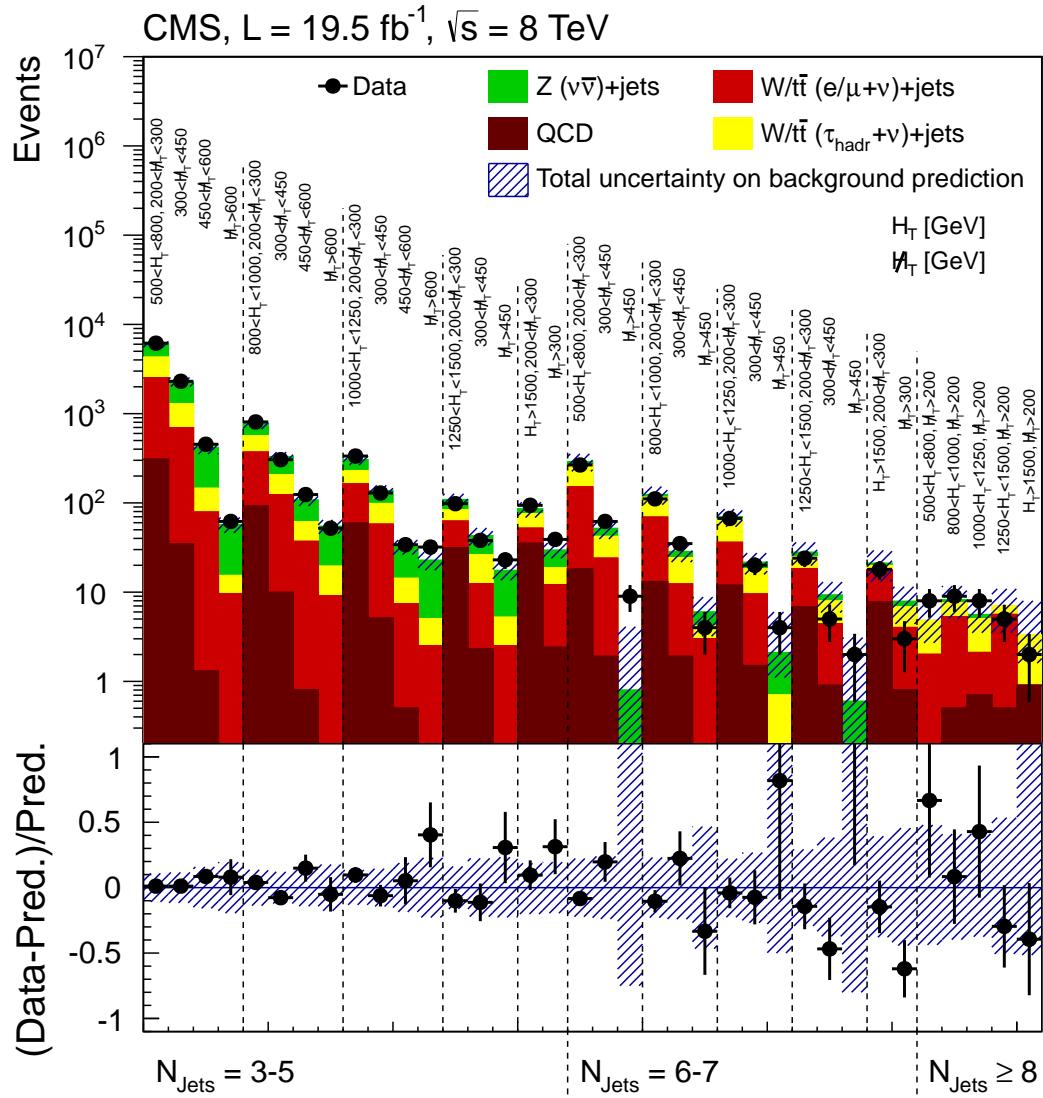


Figure 7.18: Summary of the observed number of events in each of the 36 search regions in comparison to the corresponding background prediction. The hatched region shows the total uncertainty of the background prediction. Taken from [164].

and  $\cancel{H}_T \geq 450$  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 which are

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

in case of light-flavour squarks and

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

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

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

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<sub>s</sub> [183–185] approach and denote the 95% confidence level (CL) upper limit on the production cross section of the signal. The profile likelihood ratio is used as test statistics which is derived from the combined likelihood 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 uncertainties considered for the signal acceptance and efficiency in the limit setting procedure are:

- 2.6% uncertainty on the integrated luminosity [155]
- 2% uncertainty for a possible trigger inefficiency (cf. Sec. 7.1.1)
- 3% uncertainty due to a possible mismodelling of the event cleaning efficiency
- 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 [186]
- an uncertainty considering the adjustment of the rate of initial-state radiation in simulation to match the measured rate in data [187] resulting for model parameter points with small differences between the LSP and the gluino or squark mass in an uncertainty of 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. 7.19(a)–(d), respectively. The expected (*dashed*) and observed (*solid*) 95% CL 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 [188] and MSTW2008 [189] as alternative PDF sets. By

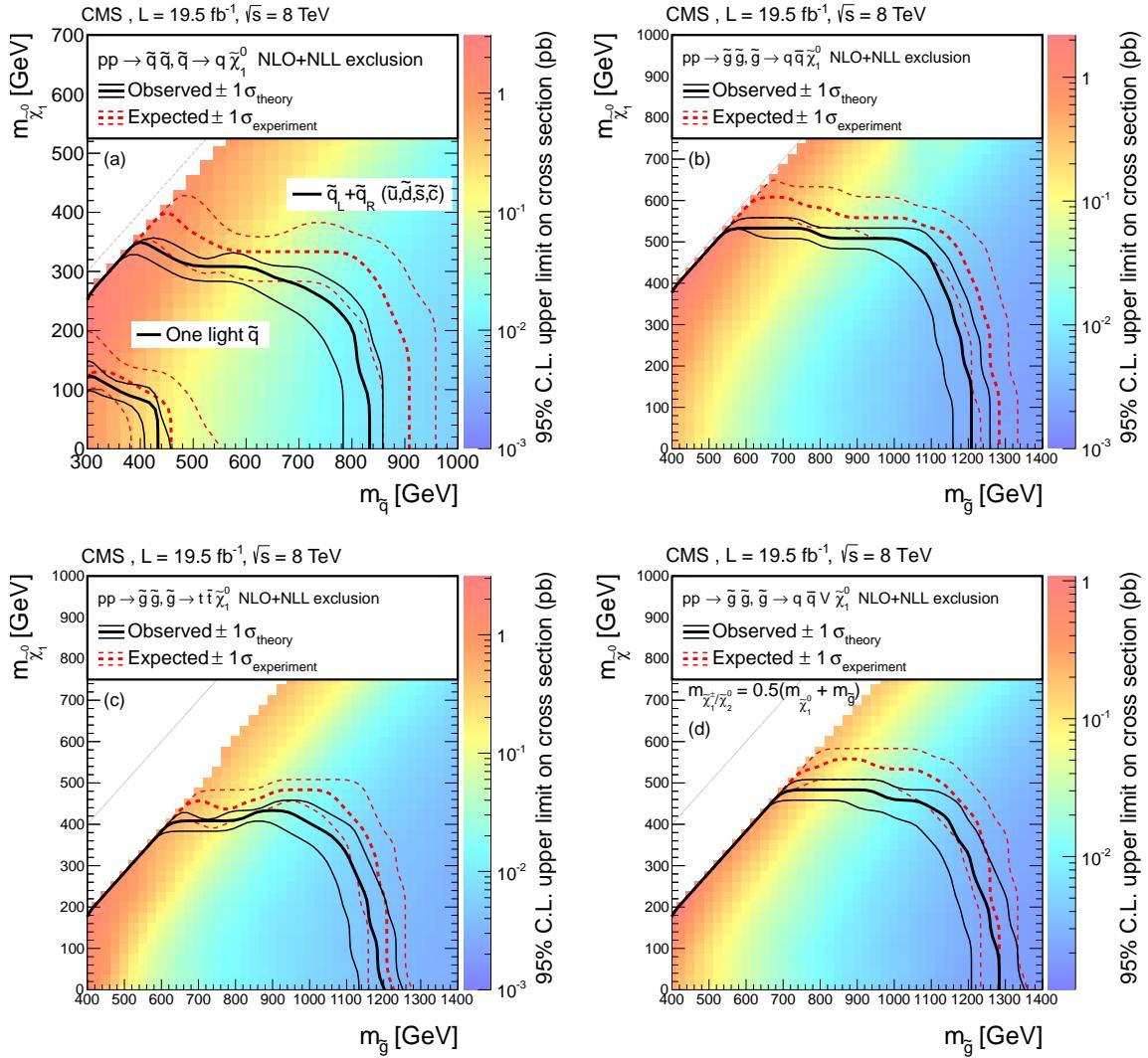


Figure 7.19: The observed and expected 95% CL upper limits on the (a) squark-squark and (b-d) gluino-gluino production cross sections in either the  $m(\text{squark})$ - $m(\text{LSP})$  or the  $m(\text{gluino})$ - $m(\text{LSP})$  plane obtained with the simplified models. For the squark-squark production the upper set of curves corresponds to the scenario when the first two generations of squarks are degenerate and light, while the lower set corresponds to only one light accessible squark [164].

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  $\cancel{H}_T$  is most efficient. For signal scenarios in which the mass difference between squark/gluino and LSP is small, softer jets and smaller values of  $\cancel{H}_T$  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  $N_{\text{Jets}}$  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. 7.20 for models (a)–(c) introduced in Sec. 7.4.

This comparison exhibits that the expected sensitivity to the models  $\tilde{g} \rightarrow q\bar{q} + \tilde{\chi}_1^0$  and  $\tilde{q} \rightarrow q + \tilde{\chi}_1^0$  is similar for the different analyses. This is not unexpected as the compared analyses all make use of the all-hadronic final state. However, the analysis published in [176] performs better especially in case of squark production, as here also search regions based on dijet events are employed.

Concerning gluino-mediated stop production it is interesting to note that the analysis presented in this thesis is also similarly sensitive to this model as other all-hadronic 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 [190–192] and the obtained exclusion limits on sparticle masses lie in a very similar mass region.

## 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 is not realised in nature. Scaling the respective branching

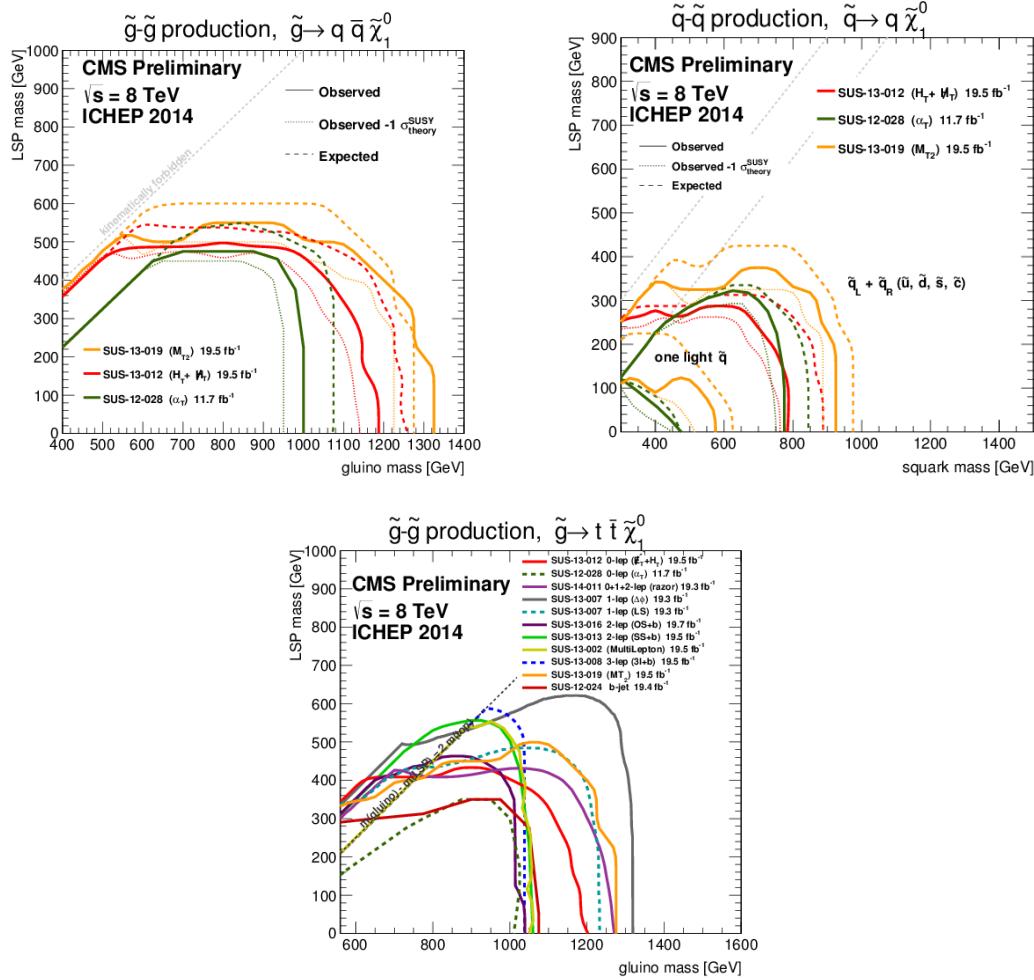


Figure 7.20: Comparison of various exclusion limits derived by different CMS analyses for the process  $\tilde{g} \rightarrow q\bar{q} + \tilde{\chi}_1^0$  (top left),  $\tilde{q} \rightarrow q + \tilde{\chi}_1^0$  (top right) and  $\tilde{g} \rightarrow t\bar{t} + \tilde{\chi}_1^0$  (bottom). Taken from [71].

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  $m_{1/2} \lesssim 800$  GeV for  $m_0 \lesssim 1$  TeV and  $m_{1/2} \lesssim 600$  GeV for  $m_0 \lesssim 6$  TeV ( $\tan \beta = 30$ ,  $A_0 = -2m_0$ ,  $\mu > 0$ ) [194], also global fits to constrain the respective CMSSM parameters are performed (cf. for instance [195, 196]). Typically, these fits do not only consider direct searches from the 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. 2.3.2 the interpretation of search results only within the

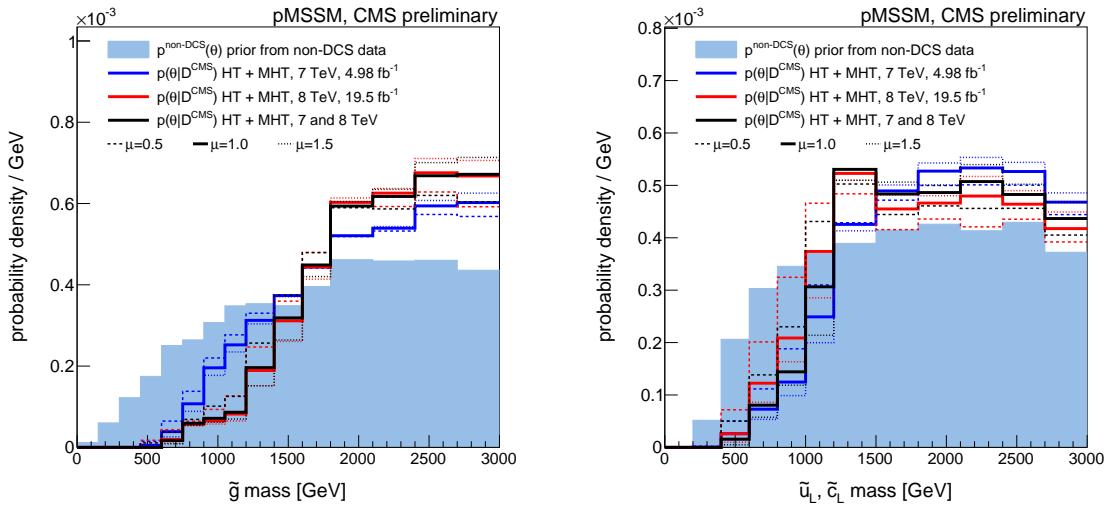


Figure 7.21: Marginalized distributions of gluino masses (*left*) and  $\tilde{u}_L, \tilde{c}_L$  masses (*right*). Filled histograms show prior distributions while line histograms illustrate posterior distributions based on the results of the analysis presented in this chapter and [74]. Solid curves denote the nominal curves while dashed lines represent systematic variations. Taken from [193].

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. 2.3.2, too. 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 [193]. In order to investigate the impact of direct searches at CMS, a global Bayesian analysis [197, 198] 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. 7.21, 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 here, 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, 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 experiment at a centre of mass energy of  $\sqrt{s} = 8$  TeV for direct production of top squark pairs, is illustrated in Fig. 7.22.

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 considered acceptable to date [32]. Consequently, a stop mass up to around

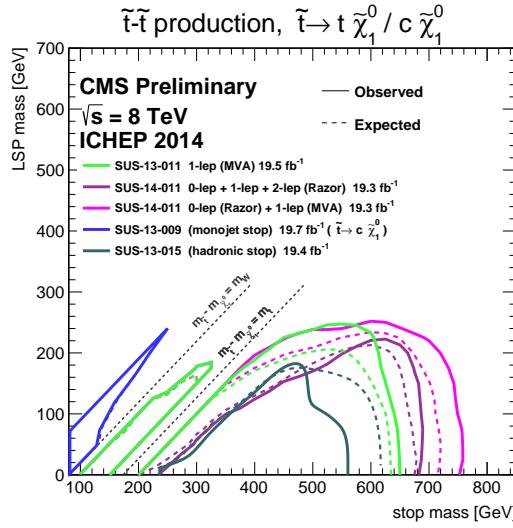


Figure 7.22: Exclusion limits derived by different CMS analyses for the process  $\tilde{t} \rightarrow t\tilde{\chi}_1^0$  in the  $m_{\tilde{t}}$  versus  $m_{\text{LSP}}$ . Taken from [71].

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  $\sqrt{s} = 13$  TeV, this mass region up to 1 TeV can be further studied. A feasibility study to search for direct stop quark pair-production at  $\sqrt{s} = 13$  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 stop quark 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  $\sqrt{s} = 13$  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 stop quarks with masses up to around 750 GeV for LSP masses below 100 GeV in case of direct stop production. Since for natural supersymmetry the stop quark mass is expected to not exceed significantly 1 TeV, it is of particular interest to probe the stop mass range  $\geq 750$  GeV during the next run period of the LHC.

The targeted process of this analysis is illustrated in a diagram in Fig. 8.1 below.

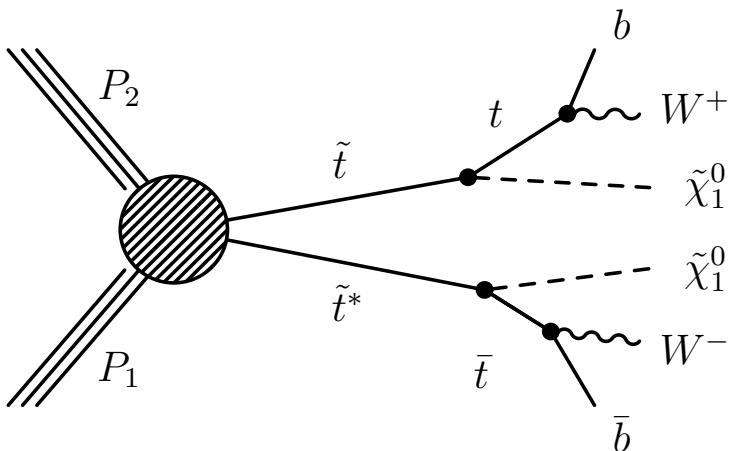


Figure 8.1: Schematic diagram of the direct pair production of stop quarks in  $pp$  collisions with subsequent decay into a top quark and the LSP. Taken from [71].

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

In this chapter, various analysis strategies for a search for top squarks at  $\sqrt{s} = 13$  TeV are discussed and compared in order to address mass scenarios with large mass differences between the stop quark mass and the LSP. Furthermore, the performance of new selections is compared to the existing all-hadronic stop search performed by the CMS experiment at  $\sqrt{s} = 8$  TeV which is published in [199] and denoted *SUS-13-015* in the following. In ad-

Table 8.1: Overview of simulated background and signal samples used in the analysis and corresponding production cross sections.

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

dition, it is interesting to study if selections that are developed for a search for direct stop production are also sensitive to gluino-mediated production of third generation quarks.

## 8.1 Data Samples

The studies presented in this chapter are based on simulated samples at  $\sqrt{s} = 13$  TeV. These are processed with the fast detector simulation and consider the pileup scenario as it was present at  $\sqrt{s} = 8$  TeV including out-of-time pileup according to a bunch spacing of 50 ns. Although these are not the pileup conditions expected for  $\sqrt{s} = 13$  TeV, the selections studied here typically are based on objects with high transverse momenta such that influences from pileup are 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 analysis sensitivity

follows in Sec. 8.9.

The processes considered in the analysis are summarized in Tab. 8.1.  $W + \text{jets}$ ,  $Z + \text{jets}$  and QCD multijet events are generated with MADGRAPH5 [110] using the PDF CTEQ6L1 [150], while  $t\bar{t}$  events are generated with POWHEG [112] and the MSTW2008 PDF [189]. For all samples, the showering process is performed with PYTHIA6 [108] Tune Z2star.

The process  $W + \text{jets}$  includes decays with  $W \rightarrow l\nu$  with up to two jets modelled in the matrix element. Similarly, also the process  $Z + \text{jets}$  includes up to two jets. Here, decays  $Z \rightarrow \nu\nu$  are modelled. For these processes, cross sections 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  $W + \text{jets}$  and  $Z + \text{jets}$ , respectively. The cross section calculation for  $t\bar{t}$  events is obtained from HATHOR [200].

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

In order to study signal events, two different processes are generated with MADGRAPH5 and showered with PYTHIA6. On the one hand direct stop-pair production is considered in which the  $\tilde{t}$  always decays to a stable neutralino  $\tilde{\chi}_0^1$  and a top quark with mass  $m_t = 172.5 \text{ GeV}$  including all decay channels of the top quark. These samples are generated with different neutralino masses of 50, 100, 150, 200, 250, 300 and 350 GeV. On the other hand gluino-pair production is generated with a decay  $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_0^1$ . For the top quark, only fully hadronic decays are simulated. The cross section given in Tab. 8.1 for gluino-pair production is corrected for the hadronic branching fraction of the top quark. The neutralino mass is fixed to 50 GeV.

All samples are normalized to an integrated luminosity of  $19.5 \text{ fb}^{-1}$  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 $\cancel{E}_T$

The targeted signature of the direct pair production of stops in the all-hadronic channel is based on jets and missing transverse momentum similar to the search presented in 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  $W + \text{jets}$  events are reduced by rejecting events containing isolated electrons or muons with  $p_T > 10 \text{ GeV}$  and  $|\eta| < 2.5$ . These are required to have a good quality track that can be associated with the primary interaction vertex [172, 173]. The isolation is given as the scalar sum of transverse momenta of PF particles (except for the lepton itself) within a cone of width  $\Delta R = 0.3$  for the electron and  $\Delta R = 0.4$  for the muon, respectively. This is required to be less than 20% of the transverse momentum of the electron and less than 15% of the transverse momentum of the muon.

