SEARCHES FOR LIGHT- AND HEAVY-FLAVOR THREE-JET RESONANCES IN PROTON-PROTON COLLISIONS WITH THE CMS DETECTOR AT $\sqrt{S} = 8$ TEV

By

CLAUDIA SEITZ

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Physics and Astronomy

written under the direction of

Prof. Eva Halkiadakis

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2014

ABSTRACT OF THE DISSERTATION

Searches for Light- and Heavy-flavor Three-jet Resonances in proton-proton Collisions with the CMS Detector at $\sqrt{s} = 8$ TeV

By CLAUDIA SEITZ

Dissertation Director:

Prof. Eva Halkiadakis

A search for three-jet hadronic resonance production in proton-proton collisions at 8 TeV center-of-mass energy has been conducted by the CMS Collaboration at the LHC. The data set corresponds to an integrated luminosity of 19.4 fb^{-1}. The search method is model-independent. Events are selected that contain a large number of jets with high transverse momentum. An ensemble of jets is used to extract a new possible signal from copious QCD background. Event selection is optimized using a benchmark model where supersymmetric gluinos are pair-produced and each of the gluinos decays exclusively into three jets. Two scenarios of this decay are considered denoted by the RPV couplings λ''_{112} and λ''_{113} or λ''_{223}. The first coupling allows for gluinos to decay into only light-flavor jets, while the latter two allow decays into one heavy-flavor and two light-flavor jets. No significant deviation is found between the selected events and the expected standard model multijet and $t\bar{t}$ background. For gluinos decaying through λ''_{112}, masses below 650 GeV are excluded at the 95% confidence level. The search including heavy-flavor jets in the final state with the couplings λ''_{113} or λ''_{223} is the first of its kind. Gluinos decaying into one heavy-flavor and two light-flavor jets are excluded for masses between 200 and 835 GeV.

Acknowledgments

First and foremost, I would like to thank my advisor Eva Halkiadakis for her continuous support, encouragement, and endless motivation during my time as a PhD and Masters student at Rutgers. Her dedication for physics has served as an inspiration and made this thesis work possible. Besides my advisor, I would like to thank Amit Lath and Scott Thomas, who played an essential role in developing the methods used as a foundation for this thesis. I would also like to extend my thanks to my other committee members Sevil Salur, Daniel Friedan, and Sung-Won Lee.

The first two years living in New Jersey were made so much more enjoyable after meeting Christian, Emmanuel, Megan, and the whole Parker and Contreras-Campana families. They gave me a sense of home while my own was so far away, which is something I will forever be grateful for.

I spent a significant amount of my PhD time at CERN and had the chance to work with outstanding postdocs and collaborate with colleagues from the Ohio State University. I would like to thank Dan Duggan, Dean Hidas, Carl Vuosalo, Khristian Kotov, and Chris Hill. Special thanks go to Dan and Carl who have been a crucial part of the analysis presented in this thesis.

My thanks also go to my fellow graduate students in the Rutgers High Energy group and at CERN Rishi, Shruti, Sanjay, A.J., Anthony, Christian, Emmanuel, and Preema.

Not just a fellow student, but the one person who has accompanied me through graduate school, while living in the US, and even during the time in France, has been Christian. On to the next endeavor, wherever life might take us!

Last but no least, I would like to thank my parents Martina and Jürgen and my grandparents Helga and Reiner.

Dedication

To life and new endeavors.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Fundamental questions about the universe and its basic building blocks have fascinated humans for centuries. The 20th century was filled with theoretical and experimental advancements in the field of particle physics. The era of modern particle physics started with the discovery of the electron in 1897, followed by Einstein's theoretical description of the photoelectric effect in 1905.

The discovery of the proton and other particles such as neutrons, neutrinos, and the heavier brother of the electron, the muon, followed. Weak interactions were introduced through the observation of β decay, where a neutron decays into a proton under the emission of an electron and a neutrino. In the 1960's the idea of even smaller constituents of protons and neutrons was introduced as the quark or parton model.

The first part of the theory known today as the Standard Model was developed in the 1960's as well. Weinberg, Glashow, and Salam proposed a theory that combine the electromagnetic and weak interactions. This theory also incorporated the idea of electroweak symmetry breaking, which predicts massive bosons as carriers of the weak force and a new scalar particle known as the Higgs boson. It took almost twenty more years to discover the force carriers of the weak interaction the charged $W^{\pm}$ and the neutral Z^0 bosons. Another prediction of the theory was the existence of a heavy partner of the bottom quark, the top quark, which was discovered in 1995 at the Tevatron accelerator at Fermilab. Both, the ATLAS and CMS collaboration at the LHC, reported the observation of a new boson with Higgs-like properties in 2012. However, despite the success of the theoretical predictions there are still unresolved questions. For example, what keeps the Higgs mass on the order of ≈ 100 GeV, or why are the scales for the fundamental forces so different, and how can gravity be incorporated?

This thesis describes a search for new physics beyond the Standard Model in an all-hadronic final state. During the last years the LHC experiments have produced a plethora of different results placing limits on new physics scenarios. However, all-hadronic final states are still difficult to handle. While the underlying theory of Quantum Chromodynamics (QCD) can be described by perturbation theory for high energies and short distances, the long distance limit is still not very well understood. Due to the properties of the strong force quarks and gluons cannot be observed as single particles. They undergo the so-called parton shower and hadronization processes resulting in a spray of hadrons, something we refer to as a jet.

In Chapter 2 we briefly discuss the field theoretical description of the Standard Model followed by one possible extension for a new physics scenario called Supersymmetry. The concept of R-parity violation is introduced. Pair-produced supersymmetric gluinos decaying through these R-parity violating couplings into three jets are used as a benchmark model throughout this thesis. Dedicated searches are performed where the final state consists of only light-flavor jets or of one heavy-flavor and two light-flavor jets.

Chapters 3 and 4 focus on the experimental setup of the LHC and the CMS experiment, as well as the data collected during the 2012 running period. The general approach for algorithms used to reconstruct the recorded events is discussed in Chapter 5.

Chapter 6 focuses on the specific analysis strategy used in this search for three-jet resonances. The jet ensemble technique is introduced, which is the main method to reconstruct the mass of the new hypothetical particle. Basic event selection criteria, as well as the optimization procedure and their results are described in detail. This chapter concludes with the definition of three distinct search regions with optimized selection requirements to improve the sensitivity to a possible new physics signal.

The background from QCD multijet production is theoretically difficult to model. Data driven methods, based on fits of parametrized function as well as binned background templates from control regions in the data, are discussed in Chapter 7. A general introduction to fitting techniques is given as well.

The search is performed by examining a smoothly falling invariant mass distribution in data for a localized Gaussian deviation. The modeling of this signal shape is presented

in Chapter 8. Systematic uncertainties affecting the signal and background models are discussed first in Chapter 9, followed by an introduction to limit setting techniques.

No significant deviation between the data and background model is found. Chapter 10 shows the experimental results interpreted as limits on the gluino pair production cross section for two different sets of R-parity violating couplings.

An additional cross-check validating the analysis technique is presented in Appendix C as a measurement of the $t\bar{t}$ cross section based on events selected for the heavy-flavor search.

Chapter 2

Theoretical overview

In this chapter we start with a theoretical overview of the Standard Model of particle physics, introducing a field theoretical description of particles and their interactions. The second part of the chapter focuses on a possible extension of the Standard Model in the form of a new symmetry, introducing a model of a new three-jet resonance. We conclude this chapter with experimental results previously obtained from searches for this new physics scenario.

2.1 The Standard Model

The Standard Model (SM) is one if the most successful theories developed during the last century. Interactions between particles are described in the context of a relativistically invariant quantum field theory containing the electromagnetic, weak, and strong forces. Gravity, the fourth fundamental force, is not included in the theory due to its expected negligible effect at the currently accessible energy scales. Each of the three forces included in the SM can be viewed as an exchange of spin-1 gauge fields represented by photons for the electromagnetic force, $W^{\pm}$ and Z^0 bosons for the weak force, and gluons for the strong force.

By means of group theory the SM is described by the symmetry group:

$$SU(3)_C \times SU(2)_L \times U(1)_Y, \tag{2.1}$$

where $SU(3)_C$ is an abstract representation for the strong force and $SU(2)_L \times U(1)_Y$ combines the electromagnetic and weak interactions. The building blocks of matter in the universe can be divided into two main categories, fermions and bosons. The field quanta

described above are bosons which have integer spin, whereas fermions have half-integer spins. This fundamental difference plays an important role when describing their behavior mathematically. Several different quantum numbers are introduced, as described below, leading to conservation laws that need to be taken into account when building a full theory of fundamental interactions. Fermions can be split up further into two categories: quarks and leptons, where both are present in three different generations displaying a large mass hierarchy. Leptons participate in the electroweak interaction and either carry integer charge (e, μ, τ) or are electrically neutral $(\nu_e, \nu_\mu, \nu_\tau)$. Independent of the charge, they are assigned a lepton number ($L = +1$ for particles and $L = -1$ for antiparticles). Quarks are subject to the electroweak as well as the strong interactions. They carry fractional charges of $\pm\frac{1}{3}$ or $\pm\frac{2}{3}$ and are assigned a baryon number ($B = +\frac{1}{3}$ for particles and $B = -\frac{1}{3}$ for antiparticles). Six flavors of quarks with corresponding quantum numbers are observed: up, down, charm, strange, top, and bottom. Additionally, quarks carry color ($r,\ b,\ g$) which can be viewed as the charge of the strong force as discussed in more detail in Section 2.1.4. The gauge bosons neither have baryon nor lepton number and they all have spin 1. The Higgs boson responsible for electroweak symmetry breaking is the only known fundamental particle with spin 0. A summary of these elementary fermions and bosons can be found in Table 2.1 and in Figure 2.1.

The SM is build upon the Lagrangian formalism of quantum field theory, which can accommodate Lorentz invariance, internal gauge symmetries, causality, and local interactions [1]. All of the above mentioned particles are described by fields representing creation and annihilation operators for particles and antiparticles. The following notation will be used throughout the rest of the discussion [1]:

- scalar fields $\phi(x)$ represent spin-0 particles
- vector fields $A_\mu(x)$ represent spin-1 particles
- spinor fileds $\psi(x)$ represent spin-1/2 fermions

Gauge invariance of the Lagrangian is the fundamental principle behind the interactions of the SM. The following sections describe in more detail our current understanding of electromagnetic, weak, and strong processes.

Name	Spin s	Baryon number B	Lepton number L	Charge Q	Mass [MeV]
leptons					
electron	$\frac{1}{2}$	0	0	-1	0.511
electron neutrino	$\frac{1}{2}$	0	0	0	$< 2.2 \times 10^{-6}$
muon	$\frac{1}{2}$	0	0	-1	105.7
muon neutrino	$\frac{1}{2}$	0	1	0	< 0.17
tau	$\frac{1}{2}$	0	0	-1	1.77×10^3
tau neutrino	$\frac{1}{2}$	0	0	0	< 15.5
quarks					
up	$\frac{1}{2}$	$\frac{1}{3}$	1	$\frac{2}{3}$	2.4
down	$\frac{1}{2}$	$\frac{1}{3}$	1	$-\frac{1}{3}$	4.8
charm	$\frac{1}{2}$	$\frac{1}{3}$	1	$\frac{2}{3}$	1.27×10^3
strange	$\frac{1}{2}$	$\frac{1}{3}$	1	$-\frac{1}{3}$	104
top	$\frac{1}{2}$	$\frac{1}{3}$	1	$\frac{2}{3}$	171.2×10^3
bottom	$\frac{1}{2}$	$\frac{1}{3}$	1	$-\frac{1}{3}$	4.2×10^3
gauge bosons					
photon	1	0	0	0	0
W± boson	1	0	0	±1	80.4×10^3
Z^0 boson	1	0	0	0	91.2×10^3
gluon	1	0	0	0	0
Higgs boson	0	0	0	0	126×10^3

Table 2.1: Particles and their quantum numbers in the Standard Model. The spin is defined in units of $\hbar$ and electrical charge in units of the elementary charge e [2].

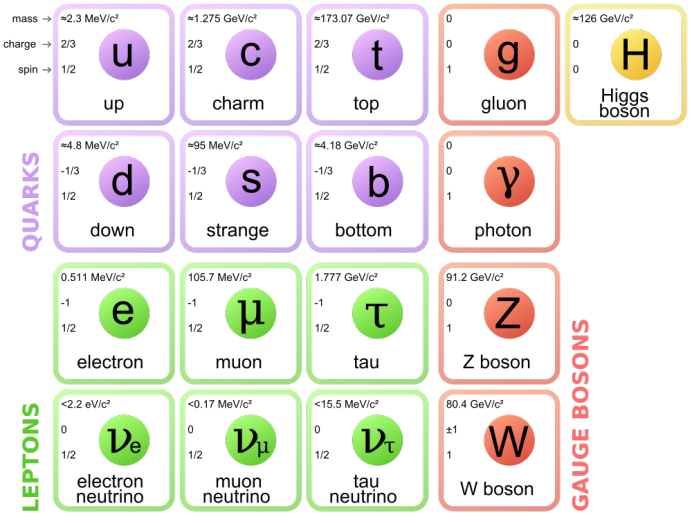

Figure 2.1: Standard Model of Particle Physics [3].

2.1.1 Quantum electrodynamics

Quantum electrodynamics (QED) describes the fundamental electromagnetic interactions between fermions mediated by massless spin-1 photons. This simple example will be used to introduce the principle of local gauge invariance. One starts from the free Lagrangian describing Dirac fermions [4]:

$$L_0^{QED} = i\overline{\psi}(x)\gamma^\mu\partial_\mu\psi(x) - m\overline{\psi}(x)\psi(x), \tag{2.2}$$

where $\psi(x)$ is the two dimensional Dirac Spinor and m represents a mass. We know already from non-relativistic quantum mechanics that the overall phase θ of the wave function has no physical effect, which results in an invariance of the Lagrangian under the transformation $\psi'(x) = e^{iQ\theta}\psi(x)$, with the charge given by Q. If the phase is global (meaning independent of the space-time coordinate x) then L_0^{QED} in the relativistic case is also invariant under this transformations described by the global symmetry group $U(1)$. However, if the phase depends on local space-time coordinates $\theta = \theta(x)$ we pick up an additional term in the Lagrangian due to the derivative of $\theta(x)$:

$$\partial_\mu(e^{iQ\theta(x)}) = iQ(\partial_\mu\theta(x))e^{iQ\theta(x)}\psi(x) + e^{iQ\theta(x)}\partial_\mu\psi. \tag{2.3}$$

The additional term $\partial_\mu\theta(x)$ has the structure of a 4-vector, to restore the invariance of L_0^{QED} a new spin-1 field $A_\mu(x)$ is added, which transforms as $A'_\mu(x) = A_\mu(x) - \frac{1}{e}\partial_\mu\theta(x)$. One then defines the covariant derivative:

$$D_\mu\psi(x) = [\partial_\mu + ieQA_\mu(x)]\psi(x), \tag{2.4}$$

and replaces ∂_μ by D_μ in Equation 2.2, which makes L_0^{QED} invariant under the local gauge transformation. Based on the principle of local gauge invariance we have now introduced interaction terms between the fermionic field $\psi(x)$ and the vector field $A_\mu(x)$ into the theory, describing the well known interaction between fermions and photons. An additional kinematic term is necessary to fully describe the vector field $A_\mu(x)$. The full Lagrangian

for electromagnetic interactions then reads:

$$L^{QED} = i\overline{\psi}(x)\gamma^{\mu}\partial_{\mu}\psi(x) - m\overline{\psi}(x)\psi(x) - eQA_{\mu}(x)\overline{\psi}(x)\gamma^{\mu}\psi(x) - \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x), \quad (2.5)$$

where $F_{\mu\nu}(x) = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ is the electromagnetic field strength tensor. A mass term of the form $\frac{1}{2}m^2A^{\mu}A_{\mu}$ is forbidden to preserve the local gauge symmetry, therefore predicting massless photons. The electric charge is related to the electromagnetic coupling constant, also called the finestructure constant, $\alpha = \frac{e^2}{4\pi} = \frac{1}{137}$.