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

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

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

- The missing transverse energy  $\cancel{E}_T$  calculated from the PF candidates is required to be  $\geq 200$  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  $\cancel{E}_T$  is not aligned with any of the leading three jets. Thus, events with

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

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

Since only simulated events are used, no dedicated event cleaning filters are applied as it is necessary for data (cf. Sec 7.1.3). The selection described here is again denoted baseline selection in the following. In Fig. 8.2, the obtained spectra of  $H_T$ ,  $\cancel{E}_T$  and  $N_{\text{Jets}}$  after applying the baseline selection are shown for the SM backgrounds and two selected signal points normalized to an integrated luminosity of  $19.5 \text{ fb}^{-1}$ . The signal points represent mass values of 600 GeV and 1100 GeV for the stop mass while the LSP mass is in both cases 50 GeV. These two signal points illustrate the difference in the kinematic properties of events for low and high stop masses. Typically, the  $H_T$  and  $\cancel{E}_T$  spectrum get harder for higher stop masses while the shape of the  $N_{\text{Jets}}$  spectrum stays nearly unchanged. After applying baseline selection criteria, the background is composed almost equally of all four SM processes ( $t\bar{t}$ : 27%,  $W + \text{jets}$ : 21%,  $Z + \text{jets}$ : 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 efficiencies  $\epsilon_{\text{sig/bg}}$  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. 8.3, the evolution of the signal versus background efficiencies is shown for the stop 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

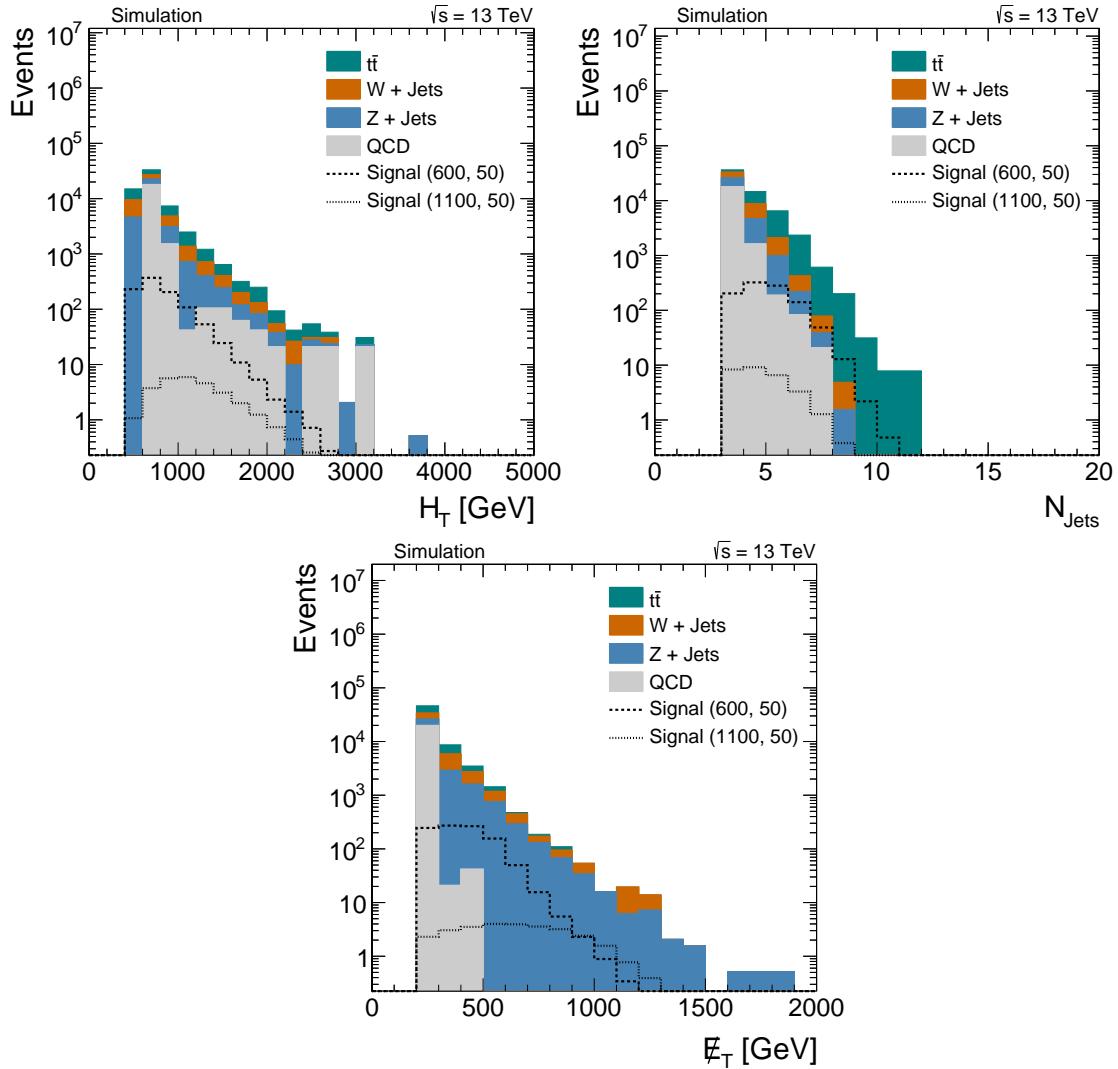


Figure 8.2: Comparison of selected  $H_T$  (top left),  $\cancel{E}_T$  (top right) and  $N_{\text{Jets}}$  (bottom) distributions in simulated events found from applying the baseline selection criteria. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

right corner. Here, the performance of  $H_T$ ,  $\cancel{E}_T$  and  $N_{\text{Jets}}$  is compared. As can be seen in Fig. 8.3,  $\cancel{E}_T$  shows in general the best performance when comparing these variables. This holds for low and high stop 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 stop mass of 600 GeV inferior to  $N_{\text{Jets}}$  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  $N_{\text{Jets}}$  spectra of signal and  $t\bar{t}$  background are almost identical. As in general the kinematic features of  $t\bar{t}$  background are closest to the signal, it is of special importance to identify selections that can reduce this background. Consequently,  $\cancel{E}_T$  and  $H_T$  are the preferred variables to distinguish signal and background processes.

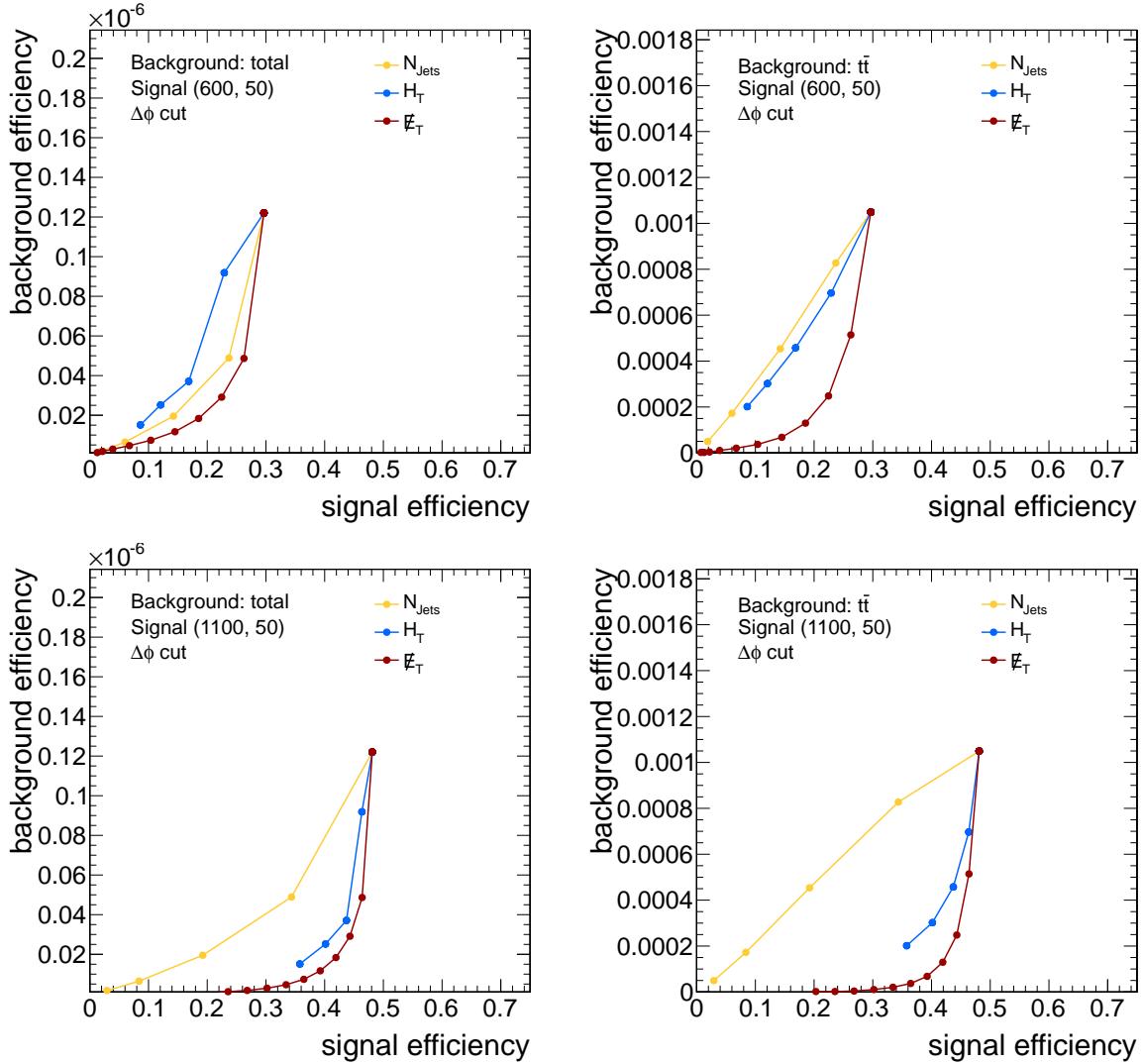


Figure 8.3: Evolution of the signal versus background efficiency for a stop mass of 600 GeV (top) and 1100 GeV (bottom) where the background is the sum of all backgrounds (left) or only the  $t\bar{t}$  background (right).

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  $\cancel{E}_T$  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  $\cancel{E}_T$  (SR2 and SR4), while the overall best sensitivity is given when combining the high  $\cancel{E}_T$  selection

Table 8.2: Exclusive search regions (SR) used in the analysis binned in  $H_T$  and  $\cancel{E}_T$ .

|     |                                |  |
|-----|--------------------------------|--|
| SR1 | $500 < H_T < 1000 \text{ GeV}$ | $200 < \cancel{E}_T < 400 \text{ GeV}$ |
| SR2 | $500 < H_T < 1000 \text{ GeV}$ | $400 \text{ GeV} < \cancel{E}_T$       |
| SR3 | $1000 \text{ GeV} < H_T$       | $200 < \cancel{E}_T < 400 \text{ GeV}$ |
| SR4 | $1000 \text{ GeV} < H_T$       | $400 \text{ GeV} < \cancel{E}_T$       |

also with the high  $H_T$  selection (SR4). However, the signal to background ratios in all search regions are so low that it is not expected to be able to really probe the considered stop 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 bins, 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 bins 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 this exclusion reach, the uncertainties of the individual background 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 considered for the different processes are

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

which are meant to be total uncertainties such that the actual statistical uncertainty of the number of simulated MC events is not considered explicitly.

Based on the selected event yields and the estimated uncertainties, the 95% confidence-level expected upper limit is calculated as an asymptotic  $\text{CL}_s$  limit [201]. The obtained exclusion curve is shown in Fig. 8.4 for the signal strength  $\mu$  which is the excluded production cross section divided by the theoretical cross section for direct stop production as a function of the stop quark mass. The LSP mass is fixed to 50 GeV in order to study the sensitivity of the selection for signal scenarios with large mass splittings between stop and neutralino mass. A particular mass point can be excluded if the expected limit drops below one. 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.

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  $19.5 \text{ fb}^{-1}$ . 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 the respective event yields.

|                   | total | [S/B]                  | SR1   | [S/B]                  | SR2  | [S/B]                  | SR3  | [S/B]                  | SR4 | [S/B]                  |
|-------------------|-------|------------------------|-------|------------------------|------|------------------------|------|------------------------|-----|------------------------|
| $t\bar{t}$        | 16461 |                        | 13525 |                        | 754  |                        | 1868 |                        | 314 |                        |
| $W + \text{jets}$ | 12481 |                        | 9660  |                        | 1516 |                        | 985  |                        | 320 |                        |
| $Z + \text{jets}$ | 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  | $[14.0 \cdot 10^{-3}]$ |

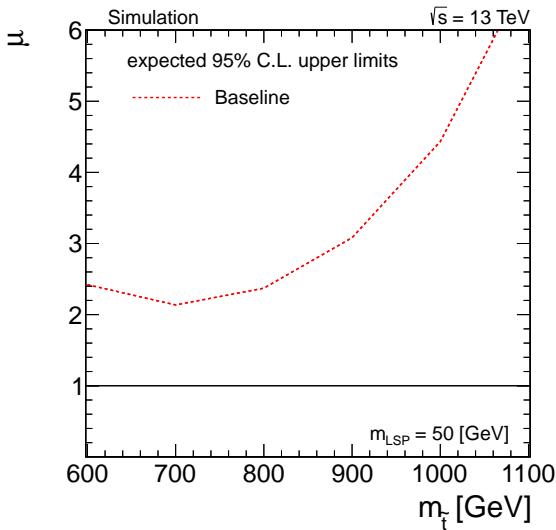


Figure 8.4: Expected 95% upper limit for signal strength versus  $m_{\tilde{t}}$ . The LSP mass is chosen to be 50 GeV.

### 8.3 Sensitivity Improvement using B Tagging

As illustrated in Fig. 8.1, 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 sensitivity of such an analysis is to employ b-tagging techniques to identify the b quarks in the final state. Typical b-tagging algorithms for the identification of b-quark jets used within the CMS experiment have been discussed in Section 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  $p_T > 30 \text{ GeV}$  in order to be not sensitive to potentially large flavour-dependent JEC uncertainties for jet transverse momenta smaller than 30 GeV. In Fig. 8.5, the b-tag multiplicity, i.e. the number of

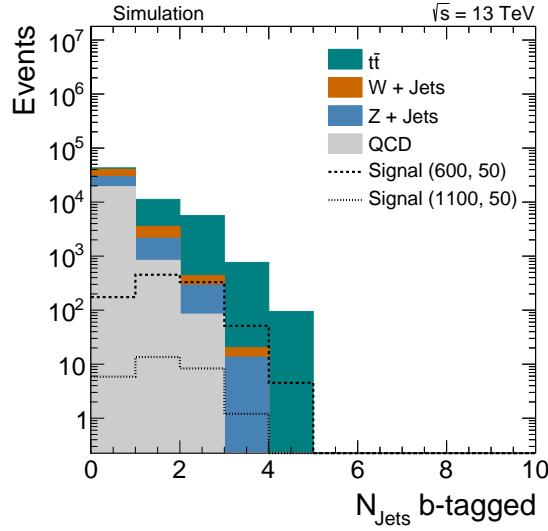


Figure 8.5: B-tag multiplicity after applying the baseline selection in simulated events. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

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 stop 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.

Due to 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 (= B.) including the additional b-tag requirement is shown in Fig 8.6.

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

## 8.4 Sensitivity Improvement using Top Tagging

In order to gain a better understanding of the kinematics of the investigated final state, the  $p_T$  spectrum of the leading generated hadronically decaying top quark is illustrated in Fig. 8.7 for  $t\bar{t}$  background and two selected signal points without applying any selection

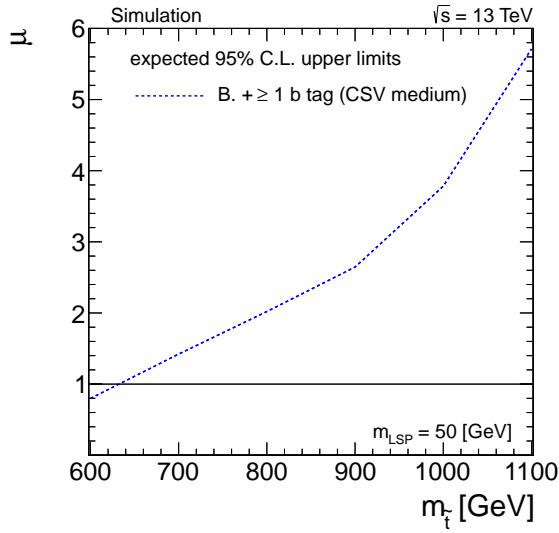


Figure 8.6: Expected 95% upper limit for signal strength versus  $m_{\tilde{t}}$ . The LSP mass is chosen to be 50 GeV.

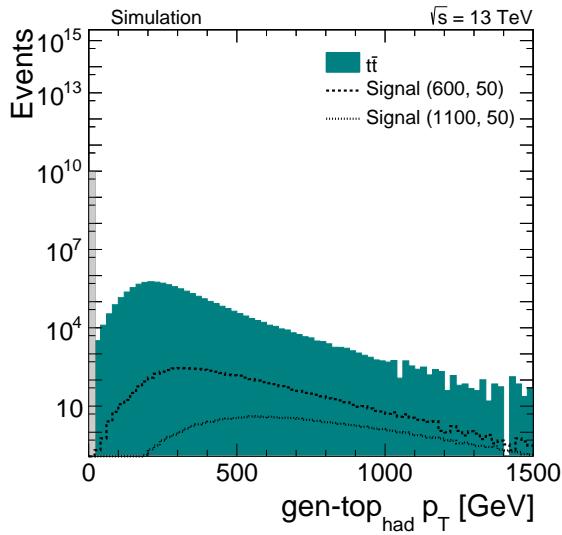


Figure 8.7: Transverse momentum spectrum of the leading generated hadronically decaying top quarks without applying any selection criteria. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

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 stop 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 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,

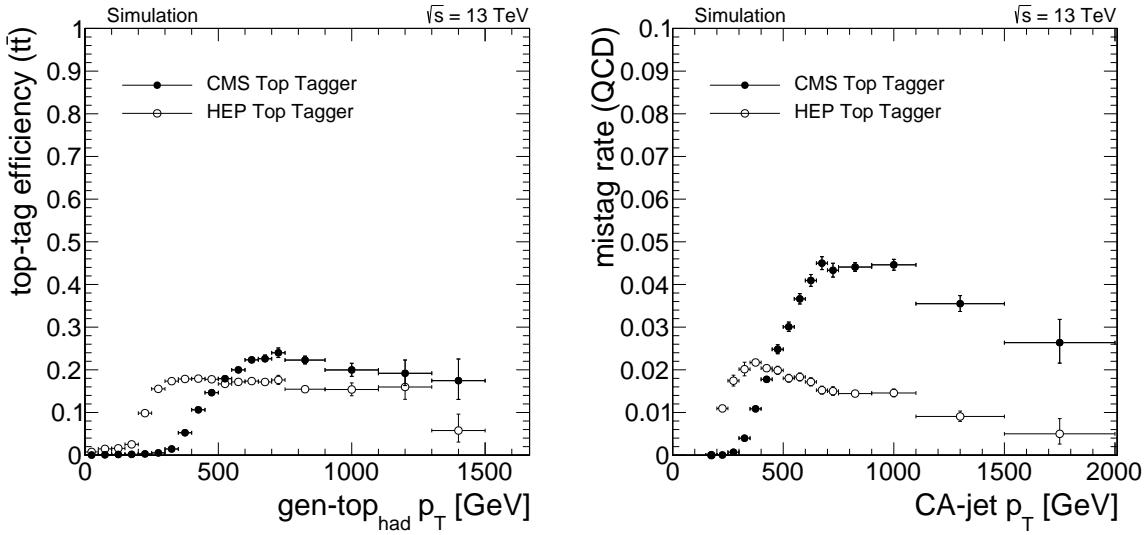


Figure 8.8: Top tag efficiency for  $t\bar{t}$  events as function of the transverse momentum of the generated top quark (*left*) and misidentification rate for QCD multijet events as a function of the CA-jet  $p_T$  (*right*). In case of the CMS top tagger CA-jets with a distance parameter of  $R = 0.8$  are used while the HEP top tagger is based on CA-jets with  $R = 1.5$ .

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-tag efficiency is defined as the number of hadronically decaying generated top quarks matched to a top-tagged CA-jet divided by the number of all generated hadronically decaying generated top quarks. A successful match is identified by requiring the  $\Delta R$  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. 8.8 for both the CMS and the HEP top tagger. It is visible that the turn-on of the HEP top tagger starts already around 200 GeV while the CMS tagger begins to become efficient not before around 400 GeV. However, the efficiency of the HEP top tagger lies around 20% in the plateau region while the plateau efficiency of the CMS top tagger is in general higher with a value of around 25%. This behaviour results mainly from the different jet sizes and selection criteria which are in case of the HEP top tagger optimized to be sensitive already in the  $p_T$  range around 200–300 GeV.

The misidentification rate can be evaluated by dividing the number of top-tagged CA-jets by the number of all CA-jets. Again, in case of the CMS top tagger the

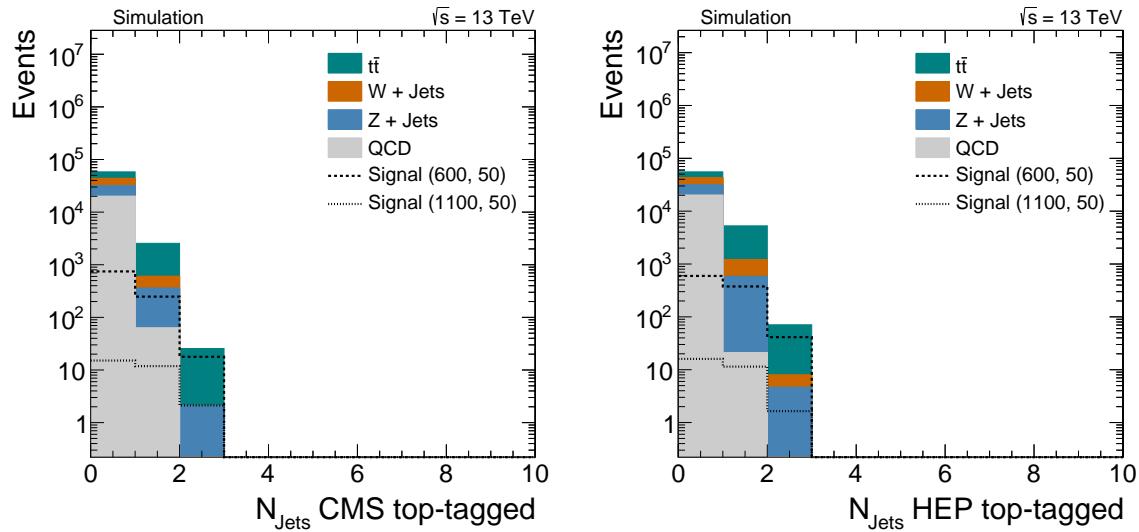


Figure 8.9: Top tag multiplicity for the CMS top tagger (*left*) and the HEP top tagger (*right*) after application of the baseline selection. Only CA-jets of the corresponding jet size with a transverse momentum above 150 GeV are considered.

jet distance parameter is  $R = 0.8$  and for the HEP top tagger it is  $R = 1.5$ . The misidentification rate as a function of the respective CA-jet transverse momentum is shown in the right part of Fig. 8.8. Here, a similar feature as for the efficiency curve is observed. The HEP top tagger shows a certain misidentification rate already for lower transverse momenta than the CMS top tagger. However, in general the misidentification rate of the HEP top tagger is significantly smaller with a plateau around 1.5% compared to the CMS top tagger which shows a misidentification rate of up to 4–5% in the plateau.

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 squarks, the more important property is an efficiency as high as possible to retain signal efficiency while the misidentification rate only plays a less important role.

The impact on the analysis when applying top tag requirements can be studied when looking for instance at the top tag multiplicity. In Fig. 8.9, the top tag multiplicity is shown for the CMS top tagger (left) and the HEP top tagger (right) after the application of the baseline selection. Here, only CA-jets of the corresponding jet size with a transverse momentum above 150 GeV are considered, since the taggers are not supposed to be efficient for smaller momenta.

The distributions exhibit that, although a significant amount of signal events does not have a CA-jet identified as top-jet by the respective algorithm, the background can be significantly reduced when requiring at least one top-tagged jet. 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:

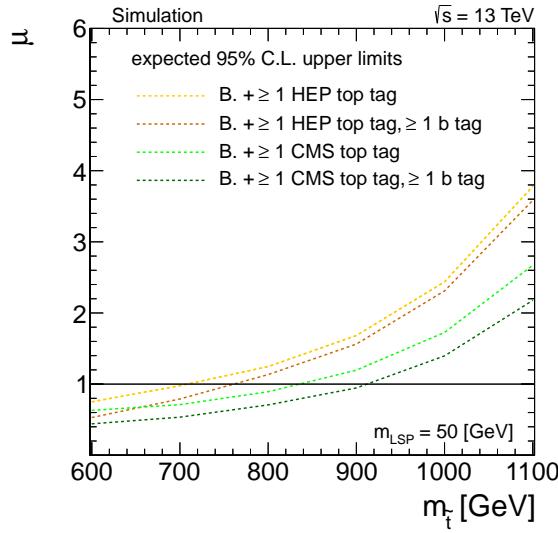


Figure 8.10: Expected 95% upper limit for signal strength versus  $m_{\tilde{t}}$ . The LSP mass is chosen to be 50 GeV.

- HEP top tagger:  
 $\epsilon_{\text{sig}}^{m_{\tilde{t}}=600 \text{ GeV}} = 12\%$ ,  $\epsilon_{\text{sig}}^{m_{\tilde{t}}=1100 \text{ GeV}} = 22\%$   
 $\epsilon_{\text{bg}}^{\text{total}} = 1.1 \cdot 10^{-6}\%$ ,  $\epsilon_{\text{bg}}^{t\bar{t}} = 0.03\%$
- CMS top tagger:  
 $\epsilon_{\text{sig}}^{m_{\tilde{t}}=600 \text{ GeV}} = 8\%$ ,  $\epsilon_{\text{sig}}^{m_{\tilde{t}}=1100 \text{ GeV}} = 23\%$   
 $\epsilon_{\text{bg}}^{\text{total}} = 0.5 \cdot 10^{-6}\%$ ,  $\epsilon_{\text{bg}}^{t\bar{t}} = 0.01\%$

In general, the signal efficiencies for low masses is larger when using the HEP top tagger than for the CMS top tagger and comparable for the higher stop 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 8.15. 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  $(m_{\tilde{t}}, m_{\text{LSP}}) = (700, 50)$  GeV or even  $(800, 50)$  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. 8.4 which allows to probe stop masses around 600 GeV for an LSP mass of 50 GeV. In general, the selection based on the CMS top tagger performs better than the selection involving the HEP top tagger, since the signal to background ratio is higher as discussed above.

Furthermore, the selections including the top tagging requirements are also compared when imposing an additional b-tag requirement, as defined in Sec. 8.4. 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 using the baseline requirements discussed in Sec. 8.2 can be improved already significantly when employing in addition b-tag and top-tag requirements, as shown in Sec. 8.4 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. 8.11, the background and signal processes are shown according to the decay mode of the process after the application of the baseline selection (left) and the baseline selection with an additional requirement of  $\geq 1$  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 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:

$\alpha_T$  : The  $\alpha_T$  variable is mainly used to reject QCD multijet events which have no intrinsic  $\cancel{E}_T$  [174, 202]. In case of dijet events, it is defined as:

$$\alpha_T = \frac{E_T^{j_2}}{M_T} \quad \text{and} \quad M_T = \sqrt{\left( \sum_{i=1}^2 E_T^{j_i} \right)^2 - \left( \sum_{i=1}^2 p_x^{j_i} \right)^2 - \left( \sum_{i=1}^2 p_y^{j_i} \right)^2} \quad (8.2)$$

with the transverse energy  $E_T^{j_2}$  of the less energetic jet and the transverse mass  $M_T$  of the dijet system. Since in case of boosted top quark decays, the two top quarks are supposed to be represented each by one fatjet, the  $\alpha_T$  variable is in this studies calculated from the two leading CA8 jets. The distribution of the  $\alpha_T$  variable after application of the baseline selection requirements is shown in Fig. 8.12 (top left). Typically,  $\alpha_T$  is 0.5 for an ideal dijet event with  $E_T^{j_1} = E_T^{j_2}$  in which each jet momentum is large compared to its mass. If an imbalance occurs due to a jet mismeasurement,  $\alpha_T$  drops below 0.5 while it is greater than 0.5 when the two jets recoil against real  $\cancel{E}_T$ , as for instance resulting from LSPs.

$M_{T2}$  : The  $M_{T2}$  variable represents a generalized version of the transverse mass [176,

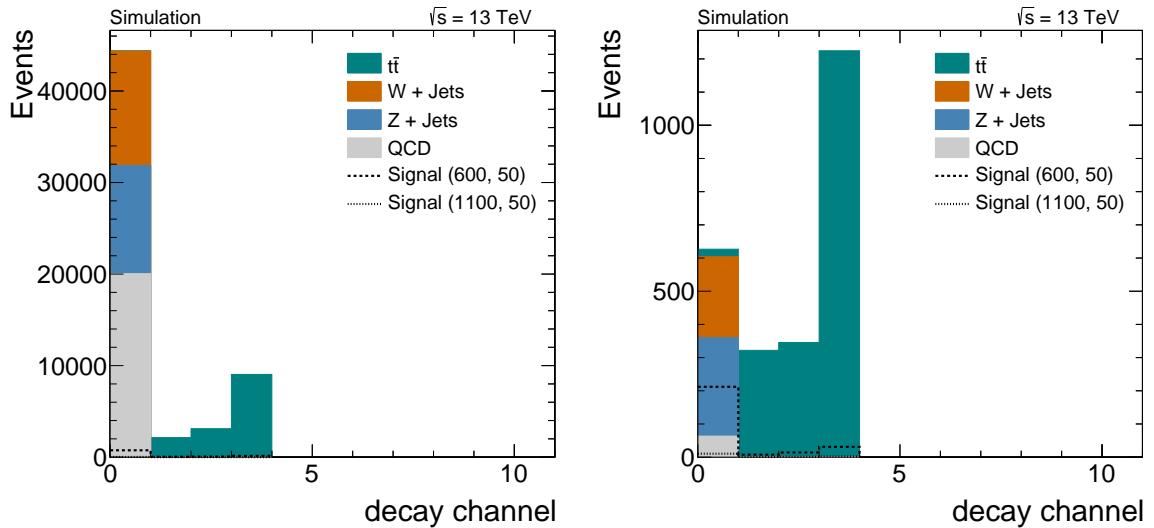


Figure 8.11: Comparison of decay modes of signal and background processes defined according to the top quark properties based on generator information after the baseline selection (*left*) and after the baseline selection and the requirement of  $\geq 1$  CMS top tag (*right*). For a definition of the decay channels see text.

203–205]. 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

$$M_{T2}(m_{\tilde{\chi}}) = \min_{\vec{p}_T^{\tilde{\chi}(1)} + \vec{p}_T^{\tilde{\chi}(2)} = \vec{p}_T^{\text{miss}}} \left[ \max \left( M_T^{(1)}, M_T^{(2)} \right) \right] \quad (8.3)$$

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

$$(M_T^{(i)})^2 = (m^{\text{vis}(i)})^2 + m_{\tilde{\chi}}^2 + 2 \left( E_T^{\text{vis}(i)} E_T^{\tilde{\chi}(i)} - \vec{p}_T^{\text{vis}(i)} \cdot \vec{p}_T^{\tilde{\chi}(i)} \right) \quad (8.4)$$

described by the transverse momenta  $\vec{p}_T^{\text{vis}(i)}$ , transverse energies  $E_T^{\text{vis}(i)}$  and masses  $m^{\text{vis}(i)}$  for the visible systems and the unknown transverse momenta of the LSPs  $\vec{p}_T^{\tilde{\chi}(i)}$ . Experimentally, the momenta  $\vec{p}_T^{\tilde{\chi}(i)}$  are not accessible separately. Thus, a minimization on trial LSP masses fulfilling the constraint given by  $\vec{p}_T^{\text{miss}}$ , the missing transverse momentum,<sup>1</sup> is performed. This minimization is carried out to make sure that  $M_{T2}$  does not exceed the mass of the parent particle. For the correct value of  $m_{\tilde{\chi}}$ , the distribution of  $M_{T2}$  is expected to have an endpoint at the mass of the parent particle.

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

The obtained distributions for  $M_{T2}$  in background and signal samples are shown in

<sup>1</sup>More commonly denoted by  $\vec{E}_T$  in this thesis.

Fig 8.12 (top right). They exhibit that the maximum of  $M_{\text{T}2}$  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 [177, 206, 207]. 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

$$M_{\text{R}} = \sqrt{\left[ (|\vec{p}^{j1}| + |\vec{p}^{j2}|)^2 - (p_z^{j1} + p_z^{j2})^2 \right]} \quad (8.5)$$

$$M_{\text{T}}^{\text{R}} = \sqrt{\frac{1}{2} \left( \cancel{E}_{\text{T}}(p_{\text{T}}^{j1} + p_{\text{T}}^{j2}) - \vec{\cancel{E}}_{\text{T}} \cdot (\vec{p}^{j1} + \vec{p}^{j2}) \right)} \quad (8.6)$$

$$R = \frac{M_{\text{T}}^{\text{R}}}{M_{\text{R}}} \quad (8.7)$$

with the transverse momenta of the two jets  $p_{\text{T}}^{j1}$  and  $p_{\text{T}}^{j2}$ . For signal events,  $M_{\text{T}}^{\text{R}}$  has an endpoint and  $R$  a maximum of approximately one.

For these studies, the two leading CA8 jets and  $\cancel{E}_{\text{T}}$  are used to compute the razor variables. In Fig. 8.12 the variable  $R^2$  is illustrated (bottom) after the application of baseline selection criteria.

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

**Transverse mass  $m_{\text{T}}$ :** As shown in Fig. 8.11, 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. Thus, the missing transverse energy and the closest b-tagged jet in  $\Delta\phi$  can be utilized to calculate a transverse mass according to

$$m_{\text{T}} = \sqrt{2p_{\text{T}}^{\text{jet}}\cancel{E}_{\text{T}} \cdot (1 - \cos(\Delta\phi(\text{jet}, \cancel{E}_{\text{T}})))}. \quad (8.8)$$

The distribution obtained after applying the baseline selection is shown for background and two selected signal samples in Fig. 8.13 (left). Here, the transverse mass is calculated from the missing transverse momentum and the closest b-tagged anti- $k_{\text{T}}$  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.

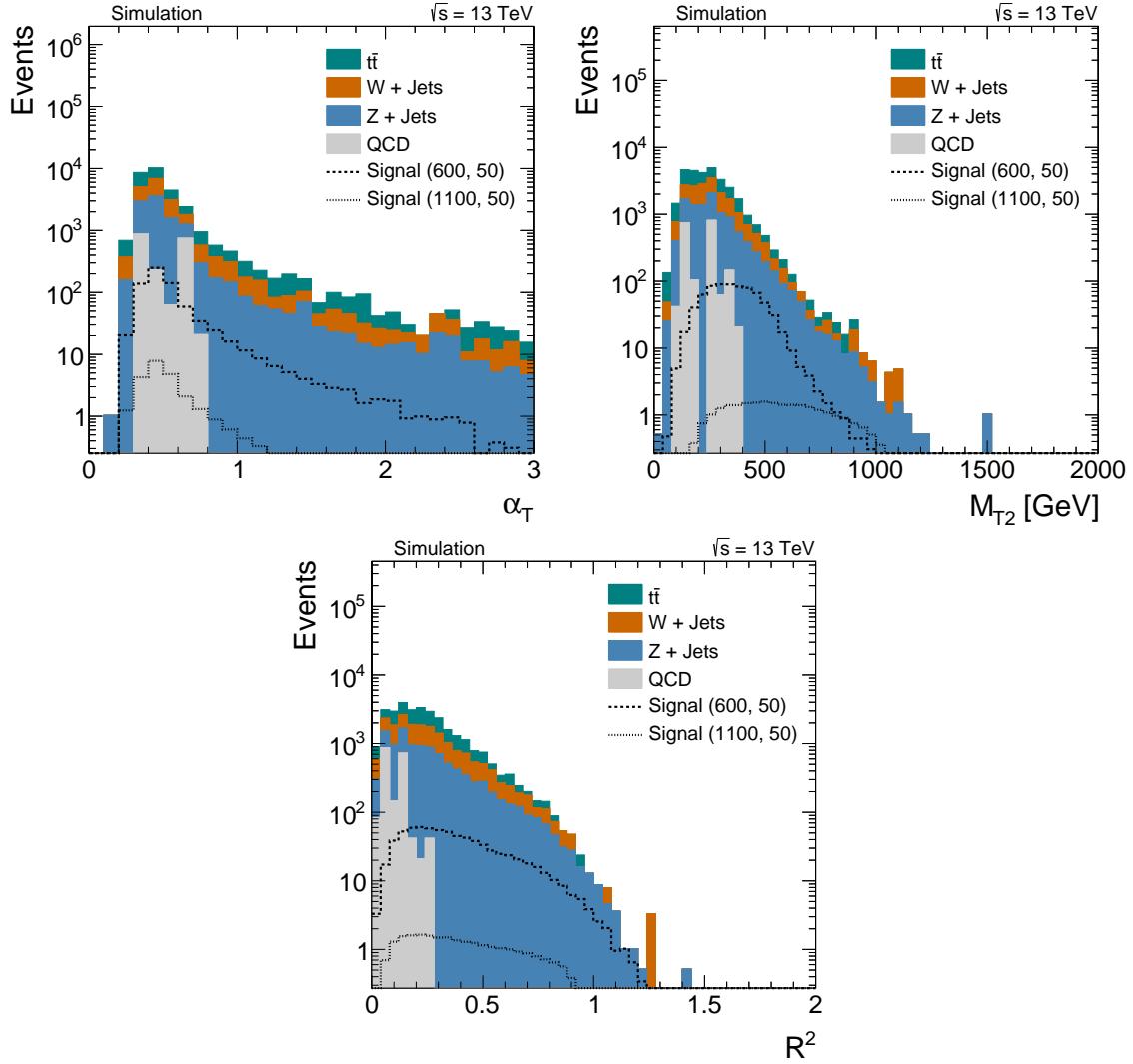


Figure 8.12: Distribution of the  $\alpha_T$  variable (top left),  $M_{T2}$  (top right) and the razor variable  $R^2$  (bottom) after application of the baseline selection requirements for background and two selected signal samples. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

**$\Delta\phi(\text{CA-jet}_1, \text{CA-jet}_2)$  :** 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 fatjets balanced against each other. For signal events however, such a topology is not expected, since the event is balanced against genuine  $\cancel{E}_T$  arising from the LSPs. The selected  $\Delta\phi$  distributions after applying baseline selection criteria are shown in Fig. 8.13 (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

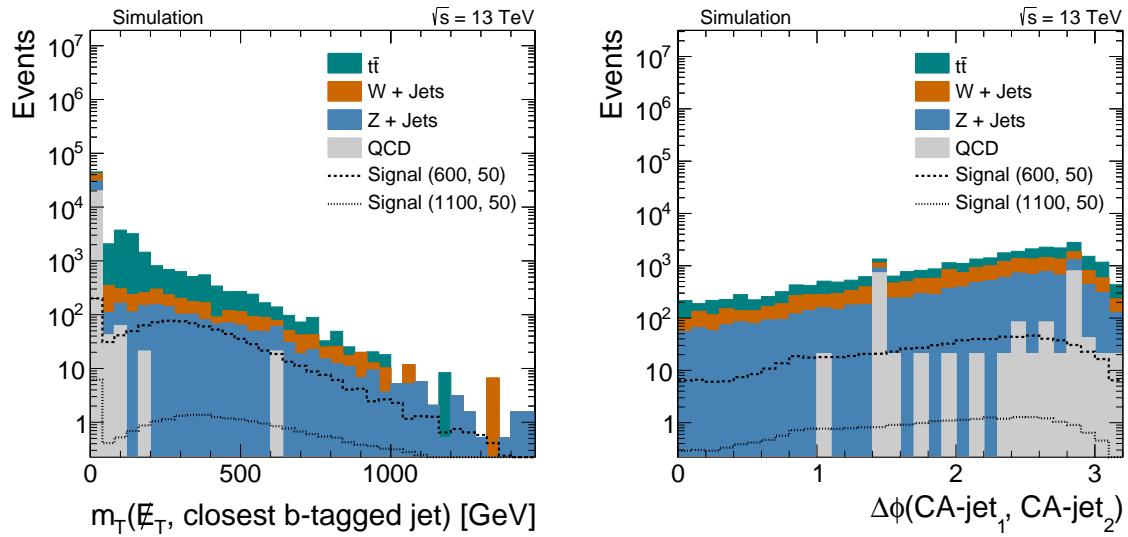


Figure 8.13: Distribution of the  $m_T$  variable (*left*) and  $\Delta\phi(\text{CA-jet}_1, \text{CA-jet}_2)$  (*right*) after application of the baseline selection requirements for background and two selected signal samples. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass.

curves as introduced in Sec. 8.2. The resulting curves after applying baseline selection criteria are shown in Fig. 8.14. For comparison, also the signal versus background efficiency curve for  $\cancel{E}_T$  is illustrated.

The calculation of  $\alpha_T$ ,  $\Delta\phi(\text{CA-jet}_1, \text{CA-jet}_2)$ ,  $M_{T2}$  and the razor variable  $R^2$  imply that there exist two CA8 jets with  $p_T > 150$  GeV in the event. Consequently, the starting point of that scan curves is different than that of the  $\cancel{E}_T$  curve which represents the signal and background efficiencies after the baseline selection. Similarly the  $m_T$  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  $m_T > 20$  GeV.

The distributions exhibit that among the variables based on the presence of two CA8 jets,  $M_{T2}$  performs best, especially for high top squark masses. This statement applies to the total background as well as to the  $t\bar{t}$  background only. Furthermore, the selection based on  $m_T$  shows a nice separation power as well and outperforms  $M_{T2}$  especially for smaller stop masses. It is visible that the curve for  $m_T$  shows a distinct kink. This corresponds to a selection keeping only events with  $m_T$  greater than the top quark mass. When imposing even tighter selection requirements on  $m_T$ , the separation power decreases rapidly.

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  $M_{T2} > 400$  GeV (*top tag + M<sub>T2</sub>*), at least one b-tagged jet (*top tag + b tag*) or  $m_T > 180$  GeV (*top tag + m<sub>T</sub>*). 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

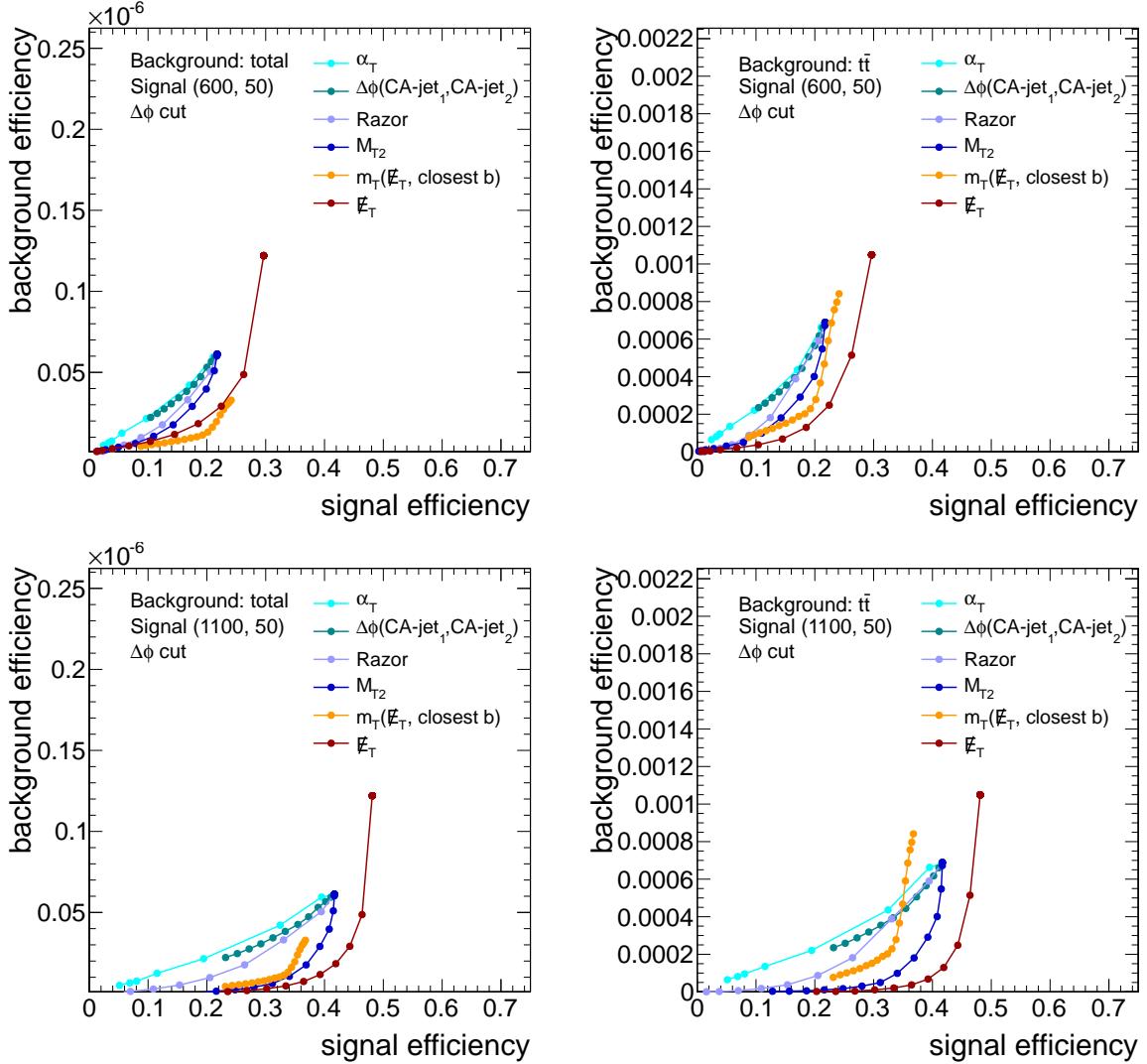


Figure 8.14: Evolution of the signal versus background efficiency for a stop mass of 600 GeV (*top*) and 1100 GeV (*bottom*) where the background is the sum of all backgrounds (*left*) or the  $t\bar{t}$  background (*right*).