2.1.2 Electroweak interactions

Two kinds of weak interactions are observed in nature: charged and neutral, both mediated by massive vector bosons the $W^{\pm}$ and Z, respectively [5]. Neutral processes, for example, include neutrino-electron or electron-electron scattering, where a neutral Z^0 is exchanged (shown on the left in Figure 2.2). Charged weak interactions are responsible for flavor changing processes, with the well known examples of the muon decay $\mu^- \rightarrow e^- + \nu_{\mu} + \overline{\nu}_e$ (shown on the right in Figure 2.2) and the β decay of the neutron $n \rightarrow p^+ + e^- + \overline{\nu}_e$. To fully

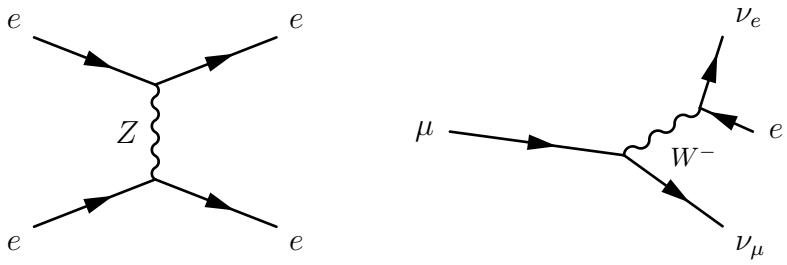

Figure 2.2: Feynman diagrams for the weak interaction. Left: neutral (electron-electron scattering), Right: charged (muon decay).

describe weak interactions a more complicated structure is needed than in QED. However, we can use again the principle of gauge invariance to combine weak and electromagnetic interactions and make use of the combined symmetry group $SU(2)_L \times U(1)_Y$ to characterize an electroweak theory. In this model leptons and quarks appear as left-handed isospin

doublets and right-handed isospin singlets, as shown below:

$$\begin{array}{cccccc} \begin{pmatrix} u \\ d \end{pmatrix}_L & \begin{pmatrix} c \\ s \end{pmatrix}_L & \begin{pmatrix} t \\ b \end{pmatrix}_L & & \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L & \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L & \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \\ u_R & c_R & t_R & & & & \\ d_R & s_R & b_R & & e_R & \mu_R & \tau_R \end{array} \tag{2.6}$$

In general, fermions are represented by spinor fields $\psi(x)$, however since the weak interaction only couples to left-handed particles, projection operators are used to separate the fields into $\psi_L(x) = \frac{1}{2}(1-\gamma_5)\psi(x)$ and $\psi_R(x) = \frac{1}{2}(1-\gamma_5)\psi(x)$. One can now start with the Lagrangian:

$$L_0^{EW} = i\overline{\psi}(x)\gamma^\mu\partial_\mu\psi(x), \tag{2.7}$$

and require that L_0^{EW} to be invariant under the local gauge transformation:

$$\psi_L(x)' = e^{ig\alpha(x)\cdot\mathbf{T}+ig'\beta(x)Y}\psi_L(x), \qquad \psi_R(x)' = e^{ig'\beta(x)Y}\psi_R(x), \tag{2.8}$$

where $\alpha(x)$ is an arbitrary three-dimensional vector and $\beta(x)$ is a one-dimensional function. $\mathbf{T}$ is the weak-isospin operator and Y is the weak hypercharge, which in combination with the third component of T relates both quantities to the physical charge $Q = T_3 + Y/2$. The invariance of the Lagrangian under this combined $SU(2)_L \times U(1)_Y$ symmetry can be restored by replacing ∂_μ by

$$D_\mu = \partial_\mu + ig\mathbf{W}_\mu \cdot \mathbf{T} + ig'\frac{1}{2}B_\mu Y. \tag{2.9}$$

The two new gauge fields are $\mathbf{W}_\mu$ and B_μ and they contain the photon, and the $W^\pm$, and Z^0 bosons. Additionally, kinematic terms of the gauge fields $-\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}\cdot B^{\mu\nu}$ can be added to the Lagrangian. The coupling constants for electromagnetic and weak interactions are denoted by g and g', respectively. Mass terms are again forbidden because they would destroy the gauge invariance, fermion mass terms are excluded as well since they would combine left- and right-handed fields. However, from experimental observation we know that $W^\pm$ and Z^0 are massive and the proposed process through which they acquire

mass is called electroweak symmetry breaking (EWSB).

2.1.3 Electroweak symmetry breaking

Since mass terms for the gauge bosons and fermions are forbidden in Equation 2.7 the symmetry has to be "spontaneously" broken for them to acquire mass. One starts with introducing a single isospin doublet of a complex scalar field $\Phi = (\phi_1, \phi_2)$ into the theory. The EWSB Lagrangian is then described by:

$$L^{EWSB} = (D_\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi), \qquad V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \frac{\lambda}{4}(\Phi^\dagger \Phi)^2, \tag{2.10}$$

where μ^2 and λ are constants that can be chosen such that a non-vanishing vacuum expectation value (vev) of $|\Phi_{vev}| = \sqrt{\frac{1}{2}\frac{\mu^2}{\lambda}}$ is created. When L^{ESWB} is added to L_0^{EW} and local gauge symmetry is required for the combined Lagrangian, three of the four gauge bosons acquire mass, while one state for the photon stays massless [6]. The gauge bosons are no longer the original fields B_μ and $\mathbf{W}_\mu$ but rather mixtures of the form $\mathrm{W}_\mu^\pm = (W_\mu^1 \pm W_\mu^2)/\sqrt{2}$. The Z^0 boson and the photon field A_μ are related through the Weinberg angle θ_W, where $\tan\theta_W = \frac{g'}{g}$. Fermion masses are generated as well through a Yukawa coupling of the lepton or quark field with the scalar field. Another additional consequence of this model is the prediction of a new scalar particle, the Higgs boson. A new particle compatible with the Higgs boson was discovered in 2012 by the two main experiments at the LHC simultaneously, CMS [7] and ATLAS [8]. The mass of this new particle given by the particle data group is $m_H = 125.9 \pm 0.4$ GeV [9].

2.1.4 Quantum chromodynamics

In this analysis the main background arises from QCD multijet production and we discuss the underlying theory of the strong interaction in more detail in this section. The same gauge principle as previously discussed applies also for QCD, however, instead of one type of electric charge, the strong interaction is described by three different color charges red (r), green (g), and blue (b). The underlying symmetry group here is $SU(3)_C$ and in addition to the Dirac spinor every quark is assigned a three-component vector giving its color. The

mediators of the strong interaction are colored gluons, where each gluon carries one unit of color and one unit of anti-color of a different type. The two quark-gluon interaction vertices are shown in the top row in Figure 2.3. Starting with three possible colors and anti-colors one obtains nine different gluon states, a color octet and one color singlet [5], one possible representation is:

$$\left\{\begin{array}{ll} (r\bar{b}+b\bar{r})/\sqrt{2} & -i(r\bar{g}-g\bar{r})/\sqrt{2} \\ (b\bar{g}+g\bar{b})/\sqrt{2} & -i(r\bar{b}-b\bar{r})/\sqrt{2} \\ (r\bar{r}-b\bar{b})/\sqrt{2} & -i(b\bar{g}-g\bar{b})/\sqrt{2} \\ (r\bar{g}+g\bar{r})/\sqrt{2} & (r\bar{r}+b\bar{b}-2g\bar{g})/\sqrt{6} \end{array}\right\} \qquad \left[(r\bar{r}+b\bar{b}+g\bar{g})/\sqrt{3}\right]. \tag{2.11}$$

The color singlet state is the colorless combination $(\bar{r}r+\bar{g}g+\bar{b}b)$ which stays unchanged under rotation in the color space, and does not participate in QCD interactions. The two possible diagrams for gluon-gluon interactions are shown in the bottom row in Figure 2.3. From a field theoretical point of view these eight gluon states correspond to the 3×3 Gell-

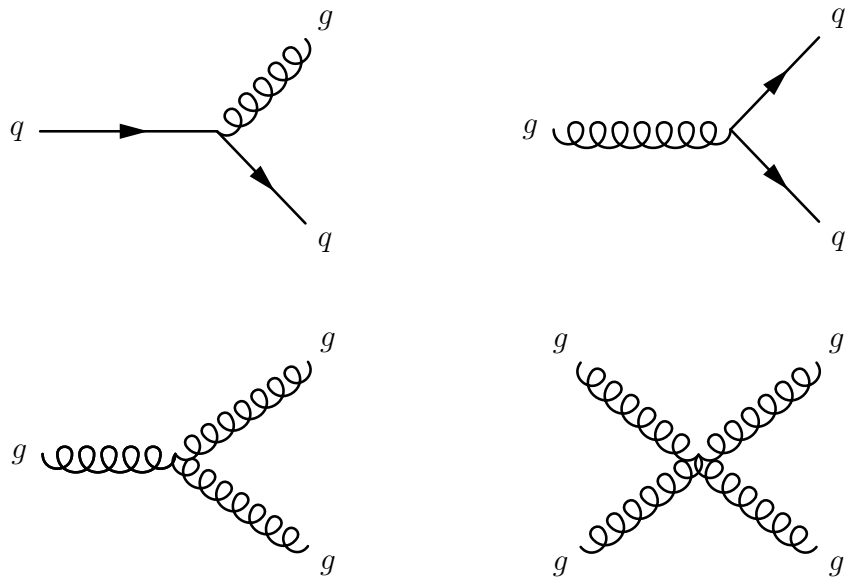

Figure 2.3: Feynman diagrams for interactions in QCD.Top: gluon emission, gluon absorption, Bottom: gluon-gluon interaction

Mann matrices λ_a, which are the generators of the group $SU(3)_C$ and $a=1,...,8$. These

matrices satisfy the commutation relation $\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}\right] = if^{abc}\frac{\lambda^c}{2}$ with f^{abc} being the structure constants for the group. Therefore, when building the Lagrangian the covariant derivative is replaced by [1]:

$$D_\mu = \partial_\mu - ig_s\frac{\lambda_a}{2}G^a_\mu. \tag{2.12}$$

The field strength tensor for the gluon fields is given by:

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f_{abc} G^b_\mu G^c_\nu. \tag{2.13}$$

The coupling constant g_s used here is related to the strong coupling constant $\alpha_s = \frac{g_s^2}{4\pi}$. If we introduce the notation of color-triplet of fermion fields $\mathbf{\Psi} = (q_1, q_2, q_3)^T$ one can write the full QCD Lagrangian as:

$$L^{QCD} = \overline{\mathbf{\Psi}}(i\gamma^\mu\partial_\mu - m)\mathbf{\Psi} + g_s\overline{\mathbf{\Psi}}\gamma^\mu\frac{\lambda_a}{2}\mathbf{\Psi}G^a_\mu - \frac{1}{4}G^a_{\mu\nu}G^{a,\mu\nu}. \tag{2.14}$$

Asymptotic freedom

As was shown in the previous sections the Lagrangians describing QED and QCD can be derived in a very similar form. However, there are a few major differences. Unlike the neutral mediator in QED, gluons carry color charge and can interact among themselves. Furthermore, we recall the coupling constant in electromagnetic interactions to be very small $\alpha = \frac{1}{137}$, this is in general not true for the strong coupling constant α_s, making the usage of perturbation theory difficult. In order to evaluate physical quantities such as decay rates or cross sections one makes use of Feynman diagrams, examples were already shown in Figures 2.3 and 2.2. For a given process, all possible Feynman diagrams need to be added up, in principle including an infinite amount of inner vertices. In QED each of these inner vertices brings a factor of α into the calculation, and because it is such a small number higher order corrections contribute less to the result [5]. A feature of the strong interaction, called asymptotic freedom, salvages the situation and makes perturbation theory applicable in QCD calculations for short distances and high energies. The strong coupling constant is not actually constant, it decreases for high momentum transfer (large q^2, short distances) and increases for low momentum transfer (small q^2, long distances). Due to this dependence

on the momentum transfer, we call α_s a running coupling constant.

2.1.5 QCD at hadron colliders

At the LHC new particles are produced through the head-on collision of two protons. Their quantum numbers and microscopic properties are well described by the assumption of a bound state of three-quarks (uud), the so-called valence quarks. When collisions occur at sufficiently high energy, a closer look at the internal structure of the proton is needed. Through the observation deep inelastic scattering (DIS) of charged electrons off protons it became clear that the three valence quarks cannot be the only constituents of the proton.

Parton distribution function

Parton distribution functions (PDF) describe the parton content of the proton and how the total momentum is distributed internally over all the constituents. They give the probability $f_a(x_a, \alpha_s, \mu)$ of finding a parton a carrying a momentum fraction x_a within the proton at a given energy scale μ. In order to determine the PDFs global fits are performed to experimental data from fixed-target experiments, the electron-proton collider HERA, and hadron colliders [10].

Cross section calculation

Cross sections in QCD are calculated through a factorization approach, where the short-distance part is described by perturbation theory and the long-distance part is approximated by the above mentioned PDFs extracted from experimental measurements. The differential cross section of an observable O of the process $ab \to F$ may be written as [11]:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}O} = \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b \sum_F \int \mathrm{d}\Phi_F f_a^{h1}(x_a, \mu_F) f_b^{h2}(x_b, \mu_F) \frac{\mathrm{d}\hat{\sigma}_{ab}}{\mathrm{d}\hat{O}} D_F(\hat{O} \to O, \mu_F), \tag{2.15}$$

where the sum runs over all partons a and b from the two protons h_1 and h_2. The PDFs $f_{a,b}(x_{a,b}, \mu_F)$ describe the parton content of the incoming protons calculated at the factorization scale μ_F, which provides the energy scale that divides short-range from long-range interactions. The fragmentation function D_F parametrizes the transition from partons in

the final state towards hadronic jets, as discussed in more detail below. The partonic cross section $\mathrm{d}\hat{\sigma}_{ab}$ depends on the two scales μ_F and the renormalization scale μ_R. Renormalization is a method frequently used to treat infinities that arise in perturbative calculations. While the choice of the two scales μ_F and μ_R is in principle arbitrary, using $\mu_F = \mu_R$ and choosing a common value corresponding to the invariant mass of the final state F, is an often used approach.

Parton shower evolution and hadronization

As discussed before due to the nature of α_s being small for high energies and small distances perturbation theory can be used for calculations in that regime. However, the long distance interaction is not very well understood. From experiments we know that quarks and gluons are not observed as free particles. They appear in bound states of quark-antiquark pairs (mesons) or three-quark states (baryons). During high energy collisions quarks and gluons are produced in great abundance, either directly through QCD processes or as decay products of heavy states like the top quark or the $\mathrm{W}^{\pm}$ boson. While these quarks are traveling outward from the hard interaction the coupling constant starts to increase.

At first they undergo the so-called parton shower, which can be approximately calculated with perturbative QCD. While moving outwards the quarks start emitting gluons, which in return emit new quarks and gluons themselves and a shower evolves.

After the parton shower the distance increases further and perturbative treatment starts to break down. Phenomenological models are used to describe how the additional quarks and gluons clump together to form hadrons, the so-called hadronization. Two models for hadronization are the string and the cluster models [12].

- String model: The color field between the outgoing partons is described by a string-like linear potential. The string is stretched as the partons move further apart and eventually breaks up into hadron-sized pieces through $q\bar{q}$ production.

- Cluster model: As a first step gluons from the parton shower are split up into colorless $q\bar{q}$ pairs. The colorless states form colorless clusters, which then decay into hadrons.

A simplified schematic description of parton shower and hadronization process is shown

in Figure 2.4. Experimentally we can observe a spray of charged and neutral particles in the direction of the outgoing parton, which we call jets. The goal is to identify these jets with specially designed algorithms to reconstruct, as close as possible, the original parton.

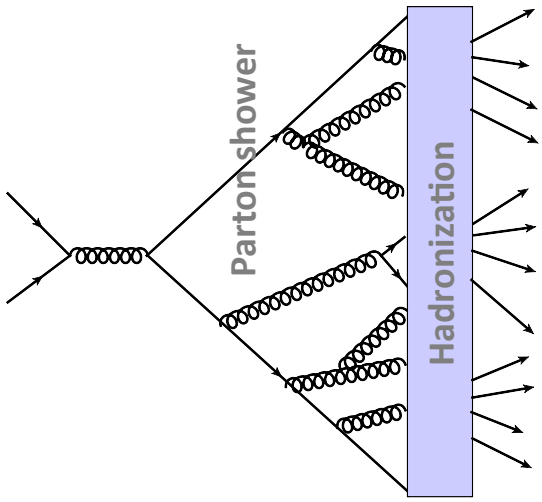

Figure 2.4: Cartoon of the hadronization model from the hard collision to final state hadrons.

2.2 Beyond the Standard Model and Supersymmetry

The SM of particles physics is one of the most precise and experimentally tested theories of the last century. With the discovery of the new boson in 2012 believed to be the SM Higgs one missing piece of the SM is has been found. Continuing efforts are necessary to determine the exact nature of this new particle. Measurements of its coupling, spin, parity, and mass should draw a conclusive picture if it is in fact the long sought-after Higgs boson predicted by EWSB. However, there are still unresolved questions, which make searches for new phenomena necessary. One of the most puzzling phenomena in particles physics is the so-called hierarchy problem, describing the vastly different scales of, for example the weak force and gravity. Another aspect arising from the hierarchy problem is the "light" mass of the Higgs boson. The Higgs mass itself depends on loop corrections from couplings to fermions pushing its value in principle up to the Planck scale. One proposed solution of the hierarchy problem is the introduction of a new symmetry, called Supersymmetry (SUSY).

As a consequence new particles are predicted, which cancel the fermion loop corrections keeping the Higgs mass on the order of ≈ 100 GeV [13].

One of the basic ideas of this new symmetry is a supersymmetric operator Q that can act on fermions as well as bosons.

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

This implies that every SM fermion (boson) has a corresponding bosonic (fermionic) superpartner. The operator Q satisfies the algebra:

$$[Q, Q^\dagger] = P^\mu, \tag{2.16}$$

$$[Q, Q] = [Q^\dagger, Q^\dagger] = 0, \tag{2.17}$$

$$[P^\mu, Q] = [P^\mu, Q^\dagger] = 0, \tag{2.18}$$

where P^μ is the four-momentum generator. A single-particle state is introduced, called a supermultiplet, which consists of fermions and bosons. Particles in the same supermultiplet have the same mass, electric charge, weak isospin, and color degree of freedom. This is due to the fact the generators $[Q, Q^\dagger]$ commute with the generators of the gauge transformations [13]. As for the naming convention, for every bosonic state an "s-" is attached to the original SM particle: sleptons and squarks. Every fermionic superpartner gets the syllable "-ino" attached to the name: gauginos and higgsinos. For example, the superpartner of the gluon is called the gluino.