$m_T$  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. 8.15, 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. 8.4, 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  $M_{T2} > 400$  GeV (pink), a b-tag requirement as discussed already in Sec. 8.4 (dark green) or a requirement of  $m_T > 180$  GeV (light blue). It turns out that the latter performs best among those described selections over the whole stop mass range, as indicated already by the achieved signal to background ratios.

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             |

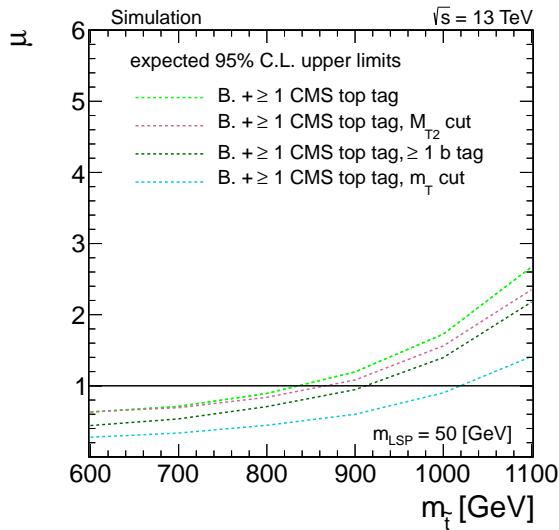


Figure 8.15: Expected 95% upper limit for signal strength versus  $m_{\tilde{t}}$ . The LSP mass is chosen to be 50 GeV.

## 8.6 Performance Comparison to Selection Based on Published Stop Analysis at $\sqrt{s} = 8 \text{ TeV}$

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 pre-selection criteria used here are:

- Events with isolated electrons and muons with  $p_T > 10 \text{ GeV}$ , as described in Sec. 8.2, are vetoed.
- Events have to have at least five jets with  $p_T > 30 \text{ GeV}$  and  $|\eta| < 2.4$ . The two highest  $p_T$  jets further have to have  $p_T > 70 \text{ GeV}$  while the next two highest jets in  $p_T$  must fulfill  $p_T > 50 \text{ GeV}$ .
- There has to be at least one b-tagged jet in the event based on the CSV algorithm with medium working point.
- The minimum azimuthal angle between the three highest jets and the missing transverse momentum has to be  $\Delta\phi(\text{jet}_n, \cancel{E}_T) > 0.5$ ,  $n = 1, 2$  and  $\Delta\phi(\text{jet}_3, \cancel{E}_T) > 0.3$ .<sup>2</sup>

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  $p_T > 5 \text{ GeV}$ , here only events with electrons and muons with  $p_T > 10 \text{ GeV}$  are vetoed. However, since the  $10 \text{ GeV}$  lepton veto is the same lepton veto requirement as for the other selections studied in this chapter, it is easier to compare the quality of the different selections applied to the sample after selecting the all-jet final state.

In addition to those pre-selection requirements, further selection criteria are imposed in order to reconstruct the hadronically-decaying top quarks. The set of five or more jets in the event is separated into all possible combinations of three jets and a remnant 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 while the other is based on the remnant system and referred to as *partially-reconstructed* top. Details on the reconstruction process of the two top quark systems can be found in [199]. In particular, this process involves that the fully-reconstructed top quark has to satisfy the criteria described in Eq. 5.6, 5.7 and 5.8 using  $R_{\min} = 0.85 \cdot (m_W/m_{\text{top}})$ ,  $R_{\max} = 1.25 \cdot (m_W/m_{\text{top}})$ ,  $m_W = 80.4 \text{ GeV}$  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  $m_{\text{top}}$  is selected. The selected remnant system is denoted  $R_{\text{sys}}$ , the fully-reconstructed top 3-jet.

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

- $\cancel{E}_T > 200 \text{ GeV}$
- The variable  $M_{\text{T2}}$ , as discussed in Sec. 8.5, 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  $\cancel{E}_T$  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, respectively, which are calculated according to

$$(M_{\text{T}}^{3\text{-jet}})^2 = (m^{3\text{-jet}})^2 + 2(E_{\text{T}}^{3\text{-jet}} \cancel{E}_T - p_{\text{T}}^{3\text{-jet}} \cancel{E}_T \cos\Delta\phi)$$

---

<sup>2</sup>The missing transverse energy in the published analysis is denoted with  $p_{\text{T}}^{\text{miss}}$ .

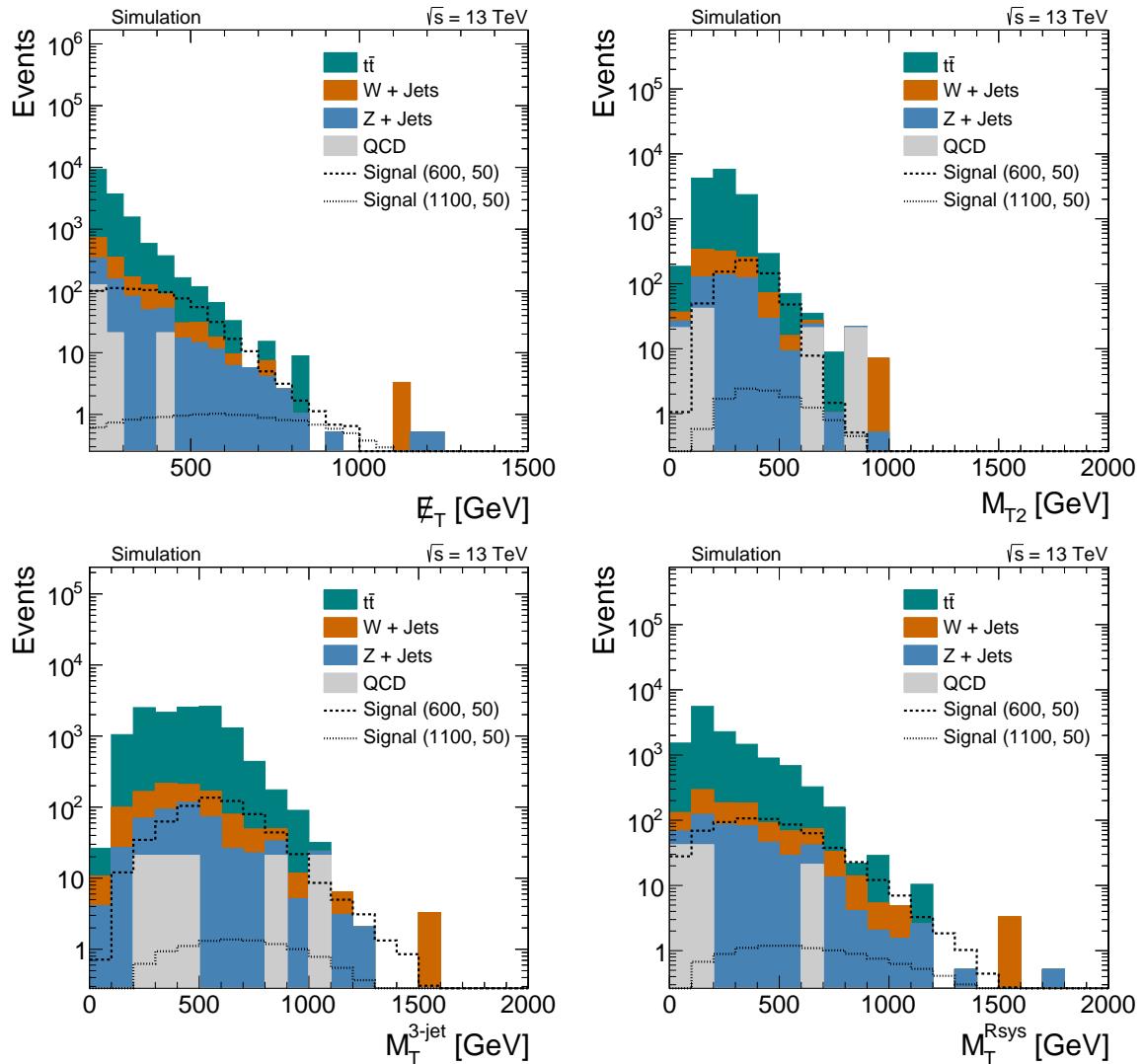


Figure 8.16: Comparison of  $E_T$ ,  $M_{T2}$ ,  $M_T^{3\text{-jet}}$  and  $M_T^{\text{Rsys}}$  distributions for signal and background events after applying pre-selection criteria and  $E_T > 200$  GeV. The signal points are labelled as (X, Y) where X is the top squark mass and Y is the LSP mass in GeV.

and

$$(M_T^{\text{Rsys}})^2 = (m^{\text{Rsys}})^2 + 2(E_T^{\text{Rsys}} E_T - p_T^{\text{Rsys}} E_T \cos \Delta\phi).$$

A comparison of the  $E_T$ ,  $M_{T2}$ ,  $M_T^{3\text{-jet}}$  and  $M_T^{\text{Rsys}}$  distributions for signal and background events after the pre-selection and a requirement of  $E_T > 200$  GeV is shown in Fig. 8.16. 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<sub>ref</sub>:  $E_T > 200$  GeV,  $N_{\text{b-jets}} \geq 1$

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 <sub>ref</sub> | Signal (600, 50)  | 0.2   |
|                    | Signal (1100, 50) | 0.005 |
| SR2 <sub>ref</sub> | Signal (600, 50)  | 0.8   |
|                    | Signal (1100, 50) | 0.02  |
| SR3 <sub>ref</sub> | Signal (600, 50)  | 0.5   |
|                    | Signal (1100, 50) | 0.01  |
| SR4 <sub>ref</sub> | Signal (600, 50)  | 1.2   |
|                    | Signal (1100, 50) | 0.03  |

- SR2<sub>ref</sub>:  $\cancel{E}_T > 350 \text{ GeV}$ ,  $N_{\text{b-jets}} \geq 1$
- SR3<sub>ref</sub>:  $\cancel{E}_T > 200 \text{ GeV}$ ,  $N_{\text{b-jets}} \geq 2$
- SR4<sub>ref</sub>:  $\cancel{E}_T > 350 \text{ GeV}$ ,  $N_{\text{b-jets}} \geq 2$

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  $\cancel{E}_T > 350 \text{ GeV}$  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 these 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 stop masses ranging from 600–1100 GeV and a LSP mass of 50 GeV, the selection requiring  $\cancel{E}_T > 350 \text{ GeV}$ ,  $N_{\text{b-jets}} \geq 2$ , as seen already from the signal over background ratios. The comparison of the expected limits is illustrated in Fig 8.17. The limit curves exhibit that the selection based on the requirements used in SUS-13-015 shows a quite good sensitivity for low stop masses while the sensitivity drops rapidly towards higher stop masses. However, the selection proposed in Sec. 8.5 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  $m_T > 180 \text{ GeV}$  is evaluated when varying the assumed  $t\bar{t}$  uncertainty.

The dependence on the assumed uncertainties is studied by increasing the uncertainties considered for  $t\bar{t}$  from 20% to 50% while others stay unchanged. Since for this selection  $t\bar{t}$

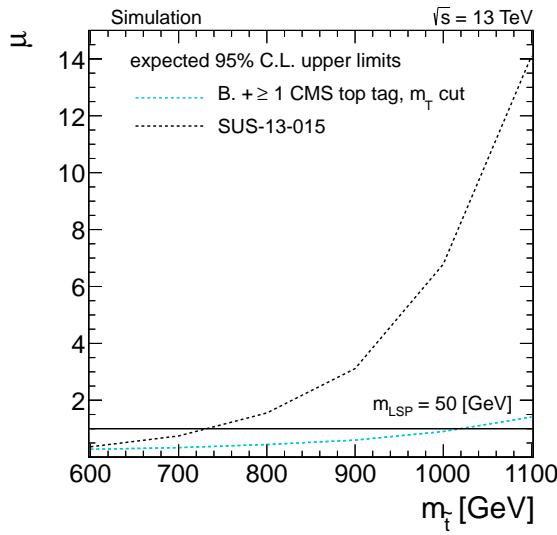


Figure 8.17: Expected 95% upper limit for signal strength versus  $m_{\tilde{t}}$ . The LSP mass is chosen to be 50 GeV.

is the main background, the variation of the  $t\bar{t}$  uncertainty is expected to cause the largest deviation. The expected limits derived using the nominal and the increased  $t\bar{t}$  uncertainty are illustrated in Fig. 8.18. It turns out that the resulting change for the sensitivity caused by this variation is small and only impacts small stop quark 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 these studied stop quark masses.

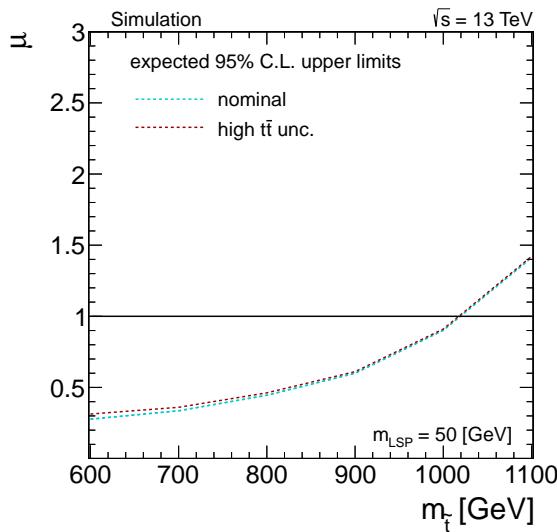


Figure 8.18: Expected 95% upper limit for signal strength versus  $m_{\tilde{t}}$ . The LSP mass is chosen to be 50 GeV.

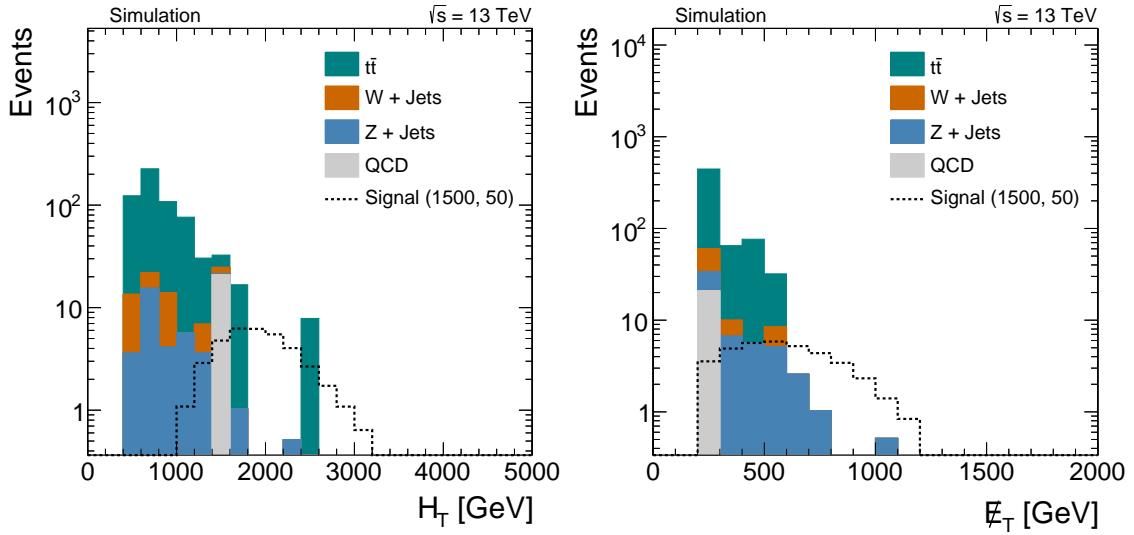


Figure 8.19: Comparison of selected  $H_T$  (*left*) and  $\cancel{E}_T$  (*right*) distributions in simulated events found from applying the baseline selection criteria, at least one CMS top tag and  $m_T > 180$  GeV. The signal points for gluino-mediated stop production are labelled as (X, Y) where X is the gluino mass and Y is the LSP mass in GeV.

## 8.8 Sensitivity to Gluino-Mediated Stop Production

The selections studied in this chapter were developed for selecting events from direct pair production of top squarks which each subsequently decay into a top quark and a LSP. Special emphasis is put on identifying mass scenarios with large mass splittings between top squark and LSP which is suitable for employing top tagging techniques. However, the studied selections possibly offer an interesting alternative to classical selections targeting gluino-mediated production of third generation squarks. In those scenarios, pair produced gluinos are considered where 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 that, the sensitivity of the best performing selection criteria based on the baseline selection, at least one CMS top-tagged jet and a transverse mass selection of  $m_T > 180$  GeV is also evaluated with respect to the gluino-mediated stop production. A comparison of  $H_T$  and  $\cancel{E}_T$  for one selected signal point and respective background events after applying these selection criteria is shown in Fig. 8.19.

The evaluation of the sensitivity towards gluino-mediated stop production is performed 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, respectively. Based on the same search strategy as before, using exclusive  $H_T$  and  $\cancel{E}_T$  search regions, the derived expected exclusion curve for gluino-mediated stop production is illustrated in Fig. 8.20.

The limit curve shows that the derived selection described in Sec. 8.5 is also sensitive to the tested mass scenarios of gluino-mediated stop production. Thus, selections involving top tagging also offer the opportunity to probe SUSY scenarios other than direct stop

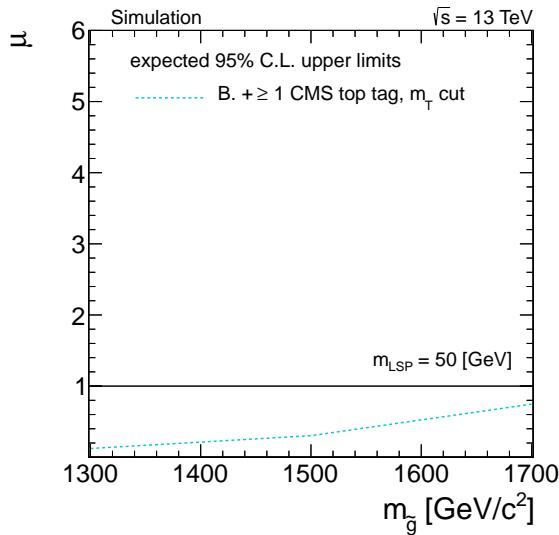


Figure 8.20: Expected 95% upper limit for signal strength versus  $m_{\tilde{g}}$ . The LSP mass is chosen to be 50 GeV.

production as long as top quarks with high transverse momenta are present in the final state. Such selections could serve as complementary approach to search strategies like those employed in chapter 7 for further supersymmetry searches at  $\sqrt{s} = 13$  TeV.

## 8.9 Results and Discussion

Selections based on different kinematic properties of direct stop pair production have been studied and analysis strategies were identified that allow to probe direct stop production up to 1 TeV in case of an LSP mass of 50 GeV (cf. Sec. 8.5). However, the sensitivity of this selection targeting large mass splittings between stop and LSP can be evaluated also for other mass scenarios. For illustration, the selected  $H_T$  and  $\cancel{E}_T$  distributions for the baseline selection, at least one CMS top tag and  $m_T > 180$  GeV are shown in Fig. 8.21 for a stop 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 stop quark 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 stop 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. 8.22 in the  $m_{\tilde{t}}$  versus  $m_{\text{LSP}}$  plane. The selections illustrated there are baseline requirements adding at least one CMS top tag and  $m_T > 180$  GeV (light blue) discussed in Sec. 8.5 and 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 stop 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

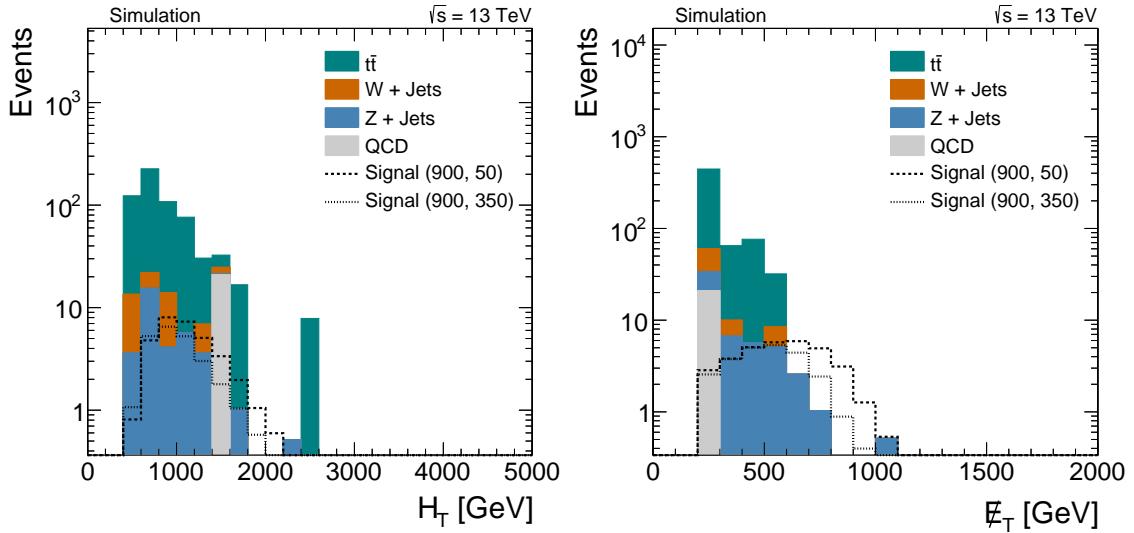


Figure 8.21: Comparison of selected  $H_T$  (*left*) and  $E_T$  (*right*) distributions in simulated events found from applying the baseline selection criteria, at least one CMS top tag and  $m_T > 180$  GeV. The signal points for direct stop production are labelled as (X, Y) where X is the stop mass and Y is the LSP mass in GeV.

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 stop masses up to 700–800 GeV also for other LSP masses than 50 GeV, the sensitivity towards larger stop masses is not achieved. These studies suggest that in order to test the whole stop 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  $\sqrt{s} = 8$  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 shown in Fig. 8.8 show a similar performance to those quantities at  $\sqrt{s} = 8$  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 [199], 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

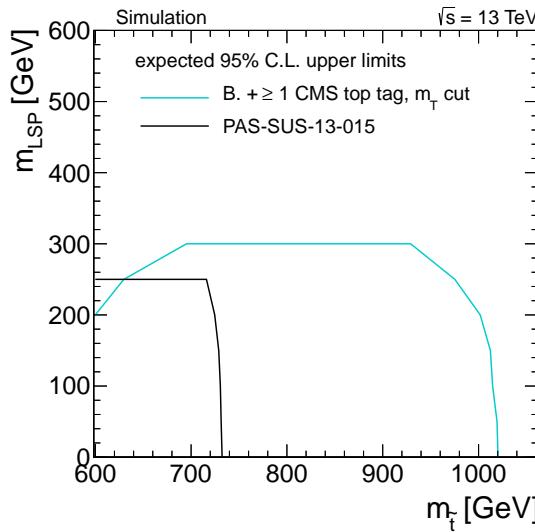


Figure 8.22: Expected 95% upper limit for  $m_{\tilde{t}}$  versus  $m_{\text{LSP}}$ .