2.2.1 Minimal Supersymmetric Standard Model

We discuss here as an example the Minimal Supersymmetric Standard Model (MSSM). In order to exactly cancel the loop corrections from SM particles one would expect that the supersymmetric partners have the same mass as their SM counterparts, however none of these new particles have been observed so far. One simple conclusion can be drawn from this experimental observation, SUSY has to be a broken symmetry. Therefore, a

theoretical description needs to be developed including a Lagrangian that preserves exact supersymmetry and introduces a spontaneous symmetry breaking. This is similar to the EWSB introduced in the SM. The superpotential for a MSSM can be written as:

$$W_{MSSM} = u\mathbf{y_u}QH_u - d\mathbf{y_d}QH_d - e\mathbf{y_e}LH_d + \mu H_u H_d, \tag{2.19}$$

where H_u, H_d, Q, L, u, d, e are chiral supermultiplets which can be seen in Table 2.2. The dimensionless Yukawa couplings $\mathbf{y_d}, \mathbf{y_u}, \mathbf{y_e}$ have the form of 3×3 matrices. The last term in equation 2.19 represents the supersymmetric Higgs boson. As said before, SUSY itself must be broken, however, the mechanism is unknown. It turns out that to introduce symmetry breaking terms into the MSSM one obtains 105 new parameters (masses, phases, and mixing angles), leading to a large arbitrariness in terms of building physical models.

names		**spin 0**	**spin $\frac{1}{2}$**	**SU(3), SU(2), U(1)**
squarks, quarks	Q	$(\widetilde{u}_L\ \widetilde{d}_L)$	$(u_L\ d_L)$	$(3,\ 2,\ \frac{1}{6})$
($\times$ 3 families)	u	$\widetilde{u}^*_R$	$u^\dagger_R$	$(\bar{3},\ 1,\ -\frac{2}{3})$
	d	$\widetilde{d}^*_R$	$d^\dagger_R$	$(\bar{3},\ 1,\ \frac{1}{3})$
sleptons, leptons	L	$(\widetilde{\nu}\ \widetilde{e}_L)$	$(\nu\ e_L)$	$(1,\ 2,\ -\frac{1}{2})$
($\times$ 3 families)	e	$\widetilde{e}^*_R$	$e^\dagger_R$	$(1,\ 1,\ 1)$
Higgs, higgsinos	H_u	$(H^+_u\ H^0_u)$	$(\widetilde{H}^+_u\ \widetilde{H}^0_u)$	$(1,\ 2,\ +\frac{1}{2})$
	H_d	$(H^0_d\ H^-_d)$	$(\widetilde{H}^0_d\ \widetilde{H}^-_d)$	$(1,\ 2,\ -\frac{1}{2})$

names	**spin $\frac{1}{2}$**	**spin 1**	**SU(3), SU(2), U(1)**
gluino, gluon	$\widetilde{g}$	g	(8,1,0)
winos, $W^\pm$ bosons	$\widetilde{W}^\pm\ \widetilde{W}^0$	$W^\pm\ W^0$	$(1,\ 3,\ 0)$
bino, B bosons	$\widetilde{B}^0$	B^0	(1,1,0)

Table 2.2: Chiral supermultiplets in the Minimal Supersymmetric Standard Model [13].

Unlike in the SM, where only one Higgs boson is included, the MSSM introduces two complex Higgs doublets $H_u = (H^+_u, H^0_u)$ and $H_d = (H^0_d, H^-_d)$, which makes EWSB more complicated. Similar to EWSB in the SM model a mixing of the gauginos and higgsinos occurs during the process. The neutral Higgsinos ($\tilde{H^0_u}$, $\tilde{H^0_d}$) and the neutral gauginos ($\tilde{B}$, $\tilde{W^0}$) build four different neutral mass eigenstates called neutralinos. The charged Higgsinos ($\tilde{H^+_u}$, $\tilde{H^-_d}$) and the charged winos ($\tilde{W^-}$, $\tilde{W^+}$) also combine into four states with charge ± 1 called chargino. In certain mass hierarchies the neutralino is a good candidate for the

Lightest Supersymmetric Particle (LSP). The partners of the SM gluons, the gluinos, form a Majorana fermion color octet, but no mixing with other particles in the MSSM occurs. The superpartners of quarks and leptons are all bosonic in nature. A wide variety of scenarios for experimentally detecting these new particles through their decays is predicted, however no significant deviation from the SM has been observed so far.

2.2.2 R-Parity

In the last section we have already introduced the term LSP, which is a consequence of a conversation law. To distinguish SM particles from SUSY particles a new quantum number called R-parity is introduced. R-parity is defined as:

$$R = (-1)^{(3B-L+2s)}, \tag{2.20}$$

where B is baryon number, L is lepton number, and s stands for the spin. All SM particles as well as the Higgs boson have even-parity R=1. On the other hand, all supersymmetric particles (sleptons, squarks, gauginos, higgsinos) have odd-parity R=-1. The aforementioned LSP is one consequence of a R-parity conserving scenario, resulting in at least one LSP in the final state of sparticle decays. Since the decay into SM particles is forbidden this LSP cannot decay any further and usually escapes detection, leaving a signature of missing energy. This missing energy can be deducted from an imbalance in energy when all other energies of visible particles in the final state are summed up. The LSP would also be a very good candidate for dark matter, which we know comprises roughly 25% of our universe. Additionally, if R-parity is conserved then sparticles can only be produced in pairs. However, there is also the possibility for non R-parity conserving processes. R-parity violation (RPV) could either be realized by baryon-number violation (BNV) or lepton-number violation (LNV). If there is no conservation of R-parity then single SUSY particles could be produced and the LSP would not necessarily be stable. Experimentally this means that a final state consisting of SM particles is possible, removing the powerful handle of missing energy in the search for SUSY. The most general R-parity odd superpotential, obeying the

gauge symmetries and consistent with the MSSM that can be added is [14]:

$$W_R = \mu_i H_u L_i + \frac{1}{2}\lambda_{ijk} L_i L_j e_k + \lambda'_{ijk} L_i Q_j d_k + \frac{1}{2}\lambda''_{ijk} u_i d_j d_k, \tag{2.21}$$

where the indices $i, j, k = 1, 2, 3$, Q and L represent the left-handed doublets, and e, u, d are the right-handed leptons and up or down type quarks. The three different couplings denoted with λ represent either LNV (λ and λ') or BNV (λ''). Figure 2.5 shows the three possible new interactions corresponding to the aforementioned couplings.

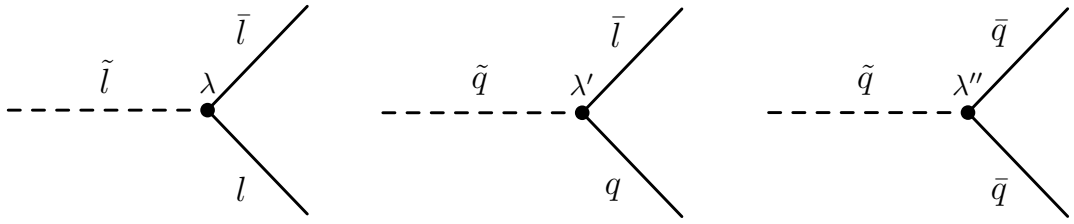

Figure 2.5: RPV vertices. From left to right λ, λ', and λ''

While the all three λ couplings are possible there is already one very strict constraint that only either lepton or baryon number is violated, but not both at the same time. If both quantum numbers would be violated then the proton could undergo the decay $p \to e^+\pi^0$ through the two vertices including λ' and λ'', making the universe unstable. Therefore, we will only focus on cases where one of the couplings is non-zero.

In the case of BNV one expects new physics signals mostly involving jets and no intrinsic missing energy, which is be the focus of this thesis. In particular, in this thesis an analysis is presented searching for pair-produced new hadronic resonances each decaying into three quarks. Gluinos decaying through RPV couplings are chosen as a benchmark process to evaluate the sensitivity to new physics in this channel. Each of the gluinos decays into a quark and off-shell squark, which then decays through λ'' into two quarks, where different choices of i, j, or k lead to different flavors in the final state. The diagram for gluino pair production and decay can be seen on the left in Figure 2.6. On the right one gluino decay is shown in more detail pointing out which baryon numbers are violated and at what vertex

this violation happens.

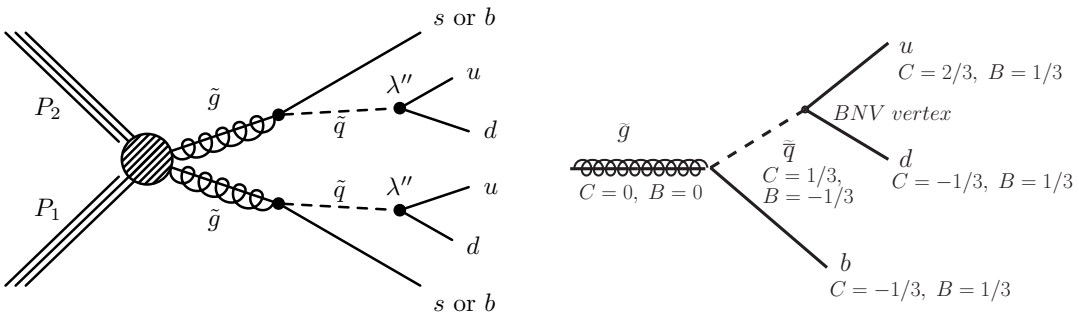

Figure 2.6: Pair-produced gluinos decaying into three jets each.

Searches for these gluinos assuming $\lambda''_{112} \neq 0$ have been performed at the Tevatron and the LHC. Gluino masses between 77 – 144 GeV [15] were excluded at CDF using 3.2 fb$^-$1 of data collected at $\sqrt{s}$ =1.96 TeV at the Tevatron. Both CMS (exclusion below 460 GeV [16, 17]) and ATLAS (exclusion below 666 GeV [18]) extended the previous existing limit based upon 5 fb^{-1} of data collected at $\sqrt{s}$ =7 TeV at the LHC. While the CMS search was performed as "bump-hunt" the ATLAS search was based on a counting experiment. Depending on which coupling is chosen the new predicted resonance can either decay into only light-flavor quarks or at least one b quark and two light-flavor quarks. The latter has not yet been searched for and a dedicated analysis for this final state is presented for the first time in this thesis. Two different scenarios for λ'' are considered:

- Scenario 1 (decay to light flavors, $\tilde{g} \to uds$): $\lambda''_{112} \neq 0$
- Scenario 2 (decay including heavy flavors, $\tilde{g} \to udb$ or csb): λ''_{113} and $\lambda''_{223} \neq 0$

The coupling is chosen such that only prompt decays are allowed. Additionally, the intermediate squark mass is set to be decoupled from the gluino mass and the branching ratio to three quarks is assumed to be 100%. ATLAS has presented a search for these two scenarios as well using the data collected in 2012 at $\sqrt{s}$ =8 TeV [19]. All of the experimental results are summarized in Table 2.3.

Experiment and luminosity	$\sqrt{s}$	Coupling	Observed mass limits at 95% C.L.	Reference
CDF				
3.2 fb^{-1}	1.96 TeV	λ''_{112}	77–144 GeV	[15, 20]
CMS				
36 pb^{-1}	7 TeV	λ''_{112}	200–280 GeV	[16]
5 fb^{-1}	7 TeV	λ''_{112}	460 GeV	[17]
19.4 fb^{-1}	8 TeV	λ''_{112}	650 GeV	[21]
		λ''_{113} or λ''_{223}	835 GeV	this thesis
ATLAS				
4.6 fb^{-1}	7 TeV	λ''_{112}	100–666 GeV	[18]
20.3 fb^{-1}	8 TeV	λ''_{112}	917 GeV	[19]
		λ''_{113} or λ''_{223}	929 GeV	

Table 2.3: Searches for gluinos decaying through RPV couplings in the all-hadronic final state by the CDF, CMS, and ATLAS collaboration.

2.3 Searching for multijet resonances

The analysis presented in this thesis is a model independent search for a pair-produced new hadronic resonance decaying into three quarks each using the CMS detector. As described in Section 2.1.4 quarks cannot be observed as single particles due to the nature of the strong force and an increasing potential energy at large distances. Experimentally we can only detect a spray of particles, a jet, mainly identified by its energy deposits in the calorimeters of the detector and tracks of charged particles. Our current experimental understanding of jets are described in more detail in Section 5.4. Since perturbation theory is only applicable for short distance interactions described by the parton shower model, large uncertainties are usually associated with multijet physics. However, measurements executed at the Tevatron and the LHC show remarkable agreement and constant improvements of our current understanding of strong interactions; for example, the inclusive jet cross section measurement of CMS [22]. Many new physics models predict strongly coupled new resonances decaying into quarks and gluons. Despite their relatively large cross sections, all-hadronic final states pose the challenge of a copious amount of QCD background, which is theoretically difficult to model. Additionally, in a search for pair-produced three-jet resonances another challenge arises from finding the correct combination of three jets out of the six jet final state.

To get around these problems, we present the so-called jet ensemble technique [23, 20]. This technique makes use of kinematic features that are present in heavy particle decays, but cannot be observed in the QCD multijet background. With the choice of trigger (described in more detail in Section 3.2.6) in 2012 at CMS and the addition of b-jet identification in the final state we are also sensitive to a known all-hadronic SM process, the production and subsequent decay of a $t\bar{t}$ pair.

The all-hadronic top quark decay is used to test the analysis technique and its contribution is included into the background model. With a mass of $m_t = 173.07 \pm 0.52 \pm 0.72$ GeV [9] it is the heaviest observed particle. The top quark has an extremely short lifetime of only $\tau = 10^{-25}s$ [24] and it decays almost 100% of the time into $\mathrm{W}^{\pm}$ and b quark before it can hadronize. At the LHC $t\bar{t}$ pairs are produced through gluon fusion (85%) or quark-antiquark annihilation (15%). The $\mathrm{W}^{\pm}$ undergoes decays into either lepton + neutrino or a $q\bar{q}$ pair. The diagram in Figure 2.7 shows the possible decay channels. As a result of the $\mathrm{W}^{\pm}$ decay we expect an all-hadronic decay of $t\bar{t}$ with six quarks, two of which are b quarks, in the final state with a probability of 46%. These all-hadronic top events are used to validate the analysis technique and dedicated studies measuring the pair production cross section are discussed in Appendix C.

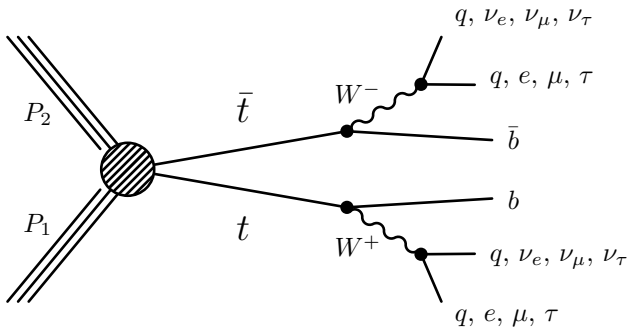

Figure 2.7: $t\bar{t}$ decay modes: leptonic (both $\mathrm{W}^{\pm}$ decay leptonically), semi-leptonic (one $\mathrm{W}^{\pm}$ decays leptonically), and all-hadronic (both $\mathrm{W}^{\pm}$ decay hadronically).

Chapter 3

The experimental setup

In this chapter we discuss the experimental setup including the accelerator complex of the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) experiment, which collected the data used in this thesis.

3.1 The Large Hadron Collider

The LHC is the world's most energetic particle accelerator with a circumference of 27 km situated at the French-Swiss border just outside of the city of Geneva, Switzerland. Design and construction were performed at the research facility of the European Organization for Nuclear Research (CERN). CERN has a long history of research in elementary particle physics and operation of large scale accelerator complexes, for example, the Large Electron-Positron collider (LEP, 1989 – 2000). The original LEP tunnel is situated an average of 100 m underground and now houses the LHC accelerator. The LHC has collided two proton beams with energies between 3.5 TeV and 4 TeV per beam leading to a center-of-mass energy $\sqrt{s}$ of 7 and 8 TeV, respectively. These are the energies that were reached so far, and the goal is to achieve the design energy of 14 TeV in the future.

Figure 3.1 shows the accelerator complex at CERN. In order to obtain free protons one starts off with hydrogen where electrons are removed through an electric field. These protons are injected into the Linac 2, a linear accelerator, where they are accelerated up to an energy of 50 MeV. Three more accelerating steps follow: the Proton Synchrotron Booster (PSB, energy up to 1.4 GeV), the Proton Synchrotron (PS, energy up to 25 GeV) and then finally the Super Proton Synchrotron (SPS) where they reach an energy up to 450 GeV [25]. Instead of having one continuous stream of protons, the LHC beams were designed to be separated into 2808 bunches with 1×10^{11} protons per bunch and a spacing of 25 ns. During

the 2012 data-taking period the nominal operating conditions used a spacing of 50 ns with a maximum number of 1278 bunches [26]. These high energy protons are then fed into the LHC ring where they are further accelerated.

The LHC is a synchrotron accelerator ring consisting of over 1200 superconducting dipole magnets (NbTi) that where designed to provide a magnetic field of up to 8 T to keep the protons on their circular path around the ring. These dipole magnets operate at a temperature below 2 K using superfluid helium for cooling. Radio frequency (RF) cavities are used to accelerate the proton bunches within the LHC, there are eight RF cavities per beam operating at 400 MHz. The beams circulate in two separate rings and are brought together for collision within the four experiments distributed around the ring ATLAS, LHCb, CMS, and ALICE. Quadrupole magnets reduce the beam size at the collision points.

An important quantity for the machine is the instantaneous luminosity [$\mathrm{cm}^{-2}\mathrm{s}^{-1}$]:

$$\mathcal{L} = \frac{N_b^2 n_b^2 f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F, \tag{3.1}$$

which depends on the following beam quantities: N_b the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalized transverse beam emittance, which describes the transverse spread in position and momentum of the protons, β^* gives a measure for the transverse width of the beam at the collision point, and F is a geometric reduction factor due to the crossing angle at the interaction point. The design luminosity for the LHC is $\mathcal{L} = 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, during the 2012 data-taking period the peak luminosity reached a value of $\mathcal{L} = 7.7 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ as shown in the left plot of Figure 3.2. The total integrated luminosity delivered to the CMS experiment in 2012 amounted to L$= \int \mathcal{L} \cdot \mathrm{dt} = 23.3$ fb^{-1} (where 1 b $\equiv$ 1 barn $\equiv 10^{-24}$ cm^2) as shown by the blue curve on the right hand side of Figure 3.2. More details about the measurement of the luminosity is given in Section 3.2.7.

3.2 The CMS experiment

The CMS experiment is one of the two general purpose particle detectors at the LHC, the second one being ATLAS. The detector itself is situated 100 m underground in the french

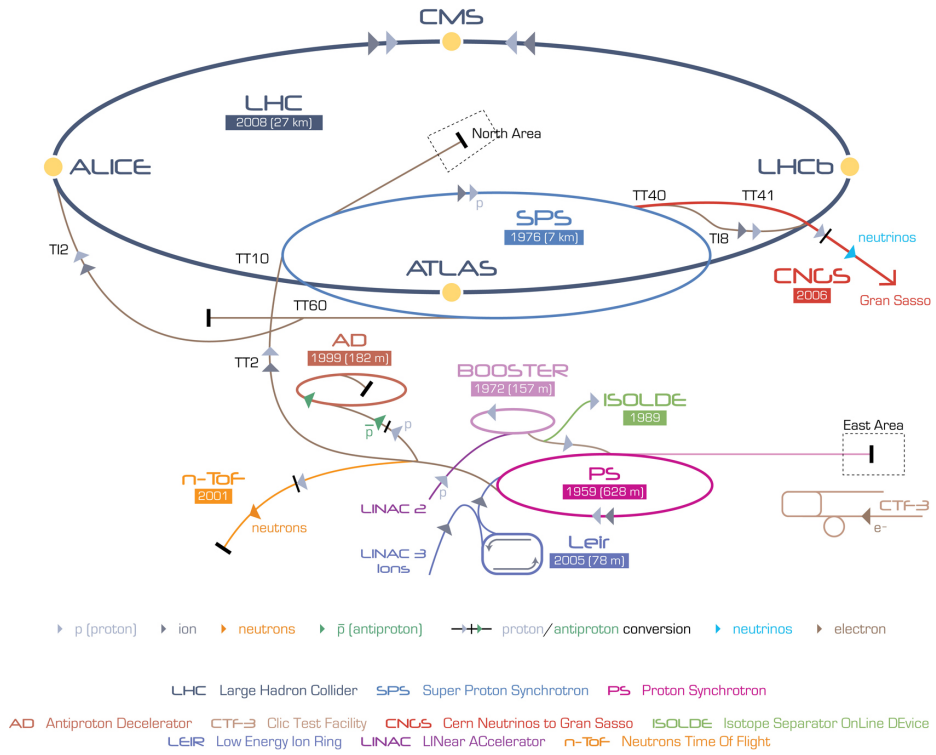

Figure 3.1: The CERN accelerator complex [27].

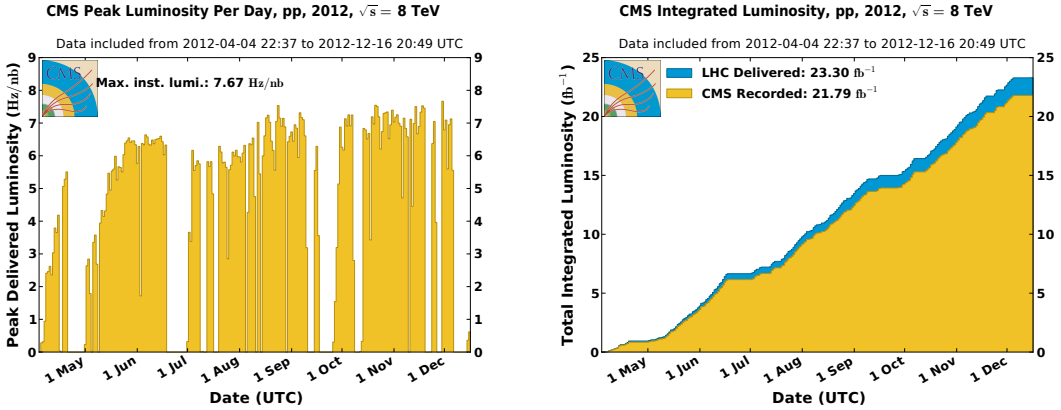

Figure 3.2: CMS peak instantaneous luminosity (left), delivered (blue) and recorded (yellow) integrated luminosity (right) [26].

village of Cessy close to the Jura mountains in the Lake Geneva region. It is 21.6 m long, has a diameter of 14.6 m and weighs over 12,500 tons. An onion-like structure covering a 2π azimuthal angle around the beamline is used to measure products of the high energy proton-proton collisions, which are delivered by the accelerator. The CMS detector consists of a superconducting solenoid, a tracking system, calorimeters, and a muon system, which are described in more detail in the following sections. A graphical overview of the CMS detector is shown in Figure 3.3. The discussion closely follows [28], where a more detailed description of the CMS detector can be found.

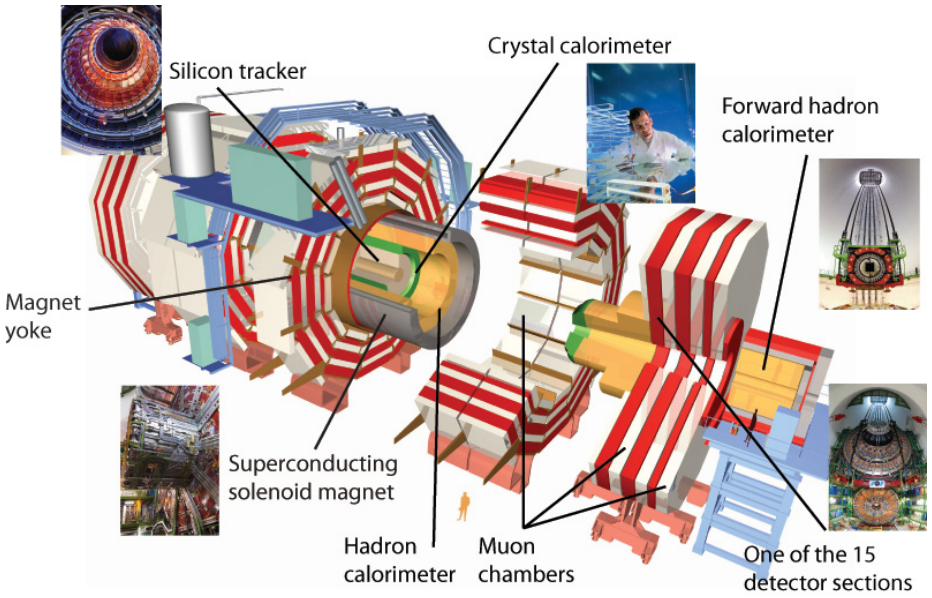

Figure 3.3: CMS detector overview [29].

3.2.1 Detector coordinates and transverse variables

The coordinate system used by CMS has its origin at the center of the detector, where the nominal collision point is expected. The y-axis points vertically upwards and the x-axis points radially inwards to the middle of the LHC ring, while the z-axis points along the beamline. Two angles are defined in this cylindrical coordinate system, the azimuthal angle ϕ and the polar angle θ, where ϕ is calculated with respect to the x-axis in the xy-plane and θ is measured from the z-axis. At a center-of-mass energy of 8 TeV we have to

consider relativistically invariant coordinates. A boost along the z-axis might occur during the collision, therefore we define the Lorentz invariant rapidity y as follows:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}, \tag{3.2}$$

where E represents the energy and p_z the z-component of the momentum vector of a particle. Additionally, a quantity called pseudorapidity η is defined:

$$\eta = -\ln \tan \frac{\theta}{2}, \tag{3.3}$$

which in the massless limit is equal to the rapidity. A value of $\eta = 0$ corresponds to a direction perpendicular to the beamline, while $\eta = 4$ points almost parallel to the beamline.

Variables defined in the transverse plane provide a framework that is independent of a boost in the z-direction. We denote the transverse momentum p_T and the transverse energy E_T of a particle by:

$$p_T = p \ \sin\theta \qquad \text{and} \qquad E_T = E \sin\theta. \tag{3.4}$$

Some particles like neutrinos and the LSP predicted in some SUSY models escape detection and, therefore, leave an energy imbalance in the transverse plane. This missing transverse energy is defined as:

$$E_T^{miss} = -\sum_i p_T^i, \tag{3.5}$$

which describes the negative sum of the p_T of all visible particles i in the event.

3.2.2 Magnet

In order to measure charged particle tracks with high precision, a large bending power of the surrounding magnet is necessary. The design of the CMS magnet system is based on the physics goal to achieve a momentum resolution of $\Delta p/p \approx 10\%$ for muons with a momentum of 1 TeV. This is needed to unambiguously identify the sign of the charge of muons at that energy and reconstruct states like the Z^0 boson decaying into two muons with opposite charge. A large superconducting solenoid with a length of 12.9 m and an inner diameter of 5.9 m is used. With this setup a constant field of 3.8 T can be produced which

accommodates the inner tracking and calorimetry systems of CMS. In order to contain the magnetic field, an iron return yoke interlaced by the muon detectors is placed outside the solenoid.

3.2.3 Tracking System

The inner tracking system of CMS consists of two major parts: the high precision pixel detector and the outer strip tracker. Both parts are fully emerged in the 3.8 T magnetic field and allow a precise reconstruction of charged particle tracks and their momenta coming from the primary interaction point, as well, as secondary vertices that are present during decays of hadrons composed of heavy-flavor quarks. The p_T resolution of the tracks is approximately 1.5% for charged particles with a p_T of 100 GeV.

The entire tracking system utilizes modern silicon technology, which can withstand the harsh radiation environment as the part of the detector closest to the proton beams. In simplified terms, a silicon detector is similar to a pn-junction (diode) run in the reverse bias mode (negative voltage at the p-doped side and positive voltage at the n-doped side). When a charged particle traverses the sensor it creates electron-hole pairs in the material, by moving electrons from the valence band into the conduction band. The negative and positive charge carriers travel then to their respective electrodes and a signal can be measured.

Inner silicon pixel detector

This part of the tracking system is closet to the proton-proton collision point and is therefore under heavy exposure to radiation, with an expected particle flux of $10^7/s$ at a radial distance of 10 cm from the beamline. Silicon pixel technology is used for this inner part of the tracking system, where each pixel has a size of 100×150 μm in (r,ϕ) and z. Three cylindrical layers with a length of 53 cm of silicon pixels are placed around the beamline at radii of 4.4 cm, 7.3 cm, and 10.2 cm. This setup is referred to as the pixel barrel detector. Each layer is divided into ladders such that the full cylinder surface is covered (20 ladders for layer 1, 32 ladders for layer 2, 44 ladders for layer 3). Each ladder consists of 8 modules, where each module has 2×8 read-out chips. The pixel barrel detector is comprised of a total of 768 pixel modules. A forward pixel detector is also installed in the form of two

disks, consisting of blades arranged in a fan-like structure, on either side of the pixel barrel. The endcap disks contain a total of 672 pixel modules with seven different modules on each blade. The pixel detector covers an $|\eta|$ range of up to a value of 2.5 with the addition of the forward detector in a range of $1.5 < |\eta| < 2.5$.

A schematic overview of the pixel layout is shown in Figure 3.4 with the barrel detector in green and the forward disks in pink. On the right hand side of Figure 3.4 a schematic drawing of a charged particle traversing the pixel barrel detector is shown. Due to the strong magnetic field the particles experience a force perpendicular to the direction of motion, which is referred to as Lorentz drift. Electrons have a higher mobility than holes and their deflection angle is larger. This has the advantage that the signal can be read out by more than one pixel leading to an improved resolution [30]. A total of almost 80 million pixels cover an area of ≈ 1 m^2 (barrel and endcap combined) .

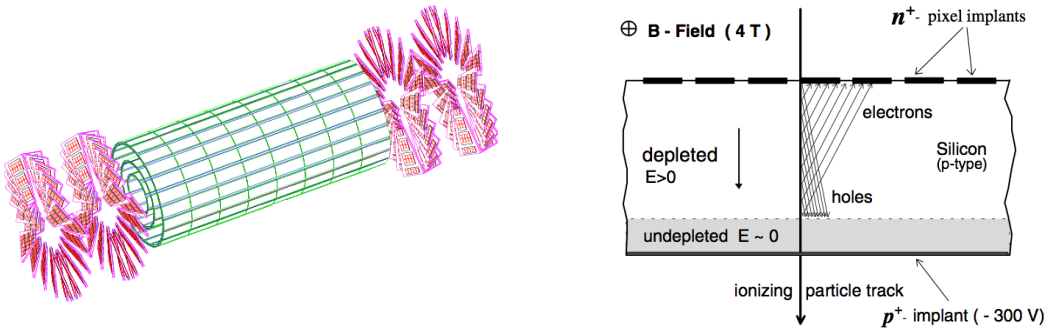

Figure 3.4: Left: Schematic view of the CMS pixel detector [28]. Right: Due to a Lorentz drift of the charge carriers the charge is distributed to neighboring pixels [30].

Outer silicon strip tracker

The outer tracker covers a much larger area than the pixel detector and uses silicon strip technology. Figure 3.5 shows an overview of the complete tracking system with all the parts of the strip tracker. The barrel region consist of two parts: the TIB (Tracker Inner Barrel) and TOB (Tracker Outer Barrel), which are arranged concentrically around the beamline. The TIB has a length of 130 cm in the z-direction, consists of four layers and uses sensors with a thickness of 320 μm and a strip pitch between 80–120 μm. The TOB has a length

of 220 cm in the z-direction and contains 6 layers using a module thickness of 500 μm and a strip pitch of 120 – 180 μm. In Figure 3.5 each line represents a silicon strip module and double lines represent so-called "stereo" modules, where two modules are mounted back-to-back. Stereo modules allow a more precise measurement in two dimensions. To cover the forwards regions two endcap systems are installed on either side of the strip barrel detector: the TID (Tracker Inner Disk) and the TEC (Tracker End Cap). Both of these detectors are disks around the beamline. The TID comprises three layers and fills the empty region between TIB and TEC. While the full strip tracker covers a range of up to $|\eta| = 2.5$, the TEC consists of 9 layers and adds additional tracking information in the region of $0.9 < |\eta| < 2.5$.

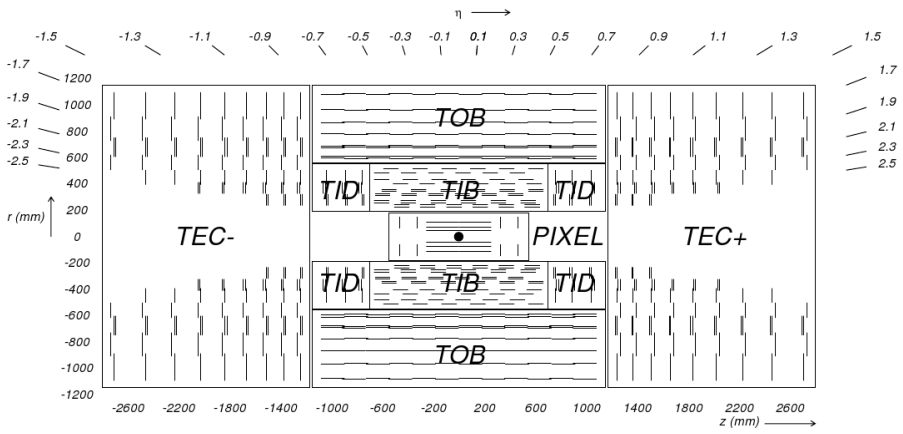

Figure 3.5: CMS silicon strip tracker system including the corresponding $|\eta|$ ranges [28]

3.2.4 Calorimetry

Besides knowing the precise track and momentum of a particle it is also of crucial importance to know its energy. The same is true for neutral particles that do not leave a signature in the tracking system. Within CMS, the calorimeter system is split up into two separate detectors: the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The goal is to have a hermetic system which captures all particles that originate from the proton-proton collisions. While the tracking system is built to have as little material as possible, the calorimeters are designed such that charged and neutral particles, apart from minimum ionizing particles such as muons, deposit all of their energy while traversing them.

Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) system is specifically designed to measure the energy of charged electromagnetic particles (mostly electrons) and neutral ones (photons). Electrons predominantly loose energy due to ionization and scattering with the electric field of the nuclei in the material and subsequent emission of bremsstrahlung. Photons interact with material through the development of an electromagnetic shower, which is started by pair production of electrons. The shower then evolves in the material producing secondary photons and electrons with lower energies until the threshold for pair production is reached. The final particles then dissipate their energy through ionization and excitation processes. The material specific quantity called radiation length, X_0, is defined as the amount of material that an electron has to traverse until its energy reaches $1/e$ of its original value [9].

CMS employs a crystal technology for its ECAL made of 61,200 separate lead tungstate ($PbWO_4$) crystals, which are arranged cylindrically around the tracking system, comprising the barrel detector up to $|\eta| < 1.479$. To cover the forward regions, additional 7,324 crystals are placed on either side of the barrel detector. The endcap regions extend the coverage up to $|\eta| < 3.0$. Lead tungstate is a scintillating crystal with a radiation length of X_0= 0.89 cm. Photodiodes are used to measure the emitted scintillator light. They operate on the basis of the photoelectric effect, the capability of a photon to free an electron from its bound state in a material, and turning the light signal into a measurable current. The initial photons hitting the photodiodes start an amplification process where secondary electrons are produced. Specially designed silicon avalanche photodiodes (APDs) and vacuum phototriode (VPT) are used for the barrel and endcap regions, respectively, to measure the signal.