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  $\sqrt{s} = 13$  TeV will shed light on the question if there is a realization of supersymmetry manifesting in top squarks below the TeV mass range.

## 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 within 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  $p_T^{\text{ave}}$  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  $1.079 \pm 0.026$  in the central region up to  $1.395 \pm 0.063$  for  $2.8 \leq |\eta| \leq 3.2$  and drops again for the outermost region  $3.2 \leq |\eta| \leq 5.0$  down to  $1.056 \pm 0.191$ . A distinct reason for that particular differences between data and simulation could not yet be identified, but in general noise effects, inhomogeneities 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 respective results is currently in preparation [208].

In the second part of this thesis, a search for supersymmetry in proton-proton collisions at  $\sqrt{s} = 8 \text{ TeV}$  in data corresponding to an integrated luminosity of  $19.5 \text{ fb}^{-1}$  has been presented. This is based on events with jet multiplicities [3, 5], [6, 7],  $\geq 8$  and large values of hadronic energy and missing transverse momentum. 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. Here, special emphasis is put on 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% CL for LSP masses not exceeding 100 GeV. The respective analysis is published in [164].

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  $\sqrt{s} = 8$  TeV. With this data, direct production 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  $\sqrt{s} = 13$  TeV. These data will open a to date unexplored parameter space also for the direct stop quark production. In order to extend the mass reach of respective analyses to top squark masses at 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  $\sqrt{s} = 13$  TeV. Selection criteria could be identified extending the mass reach of direct stop searches up to roughly 1 TeV for LSP masses below 100–200 GeV. It has been shown that the proposed selections are even more sensitive than a selection following closely the search for direct stop production at  $\sqrt{s} = 8$  TeV [199]. 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

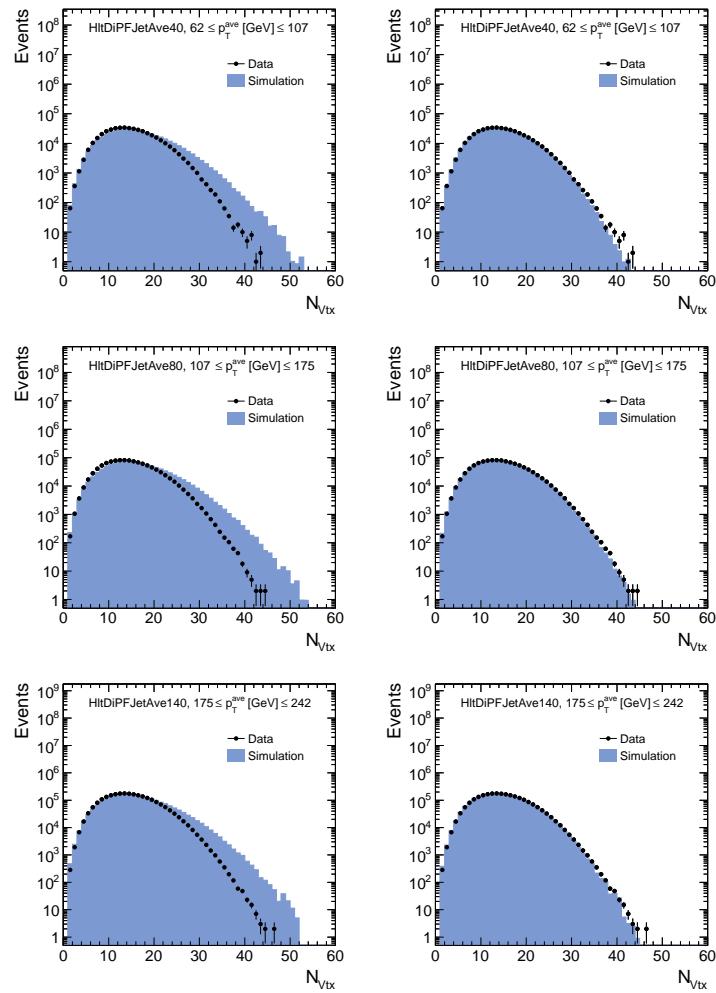


Figure A.1: Distribution of number of primary vertices in data (black dots) and simulation (blue histogram) before (left) and after (right) reweighting of the pile-up scenario in simulation for trigger paths HltDiPFJetAve40 (top), HltDiPFJetAve80 (middle) and HltDiPFJetAve140 (bottom).

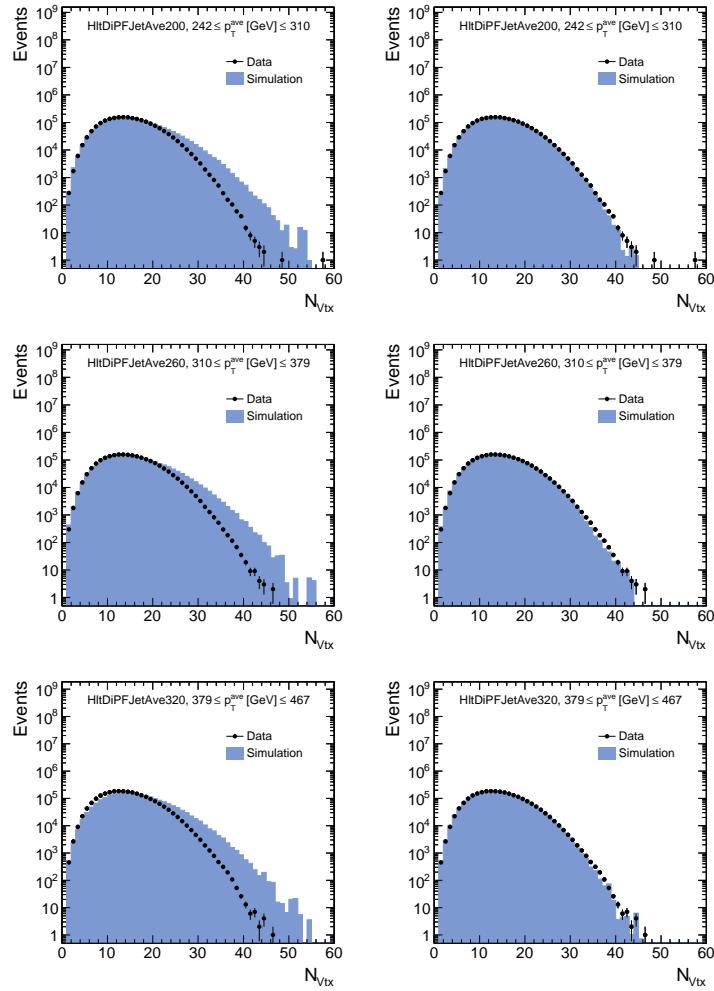


Figure A.2: Distribution of number of primary vertices in data (black dots) and simulation (blue histogram) before (*left*) and after (*right*) reweighting of the pile-up scenario in simulation for trigger paths HltDiPFJetAve200 (*top*), HltDiPFJetAve260 (*middle*) and HltDiPFJetAve320 (*bottom*).

## A.2 Extrapolation Graphs

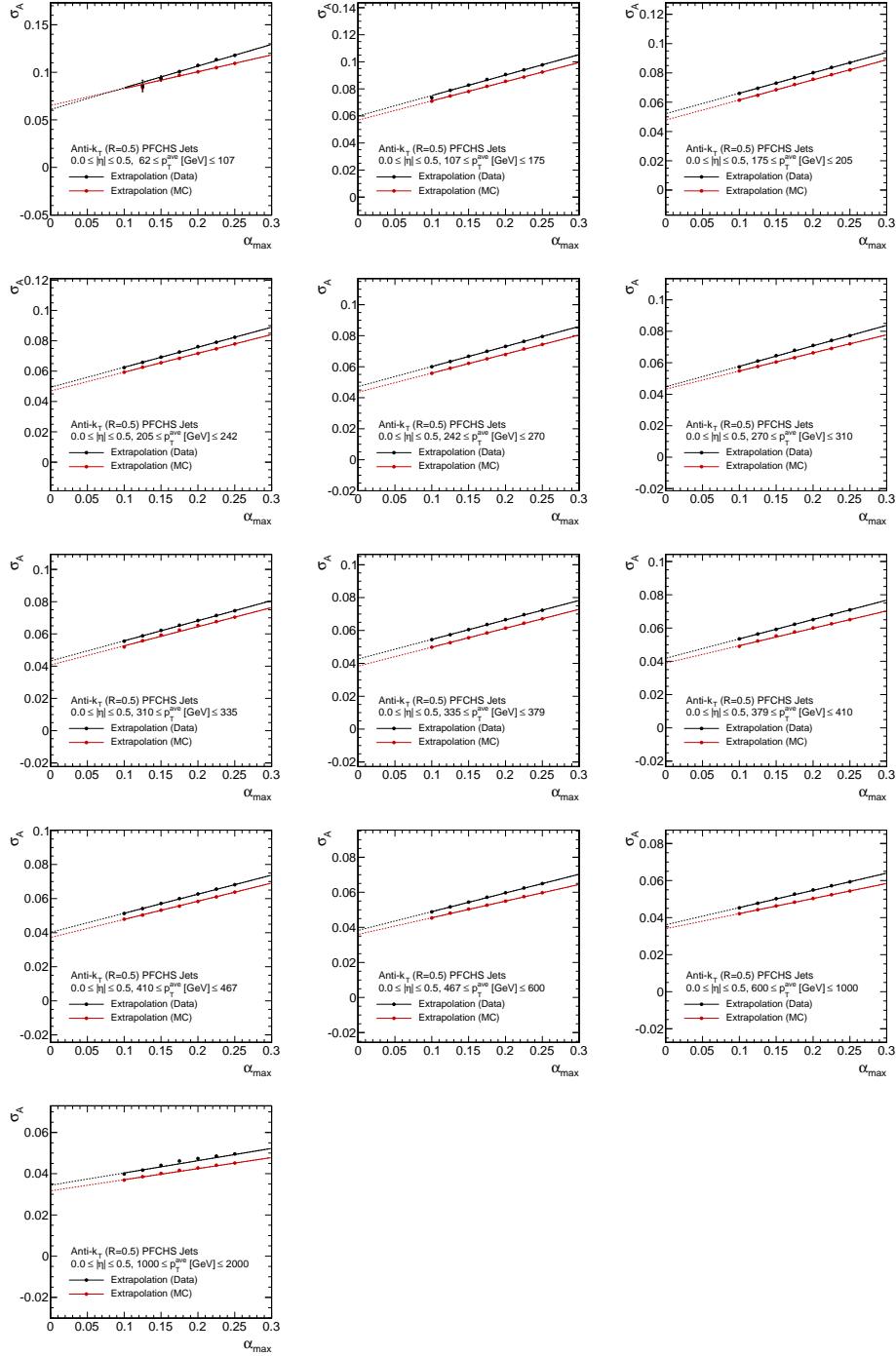


Figure A.3: Extrapolations of measured values for  $\sigma_A$  in data and simulation to obtain the result at zero additional jet activity for  $0.0 \leq |\eta| \leq 0.5$ .

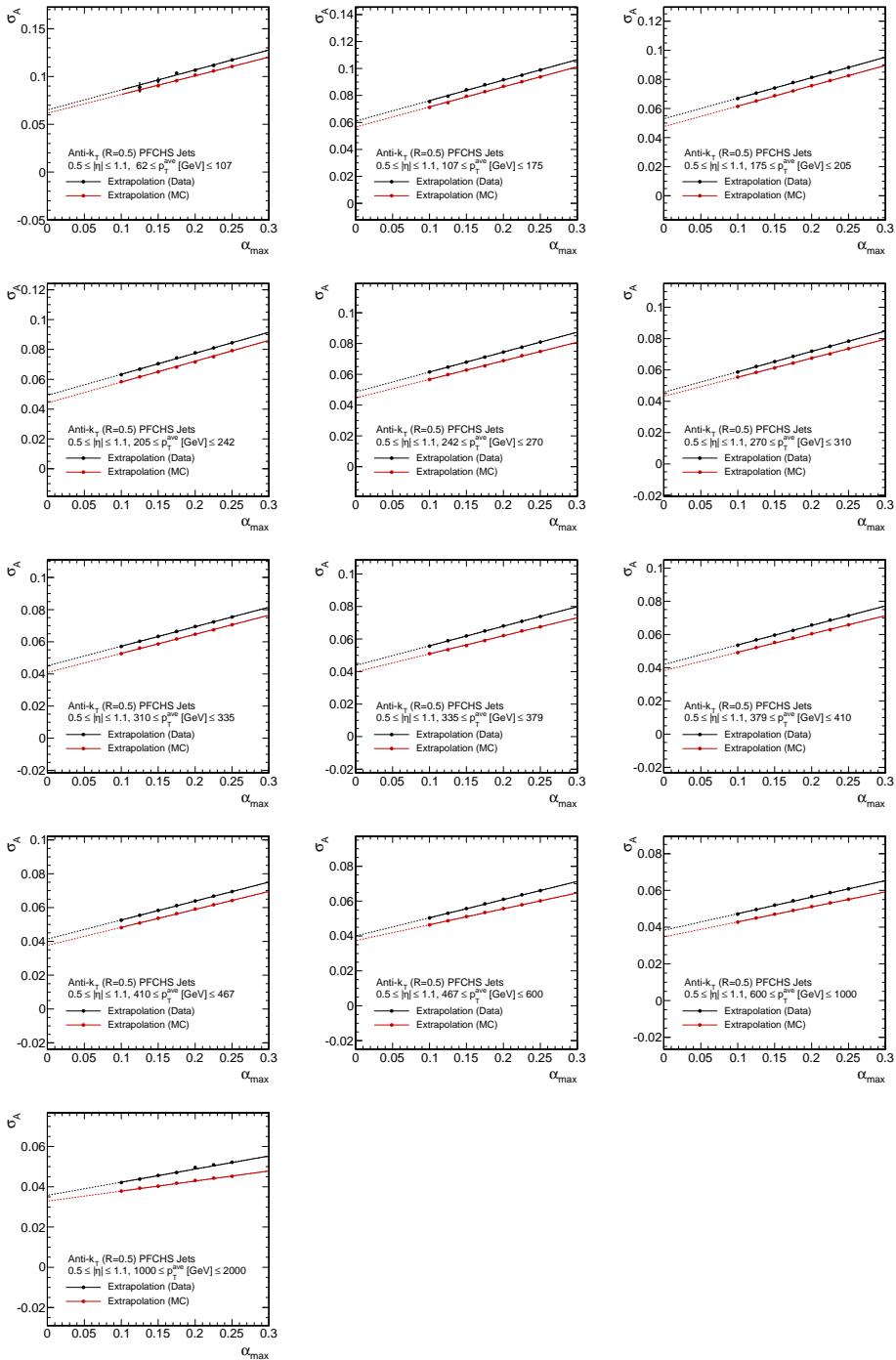


Figure A.4: Extrapolations of measured values for  $\sigma_A$  in data and simulation to obtain the result at zero additional jet activity for  $0.5 \leq |\eta| \leq 1.1$ .

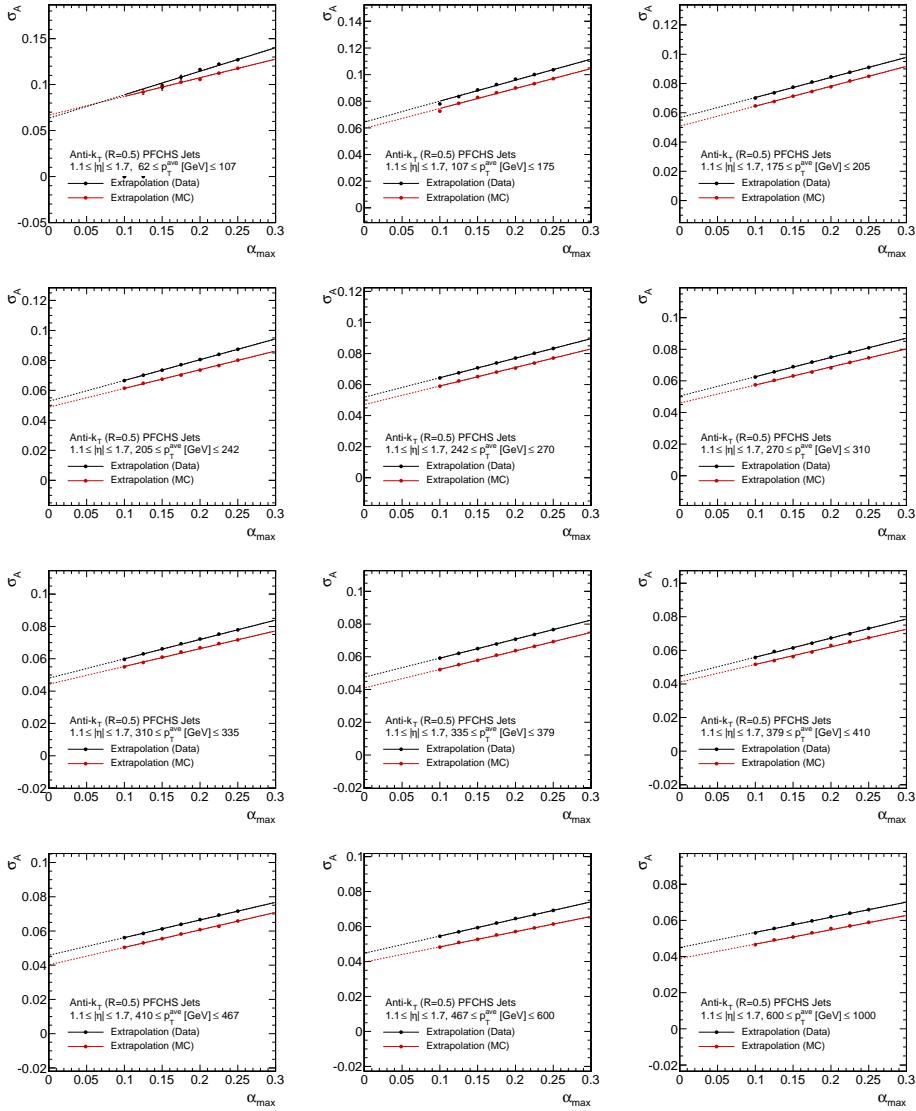


Figure A.5: Extrapolations of measured values for  $\sigma_A$  in data and simulation to obtain the result at zero additional jet activity for  $1.1 \leq |\eta| \leq 1.7$ .

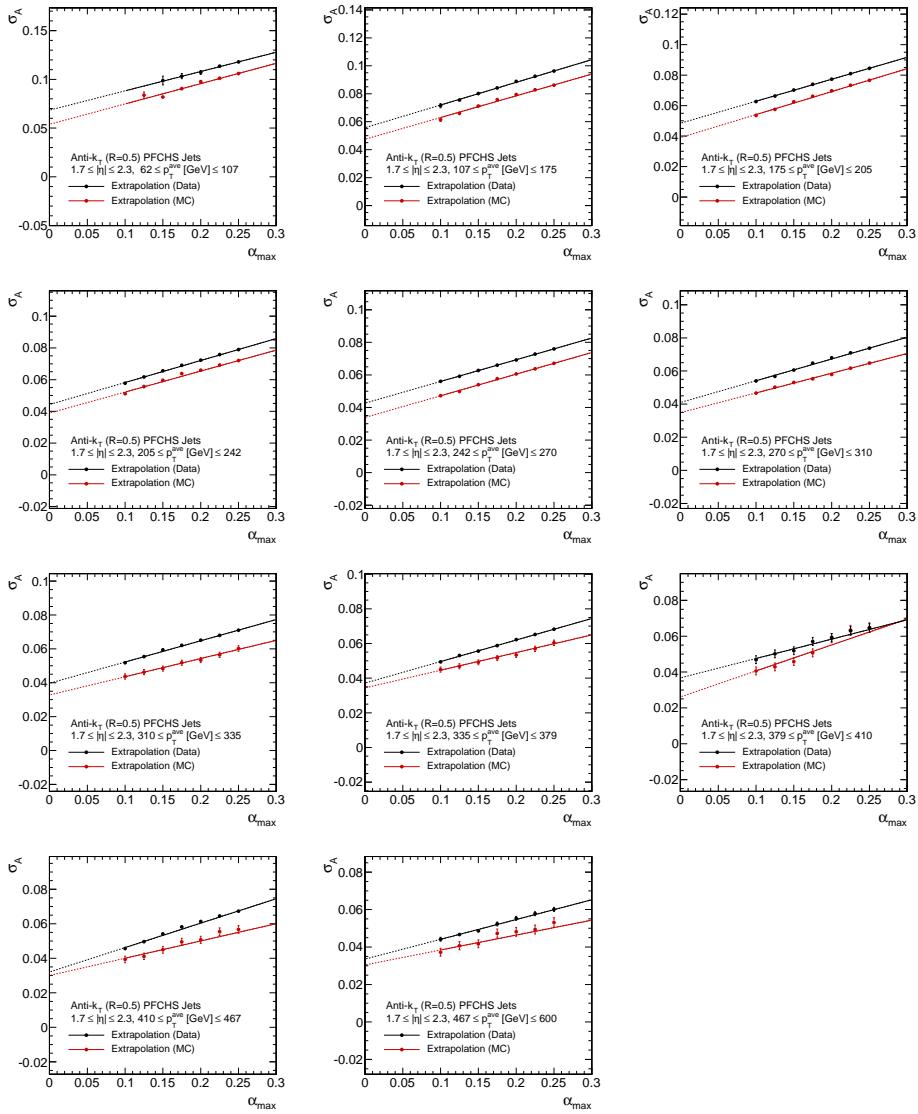


Figure A.6: Extrapolations of measured values for  $\sigma_A$  in data and simulation to obtain the result at zero additional jet activity for  $1.7 \leq |\eta| \leq 2.3$ .

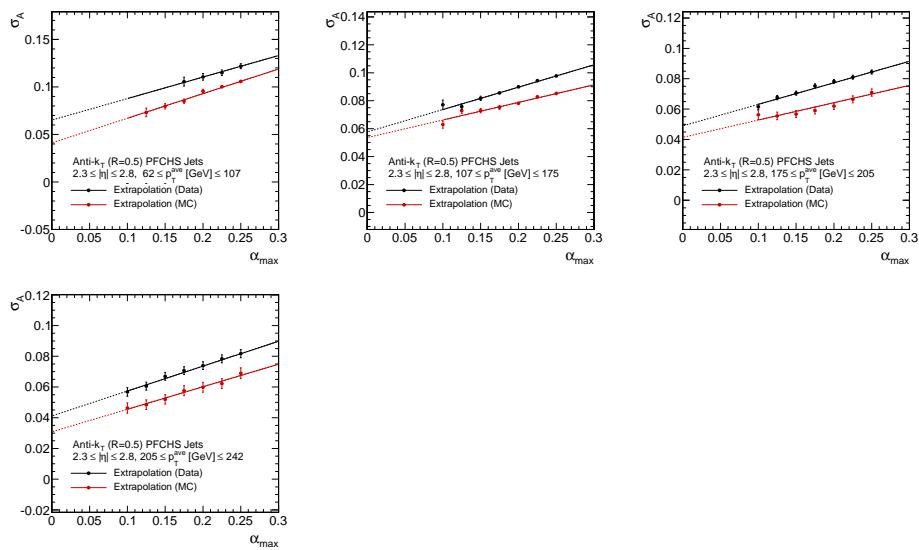


Figure A.7: Extrapolations of measured values for  $\sigma_A$  in data and simulation to obtain the result at zero additional jet activity for  $2.3 \leq |\eta| \leq 2.8$ .