The ECAL energy (E) resolution is given by:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{3.6}$$

where S stands for the stochastic term, which describes statistics-related fluctuations, N represents the contribution from electronic noise, and C is a constant term, which describes detector non-uniformity and energy calibration uncertainties [9]. The energy resolution for photons measured from Higgs boson decays ranges from 1.1–2.6% in the barrel and from

2.2–5% in the endcap. For electrons originating from Z^0 boson decays the energy resolution is determined to range from 2–5% from the barrel region to the endcap [31]. Figure 3.6 shows the CMS ECAL system.

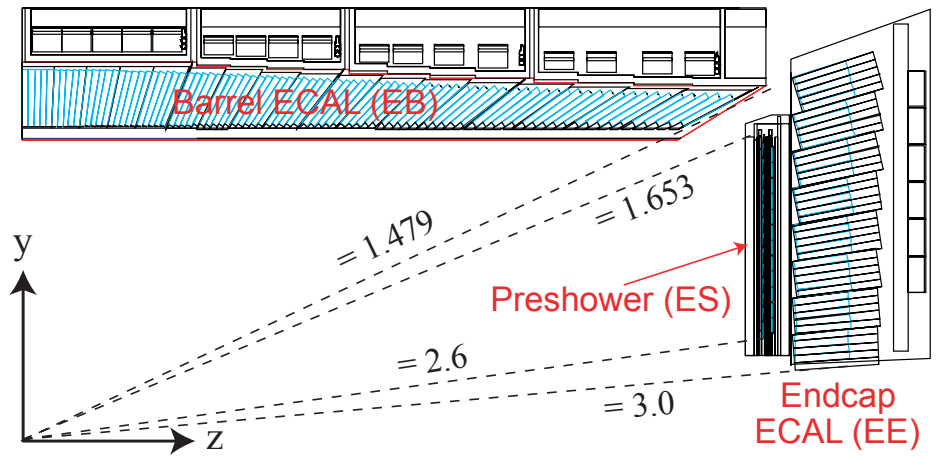

Figure 3.6: CMS the electromagnetic calorimeter system including the respective $|\eta|$ ranges [32].

Hadronic calorimeter

The hadronic calorimeter (HCAL) is important to measure the energy of heavy particles such as protons, neutrons, and pions, which do not deposit all of their energy in the ECAL. These particles account for the energy measurements of jets, which are discussed in more detail in Section 5.4. Another crucial aspect is the indirect measurement of E_T^{miss} as defined in Equation 3.5 through an imbalance of the total energy in an event. When heavy particles traverse a dense material they loose energy due to inelastic collisions with the electrons of the material causing excitation and ionization or interact through the strong force with the nuclei of the material. A characteristic quantity of hadronic calorimeters, the interaction length λ_I, defines the distance after which the incident particle has reduced its energy to $1/e$.

The CMS hadronic calorimeter is situated radially between the outer radius of the ECAL ($R = 1.77$ m) and the superconducting solenoid ($R = 2.95$ m). With the constraint to have most of the calorimeter system within the magnetic field, the goal is to maximize

the amount of material, meaning the number of interactions lengths λ_I, in that region. A sampling structure is chosen using brass (mixture of 70% copper and 30% zinc) as the absorber material, interlaced with plastic scintillator tiles as the active material. When particles go through the absorber they interact with the material and produce a shower of secondary particles. These particles cause a scintillation in the active material, which emits blue-violet light. Wavelength shifting fibers are used to shift the emitted wavelength towards green light which is then transported with optical fibers to the readout electronics consisting of hybrid photodiodes (HPDs). When photons hit the photocathode of the HPD they free electrons, which are then accelerated towards a pixelated silicon diode to amplify the signal through secondary electrons.

Figure 3.7 shows the complete hadronic calorimeter system of CMS. The barrel region of the HCAL (HB) consists of 36 identical azimuthal wedges. The barrel is split up into two half-barrels (HB+ and HB-) and covers a range of $|\eta| < 1.4$ with a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. For structural strength, the innermost (thickness 40 mm) and outermost (thickness 75 mm) layers are made of stainless steel. The brass absorber layers vary in thickness between 50.5 mm and 56.5 mm. Perpendicular to the beamline this amounts to 5.82 λ_I which increases to 10.6 λ_I at $|\eta| = 1.3$. The ECAL material adds another 1.1 λ_I. The hadronic endcap region of the HCAL (HE) covers an additional pseudorapidity range of $1.3 < |\eta| < 3.0$ and employs a similar sampling structure as the HB. In order to reduce the amount of energy that escapes detection, the HCAL extends beyond the solenoid of the magnet in both the barrel and the endcap. This outer hadronic calorimeter (HO) uses the iron return yoke as passive absorber material and scintillator plates as active material to detected the emitted light. The $|\eta|$ coverage is further increased up to a value of $|\eta| < 5.0$ through the forward hadronic calorimeter (HF), which is situated 11.2 m away from the z-coordinate of the interaction point. When charged particles travel faster than the speed of light in a specific medium they emit so-called Cherenkov radiation, which is used to measure the signal in the HF.

When the HCAL is combined with ECAL measurements, jets energies are reconstructed with an energy resolution of $(\sigma/E)^2 = (100\%/\sqrt{E})^2 + (5\%)^2$.

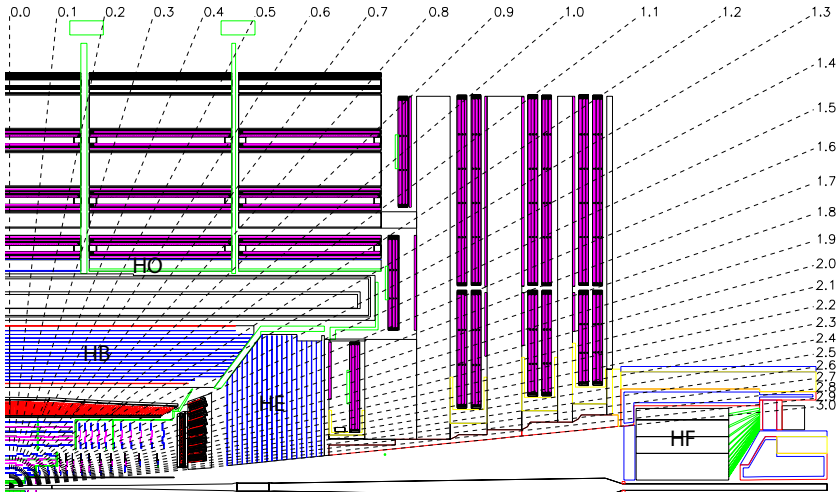

Figure 3.7: CMS the hadronic calorimeter system including the respective $|\eta|$ ranges [28].

3.2.5 Muon system

Muons belong to the second generation of fermions and are roughly 200 times heavier than electrons, but otherwise carry the same quantum numbers. As charged particles they leave a signal in the tracking system, however they interact very little with the material of the electromagnetic and hadronic calorimeters. Therefore, the muon system is the outermost part of the CMS detector. Different types of gaseous detectors are used: drift tube chambers (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC). When a muon travels through the gas in the detector it ionizes the gas atoms. Once an electric field is generated through a voltage difference between the anode and cathode (for example realized through wires), the negatively charged electrons and positively charged ions are used to measure a current indicating the signal.

The different components of the muon system are shown in Figure 3.8. As can be seen from the figure, they are interlaced with the iron return yoke. Drift tubes are installed in the barrel region and cover an $|\eta|$ range of 1.2 with four concentric cylinders and a total of about 250 chambers. Over 172,000 sensitive wires are used in this part of the detector. To cover

the region of $0.9 < |\eta| < 2.4$, CSCs are placed on either side of the CMS detector consisting of four stations with a total of 468 chambers. Each of the chambers is composed of six anode wire planes and seven cathode plates, which measure the position of the traversing muon based on ionization of the surrounding gas. One crucial aspect of the CMS muon system is that it should be able to "trigger" on the p_T of a muon with high efficiency. The concept of triggers within CMS is discussed in more detail in Section 3.2.6. In order to achieve this, a dedicated additional muon system is installed in the barrel and endcap regions covering an $|\eta|$ region up to 1.6, consisting of RPCs shown in red in Figure 3.8. The RPCs are gaseous parallel-plate detectors, where one plate functions as the anode and another one as the cathode.

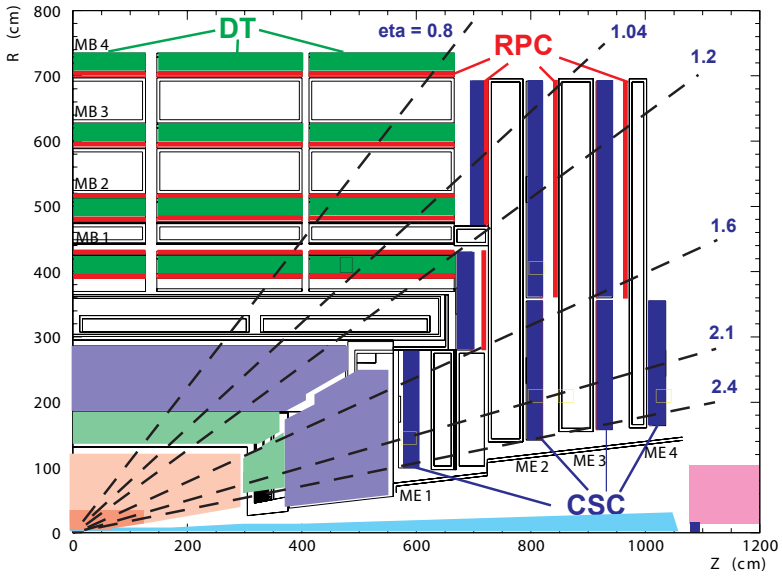

Figure 3.8: CMS the muon system [28].

3.2.6 Trigger system

During 2012, the peak luminosity delivered by the LHC was $8{\times}10^{33}$ cm^{-2}s^{-1}. This multiplied by a total inelastic cross section of about 70 mb leads to a maximum event rate of $5.6{\times}10^{7}$ s^{-1}. It is impossible to store every single event, therefore the rate needs to be reduced significantly in order to allow for processing of the delivered collisions. This rate reduction is motivated by physics aspects in order to single out interesting events that are

worth keeping for further analysis, for example, precision measurements or searches for new physics. In CMS the so-called trigger system is responsible for the rate reduction based on interesting physics phenomena. The trigger system is designed with two separate steps, the level-1 (L1) trigger and the high level trigger (HLT), and both are described in this section in more detail.

Level-1 trigger

The L1 trigger is based on customized and programmable electronics and is designed to decrease the rate of events down to 100 kHz. The decision whether to accept or reject an event occurs within 3.2 μs. This part of the trigger uses coarse information from the calorimeters and the muon system and keeps the high-resolution data in memory. The two major components of the L1 system, the muon trigger and calorimeter triggers, can be further subdivided into local, regional, and global components, ordered by increasing complexity. A schematic setup of the L1 trigger system is shown in Figure 3.9.

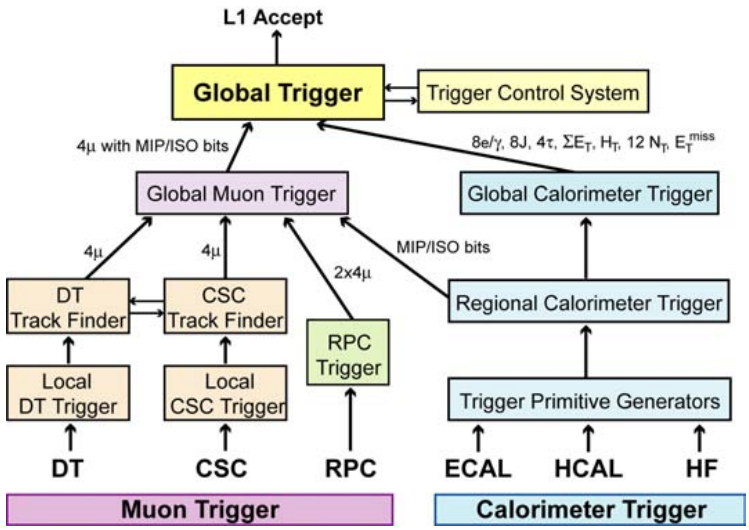

Figure 3.9: Schematic overview of the L1 trigger system layout [28].

Calorimeter trigger at L1

The lowest layer of the calorimeter trigger is called Trigger Primitive Generator (TPG)

and defines a list of trigger towers. A trigger tower in the barrel region is defined by mapping single HCAL cells on to 5×5 ECAL crystal arrays and summing up their respective energies [33] with an (η, ϕ)-coverage of 0.087×0.087. These trigger towers are sent to the Regional Calorimeter Trigger (RCT) which finds electron/photon candidates and sums the energy deposits further into regions, where a region is defined as 4×4 towers. Electrons/photons are identified by finding the trigger towers with the largest energy deposits and applying quality selection criteria depending on the ratio of energy deposited in the HCAL and ECAL (typically up to 5%) and the extension of the electromagnetic shower. An isolated electron is found when all eight neighboring towers also satisfy the same criteria on the ratio of the HCAL and ECAL energy deposits. Four isolated and four non-isolated electron/photon candidates are sent to the Global Calorimeter Trigger (GCT). The GCT provides an ordering of the electron/photon candidates by their transverse energies and determines jets, H_T (transverse sum of jets above a threshold), and E_T^{miss}. At this level, jets are reconstructed by a four-stage clustering algorithm described in detail in Reference [34] and the result is a 3×3 cluster of region sums (seeds) provided by the RCT. The variable H_T, which is a useful L1 seed for multijet triggers, is calculated from the previously found jets with $|\eta| < 3.0$ and E_T >10 GeV. The result from the GCT is then forwarded to the Global Trigger (GT) which makes the final L1 decision. During the 2012 data-taking period a strong dependence of the H_T trigger rate was observed on the instantaneous luminosity. To mitigate this effect a 5 GeV threshold was introduced on the seeds used in the jet-finding algorithm in June 2012.

Muon trigger at L1

All three muon subsystems, DT, CSC, and RPC, participate in identifying muons for trigger purposes. The first local level of reconstruction comes from the electronics in the DT (barrel region) and CSC (endcap region) systems delivering information for each of the muon chambers separately. This information is collected by the so-called Track Finders (TF), which combine the hits in the different chambers, reconstruct the track of the muon, and assign a transverse momentum to the candidate. Each TF sends the four highest p_T candidates to the Global Muon Trigger (GMT). The

RPC system does not use the first local steps, but instead reconstructs the track of the muon candidate directly and sends a total of eight candidates (four for barrel and four for endcap) to the GMT. The GMT tries to match the DT and CSC candidates with the information from the RPC. It also uses information about the calorimeters from the RCT to determine whether the muon candidates are isolated. Finally the four best muon candidates, based on p_T, isolation criteria, and correlations of the different muon subsystems, are sent to the GT, which makes the final decision to keep an event or reject it, before passing it to the HLT.

High Level Trigger

The HLT is the second level of the CMS trigger system and is based on software that is run on a farm of commercial computers with over 13,000 CPU cores [35]. The goal is to further reduce the output rate of the L1 trigger (≈100 kHz) down to a manageable rate of ≈100Hz. The event reconstruction performed by the HLT is referred to as "online reconstruction", whereas the event reconstruction performed further downstream is called "offline reconstruction", and is described in more detail in Chapter 5. The online reconstruction is designed to minimize CPU usage and time needed to make a decision whether or not to keep an event. Each trigger path consists of certain L1 seeds, which could be, for example, muons, jets, or H_T, followed by selection criteria applied to the objects reconstructed by the HLT system. The name of the trigger gives an indication of the HLT selection. For instance, *HLT_Mu20_eta2p1* refers to events that contain a muon above 20 GeV and $|\eta| < 2.1$. In order to control the rate of certain trigger paths, without having to alter the HLT setup, the concept of a prescale is introduced. The prescale value gives the suppression factor for a given trigger path. For example, a prescale value of 100 means that only one event is recorded for every 100 where the trigger would have resulted in a positive decision. This reduction factor is adjusted dynamically depending on the instantaneous luminosity delivered by the LHC in order to keep the output rate roughly constant. For physics analysis un-prescaled triggers are used when the integrated luminosity is important. Prescaled triggers can be used for trigger efficiency studies or measurements of branching ratios, where the overall integrated luminosity is not important since ratios of quantities are determined.

For this analysis we are mostly interested in events with multiple jets, namely the trigger path *HLT_QuadJet60_DiJet20* where events are required to have four jets above 60 GeV and two additional jets above 20 GeV. Jets are reconstructed at the HLT with the anti-k_t algorithm based on calorimeter information only. This is similar to the offline reconstruction and a description of different jet algorithms can be found in Section 5.4.1. This trigger was un-prescaled for the duration of the 2012 data-taking. As mentioned before, each HLT path is seeded by a L1 requirement. At first the *HLT_QuadJet60_DiJet20* was seeded by requiring four jets above 32 GeV at L1, collecting about 12 fb^{-1} of data. For the last 7.4 fb^{-1}, the seed was extended to include an "or" of H_T and dijet (two jets above 52 GeV) L1 seeds in addition to the previously mentioned four-jet seed. The choice of trigger and measurements of its efficiency are discussed in Section 6.1.

Depending on the HLT path events are categorized and stored in different primary datasets.

3.2.7 Luminosity measurement

The measurement of the luminosity delivered to and recorded by CMS is of great importance to most physics analyses and the recent results for the 2012 data-taking are presented in Reference [36]. Whenever the LHC goes into the mode of "stable beams", meaning proton-proton collisions for physics measurements are delivered, the CMS detector needs to be in a state where it is able to record events. After all subdetectors are configured, a new "run" is started manually by the shift crew in the CMS control room. At this point data is recorded with a granularity that is referred to as "luminosity section" (LS), which is defined as the time interval $t_s = 23.31$ s. Two parts of the CMS detector can be used to measure the luminosity, the hadronic forward calorimeter and the pixel detector. The latter method is chosen due to its smaller dependence on multiple interactions and other beam conditions. The luminosity is then evaluated from the average number of pixel clusters $\langle n \rangle$, groups of sufficiently charged adjacent pixels in zero-bias events (zero-bias triggers only require the occurrence of a bunch crossing), and given by:

$$\mathcal{L} = \frac{\langle n \rangle f_{rev}}{\sigma_{vis}}, \tag{3.7}$$

where $f_{rev} = 11246$ Hz is the beam revolution frequency, and σ_{vis} the visible cross section, which is calibrated through a Van der Meer (VdM) scan. During a VdM scan the two beams are moved with respect to each other in the horizontal and vertical plane. This allows one to measure the beam overlap in the x- and y-directions, which is needed to determine σ_{vis}. To estimate the integrated luminosity for a given dataset the average number of pixel clusters per event is calculated for a given LS and multiplied by t_s. The total integrated luminosity is then the sum of all LS that are used in the analysis. The uncertainty on the measurement is estimated to be about 2.6% [36]. Figure 3.2 shows the delivered (blue) and recorded (yellow) integrated luminosity for the 2012 data-taking period.

Chapter 4

Data and simulated samples

A general overview is given of the collected data and quality criteria that are applied before further physics analysis is performed. The chapter ends with a discussion of the simulated signal and background samples important for this analysis.

4.1 Data sample

This analysis uses a total integrated luminosity of 19.4 fb^{-1} of proton-proton collisions collected with the CMS experiment during the 2012 run of the LHC at $\sqrt{s} = 8$ TeV. After a positive trigger decision is made, as described in Section 3.2.6, events are stored in primary datasets and reconstructed following algorithms described in Chapter 5. For this analysis we use the "Multijet primary dataset", which contains several different trigger paths requiring at least two or more jets in the event. A dedicated Data Quality Monitoring (DQM) system is in place to ensure a high efficiency and quality of the recorded data. Real-time online monitoring of all parts of the detector is available during data-taking, to spot problems quickly and act accordingly. The second step of the monitoring occurs simultaneously with the reconstruction of events. Characteristic distributions, for example of pixel clusters and their charge, or track parameters, are compared to reference distributions. Depending on the agreement with the reference distributions a quality flag is set to "good" or "bad". The data is split up into different runs and further subdivided into several LS, as mentioned in Section 3.2.7. Ideally all subdetectors have a "good" flag set for a given LS. This information is stored and a list is created containing all good runs and LS, which is used to select data for further physics analyses. Events that are reconstructed shortly after they have been recorded are referred to as prompt reconstruction. These datasets are used to derive measurements of quantities like jet energy corrections (see Section 5.4.2) or general updates on detector

alignment conditions, which then in return can be included when the data is reprocessed at a later stage, referred to as re-reconstruction. In this analysis, we use a combination of prompt and re-reconstructed data. Table 4.1 shows the different datasets and the corresponding integrated luminosity for each dataset based upon the aforementioned list of good runs.

Run_{begin}	Run_{end}	$\int \mathcal{L}$ [pb]	Dataset
190782	190949	81	/MultiJet/Run2012A-recover-06Aug2012-v1
190645	193621	796	/MultiJet/Run2012A-13Jul2012-v1
194305	196432	4387	/MultiJet/Run2012B-13Jul2012-v1
198049	198522	470	/MultiJet/Run2012C-PromptReco-v1
198941	202016	6320	/MultiJet/Run2012C-PromptReco-v2
205515	208686	7295	/MultiJet/Run2012D-PromptReco-v1

Table 4.1: Multijet primary dataset.

4.2 Simulated samples

Simulations of physics processes and expected detector performance are of crucial importance in the field of high energy physics and used to understand the measured detector responses, make comparisons of collision data with SM processes, and to aid in the search for new physics signals. These simulations are based on so-called Monte Carlo (MC) techniques to numerically solve complex matrix elements, model parton showering, and produce events similar to recorded data. These events are then passed through the CMS detector simulation based on the GEANT4 [37] package and the output is treated almost the same way as are the data. The two event generators dominantly used in this analysis are the general purpose generator PYTHIA [38] and the matrix element calculator MADGRAPH [39]. The explicit version number is given when the different samples are discussed. While PYTHIA is preferably used to model parton shower evolution, MADGRAPH can be used to easily implement new physics models based on Feynman diagrams implemented in FeynRules [40]. A common procedure is to interface both generators, making use of the matrix element calculation in MADGRAPH and the hadronic shower evolution in PYTHIA.

MLM matching scheme between MADGRAPH and PYTHIA

When a matrix element generator is interfaced with the parton shower program both can contribute additional partons in the final state. In order to avoid double counting of final state particles a matching scheme is implemented following the MLM [41] prescription using k_t based clustering, as described in Section 5.4.1. For this approach, two values denoted with:

- *xqcut*: defines the minimum k_t measure for additional partons in MADGRAPH
- Q_{cut}: the matching scale parameter, defines the scale of the transition region between matrix element and parton shower generator

are introduced. After the showering has been performed jets are reconstructed with a k_t algorithm using the final state partons in the event. These jets are compared to original partons generated by the matrix element calculation. If any jet remains without a matched partner the event is rejected, resulting in samples of exclusive jet multiplicities. For the highest multiplicity, extra jets are allowed and, in combination with the exclusive jet multiplicity samples, one obtains a final inclusive sample [42].

4.2.1 Benchmark MC signal samples

The new physics model of pair-produced gluinos each decaying through RPV couplings into three quarks has already been discussed in detail in Section 2.2.2. This signal is modeled using the PYTHIA event generator (v6.424). The baryon number violating decay parameters (RVLAMB) are set to non-zero values to allow gluino decays into three quarks, i.e., $\tilde{g} \to u_i d_j d_j$, where u represents up-type quarks and d represents down-type quarks, and the indices i, j, k run over the three generations. Two different scenarios of this decay are considered and the respective PYTHIA settings are given below. The CTEQ6L1 [43] PDF set is used for the generation of events. More details of the PYTHIA input parameters for the signal generation can be found in Appendix A.

- Scenario 1 (decay to three light-flavor quarks):
  - $\tilde{g} \to uds$

  - PYTHIA : RVLAMB(1,1,2) = 0.005

- Scenario 2 (decay to at least one heavy-flavor and two light-flavor quarks):
  - $\tilde{g} \to udb$ or csb
  - PYTHIA : RVLAMB(1,1,3) = 0.005
  - PYTHIA : RVLAMB(2,2,3) = 0.005

We produce samples with gluino masses of 200 – 500 GeV in 50 GeV increments, and from 750 – 1500 GeV in 250 GeV increments. The invariant mass of the intermediate squark in the decay is taken to be decoupled from the gluino, set to 2 TeV for gluino masses up to 500 GeV and to 7.5 TeV for higher gluino masses. This leads to prompt decays, meaning the gluino decays immediately after its production through a virtual squark. The intrinsic width of the gluino, however, is such that it is negligible with respect to the mass resolution resulting from the reconstructed jets. In addition to the decoupled mass spectrum we assume a 100% branching ratio of the gluino decaying into quarks. Several other processes are proposed in the context of these so-called simplified models, which were designed to reduce the number of free parameters and involve usually only a few particles and their interactions [44].

The production cross section for some of these models is determined from next-to-leading-order (NLO) and next-to-leading-logarithm (NLL) calculations [45, 46, 47, 48, 49] as a function of the particle mass and shown in Figure 4.1. Most notably the gluino pair production cross section in grey is the highest followed by the squark-antisquark production cross section in red. The shaded bands around the solid line represent uncertainties associated with the calculation due to the choice of PDF and renormalization and factorization scale [50]. Two sets of PDFs are used CTEQ6.6 and MSTW2008 and the average of the two calculations determines the nominal value. Table 4.2 summarizes the gluino pair production cross sections with corresponding uncertainties for the mass points considered in this analysis. The uncertainties range from 15% to 43% for the highest masses.

4.2.2 Background samples: QCD and $t\bar{t}$

The main background in this analysis arises from QCD multijet production, which is theoretically difficult to model. Especially final states with six or more jets are computationally

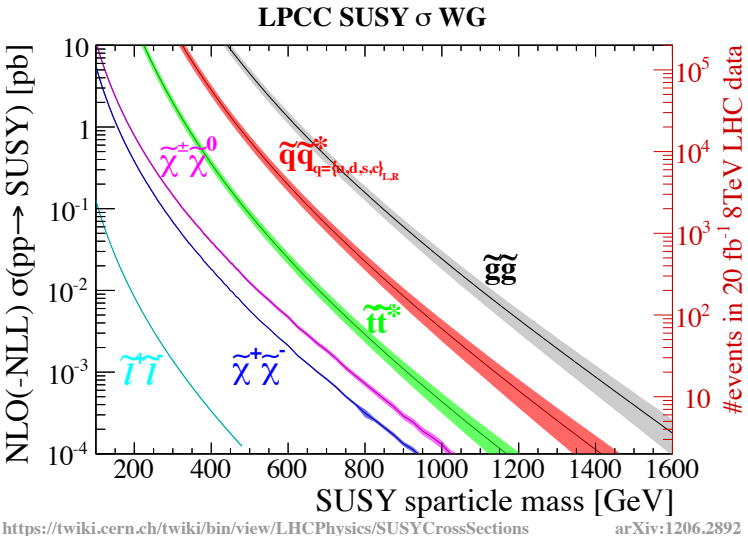

Figure 4.1: NLO+NLL theory cross section simplified topology used for SUSY searches at the LHC. The gluino pair production cross section and its uncertainties are shown in grey as a function of gluino mass [51].

Gluino mass [GeV]	Cross section [pb]
200	1010 ± 15%
250	302 ± 15%
300	106 ± 15%
350	43 ± 15%
400	19 ± 15%
450	8.9 ± 15%
500	4.5 ± 16%
750	0.26 ± 19%
1000	0.024 ± 26%
1250	0.0029 ± 34%
1500	0.00039 ± 43%

Table 4.2: NLO + NLL cross sections for pair-produced gluinos with all squark masses decoupled [45, 46, 47, 48, 49].

very time intensive. The samples used here to compare QCD multijet background and data are produced with the MADGRAPH event generator (v5.1.3.30) which is interfaced with the PYTHIA generator (v.6.424) to model the parton shower. The CTEQ6L1 PDF set is used for the production and at the matrix element level up to four partons are included in the final state. During the parton shower evolution and subsequent hadronization PYTHIA adds additional jets to the event. The matching procedure follows the MLM prescription with the parameters $xqcut = 40$ GeV and $Q_{cut} = 60$ GeV. Millions of events are necessary in these samples, since the cross section decreases drastically for high jet multiplicities, roughly by one order of magnitude going from one jet multiplicity (n jets) to the next (n+1 jets). Due to the size of the samples several different ones are produced containing events where the sum of the p_T of all particles in the event H_T lies between 250–500 GeV, 500-1000 GeV, and higher than 1000 GeV.

Another background that becomes important when events are required to contain b jets, is the contribution from all-hadronic $t\bar{t}$ events. These samples are also produced with MADGRAPH and showered with PYTHIA, with the same version numbers and choice of PDF as noted above for the QCD multijet samples. The top mass is set to $m_t = 172.5$ GeV and up to three additional partons in the final state are included by the matrix element calculation. The matching scheme is again following the MLM prescription with the parameters set to $xqcut = 20$ GeV and Q_{cut}=40 GeV. In these $t\bar{t}$ MADGRAPH samples the scale of the hard interaction is set to $Q^2 = m_t^2 + \sum p_T^2$, with m_t being the mass of the top quark and the sum p_T of all additional partons from the matrix element calculation.

Table 4.3 shows a summary of the QCD and $t\bar{t}$ samples for this analysis. The first column gives the number of events generated and the second column represents the theoretical cross section that is used to estimate the expected number of events of a given sample in data. The cross section for QCD is calculated at leading-order LO and the one $t\bar{t}$ is determined at next-to-next-to-leading-order NNLO [52]. Theory uncertainties associated with the modeling of events are discussed in more detail later in Section 9.3, however we also list in Table 4.3 alternative simulated $t\bar{t}$ samples for the evaluation of some of these uncertainties.

# of Events	Cross section [pb]	Dataset name
QCD multijet samples (MADGRAPH+PYTHIA, TuneZ2star)		
26 292 779	276000	250 GeV $< H_T <$ 500 GeV
31 302 905	8426	500 GeV $< H_T <$ 1000 GeV
13 879 218	204	1000 GeV $< H_T <$ Infinity
$t\bar{t}$+ jets (MADGRAPH+PYTHIA, Tauola, TuneZ2star)		
6 923 750	245.8	$t\bar{t}$+ jets $xqcut$ = 20 GeV and Q_{cut}=40 GeV nominal scale $\rightarrow Q^2 = m_t^2 + \sum p_T^2$
5 476 728	245.8	$t\bar{t}$+ jets matching scale down $xqcut$ = 10 GeV and Q_{cut}=30 GeV
5 415 010	245.8	$t\bar{t}$+ jets matching scale up $xqcut$ = 40 GeV and Q_{cut}=60 GeV
5 387 181	245.8	$t\bar{t}$+ jets renormalization/factorization scale down nominal scale $\times$ 0.5
5 009 488	245.8	$t\bar{t}$+ jets renormalization/factorization scale up nominal scale $\times$ 2.0

Table 4.3: Monte Carlo samples for QCD multijet events (top three rows) and $t\bar{t}$ pair production including dedicated samples to study the effect of modeling uncertainties.

Chapter 5

Event Reconstruction

After a positive decision from the HLT the events are sent to the so-called storage manager that writes the data to disks at P5 (the location of CMS in Cessy). Afterwards the data is transferred to CERN for further processing, where the offline event reconstruction is performed after 48 hours of recording the data. In this chapter we discuss the algorithms used by CMS to reconstruct objects such as tracks, vertices, jets, and leptons, with a focus on jets, which are important for this analysis.

5.1 Particle-flow algorithm

Event reconstruction in CMS is performed with the so-called particle-flow algorithm, which attempts a description of the full event based on single identified particles [53]. The goal is to reconstruct all stable particles like electrons, muons, photons, charged hadrons, and neutral hadrons, using a combination of all subdetectors of CMS. Each particle traversing the detector leaves either a track (if charged), deposits energy in the calorimeters (charged or neutral), or leaves hits in the muon chambers (charged). The challenge is to combine the information from the different detector parts and to link all elements together properly to identify each of the aforementioned particle types. Since charged hadrons leave signals in the pixel and strip trackers, their momenta can be measured with very high precision. The final result is a list of particles for each event.

5.2 Charged particle tracks

A crucial part of the particle-flow algorithm is to identify tracks for charged particles with high precision. Particles moving outwards from the collision point have to traverse the CMS detector following a helix path, due to the strong magnetic field. An iterative tracking

Iteration	Seeding Layers	p_T (GeV)	d_0 (cm)	$\lvert z_0 \rvert$
0	three pixel hits	> 0.8	< 0.2	$< 3\sigma$
1	two hits + primary vertex	> 0.6	< 0.2	< 2 cm
2	three pixel hits	> 0.075	< 0.2	$< 3.3\sigma$
3	three hits in pixel or strip tracker	> 0.35	< 1.2	< 10 cm
4	TIB 1+2 & TID/TEC ring 1+2	> 0.5	< 2.0	< 10 cm
5	TOB 1+2 & TEC ring 5	> 0.6	< 5.0	< 30 cm

Table 5.1: Requirements on seeds used for the six iterations of the track finding algorithm [57]. From left to right for each iteration the seed layers are defined followed by the minimum p_T requirement, and the maximum transverse (d_0) and longitudinal (z_0) distance from the beamspot (σ here denotes the Gaussian width of the beamspot in z-direction).

algorithm called the combinatorial track finder (CTF) is used in CMS [54], where the seeding and p_T requirements are changed for each iteration. A seed is an initial estimate of a track and its uncertainty based only on two or three hits in the tracking system. These seeds are propagated outwards by a Kalman filter [55] to find hits that coincide with the predicted trajectory of the charged particle. A Kalman filter is a recursive algorithm that can be used to predict the evolution of dynamical systems [56], which in this case is the trajectory. It combines information about the current state of the trajectory with uncertainties, the statistical noise, and the underlying physical process of a particle moving in a magnetic field. With each new hit the track and its uncertainty is recalculated until either the outer layer of the tracker is reached or no additional compatible hits can be found. Six iterations are performed ordered by how difficult it is to identify the track; iteration 0, for example, looks for tracks with $p_T > 0.8$ that originate close to the interaction point (prompt tracks) and have at least three hits in the pixel detector. After each iteration tracks that satisfy certain selection criteria, such as the number of layers that have hits, the χ^2 per degree of freedom for the fitted track and the distance to the primary vertex, are marked as "high purity". The associated hits from these tracks are removed from the hit collection used in subsequent iterations. Iterations 1 and 2 try to find tracks with only two hits in the pixel detector or have lower p_T than in iteration 0. Iterations 3-5 then try to reconstruct non-prompt tracks that originate further away from the primary interaction point [57]. Table 5.1 shows the different seeding requirements for all six iterations.