### A.3 Correlation of Asymmetry Widths

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

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

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

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

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

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

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

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

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

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

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

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

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

with the variance-covariance matrix  $\Sigma^f$  for function  $f$  and the variance-covariance matrix  $\Sigma^x$  for variables  $x$ .

Applying this to the function

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

with

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

gives the relation

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

Consequently, it is

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

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

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

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

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

the result is determined to be

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

which is the same as expressed in Eq. 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  $c(\text{Data}/\text{MC})$  with absolute statistical uncertainty and systematic uncertainty for each uncertainty source in different  $|\eta|$  regions.

|                            | $ \eta_{\text{ref}}  \in [0.0, 0.5]$<br>$ \eta_{\text{probe}} $ |             |             |             |             |             |             |
|----------------------------|---|-------------|-------------|-------------|-------------|-------------|-------------|
|                            | 0.0–0.5   | 0.5–1.1     | 1.1–1.7     | 1.7–2.3     | 2.3–2.8     | 2.8–3.2     | 3.2–5.0     |
| $c(\text{Data}/\text{MC})$ | —   | 1.106       | 1.133       | 1.227       | 1.253       | 1.410       | 1.171       |
| Stat. uncertainty          | —   | $\pm 0.008$ | $\pm 0.009$ | $\pm 0.025$ | $\pm 0.047$ | $\pm 0.068$ | $\pm 0.116$ |
| 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       |
| $\alpha$ -spectrum         | —   | 0.010       | 0.006       | 0.011       | 0.006       | 0.009       | 0.012       |
| $\alpha$ -range            | —   | 0.010       | 0.014       | 0.046       | 0.089       | 0.015       | 0.018       |
| Non-Gaussian tails         | —   | 0.001       | 0.007       | 0.005       | 0.044       | 0.046       | 0.029       |
| Jet Flavour                | —   | 0.004       | 0.004       | 0.001       | 0.021       | 0.013       | 0.008       |
| Ratio shape                | —   | 0.022       | 0.023       | 0.025       | 0.025       | 0.028       | 0.023       |
| Total syst. uncertainty    | —   | $\pm 0.028$ | $\pm 0.030$ | $\pm 0.058$ | $\pm 0.112$ | $\pm 0.067$ | $\pm 0.079$ |

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

|                            | $ \eta_{\text{ref}}  \in [0.5, 1.1]$<br>$ \eta_{\text{probe}} $ |         |             |             |             |             |             |
|----------------------------|---|---------|-------------|-------------|-------------|-------------|-------------|
|                            | 0.0–0.5   | 0.5–1.1 | 1.1–1.7     | 1.7–2.3     | 2.3–2.8     | 2.8–3.2     | 3.2–5.0     |
| $c(\text{Data}/\text{MC})$ | 1.081   | —       | 1.111       | 1.206       | 1.300       | 1.356       | 0.829       |
| Stat. uncertainty          | $\pm 0.008$   | —       | $\pm 0.009$ | $\pm 0.023$ | $\pm 0.047$ | $\pm 0.058$ | $\pm 0.082$ |
| 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       |
| $\alpha$ -spectrum         | 0.007   | —       | 0.004       | 0.003       | 0.008       | 0.036       | 0.002       |
| $\alpha$ -range            | 0.004   | —       | 0.012       | 0.016       | 0.043       | 0.048       | 0.041       |
| Non-Gaussian tails         | 0.004   | —       | 0.008       | 0.010       | 0.061       | 0.023       | 0.012       |
| Jet Flavour                | 0.005   | —       | 0.007       | 0.009       | 0.013       | 0.042       | 0.007       |
| Ratio shape                | 0.022   | —       | 0.022       | 0.024       | 0.026       | 0.027       | 0.017       |
| Total syst. uncertainty    | $\pm 0.026$   | —       | $\pm 0.030$ | $\pm 0.039$ | $\pm 0.082$ | $\pm 0.105$ | $\pm 0.149$ |

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

|                            | $ \eta_{\text{ref}}  \in [1.1, 1.7]$<br>$ \eta_{\text{probe}} $ |             |         |             |             |             |             |
|----------------------------|---|-------------|---------|-------------|-------------|-------------|-------------|
|                            | 0.0–0.5   | 0.5–1.1     | 1.1–1.7 | 1.7–2.3     | 2.3–2.8     | 2.8–3.2     | 3.2–5.0     |
| $c(\text{Data}/\text{MC})$ | 1.084   | 1.082       | —       | 1.189       | 1.250       | 1.432       | 1.137       |
| Stat. uncertainty          | $\pm 0.012$   | $\pm 0.012$ | —       | $\pm 0.031$ | $\pm 0.051$ | $\pm 0.066$ | $\pm 0.105$ |
| 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       |
| $\alpha$ -spectrum         | 0.006   | 0.008       | —       | 0.019       | 0.006       | 0.019       | 0.009       |
| $\alpha$ -range            | 0.016   | 0.022       | —       | 0.036       | 0.032       | 0.016       | 0.048       |
| Non-Gaussian tails         | 0.010   | 0.010       | —       | 0.005       | 0.018       | 0.012       | 0.047       |
| Jet Flavour                | 0.005   | 0.007       | —       | 0.012       | 0.015       | 0.031       | 0.054       |
| Ratio shape                | 0.022   | 0.022       | —       | 0.024       | 0.025       | 0.029       | 0.023       |
| Total syst. uncertainty    | $\pm 0.033$   | $\pm 0.036$ | —       | $\pm 0.063$ | $\pm 0.065$ | $\pm 0.077$ | $\pm 0.096$ |

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

### B.1 Signal Trigger Efficiencies

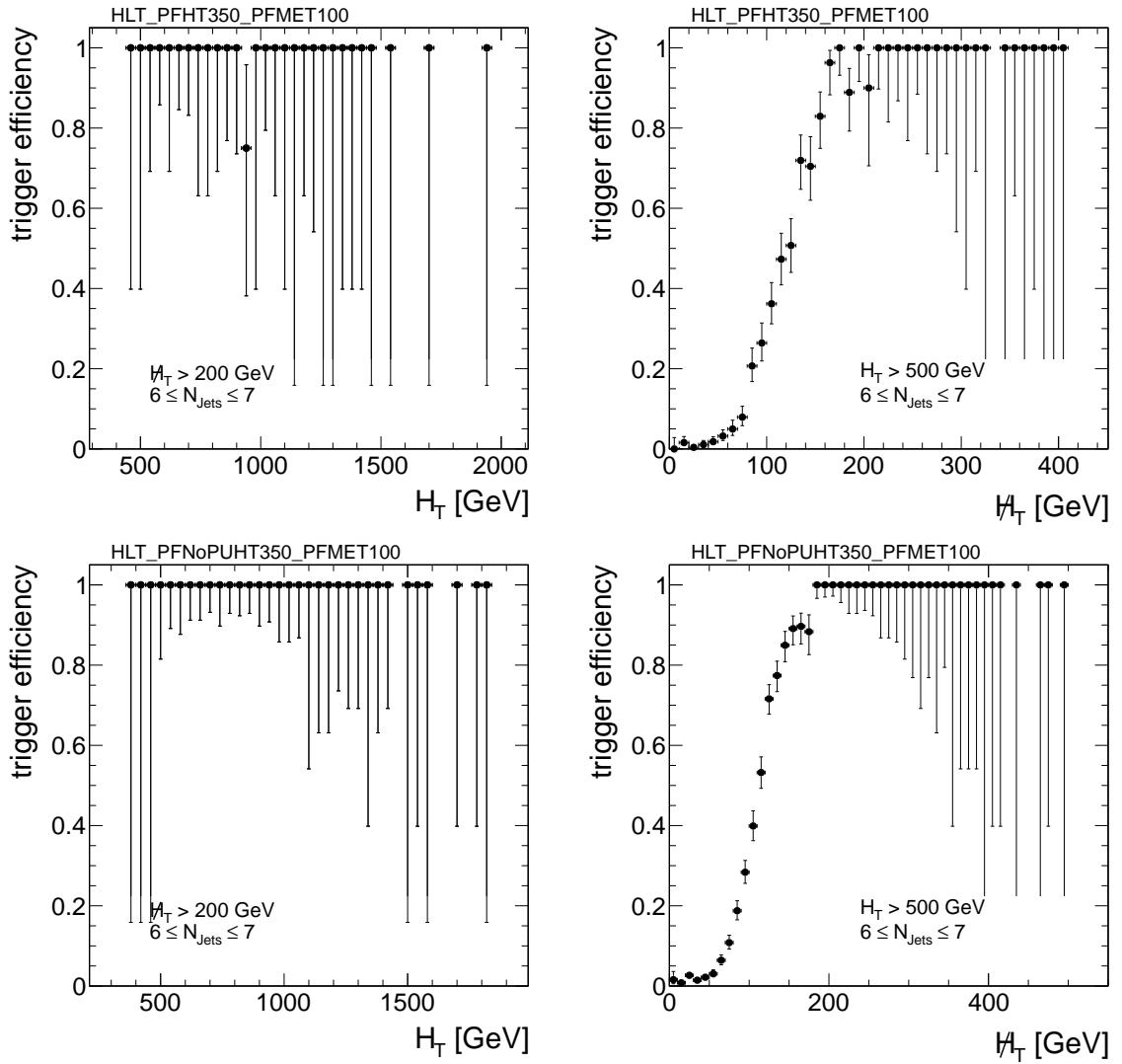


Figure B.1: Measured relative trigger efficiency for paths HLT\_PFHT350\_PFMET100 (*top*) and HLT\_PFNNoPUHT350\_PFMET100 (*bottom*) as a function of  $H_T$  (*left*) and  $\cancel{H}_T$  (*right*) shown for  $6 \leq N_{\text{jets}} \leq 7$ .

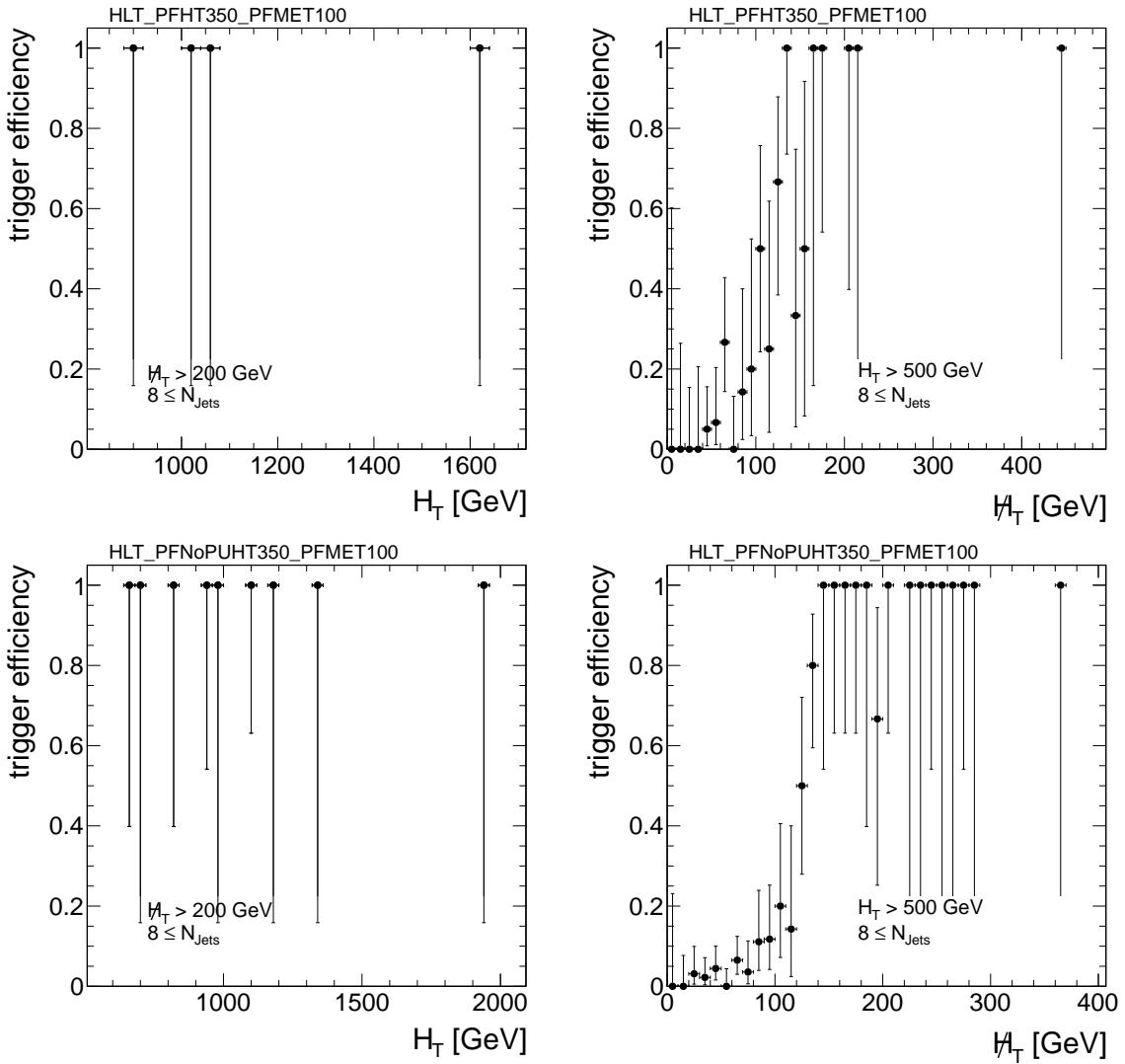


Figure B.2: Measured relative trigger efficiency for paths `HLT_PFT350_PFMET100` (*top*) and `HLT_PFNopUHT350_PFMET100` (*bottom*) as a function of  $H_T$  (*left*) and  $\cancel{H}_T$  (*right*) shown for  $8 \leq N_{\text{jets}}$ .

## B.2 Data-to-Simulation Ratios: Core

Table B.1: Data-to-simulation ratio in various  $|\eta|$  regions with total uncertainty as used in the QCD background prediction with the R+S method for the correction of the core response.

| $ \eta $  | $c(\text{Data}/\text{MC})$ |
|-----------|----------------------------|
| 0.0 – 0.5 | $1.052^{+0.063}_{-0.062}$  |
| 0.5 – 1.1 | $1.057^{+0.057}_{-0.056}$  |
| 1.1 – 1.7 | $1.096^{+0.065}_{-0.064}$  |
| 1.7 – 2.3 | $1.134^{+0.094}_{-0.092}$  |
| 2.3 – 5.0 | $1.288^{+0.020}_{-0.020}$  |

### B.3 Data-to-Simulation Ratios: Tails

Table B.2: Data-to-simulation ratio  $\rho_{\text{tail}}$  of the fractional tail-size together with the total uncertainty in different  $|\eta| \times p_T^{\text{ave}}$  intervals for the tail regions derived from the dijet asymmetry parts ( $A > 2\sigma_c$ ) used in the QCD background prediction with the R+S method for the correction of the tail parts of the response [144].

| $ \eta $  | $p_T^{\text{ave}} [\text{GeV}]$ | $\rho_{\text{tail}} (\mathcal{A} = 2\sigma_c)$ |
|-----------|---------------------------------|--|
| 0.0 – 0.5 | 45 – 220                        | $0.953 \pm 0.283$                              |
| 0.0 – 0.5 | 220 – 270                       | $1.418 \pm 0.502$                              |
| 0.0 – 0.5 | 270 – 312                       | $1.156 \pm 0.349$                              |
| 0.0 – 0.5 | 312 – 360                       | $1.305 \pm 0.350$                              |
| 0.0 – 0.5 | 360 – 498                       | $1.342 \pm 0.393$                              |
| 0.0 – 0.5 | 498 – 1500                      | $1.353 \pm 0.350$                              |
| 0.5 – 1.1 | 45 – 220                        | $1.096 \pm 0.374$                              |
| 0.5 – 1.1 | 220 – 294                       | $1.083 \pm 0.372$                              |
| 0.5 – 1.1 | 294 – 360                       | $1.195 \pm 0.325$                              |
| 0.5 – 1.1 | 360 – 1500                      | $1.248 \pm 0.424$                              |
| 1.1 – 1.7 | 45 – 220                        | $0.965 \pm 0.333$                              |
| 1.1 – 1.7 | 220 – 335                       | $1.035 \pm 0.295$                              |
| 1.1 – 1.7 | 335 – 1500                      | $1.358 \pm 0.327$                              |
| 1.7 – 2.3 | 45 – 220                        | $0.938 \pm 0.286$                              |
| 1.7 – 2.3 | 220 – 1500                      | $1.196 \pm 0.425$                              |
| 2.3 – 5.0 | 45 – 220                        | $1.069 \pm 0.77$                               |

## B.4 Input Truth Resolutions for Rebalancing in Kinematic Fit

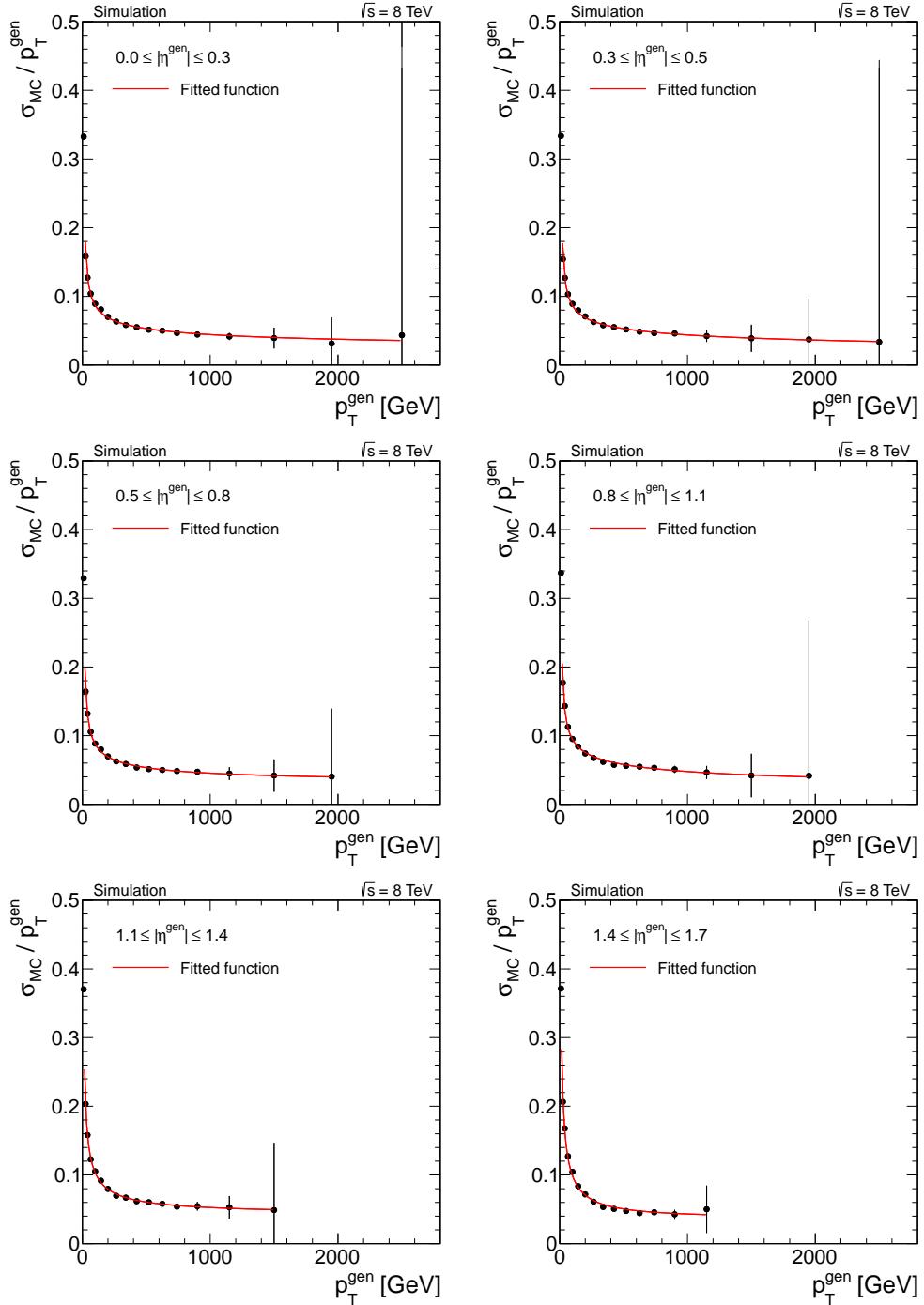


Figure B.3: Relative truth- $p_T$  resolution derived from simulated events shown as a function of  $p_T^{\text{gen}}$ . The distribution is fitted with a function as described in the next used as input for the kinematic fit employed to gain a balanced seed sample.

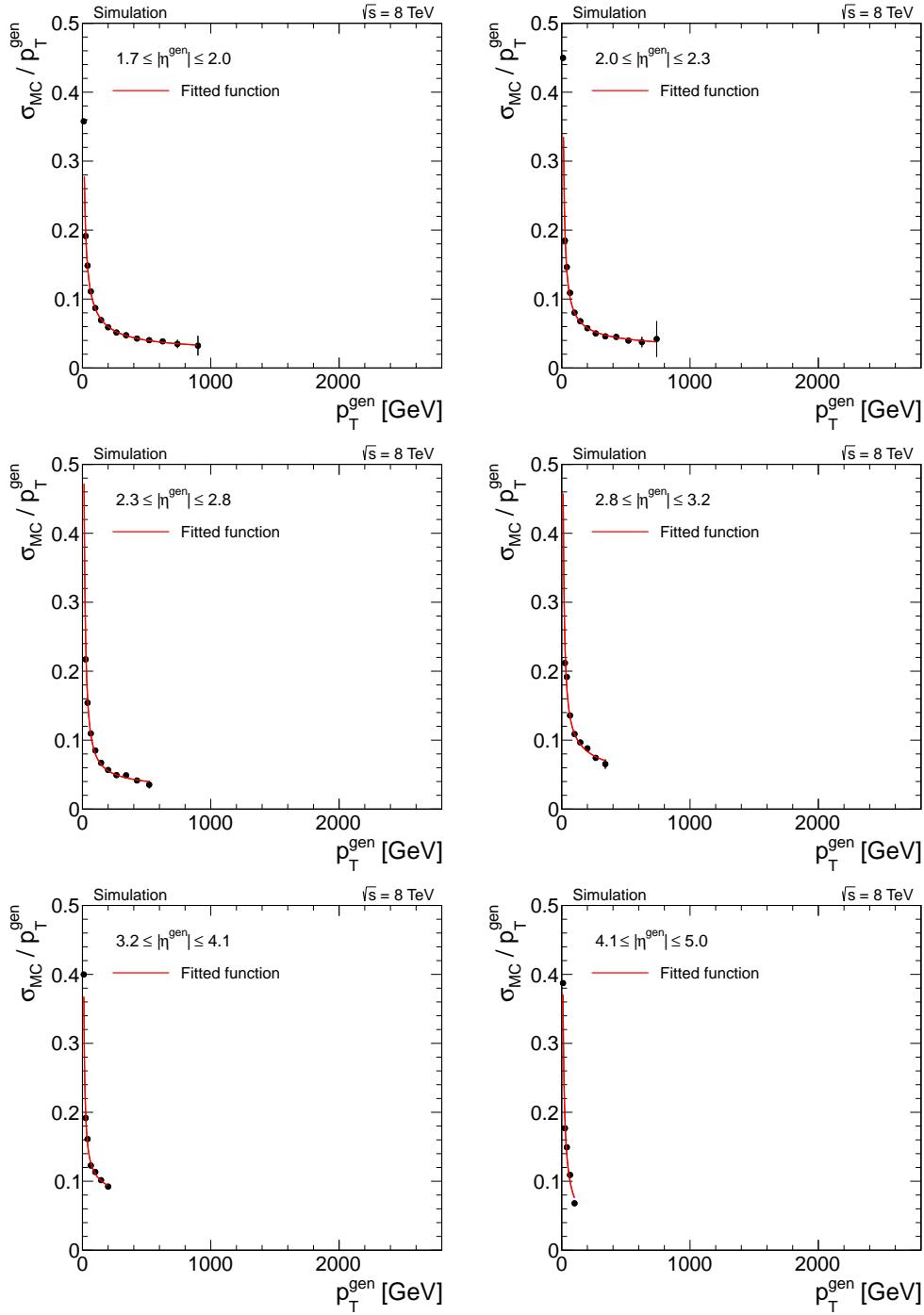


Figure B.4: Continued from Fig. B.3.

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

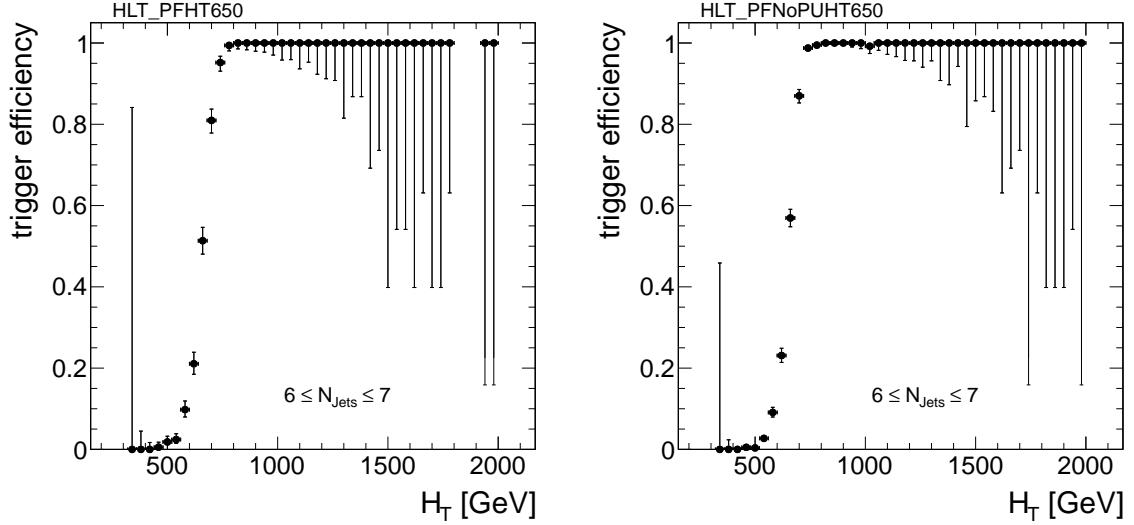


Figure B.5: Measured trigger efficiency for paths HLT\_PFHT650 (*left*) and HLT\_PFNopUHT650 (*right*) as a function of  $H_T$  illustrated for  $6 \leq N_{\text{Jets}} \leq 7$ .

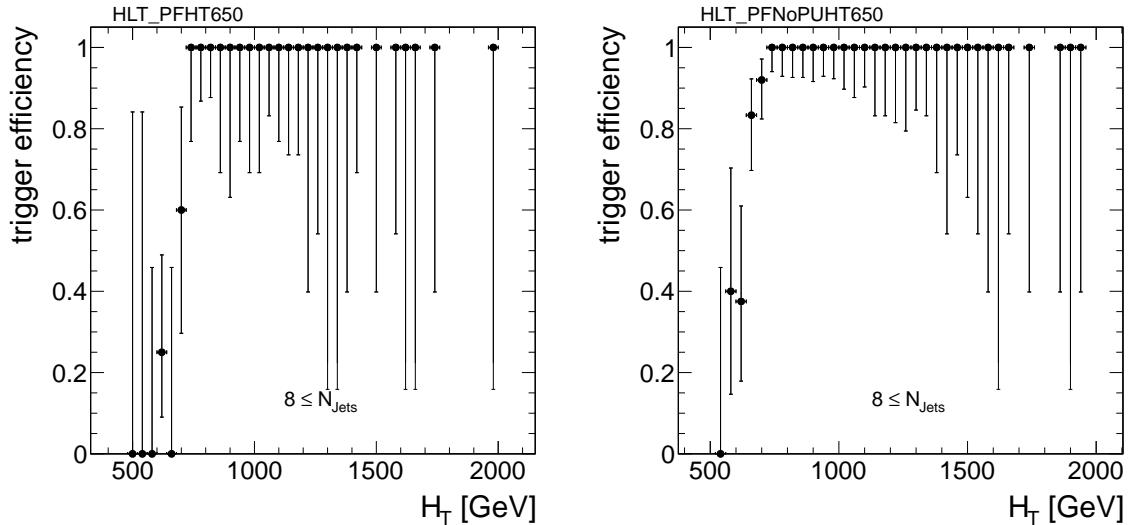


Figure B.6: Measured trigger efficiency for paths HLT\_PFHT650 (*left*) and HLT\_PFNopUHT650 (*right*) as a function of  $H_T$  illustrated for  $8 \leq N_{\text{Jets}}$ .



## 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](https://doi.org/10.1088/1674-1137/38/9/090001).
- [2] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10** (Jun, 1963) 531 – 533. doi:[10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531).
- [3] M. Kobayashi and T. Maskawa, “ $CP$ -Violation in the Renormalizable Theory of Weak Interaction”, *Progress of Theoretical Physics* **49** (1973), no. 2, 652–657. doi:[10.1143/PTP.49.652](https://doi.org/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](https://doi.org/10.1143/PTP.28.870).
- [5] B. Pontecorvo, “Neutrino Experiments and the Problem of Conservation of Leptonic Charge”, *Sov.Phys.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](https://arxiv.org/abs/hep-ex/9807003). doi:[10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562).
- [7] R. Alkofer and J. Greensite, “Quark Confinement: The Hard Problem of Hadron Physics”, *J.Phys.* **G34** (2007) S3, arXiv:[hep-ph/0610365](https://arxiv.org/abs/hep-ph/0610365). doi:[10.1088/0954-3899/34/7/S02](https://doi.org/10.1088/0954-3899/34/7/S02).
- [8] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Phys. Rev. Lett.* **30** (Jun, 1973) 1346 – 1349. doi:[10.1103/PhysRevLett.30.1346](https://doi.org/10.1103/PhysRevLett.30.1346).
- [9] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Non-Abelian Gauge Theories”, *Phys. Rev. Lett.* **30** (Jun, 1973) 1343 – 1346. doi:[10.1103/PhysRevLett.30.1343](https://doi.org/10.1103/PhysRevLett.30.1343).
- [10] S. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl.Phys.* **22** (1961) 579–588. doi:[10.1016/0029-5582\(61\)90469-2](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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.* **B716** (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.* **B716** (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.* **B726** (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”, [arXiv:1408.5191](#).
- [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”, [arXiv:1408.7084](#).
- [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”. Lecture Notes in Physics. Springer, 2010. 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”, *Nuclear Physics B* **70** (1974), no. 1, 39 – 50. `doi:10.1016/0550-3213(74)90355-1.`
- [25] I. J. Aitchison, “Supersymmetry and the MSSM: An Elementary introduction”, [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  $p \rightarrow \nu K^+$  using 260kiloton · year data of Super-Kamiokande”, *Phys. Rev. D* **90** (Oct, 2014) 072005.  
`doi:10.1103/PhysRevD.90.072005.`

- [28] T. Ibrahim and P. Nath, “CP Violation From Standard Model to Strings”, *Rev.Mod.Phys.* **80** (2008) 577–631, [arXiv:0705.2008](https://arxiv.org/abs/0705.2008).  
[doi:10.1103/RevModPhys.80.577](https://doi.org/10.1103/RevModPhys.80.577).
- [29] J. Scherk, “An introduction to the theory of dual models and strings”, *Rev. Mod. Phys.* **47** (Jan, 1975) 123–164. [doi:10.1103/RevModPhys.47.123](https://doi.org/10.1103/RevModPhys.47.123).
- [30] M. J. Veltman, “The infrared - ultraviolet connection”, *Acta Phys. Polon. B* **12** (1981) 437.
- [31] R. Barbieri and G. Giudice, “Upper bounds on supersymmetric particle masses”, *Nuclear Physics B* **306** (1988), no. 1, 63 – 76.  
[doi:\[http://dx.doi.org/10.1016/0550-3213\\(88\\)90171-X\]\(http://dx.doi.org/10.1016/0550-3213\(88\)90171-X\)](http://dx.doi.org/10.1016/0550-3213(88)90171-X)
- [32] N. Craig, “The State of Supersymmetry after Run I of the LHC”, [arXiv:1309.0528](https://arxiv.org/abs/1309.0528).
- [33] A. Djouadi, “The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model”, *Phys.Rept.* **459** (2008) 1–241, [arXiv:hep-ph/0503173](https://arxiv.org/abs/hep-ph/0503173). [doi:10.1016/j.physrep.2007.10.005](https://doi.org/10.1016/j.physrep.2007.10.005).
- [34] S. Dimopoulos and D. W. Sutter, “The Supersymmetric flavor problem”, *Nucl.Phys.* **B452** (1995) 496–512, [arXiv:hep-ph/9504415](https://arxiv.org/abs/hep-ph/9504415).  
[doi:10.1016/0550-3213\(95\)00421-N](https://doi.org/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. [doi:10.1103/PhysRevLett.49.970](https://doi.org/10.1103/PhysRevLett.49.970).
- [36] L. Alvarez-Gaume, J. Polchinski, and M. B. Wise, “Minimal Low-Energy Supergravity”, *Nucl.Phys.* **B221** (1983) 495.  
[doi:10.1016/0550-3213\(83\)90591-6](https://doi.org/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.* **D49** (1994) 6173–6210, [arXiv:hep-ph/9312272](https://arxiv.org/abs/hep-ph/9312272).  
[doi:10.1103/PhysRevD.49.6173](https://doi.org/10.1103/PhysRevD.49.6173).
- [38] H. Baer, C. Balazs, A. Belyaev et al., “Updated constraints on the minimal supergravity model”, *JHEP* **0207** (2002) 050, [arXiv:hep-ph/0205325](https://arxiv.org/abs/hep-ph/0205325).
- [39] M. Dine and W. Fischler, “A Phenomenological Model of Particle Physics Based on Supersymmetry”, *Phys.Lett.* **B110** (1982) 227.  
[doi:10.1016/0370-2693\(82\)91241-2](https://doi.org/10.1016/0370-2693(82)91241-2).
- [40] L. Alvarez-Gaume, M. Claudson, and M. B. Wise, “Low-Energy Supersymmetry”, *Nucl.Phys.* **B207** (1982) 96. [doi:10.1016/0550-3213\(82\)90138-9](https://doi.org/10.1016/0550-3213(82)90138-9).
- [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, [arXiv:1302.3415](https://arxiv.org/abs/1302.3415).
- [42] J. Erler, “Tests of the Electroweak Standard Model”, *J.Phys.Conf.Ser.* **485** (2014) 012010, [arXiv:1209.3324](https://arxiv.org/abs/1209.3324). [doi:10.1088/1742-6596/485/1/012010](https://doi.org/10.1088/1742-6596/485/1/012010).

- [43] M. Ciuchini, E. Franco, S. Mishima et al., “Electroweak Precision Observables, New Physics and the Nature of a 126 GeV Higgs Boson”, *JHEP* **1308** (2013) 106, [arXiv:1306.4644](https://arxiv.org/abs/1306.4644). doi:10.1007/JHEP08(2013)106.
- [44] M. Baak et al., “The global electroweak fit at NNLO and prospects for the LHC and ILC”, *Eur.Phys.J.* **C74** (2014) 3046, [arXiv:1407.3792](https://arxiv.org/abs/1407.3792). doi:10.1140/epjc/s10052-014-3046-5.
- [45] CMS Collaboration, “Measurement of the  $B(s)$  to  $\mu^+ \mu^-$  branching fraction and search for  $B_0$  to  $\mu^+ \mu^-$  with the CMS Experiment”, *Phys.Rev.Lett.* **111** (2013) 101804, [arXiv:1307.5025](https://arxiv.org/abs/1307.5025). doi:10.1103/PhysRevLett.111.101804.
- [46] LHCb Collaboration, “Measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction and search for  $B^0 \rightarrow \mu^+ \mu^-$  decays at the LHCb experiment”, *Phys.Rev.Lett.* **111** (2013) 101805, [arXiv:1307.5024](https://arxiv.org/abs/1307.5024). doi:10.1103/PhysRevLett.111.101805.
- [47] CMS and LHCb Collaboration, “Combination of results on the rare decays  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  from the CMS and LHCb experiments”, Technical Report CMS-PAS-BPH-13-007. CERN-LHCb-CONF-2013-012, CERN, Geneva, Jun, 2014.
- [48] Muon G-2 Collaboration, “Final report of the muon E821 anomalous magnetic moment measurement at BNL”, *Phys. Rev.* **D73** (2006) 072003, [arXiv:hep-ex/0602035](https://arxiv.org/abs/hep-ex/0602035). doi:10.1103/PhysRevD.73.072003.
- [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](https://arxiv.org/abs/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(MZ)”, *Eur.Phys.J.* **C71** (2011) 1515, [arXiv:1010.4180](https://arxiv.org/abs/1010.4180). doi:10.1140/epjc/s10052-012-1874-8, 10.1140/epjc/s10052-010-1515-z.
- [51] D. G. Cerdeno and A. M. Green, “Direct detection of WIMPs”, *arXiv preprint arXiv:1002.1912* (2010).
- [52] 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](https://arxiv.org/abs/1012.4515). doi:10.1088/1475-7516/2012/10/E01, 10.1088/1475-7516/2011/03/051.
- [53] C. L. Bennett et al., “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results”, *Astrophysical Journal Supplement Series* (2012) [arXiv:1212.5226](https://arxiv.org/abs/1212.5226).
- [54] H1 Collaboration, “A Search for Selectrons and Squarks at HERA”, *Phys. Lett.* **B380** (1996) 461, [arXiv:hep-ex/9605002](https://arxiv.org/abs/hep-ex/9605002). doi:10.1016/0370-2693(96)00640-5.
- [55] ZEUS Collaboration, “Search for selectron and squark production in  $e^+ p$  collisions at HERA”, *Phys.Lett.* **B434** (1998) 214–230, [arXiv:hep-ex/9806019](https://arxiv.org/abs/hep-ex/9806019). doi:10.1016/S0370-2693(98)00817-X.

- [56] ALEPH Collaboration, “Absolute Mass Lower Limit for the Lightest Neutralino of the MSSM from  $e^+e^-$  data at  $\sqrt{s}$  up to 209 GeV”, *Phys. Lett. B* **583** (2004) 247. See also references therein. doi:10.1016/j.physletb.2003.12.066.
- [57] DELPHI Collaboration, “Searches for Supersymmetric Particles in  $e^+e^-$  Collisions up to 208 GeV and Interpretation of the Results within the MSSM”, *Eur. Phys. J. C* **31** (2003) 421, arXiv:hep-ex/0311019. See also references therein. doi:10.1140/epjc/s2003-01355-5.
- [58] L3 Collaboration, “Search for Scalar Leptons and Scalar Quarks at LEP”, *Phys. Lett. B* **580** (2004) 37, arXiv:hep-ex/0310007. See also references therein. doi:10.1016/j.physletb.2003.10.010.
- [59] OPAL Collaboration, “Search for Chargino and Neutralino Production at  $\sqrt{s} = 192$  GeV to 209 GeV at LEP”, *Eur. Phys. J. C* **35** (2004) 1, arXiv:hep-ex/0401026. See also references therein. doi:10.1140/epjc/s2004-01780-x.
- [60] ALEPH, DELPHI, L3 and OPAL Collaborations, “Joint SUSY Working Group”, *LEPSUSYWG/02-06-2* (2004). <http://lepsusy.web.cern.ch/lepsusy>.
- [61] 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.
- [62] DØ Collaboration, Abazov et al., “Search for squarks and gluinos in events with jets and missing transverse energy using  $2.1 \text{ fb}^{-1}$  of  $p\bar{p}$  collision data at  $\sqrt{s} = 1.96$  TeV”, *Phys. Lett. B* **660** (2008) 449, arXiv:0712.3805.
- [63] DØ Collaboration, “Search for associated production of charginos and neutralinos in the trilepton final state using  $2.3 \text{ fb}^{-1}$  of data”, *Phys. Lett. B* **680** (2009), no. 1, 34 – 43. doi:DOI: 10.1016/j.physletb.2009.08.011.
- [64] G. L. Kane and J. Leveille, “Experimental Constraints on Gluino Masses and Supersymmetric Theories”, *Phys.Lett. B* **112** (1982) 227. doi:10.1016/0370-2693(82)90968-6.
- [65] P. Harrison and C. Llewellyn Smith, “Hadroproduction of Supersymmetric Particles”, *Nucl.Phys. B* **213** (1983) 223. doi:10.1016/0550-3213(83)90510-2.
- [66] E. Reya and D. Roy, “Supersymmetric-particle production at p anti-p collider energies”, *Phys.Rev. D* **32** (1985) 645. doi:10.1103/PhysRevD.32.645.
- [67] S. Dawson, E. Eichten, and C. Quigg, “Search for Supersymmetric Particles in Hadron - Hadron Collisions”, *Phys.Rev. D* **31** (1985) 1581. doi:10.1103/PhysRevD.31.1581.
- [68] H. Baer and X. Tata, “Component Formulae for Hadroproduction of Left-handed and Right-handed Squarks”, *Phys.Lett. B* **160** (1985) 159. doi:10.1016/0370-2693(85)91484-4.
- [69] M. Kramer, A. Kulesza, R. van der Leeuw et al., “Supersymmetry production cross sections in  $pp$  collisions at  $\sqrt{s} = 7$  TeV”, arXiv:1206.2892.

- [70] 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”, [arXiv:1407.5066](https://arxiv.org/abs/1407.5066).
- [71] CMS Collaboration, “CMS Supersymmetry Physics Results”, *Public CMS Wiki* (2014).  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>, Topic revision: r279.
- [72] ATLAS, CDF, CMS and D0 Collaboration, “First combination of Tevatron and LHC measurements of the top-quark mass”, [arXiv:1403.4427](https://arxiv.org/abs/1403.4427).
- [73] CMS Collaboration, “Search for new physics with jets and missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 7$  TeV”, *Journal of High Energy Physics* **2011** (2011) 1–46, [arXiv:1106.4503](https://arxiv.org/abs/1106.4503). doi:10.1007/JHEP08(2011)155.
- [74] 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](https://arxiv.org/abs/1207.1898). doi:10.1103/PhysRevLett.109.171803.
- [75] N. Arkani-Hamed, P. Schuster, N. Toro et al., “MAMOSET: The Path from LHC Data to the New Standard Model via On-Shell Effective Theories”, [arXiv:hep-ph/0703088](https://arxiv.org/abs/hep-ph/0703088).
- [76] J. Alwall, P. Schuster, and N. Toro, “Simplified Models for a First Characterization of New Physics at the LHC”, *Phys.Rev.* **D79** (2009) 075020, [arXiv:0810.3921](https://arxiv.org/abs/0810.3921). doi:10.1103/PhysRevD.79.075020.
- [77] J. Alwall, M.-P. Le, M. Lisanti et al., “Model-Independent Jets plus Missing Energy Searches”, *Phys.Rev.* **D79** (2009) 015005, [arXiv:0809.3264](https://arxiv.org/abs/0809.3264). doi:10.1103/PhysRevD.79.015005.
- [78] CMS Collaboration, “Interpretation of Searches for Supersymmetry with simplified Models”, *Phys.Rev.* **D88** (2013), no. 5, 052017, [arXiv:1301.2175](https://arxiv.org/abs/1301.2175). doi:10.1103/PhysRevD.88.052017.
- [79] MSSM Working Group Collaboration, “The Minimal supersymmetric standard model: Group summary report”, [arXiv:hep-ph/9901246](https://arxiv.org/abs/hep-ph/9901246).
- [80] CMS Collaboration, “Phenomenological MSSM interpretation of the CMS 2011 5fb $^{-1}$  results”, *CMS Physics Analysis Summary CMS-PAS-SUS-12-030* (2013).
- [81] O. S. Brüning, P. Collier, P. Lebrun et al., “LHC Design Report”, volume 1: The LHC Main Ring. CERN, Geneva, 2004. CERN-2004-003-V-1.
- [82] L. Evans and P. Bryant, “LHC Machine”, *Journal of Instrumentation* **3** (2008), no. 08, S08001. doi:10.1088/1748-0221/3/08/S08001.
- [83] CERN, “LEP Design Report”, volume 2: The LEP Main Ring. CERN, Geneva, 1984. CERN-LEP-84-01.
- [84] CERN, “LHC: the guide (English version)”. CERN-Brochure-2009-003-Eng, 2009.

- [85] W. Stirling, “Parton luminosity and cross section plots”, Nov. 2014.  
<http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>.
- [86] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003. doi:10.1088/1748-0221/3/08/S08003.
- [87] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.
- [88] CMS Collaboration, “CMS Detector Performance and Software”, volume 1 of *CMS Physics Technical Design Report*. CERN, Geneva, 2006. CMS-TDR-008-1. CERN-LHCC-2006-001.
- [89] LHCb Collaboration, “The LHCb Detector at the LHC”, *JINST* **3** (2008) S08005. doi:10.1088/1748-0221/3/08/S08005.
- [90] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002. doi:10.1088/1748-0221/3/08/S08002.
- [91] CMS Collaboration, “The CMS tracker system project”, volume 5 of *Technical Design Report CMS*. CERN, Geneva, 1997. CMS-TDR-005. CERN-LHCC-98-006.
- [92] CMS HCAL Collaboration, “Studies of the response of the prototype CMS hadron calorimeter, including magnetic field effects, to pion, electron, and muon beams”, *Nucl.Instrum.Meth.* **A457** (2001) 75–100, arXiv:hep-ex/0007045. doi:10.1016/S0168-9002(00)00711-7.
- [93] CMS ECAL/HCAL Collaboration, “The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c”, *The European Physical Journal C - Particles and Fields* **60** (2009) 359–373. doi:10.1140/epjc/s10052-009-0959-5.
- [94] CMS Collaboration, “Electromagnetic calorimeter calibration with 7 TeV data”, *CMS Physics Analysis Summary CMS-PAS-EGM-10-003* (2010).
- [95] CMS Collaboration, “Luminosity Information for the 2012 Proton Run”, 2012.  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [96] “LHC machine outreach”, 2013.  
<http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/>.
- [97] M. Lamont, “Status of the LHC”, *J.Phys.Conf.Ser.* **455** (2013) 012001. doi:10.1088/1742-6596/455/1/012001.
- [98] N. Metropolis and S. Ulam, “The Monte Carlo Method”, *Journal of the American Statistical Association* **44** (1949), no. 247, 335–341. doi:10.2307/2280232.
- [99] M. H. Seymour and M. Marx, “Monte Carlo Event Generators”, arXiv:1304.6677.
- [100] 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.