5.3 Vertex reconstruction

With the increasing instantaneous luminosity and center-of-mass energy the effects of pile-up, multiple interactions of different protons in one bunch crossing, become increasingly difficult to handle. Additionally, there is a finite probability that two or more hard interactions occur between partons in the same proton-proton collision. For many physics analyses it is important to determine the position of these interactions. The vertex reconstruction is performed in two steps in CMS. First, tracks are selected that are produced promptly and grouped together in clusters based on their z-coordinate. This first step of grouping the tracks is performed with a deterministic annealing (DA) algorithm [58]. For each track the z-coordinate of the point of closest approach to the beamline is denoted by z_i with a measured uncertainty of σ_i. The goal is to assign these tracks to an unknown number of vertices denoted by a z-coordinate of z_k. As an example, a χ^2 is defined as the figure of merit:

$$\chi^2 = \sum_{ik} p_{ik} \frac{(z_i - z_k)^2}{\sigma_i^2}, \tag{5.1}$$

where p_{ik} can be interpreted as a probability with values between 0 and 1. Instead of just finding the setup that minimizes this χ^2, the DA algorithm finds the most likely distribution for a given value χ_0^2. This χ_0^2 is then decreased until it finds a good reliable minimum [58]. Once the tracks are assigned to the different vertices a three dimensional fit is performed using the AdaptiveVertexFitter [59]. In this fit each track is assigned a weight between 0 and 1. After the last step of the fitting most tracks are either categorized as 1 (good) or close to 0 (outliers). Then the sum of the weights corresponds roughly to the number of tracks associated with the vertex. Based on these track weights each vertex is assigned a number of degrees of freedom $n_{dof} = 2\sum w_i - 3$. A quality selection is applied to only consider vertices with $n_{dof} > 4$, which corresponds here to having at least four tracks assigned to the vertex [58]. All reconstructed vertices are ordered by the sum of the p_T^2 of the tracks associated with them. The average number of reconstructed vertices for the dataset of this analysis is around 15.

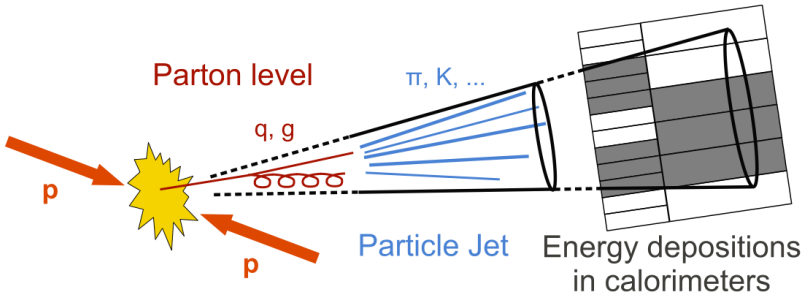

Figure 5.1: Graphical representation of a jet [60].

5.4 Jets

As discussed in Section 2.1.4, quarks and gluons cannot be observed as single particles, since due to the nature of the strong force they undergo the process of parton showering and hadronization. What we observe is a spray of hadrons that are clustered together by algorithms which then define what we call a "jet". Experimentally these hadrons leave certain signatures in the detector by depositing their energy in the electromagnetic and hadronic calorimeters, and if charged, leave tracks in the pixel and silicon strip tracker. The graphic in Figure 5.1 shows schematically how a quark or gluon leaves the hard interaction, undergoes the processes of parton showering and hadronization, and finally how the hadrons deposit their energy in the calorimeters. Within CMS three different methods, making use of information from different parts of the detector, are used as input to the jet reconstruction algorithm. Jets can be reconstructed based solely on calorimeter information, or based on tracking and calorimeter information ("jets-plus-tracks"), or the particle-flow approach discussed in Section 5.1. With the particle-flow algorithm one obtains a list of all charged and neutral particles in the event, which is used to form the jet. We mostly focus on particle-flow jets in this thesis, since they have the advantage of an improved jet energy resolution and smaller uncertainties with respect to the other two approaches. Much theoretical work has gone into studying jet algorithms, their properties, and how closely they can represent the original parton that was produced.

5.4.1 Jet algorithms

To understand processes that involve quarks and gluons in the final state, a set of rules, in the form of a jet algorithm, needs to be defined on how particles are clustered together. Theoretical interests, as well as experimental constraints in terms of computing time, need to be considered for this task. In this section, we briefly discuss different types of jet algorithms emphasizing the ones important for CMS and for this thesis.

Cone algorithms

Some of the first jet algorithms involved using fixed "cones" in $y - \phi$ space to define jets. In case of "iterative cone" algorithms one starts with a seed i, for example, the highest energy ECAL tower or highest p_T particle in the event. Iteratively all four-vectors of the particles j within a radius of:

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 < R^2, \tag{5.2}$$

are added to the four-vector of the original seed particle i. The variable R in Equation 5.2 is called the jet radius, and y and ϕ are the rapidity and azimuthal angle respectively, of the particles i and j. This basic approach has multiple drawbacks, for example, it is not clear what to do with overlapping cones. The more serious issue is that it is infrared and collinear unsafe (IRC unsafe). Having an algorithm that is IRC safe is an important feature since it ensures that the final set of hard jets is unchanged when collinear splitting of hard particles or soft emission of radiation occurs. One way of avoiding these issues is the use of a seedless algorithm which identifies all stable cones in an event and uses a split-merge approach when overlapping cones are found. This split-merge step either merges two jets or splits them if the ratio of the p_T associated with both cones is larger or smaller than some fraction f [61].

Sequential recombination algorithms

These types of algorithms do not rely on a fixed cone but rather perform the clustering of particles based on a minimum distance measure between the particles i and j and between

the beamline B and particle i:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \tag{5.3}$$

$$d_{iB} = k_{ti}^{2p}, \tag{5.4}$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_t is the transverse momentum, y the rapidity, and ϕ the azimuthal angle of the i'th or j'th particle [62]. The distance parameter R is similar to the one introduced in the previous section. The variable p can be either 1 (k_t algorithm [63]), 0 (Cambridge/Aachen algorithm [64]), or -1 (anti-k_t algorithm [62])), and defines different types of clustering algorithms. Sequential clustering algorithms are infrared and collinear safe and are therefore favored from a theoretical point of view. For each particle and pair of particles the two aforementioned values are calculated and the minimal d_{ij} or d_{iB} is determined. If the minimum is the pair d_{ij}, both four-vectors are combined into a new single particle four-vector and the procedure starts from the beginning. If d_{iB} is the minimal value then this particle i is called a jet and removed from the available list of particles for clustering. These steps are repeated until no particles remain [61]. Figure 5.2 shows a comparison of the reconstructed jets when the same event is clustered with the three different choices of p. While the k_t and Cambridge/Aachen jets show a rather irregular shape, the anti-k_t algorithm produces nearly conical jets. This is attributed to the fact that soft particles tend to cluster with the harder ones first, since the distance to the d_{ij} is mostly defined by the transverse momentum of the hard particle [62]. For this analysis we focus on jets clustered with the anti-k_t algorithm and a distance parameter of $R = 0.5$. Cambridge/Aachen jets offer the possibility to make use of newly developed substructure techniques and are in addition to anti-k_t jets used within the CMS collaboration, for example, in searches for decays of heavy particles in boosted topologies [65].

5.4.2 Jet energy scale corrections

The reconstructed jets and their energy need to be corrected to take into account the nonlinear and nonuniform response of the calorimeters, to reduce the effect of pile-up, and

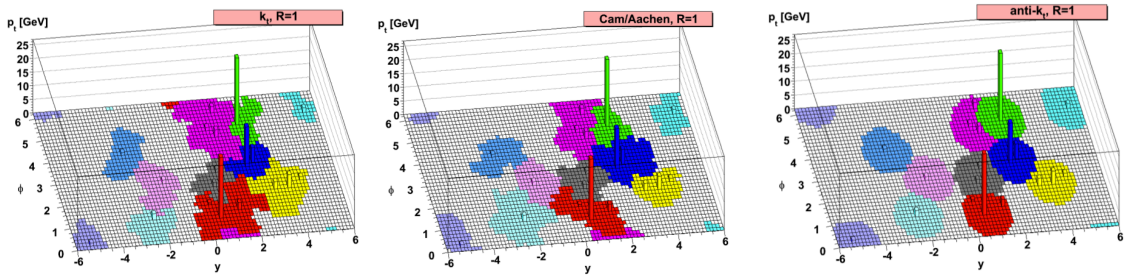

Figure 5.2: The same event is clustered with three different algorithms, from left to right k_t, Cambridge/Aachen, and anti-k_t [62].

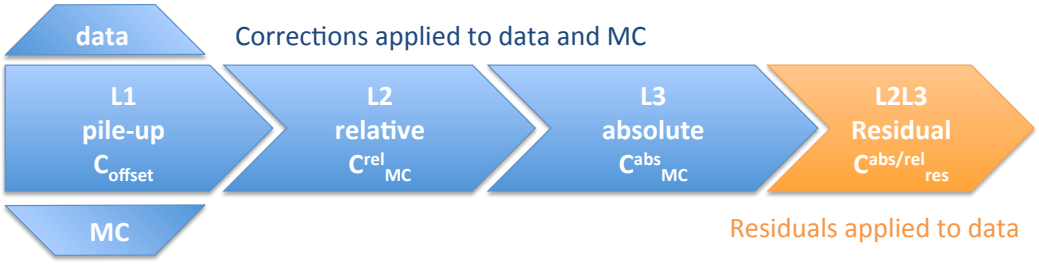

Figure 5.3: Factorized approach for jet energy scale corrections.

to correct for small residual effects in the data after corrections based on MC simulation are applied. The precise knowledge of the jet energy scale is of crucial importance for many physics analyses and the uncertainty in the corrections is one of the leading uncertainties for the analysis presented in this thesis. The discussion here follows mainly Reference [66], where a more detailed description can be found. A factorized approach was developed by CMS to apply all of these corrections, labeled from L1 to L3, as shown in a schematic overview in Figure 5.3. Equation 5.5 shows the factorized approach mathematically, where the correction factors C are applied to each of the four-vector components of the uncorrected reconstructed momentum p_μ^{raw}, to obtain a calibrated jet denoted with p_μ^{corr}. The corrected momentum is given by:

$$p_\mu^{corr} = C^{\text{offset}}(p_T^{raw}) \cdot C_{\text{MC}}^{abs,rel}(p_T', \eta) \cdot C_{res}^{\text{rel}}(\eta) \cdot C_{res}^{\text{abs}}(p_T'') \cdot p_\mu^{raw} , \tag{5.5}$$

where p'_T and p''_T are the transverse momenta after applying the offset and relative/absolute corrections, respectively. First, an offset correction (L1) is applied to subtract energy that is not associated with the hard scattering, and originates from electronic noise, pile-up, or the underlying event. Afterwards corrections are derived from MC simulation (L2, L3) to match the energy of reconstructed jets to the energy of simulated particle jets, these corrections are then applied to simulation and data. The last step in the standard chain is applied in data only, where residual corrections are determined with data driven techniques. The different corrections are discussed in more detail below and we will focus on jets reconstructed with the anti-k_t algorithm with $R = 0.5$ based on inputs from the particle-flow event description. Additionally, a procedure is applied to the jets where charged particles are removed that are not compatible with the primary vertex. This is referred to as charged hadron subtraction (CHS).

- **L1 offset correction:** To estimate and subtract energy that does not originate from the hard scattering an average p_T density (ρ) per unit area is calculated in data as well as in simulated samples. For each event a large number of infinitely soft four-vectors are added artificially, clustered by the jet algorithm, and used to define an active jet area A_j for each jet j. The average p_T density ρ is then defined as the median of all $\rho_i = p_{Ti}/A_i$ where i runs over all jets. Additionally, the η dependence for this correction is taken into account. The p_T offset as a function of η is shown in Figure 5.4 for data and MC simulation as well as different number of primary vertices.

- **L2, L3 relative/absolute MC correction:** These corrections are derived from simulated QCD events enriched in low-p_T gluon jets, where each reconstructed jet is matched in $\eta - \phi$ space to a simulated particle jet within $\Delta R < 0.25$. The response variable $\mathcal{R} = p_T^{reco}/p_T^{gen}$ is then determined from the detector jet p_T^{reco} and MC particle jet p_T^{gen} for each bin of p_T^{gen}. The average correction factor is defined as $C_{MC}^{rel,abs}(p_T^{reco}) = \frac{1}{<\mathcal{R}>}$ and described as a function of the average detector jet $p_T < p_T^{reco} >$. The correction factor is shown on the right of Figure 5.4 for different corrected jet p_T as a function of η. The large variations as a function of η arise due to non-linear response of the detector, for example, the feature around $|\eta| \approx 1.3$ is due to the barrel-endcap

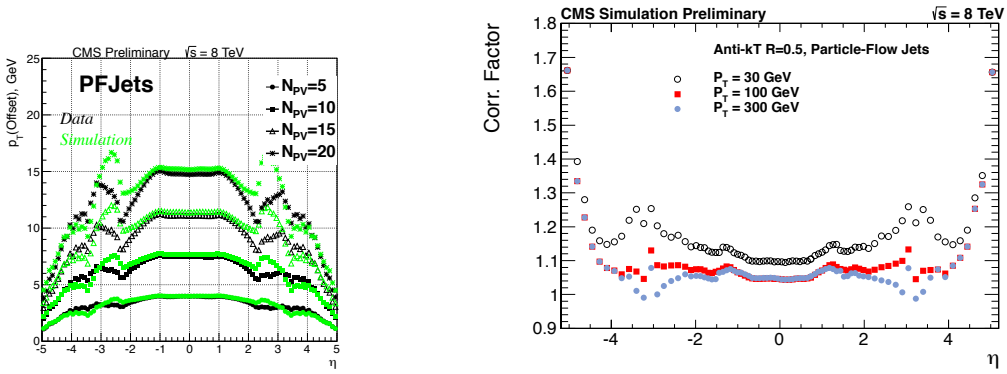

Figure 5.4: Left: L1 offset correction shown for PF jets for data and MC simulation, Right: L2, L3 correction shown for three different jet p_T bins as a function of pseudorapidity η [67].

boundary and a large amount of tracker material in this region.

- **absolute/residual correction:** After the L2 and L3 corrections derived from simulation are applied to the data, additional measurements are necessary to validate the corrections and derive residual correction factors. The absolute scale is measured using γ+ jets or Z^0 + jets events in data, where the Z^0 boson or γ are measured more precisely than the jet. Under the assumption of a balanced two-object event the jet energy is calibrated with respect to the energy of the other object. The relative energy scale can be determined with the dijet p_T-balancing method, where one jet (central jet) is required to lie in the barrel region with $|\eta| < 1.3$ and the other jet (probe jet) can have an arbitrary η.

As mentioned before, the current corrections are derived from simulated QCD events with a flavor compositions that is enriched in low-p_T gluon jets. Gluons and heavier quarks result in jets with a higher particle multiplicity with a softer p_T spectrum, resulting in a different energy response than jets from light-flavor quarks. Flavor dependent corrections could be added in the future as an additional step in the multiplicative chain described above.

Jet energy scale corrections uncertainties

The correction factors derived from data and MC simulation are subject to effects of systematic uncertainties, which need to be taken into account for physics analyses. Each of

the steps in the jet energy scale correction chain has its own uncertainties. Figure 5.5 shows from left to right: the uncertainty as a function of p_T for central jets $|\eta| = 0$, the uncertainty as a function of p_T for jets with $|\eta| = 2.7$, and the uncertainty as a function $|\eta|$ for a jet p_T of 100 GeV. As can be seen, low-p_T jets are more affected by uncertainties where the pile-up corrections carry more weight. In the p_T and $|\eta|$ ranges important for this analysis the effect on the jet corrections is of the order of 2%. The uncertainty on the *absolute scale* arises mainly from the reference energy scale for γ and Z^0 bosons and corrections for initial and final state radiation. The dominant contribution to the *relative scale* originates from the modeling of the jet energy resolution in MC simulation (discussed in the following section). In general, the dijet sample includes jets with much higher p_T than available in the γ+ jets sample. For the measurement of the residual absolute correction a comparison between data and MC is performed in the energy range where γ+ jets are available. The result is then *extrapolated* for higher p_T jets in the dijet sample. The uncertainty on the *pile-up* corrections is large for low-p_T jets but significantly decreases for $p_T > 30$ GeV. The *jet flavor* uncertainty arises from the modeling of the flavor content of the QCD MC samples used to derive the L2/L3 corrections. A small effect for central jets is the *time stability* uncertainty, which is presumed to originate from radiation damage in the forward regions of the detector.

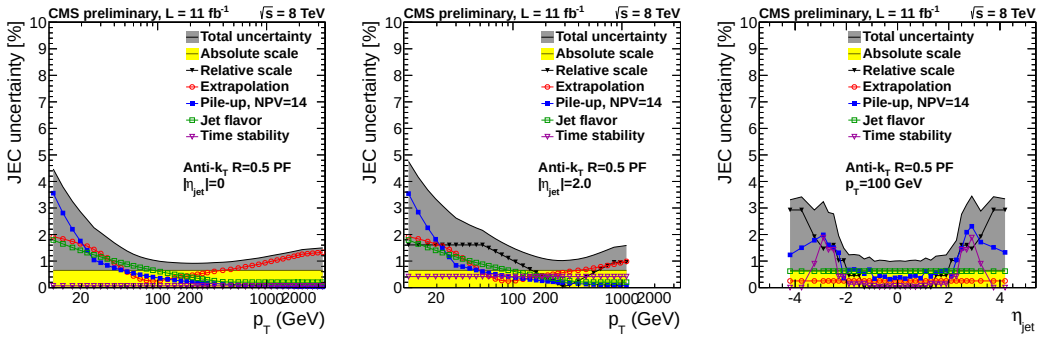

Figure 5.5: Uncertainties on the jet energy scale corrections. From left to right: uncertainty as a function of p_T for central jets $|\eta| = 0$, uncertainty as a function of p_T for jets with $|\eta| = 2.7$, and uncertainty as a function $|\eta|$ for a jet p_T of 100 GeV [67].