- [101] M. Dobbs and J. Hansen, “HepMC 2 — a C++ Event Record for Monte Carlo Generators: User Manual Version 2.06”, *Comput. Phys. Commun.* **134** (2001) 41. <http://lcgapp.cern.ch/project/simu/HepMC>.
- [102] H1 and ZEUS Collaboration, “Combined Measurement and QCD Analysis of the Inclusive  $e^\pm p$  Scattering Cross Sections at HERA”, *JHEP* **01** (2010) 109, [arXiv:0911.0884](https://arxiv.org/abs/0911.0884). doi:10.1007/JHEP01(2010)109.
- [103] J. C. Collins and D. E. Soper, “Theorems of Perturbative QCD”, *Ann. Rev. Nucl. Part. Sci.* **37** (1987) 383–409. doi:10.1146/annurev.ns.37.120187.002123.
- [104] J. C. Collins, D. E. Soper, and G. F. Sterman, “Factorization of Hard Processes in QCD”, *Adv. Ser. Direct. High Energy Phys.* **5** (1988) 1–91, [arXiv:hep-ph/0409313](https://arxiv.org/abs/hep-ph/0409313).
- [105] B. Andersson, G. Gustafson, G. Ingelman et al., “Parton fragmentation and string dynamics”, *Physics Reports* **97** (1983), no. 2 - 3, 31 – 145. doi:[http://dx.doi.org/10.1016/0370-1573\(83\)90080-7](http://dx.doi.org/10.1016/0370-1573(83)90080-7).
- [106] D. Amati and G. Veneziano, “Preconfinement as a property of perturbative QCD”, *Physics Letters B* **83** (1979), no. 1, 87 – 92. doi:[http://dx.doi.org/10.1016/0370-2693\(79\)90896-7](http://dx.doi.org/10.1016/0370-2693(79)90896-7).
- [107] A. Bassutto, M. Ciafaloni, and G. Marchesini, “Color singlet distributions and mass damping in perturbative QCD”, *Physics Letters B* **83** (1979), no. 2, 207 – 212. doi:[http://dx.doi.org/10.1016/0370-2693\(79\)90687-7](http://dx.doi.org/10.1016/0370-2693(79)90687-7).
- [108] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **0605** (2006) 026, [arXiv:hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175). doi:10.1088/1126-6708/2006/05/026.
- [109] M. Bahr, S. Gieseke, M. Gigg et al., “Herwig++ Physics and Manual”, *Eur. Phys. J. C* **58** (2008) 639–707, [arXiv:0803.0883](https://arxiv.org/abs/0803.0883). doi:10.1140/epjc/s10052-008-0798-9.
- [110] J. Alwall, P. Demin, S. de Visscher et al., “MadGraph/MadEvent v4: The New Web Generation”, *JHEP* **0709** (2007) 028, [arXiv:0706.2334](https://arxiv.org/abs/0706.2334). doi:10.1088/1126-6708/2007/09/028.
- [111] J. Alwall, R. Frederix, S. Frixione et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **1407** (2014) 079, [arXiv:1405.0301](https://arxiv.org/abs/1405.0301). doi:10.1007/JHEP07(2014)079.
- [112] C. Oleari, “The POWHEG-BOX”, *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 36–41, [arXiv:1007.3893](https://arxiv.org/abs/1007.3893). doi:10.1016/j.nuclphysbps.2010.08.016.
- [113] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **0411** (2004) 040, [arXiv:hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146). doi:10.1088/1126-6708/2004/11/040.

- [114] S. Agostinelli, J. Allison, K. Amako et al., “G4 – a simulation toolkit”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250 – 303.  
[doi:10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [115] 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](https://doi.org/10.1088/1742-6596/331/3/032049).
- [116] CMS Collaboration, “Comparison of the Fast Simulation of CMS with the first LHC data”, Technical Report CMS-DP-2010-039, CERN, Geneva, Oct, 2010.
- [117] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and  $E_T^{\text{miss}}$ ”, *CMS Physics Analysis Summary CMS-PAS-PFT-09-001* (2009).
- [118] W. Adam, B. Mangano, T. Speer et al., “Track Reconstruction in the CMS tracker”, Technical Report CMS-NOTE-2006-041, CERN, Geneva, Dec, 2006.
- [119] G. P. Salam, “Towards Jetography”, *Eur.Phys.J.* **C67** (2010) 637–686, [arXiv:0906.1833](https://arxiv.org/abs/0906.1833). [doi:10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6).
- [120] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, [arXiv:1111.6097](https://arxiv.org/abs/1111.6097).
- [121] M. Cacciari and G. P. Salam, “Dispelling the  $N^3$  myth for the  $k_t$  jet-finder”, *Phys.Lett.* **B641** (2006) 57–61, [arXiv:hep-ph/0512210](https://arxiv.org/abs/hep-ph/0512210).  
[doi:10.1016/j.physletb.2006.08.037](https://doi.org/10.1016/j.physletb.2006.08.037).
- [122] G. P. Salam and G. Soyez, “A Practical Seedless Infrared-Safe Cone jet algorithm”, *JHEP* **0705** (2007) 086, [arXiv:0704.0292](https://arxiv.org/abs/0704.0292).  
[doi:10.1088/1126-6708/2007/05/086](https://doi.org/10.1088/1126-6708/2007/05/086).
- [123] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev. D* **48** (Oct, 1993) 3160 – 3166.  
[doi:10.1103/PhysRevD.48.3160](https://doi.org/10.1103/PhysRevD.48.3160).
- [124] Y. L. Dokshitzer, G. Leder, S. Moretti et al., “Better jet clustering algorithms”, *JHEP* **9708** (1997) 001, [arXiv:hep-ph/9707323](https://arxiv.org/abs/hep-ph/9707323).  
[doi:10.1088/1126-6708/1997/08/001](https://doi.org/10.1088/1126-6708/1997/08/001).
- [125] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross-sections in deep inelastic scattering”, [arXiv:hep-ph/9907280](https://arxiv.org/abs/hep-ph/9907280). DESY-THESIS-2000-049.
- [126] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_T$  jet clustering algorithm”, *Journal of High Energy Physics* **2008** (2008), no. 04, 063, [arXiv:0802.1189](https://arxiv.org/abs/0802.1189).  
[doi:10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).
- [127] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *Journal of Instrumentation* **6** (2011) P11002, [arXiv:1107.4277](https://arxiv.org/abs/1107.4277). [doi:10.1088/1748-0221/6/11/P11002](https://doi.org/10.1088/1748-0221/6/11/P11002).
- [128] CMS Collaboration, “Jet Energy Corrections and Uncertainties. Detector Performance Plots for 2012”, Technical Report CMS-DP-2012-012, CERN, Geneva, Jul, 2012.

- [129] CMS Collaboration, “The Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS”, *CMS Physics Analysis Summary CMS-PAS-JME-09-002* (2009).
- [130] CMS Collaboration, “8 TeV Jet Energy Corrections and Uncertainties based on  $19.8 \text{ fb}^{-1}$  of data in CMS”, Technical Report CMS-DP-2013-033, CERN, Geneva, Oct, 2013.
- [131] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Physics Letters B* **659** (2008), no. 1-2, 119 – 126. doi:10.1016/j.physletb.2007.09.077.
- [132] 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.
- [133] CMS Collaboration, “Performance of b tagging at  $\sqrt{s}=8 \text{ TeV}$  in multijet, ttbar and boosted topology events”, *CMS Physics Analysis Summary CMS-PAS-BTV-13-001* (2013).
- [134] J. M. Butterworth, A. R. Davison, M. Rubin et al., “Jet substructure as a new Higgs search channel at the LHC”, arXiv:0810.0409.
- [135] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Techniques for improved heavy particle searches with jet substructure”, *Phys.Rev.* **D80** (2009) 051501, arXiv:0903.5081. doi:10.1103/PhysRevD.80.051501.
- [136] D. Krohn, J. Thaler, and L.-T. Wang, “Jet Trimming”, *JHEP* **1002** (2010) 084, arXiv:0912.1342. doi:10.1007/JHEP02(2010)084.
- [137] CMS Collaboration, “Boosted Top Jet Tagging at CMS”, *CMS Physics Analysis Summary CMS-PAS-JME-13-007* (2014).
- [138] CMS Collaboration, “Boosted Top Jet Tagging at CMS”, Technical Report CMS-DP-2014-036, CERN, Geneva, 2014.
- [139] 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).
- [140] 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.
- [141] T. Plehn, M. Spannowsky, M. Takeuchi et al., “Stop Reconstruction with Tagged Tops”, *JHEP* **1010** (2010) 078, arXiv:1006.2833. doi:10.1007/JHEP10(2010)078.
- [142] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP* **1103** (2011) 015, arXiv:1011.2268. doi:10.1007/JHEP03(2011)015.
- [143] J. Thaler and K. Van Tilburg, “Maximizing Boosted Top Identification by Minimizing N-subjettiness”, *JHEP* **1202** (2012) 093, arXiv:1108.2701. doi:10.1007/JHEP02(2012)093.

- [144] M. Schröder, “Quality of Jet Measurements and Impact on a Search for New Physics at CMS”. PhD thesis, Universität Hamburg, Germany, 2012.  
DESY-THESIS-2012-042.
- [145] 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.
- [146] 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.
- [147] 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.
- [148] K. Goebel, J. Haller, J. Ott et al., “Jet Transverse Momentum Resolution Measurement using Dijet Events at  $\sqrt{s} = 8$  TeV”, *CMS Analysis Note CMS-AN-13-416* (2013). Internal documentation.
- [149] CMS Collaboration, “Measurement of the Underlying Event Activity at the LHC with  $\sqrt{s} = 7$  TeV and Comparison with  $\sqrt{s} = 0.9$  TeV”, *JHEP* **1109** (2011) 109,  
[arXiv:1107.0330](https://arxiv.org/abs/1107.0330). doi:[10.1007/JHEP09\(2011\)109](https://doi.org/10.1007/JHEP09(2011)109).
- [150] J. Pumplin, D. Stump, J. Huston et al., “New generation of parton distributions with uncertainties from global QCD analysis”, *JHEP* **0207** (2002) 012,  
[arXiv:hep-ph/0201195](https://arxiv.org/abs/hep-ph/0201195). doi:[10.1088/1126-6708/2002/07/012](https://doi.org/10.1088/1126-6708/2002/07/012).
- [151] D0 Collaboration, “High- $p_T$  jets in  $\bar{p}p$  collisions at  $\sqrt{s} = 630$  GeV and 1800 GeV”, *Phys.Rev.* **D64** (2001) 032003, [arXiv:hep-ex/0012046](https://arxiv.org/abs/hep-ex/0012046).  
doi:[10.1103/PhysRevD.64.032003](https://doi.org/10.1103/PhysRevD.64.032003).
- [152] D0 Collaboration, “Measurement of the inclusive jet cross section in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV”, *Phys.Rev.* **D85** (2012) 052006, [arXiv:1110.3771](https://arxiv.org/abs/1110.3771).  
doi:[10.1103/PhysRevD.85.052006](https://doi.org/10.1103/PhysRevD.85.052006).
- [153] ATLAS Collaboration, “Jet energy resolution in proton-proton collisions at  $\sqrt{s} = 7$  TeV recorded in 2010 with the ATLAS detector”, *Eur.Phys.J.* **C73** (2013) 2306,  
[arXiv:1210.6210](https://arxiv.org/abs/1210.6210). doi:[10.1140/epjc/s10052-013-2306-0](https://doi.org/10.1140/epjc/s10052-013-2306-0).
- [154] CMS Collaboration, “Measurement of Differential Jet Cross Sections at  $\sqrt{s} = 7$  TeV with the CMS Detector”, *CMS Physics Analysis Summary CMS-PAS-QCD-11-004* (2012).
- [155] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update”, *CMS Physics Analysis Summary CMS-PAS-LUM-13-001* (2013).
- [156] D. Rathjens, “PhD thesis in preparation”, 2014. University of Hamburg.
- [157] CMS Collaboration, “Calorimeter Jet Quality Criteria for the First CMS Collision Data”, *CMS Physics Analysis Summary CMS-PAS-JME-09-008* (2008).

- [158] CMS Collaboration, “Jet Performance in  $pp$  Collisions at 7 TeV”, *CMS Physics Analysis Summary CMS-PAS-JME-10-003* (2010).
- [159] 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.
- [160] CMS Collaboration, “Measurement of  $B\bar{B}$  Angular Correlations based on Secondary Vertex Reconstruction at  $\sqrt{s} = 7$  TeV”, *JHEP* **1103** (2011) 136, arXiv:1102.3194. doi:10.1007/JHEP03(2011)136.
- [161] CMS Collaboration, “Measurement of the top-quark mass in t t-bar events with lepton+jets final states in  $pp$  collisions at  $\text{sqrt}(s)=8$  TeV”, Technical Report CMS-PAS-TOP-14-001, CERN, Geneva, 2014.
- [162] D0 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.
- [163] 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.
- [164] 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* **1406** (2014) 055, arXiv:1402.4770. doi:10.1007/JHEP06(2014)055.
- [165] N. Kidonakis, “Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution”, *Phys.Rev.* **D82** (2010) 114030, arXiv:1009.4935. doi:10.1103/PhysRevD.82.114030.
- [166] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through O( $\alpha(s)^{**2}$ )”, *Phys.Rev.* **D74** (2006) 114017, arXiv:hep-ph/0609070. doi:10.1103/PhysRevD.74.114017.
- [167] W. Beenakker, R. Hopker, M. Spira et al., “Squark and gluino production at hadron colliders”, *Nucl.Phys.* **B492** (1997) 51–103, arXiv:hep-ph/9610490. doi:10.1016/S0550-3213(97)80027-2.
- [168] A. Kulesza and L. Motyka, “Threshold Resummation for Squark-Antisquark and Gluino-Pair Production at the LHC”, *Phys. Rev. Lett.* **102** (Mar, 2009) 111802. doi:10.1103/PhysRevLett.102.111802.
- [169] A. Kulesza and L. Motyka, “Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC”, *Phys. Rev. D* **80** (Nov, 2009) 095004. doi:10.1103/PhysRevD.80.095004.
- [170] W. Beenakker, S. Breuning, M. Kramer et al., “Soft-gluon resummation for squark and gluino hadroproduction”, *JHEP* **0912** (2009) 041, arXiv:0909.4418. doi:10.1088/1126-6708/2009/12/041.

- [171] W. Beenakker, S. Brensing, M. Kramer et al., “Squark and Gluino Hadroproduction”, *Int.J.Mod.Phys. A* **26** (2011) 2637–2664, [arXiv:1105.1110](https://arxiv.org/abs/1105.1110).  
[doi:10.1142/S0217751X11053560](https://doi.org/10.1142/S0217751X11053560).
- [172] CMS Collaboration, “Electron reconstruction and identification at  $\sqrt{s} = 7 \text{ TeV}$ ”, *CMS Physics Analysis Summary CMS-PAS-EGM-10-004* (2010).
- [173] CMS Collaboration, “Performance of muon identification in  $pp$  collisions at  $\sqrt{s} = 7 \text{ TeV}$ ”, *CMS Physics Analysis Summary CMS-PAS-MUO-10-002* (2010).
- [174] CMS Collaboration, “Search for supersymmetry in hadronic final states with missing transverse energy using the variables  $\alpha_T$  and b-quark multiplicity in  $pp$  collisions at  $\sqrt{s} = 8 \text{ TeV}$ ”, *Eur.Phys.J. C* **73** (2013) 2568, [arXiv:1303.2985](https://arxiv.org/abs/1303.2985).  
[doi:10.1140/epjc/s10052-013-2568-6](https://doi.org/10.1140/epjc/s10052-013-2568-6).
- [175] 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](https://arxiv.org/abs/1305.2390). [doi:10.1016/j.physletb.2013.06.058](https://doi.org/10.1016/j.physletb.2013.06.058).
- [176] CMS Collaboration, “Search for supersymmetry in hadronic final states using MT2 with the CMS detector at  $\sqrt{s} = 8 \text{ TeV}$ ”, *CMS Physics Analysis Summary CMS-PAS-SUS-13-019* (2014).
- [177] 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).
- [178] S. Ask, M. Parker, T. Sandoval et al., “Using gamma+jets Production to Calibrate the Standard Model Z(nunu)+jets Background to New Physics Processes at the LHC”, *JHEP* **1110** (2011) 058, [arXiv:1107.2803](https://arxiv.org/abs/1107.2803).  
[doi:10.1007/JHEP10\(2011\)058](https://doi.org/10.1007/JHEP10(2011)058).
- [179] Z. Bern, G. Diana, L. Dixon et al., “Missing Energy and Jets for Supersymmetry Searches”, *Phys.Rev. D* **87** (2013), no. 3, 034026, [arXiv:1206.6064](https://arxiv.org/abs/1206.6064).  
[doi:10.1103/PhysRevD.87.034026](https://doi.org/10.1103/PhysRevD.87.034026).
- [180] 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](https://arxiv.org/abs/1106.1423).  
[doi:10.1103/PhysRevD.84.114002](https://doi.org/10.1103/PhysRevD.84.114002).
- [181] B. Efron, “The Jackknife, the bootstrap and other resampling plans”. Number 38 in Regional Conference Series in applied mathematics. Society for Industrial and applied mathematics, Philadelphia, Pa., 1982.
- [182] J. D’Hondt, S. Lowette, O. L. Buchmüller et al., “Fitting of Event Topologies with External Kinematic Constraints in CMS”, Technical Report CMS-NOTE-2006-023, CERN, Geneva, Jan, 2006.
- [183] A. L. Read, “Presentation of search results: the  $CL_s$  technique”, *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10, 2693 – 2704.  
[doi:10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).

- [184] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **434** (1999), no. 2-3, 435 – 443. doi:10.1016/S0168-9002(99)00498-2.
- [185] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report ATL-PHYS-PUB-2011-011. CMS-NOTE-2011-005, CERN, Geneva, 2011.
- [186] M. Botje, J. Butterworth, A. Cooper-Sarkar et al., “The PDF4LHC Working Group Interim Recommendations”, arXiv:1101.0538.
- [187] 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.
- [188] 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.
- [189] 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.
- [190] 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* **1409** (2014) 176, arXiv:1405.7875. doi:10.1007/JHEP09(2014)176.
- [191] ATLAS Collaboration, “Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at  $\text{sqrt}(s)=8$  TeV proton-proton collisions using the ATLAS experiment”, *JHEP* **1310** (2013) 130, arXiv:1308.1841. doi:10.1007/JHEP10(2013)130.
- [192] 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* **1410** (2014) 24, arXiv:1407.0600. doi:10.1007/JHEP10(2014)024.
- [193] CMS Collaboration, “Phenomenological MSSM interpretation of the CMS 7 and 8 TeV results”, *CMS Physics Analysis Summary CMS-PAS-SUS-13-020* (2014).
- [194] ATLAS Collaboration, “ATLAS Supersymmetry (SUSY) searches”, *ATLAS Wiki* (2014). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults>, Topic revision: r437.
- [195] P. Bechtle, T. Bringmann, K. Desch et al., “Constrained Supersymmetry after two years of LHC data: a global view with Fittino”, arXiv:1204.4199.

- [196] O. Buchmueller, R. Cavanaugh, A. De Roeck et al., “The CMSSM and NUHM1 after LHC Run 1”, *Eur.Phys.J.* **C74** (2014) 2922, [arXiv:1312.5250](https://arxiv.org/abs/1312.5250).  
[doi:10.1140/epjc/s10052-014-2922-3](https://doi.org/10.1140/epjc/s10052-014-2922-3).
- [197] C. P. Robert, “The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation (Springer Texts in Statistics)”. Springer-Verlag New York, 2001.
- [198] A. O’Hagan, “Bayesian inference”. Number Vol. 2B in Kendall’s Advanced Theory of Statistics. Edward Arnold, London 1994.
- [199] 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).
- [200] M. Aliev et al., “HATHOR – HAdronic Top and Heavy quarks crOss section calculator”, *Comput. Phys. Commun.* **182** (2011) 1034–1046, [arXiv:1007.1327](https://arxiv.org/abs/1007.1327).  
[doi:10.1016/j.cpc.2010.12.040](https://doi.org/10.1016/j.cpc.2010.12.040).
- [201] J. Ott, “Theta framework”, Nov. 2014. [www.theta-framework.de](http://www.theta-framework.de).
- [202] CMS Collaboration, “Search for supersymmetry in final states with missing transverse energy and 0, 1, 2, or at least 3 b-quark jets in 7 TeV  $pp$  collisions using the variable alphaT”, *JHEP* **1301** (2013) 077, [arXiv:1210.8115](https://arxiv.org/abs/1210.8115).  
[doi:10.1007/JHEP01\(2013\)077](https://doi.org/10.1007/JHEP01(2013)077).
- [203] C. G. Lester and D. J. Summers, “Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders”, *Phys. Lett.* **B 463** (1999) 99–103, [arXiv:hep-ph/9906349](https://arxiv.org/abs/hep-ph/9906349). [doi:10.1016/S0370-2693\(99\)00945-4](https://doi.org/10.1016/S0370-2693(99)00945-4).
- [204] A. Barr, C. Lester, and P. Stephens, “ $m_{T2}$ : the Truth behind the glamour”, *J. Phys. G* **29** (2003) 2343–2363, [arXiv:hep-ph/0304226](https://arxiv.org/abs/hep-ph/0304226).  
[doi:10.1088/0954-3899/29/10/304](https://doi.org/10.1088/0954-3899/29/10/304).
- [205] CMS Collaboration, “Search for supersymmetry in hadronic final states using MT2 in  $pp$  collisions at  $\sqrt{s} = 7$  TeV”, *JHEP* **1210** (2012) 018, [arXiv:1207.1798](https://arxiv.org/abs/1207.1798).  
[doi:10.1007/JHEP10\(2012\)018](https://doi.org/10.1007/JHEP10(2012)018).
- [206] CMS Collaboration, “Inclusive search for supersymmetry using the razor variables in  $pp$  collisions at  $\sqrt{s} = 7$  TeV”, *Phys.Rev.Lett.* **111** (2013), no. 8, 081802, [arXiv:1212.6961](https://arxiv.org/abs/1212.6961). [doi:10.1103/PhysRevLett.111.081802](https://doi.org/10.1103/PhysRevLett.111.081802).
- [207] CMS Collaboration, “Search for supersymmetry with razor variables in  $pp$  collisions at  $\sqrt{s}=7$  TeV”, [arXiv:1405.3961](https://arxiv.org/abs/1405.3961).
- [208] CMS Collaboration, “Jet energy calibration in the 8 TeV  $pp$  data”, 2014. Paper in preparation.