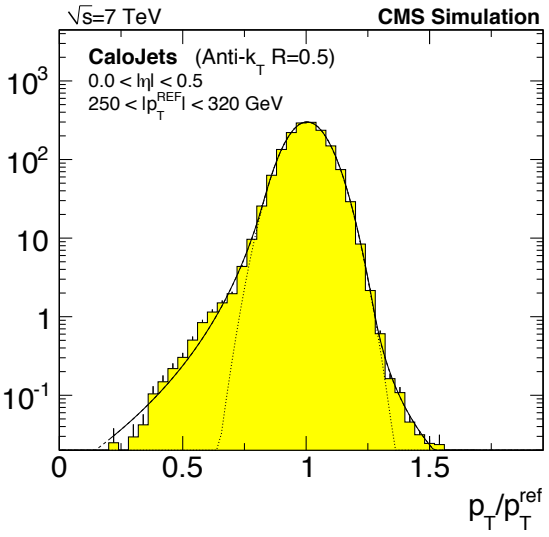

Figure 5.6: Jet energy resolution in MC. The distribution shows the ratio p_T^{reco}/p_T^{gen} for simulated QCD events and calorimeter jets with $|\eta| < 0.5$ and $250 < p_T^{gen} < 320$ GeV. The distribution is modeled by a Gaussian core and a Crystal Ball function to describe the tails [66].

5.4.3 Jet energy resolution

The jet energy resolution gives a measure of how accurately we can measure the p_T of a jet. In MC simulation this resolution is derived from generator level jets which are spatially matched to reconstructed jets after the full detector simulation and reconstruction have been performed. The p_T response is then defined as the ratio of reconstructed p_T^{reco} over generator jet p_T^{gen}. The resulting distribution is shown in Figure 5.6 for calorimeter jets with $|\eta| < 0.5$ and $250 < p_T^{gen} < 320$ GeV and consists of a Gaussian core and a Crystal Ball function to model the tails. The Gaussian width of this distribution gives a measure for the jet energy resolution.

In data the resolution is measured with the previously mentioned dijet or γ+jet balancing methods. The resolution in data is slightly worse than it is in MC and correction factors are derived. Figure 5.7 shows the measured resolution for 36 pb^{-1} of data collected at $\sqrt{s} =$ 7 TeV in black for central jets reconstructed from calorimeter information only (left) and reconstructed using the particle-flow approach (right). The original MC resolution is shown

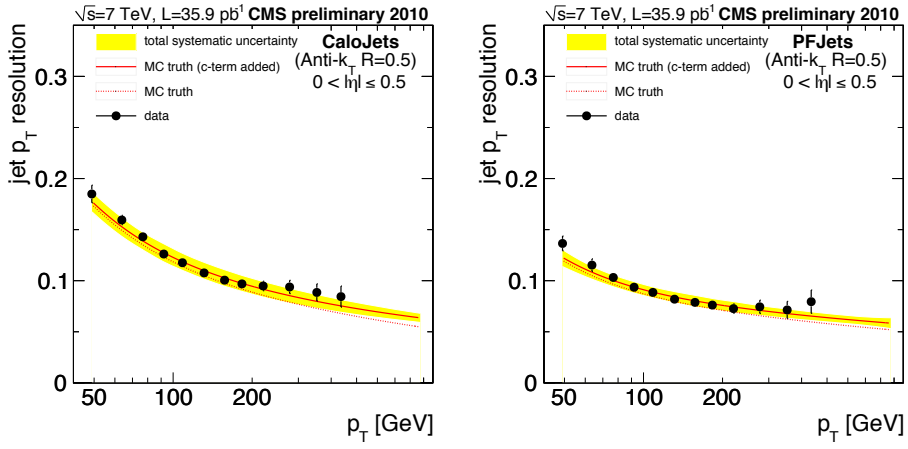

Figure 5.7: Jet energy resolution measured in data with the dijet balancing method in black compared to the uncorrected (red dashed) and corrected resolution (red solid) in MC. Systematic uncertainties are shown in yellow. The plot on the left shows jets reconstructed from calorimeter information only while the right one uses jets reconstructed with the particle-flow algorithm [66].

by a red dashed line and the corrected resolution by a solid red line. As mentioned before the particle-flow approach yields an improved energy resolution, which is about 10% for central jets of $p_T \approx 100$ GeV.

5.4.4 b-jet identification

For many physics analyses it is important to distinguish between jets that originate from light- or heavy-flavor quarks. During the hadronization process b quarks produce mostly B hadrons, which have a lifetime that allows them to travel a few millimeters before they decay, leading to a so-called secondary vertex within the jet. This and other quantities allows us to "tag" jets as b jets and the current algorithms employed by CMS are discussed in this section following Reference [68].

Two major b-tagging algorithms, called taggers, are used in CMS: the jet probability tagger (JP) and the combined secondary vertex tagger (CSV). Both of these algorithms produce a discriminating variable that can successfully distinguish light- from heavy-flavor jets. Several selection criteria, referred to as working points, are denoted by loose "L", medium "M", and tight "T", based on a misidentification probability of 10%, 1%, and

0.1%, respectively. For each jet the impact parameter (IP) is calculated in three dimensions and describes the distance of the jet from the primary vertex (the vertex with the highest sum p_T^2 of all associated tracks). This signed quantity is more likely to be positive for decays of particles that travel along the jet axis. The impact parameter significance S_{IP} is given by the IP divided by its uncertainty and is used as a discriminating variable to distinguish between light- and heavy-flavor jets, since it has similar properties as the IP. The JP tagger makes use of the IP information of several tracks in one jet and estimates a likelihood that they originated from the same primary vertex.

Given the very fine granularity of the CMS pixel detector it is also possible to reconstruct secondary vertices within jets. A neural network approach is used for the CSV tagger, combining track based lifetime information with the secondary vertex reconstruction. For a medium working point (CSVM) a b-tagging efficiency of around 70% for central jets with p_T between 100 and 200 GeV as is shown on the left in Figure 5.8 is achieved. For higher transverse momenta beyond 800 GeV this efficiency reduces to about 55%. Several methods are employed by CMS to measure these efficiencies in data. For example, in data jets can be selected that have a muon close by (within $\Delta R = 0.4$), due to the large semi-leptonic branching fraction of B hadrons into muons these jets are more likely to originate from the hadronization products of a b quark. Other methods are discussed in more detail in Reference [68] and all yield comparable results. The misidentification rate for the CSVM working point is shown on the right in Figure 5.8 and lies between 1-2% for jets up to 300 GeV. One can clearly see in Figure 5.8 that the efficiency in data and simulation is different, therefore scale factors are derived and applied to the MC to match the measured efficiency in data. For this analysis we tested different taggers as well as working points, to determine which is most optimal for the analysis. This optimization procedure is described in more detail in Section 6.4.3.

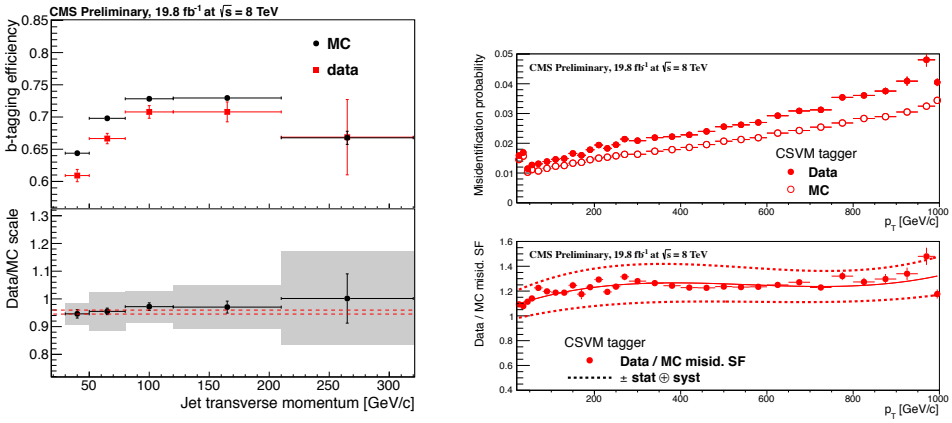

Figure 5.8: Left: Efficiency to identify b jets with the CSV tagger and medium working point for data and simulation Right: Misidentification rate for the CSVM tagger [69].

5.5 Leptons, photons, and missing transverse energy

While this analysis heavily relies on jets and their reconstruction, we briefly discuss how other objects such as electrons, muons, taus, and photons are identified [32]. The particles build the foundation of a variety of physics analyses performed by the CMS collaboration.

5.5.1 Electrons and photons

Electrons and photons both deposit a large fraction of their energy into the ECAL with $\approx$ 94% of their energy contained in 3×3 crystals. Due to the tracker material in front of the ECAL electrons and photons start producing electromagnetic showers before reaching the ECAL. This shower spreads in ϕ-direction, because of the strong magnetic field. In order to account for that effect, "superclusters" are formed that extend in the ϕ-direction. The energy of the supercluster is corrected in order to account for non-linear responses of the detector and to compensate for the interaction occurring before the ECAL is reached.

- **Photons** are reconstructed from the superclusters and a momentum is assigned based on the supercluster position and the position of the primary reconstructed vertex. Quality criteria are applied in order to distinguish photons from electrons, for example, by requiring them not to match pixel hits consistent with a track from the interaction region. Additionally, the photon should be isolated, where several isolation variables

are calculated based on the ECAL and HCAL energy deposits and sum of track p_T in a cone of radius 0.4 around the photon candidate [70].

- **Electrons** are reconstructed by two complementary algorithms either starting with ECAL superclusters or with hits measured in the tracking detector. The first method is geared towards isolated electrons in the p_T range of $W^{\pm}$ and Z^0 boson decays, whereas the second one is more suitable for low-p_T non-isolated electrons. Electron candidate tracks are seeded similarly as general tracks described in Section 5.2 and the reconstructed tracks are spatially matched to the ECAL cluster. The energy sum of all bremstrahlung photons associated with the track are included in the electron energy. Similar criteria on the isolation as discussed for photons is also applied to electron candidates.

5.5.2 Muons

Muon reconstruction has already been mentioned in the discussion of the L1 muon trigger system. Candidates found in the L1 reconstruction define a region of interest in the muon system, which is used to perform the offline reconstruction in three steps. The first step tries to align hits in the DT or CSC chambers and form an initial track segment. At the second step, the previously found track segments are linked together to form a full muon track in the muon system only (standalone tracks). Finally in the last stage, these standalone tracks are combined with hit information from the silicon tracker building a so-called global muon.

5.5.3 Tau leptons

Tau leptons can be identified through their decay products, which include semi-leptonic decays into a neutrino and electron or muon (in 35% of the cases) or hadronic decays (65%) resulting in a so-called tau-jet [9]. The hadronic decays are further subdivided into one-prong (one charged hadron and multiple neutral pions) or three-prong decays (three charged hadrons) and produce an experimental signature of narrow jets with low particle multiplicity. The reconstruction algorithm starts with a particle-flow jet with a distance parameter of $R = 0.5$ and identifies first neutral pions and then charged hadrons within

that jet. These components are used to calculate the hadronic tau (one- and three-prong) and its four-vector.

5.5.4 Missing transverse energy

The CMS detector is designed in such a way that all particles that interact with matter and are produced in a proton-proton collision deposit their energy completely within the calorimeters or are measured by the muon system. Many electroweak decays include neutrinos, which do not interact with the detector material. This results in an imbalance of energy in the transverse plane. Also non-interacting particles are predicted by new physics models like SUSY where the LSP escapes detection. E_T^{miss} is defined as the negative of the vector sum of the transverse momenta of all reconstructed particles. Three different definitions of E_T^{miss} are used in CMS based on the either the particle-flow algorithm, the sum of the calorimeter energies, or the sum of the calorimeter energies plus tracks reconstructed in the tracking system.

Chapter 6

Event selection

While the reconstruction of events is performed centrally, every physics analysis defines criteria for the specific physics goal to be achieved. In this analysis we are are looking for pair-produced three-jet resonances, which indicates, for example, the minimum number of jets we are looking for. Other aspects of the event selection include choices like the trigger path, p_T thresholds, and optimization procedures. In this chapter we discuss these general choices as well as the analysis technique and selection optimization based on the model of a heavy resonances decaying 100% into three jets.

6.1 Trigger selection

Events are recorded using a two-tiered trigger system as described in Section 3.2.6. First, they are selected by the L1 trigger and then passed to the HLT system. At the HLT jets are reconstructed only from calorimeter information and we use the *HLT_QuadJet60_DiJet20* trigger path for this analysis. For this trigger four jets with p_T above 60 GeV and two jets with p_T above 20 GeV are required. This six-jet trigger was unprescaled during the course of the 2012 data-taking period, making it suitable for this analysis. In order to measure the efficiency of the chosen trigger path we use fully reconstructed events as described in Chapter 5. At the analysis stage jets are reconstructed from particle-flow candidates and not only from calorimeter information, as is the case at the trigger level.

Given that there are two separate p_T thresholds we measure the trigger efficiency as a function of the fourth-jet and sixth-jet p_T simultaneously with respect to a trigger with lower thresholds or a different L1 requirement. For this measurement events are selected with at least six particle flow jets above 35 GeV. Figure 6.1 shows the trigger efficiency of the *HLT_QuadJet60_DiJet20* trigger path with respect to the *HLT_QuadJet50* path, where

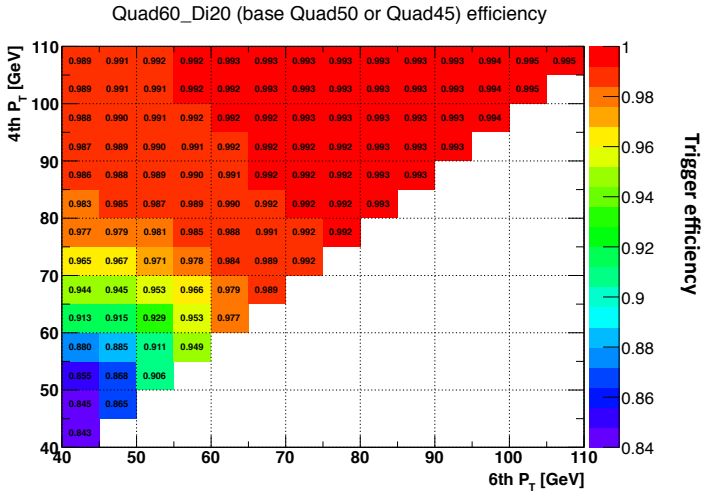

Figure 6.1: Trigger efficiency for the *HLT_QuadJet60_DiJet20* trigger. The efficiency is measured as a function of 4th (y-axis) and 6th (x-axis) p_T threshold. We find that requiring the 4th jet p_T to be above 80 GeV and the 6th jet p_T to be above 60 GeV gives an efficiency greater than 99%, which remains flat as a function of jet p_T.

only four jets above 50 GeV are required at the HLT. We find that for a selection of four jets above 80 GeV and two additional jets above 60 GeV the trigger is more than 99% efficient and the efficiency remains flat as a function of jet p_T. When additional jets are present in the events we require them to have at least a $p_T > 35$ GeV, to reduce the effect of low p_T jets from pile-up.

6.2 Basic event selection

Events that pass the trigger requirement must satisfy further selection criteria as discussed in this section. We require events to have at least one primary vertex that has at least four degrees of freedom $n_{dof} > 4$. Additionally, the vertex position in z-direction z_{vtx} is required to be less than 24 cm away form the geometric center of the detector. The maximum distance in the transverse plane measured from the same reference point d_o^{tvx} has to be less than 2 cm.

Jets are reconstructed based on the particle-flow algorithm, as described in Section 5.1. The default CMS anti-k$_t$ [62] jet-clustering algorithm is used with a distance parameter

of $R = 0.5$ and a list of particle-flow candidates as input. Further jet identification (ID) criteria are imposed in order to reduce the effect of electronic noise in the ECAL and HCAL system and other particles that are misidentified as jets. The energy composition of each jet is attributed to neutral or charged particles depositing their energy in the ECAL and HCAL. The ratio of these deposits with respect to the total energy can be used as a measure to identify jets. A large fraction of hadronic energy from neutral particles is an indication of noise in the HCAL, whereas a large fraction of electromagnetic energy from charged particles indicate electrons that are misidentified as jets. Each jet has to have more than one constituent and for jets in the central region $|\eta| < 2.4$ at least one of the constituents has to be charged. In summary, the following specific criteria are applied to uncorrected jets [71]:

- General selection
  - Neutral hadron energy fraction< 0.99
  - Neutral electromagnetic energy fraction < 0.99
  - Number of constituents > 1
- for $|\eta| < 2.4$
  - Charged electromagnetic energy fraction < 0.99
  - Charged hadron energy fraction > 0.0
  - Charged multiplicity > 0

Figure 6.2 shows the different variables for the jet ID in data compared with simulated QCD MC events, where at least six jets are selected in the event with the fourth-jet above 80 GeV and the sixth-jet above 60 GeV. In these distributions jets are shown where the jet p_T has been corrected to compensate for inefficiencies in the detector response. The corrections are discussed in more detail in Section 5.4.2 and applied on a jet-by-jet basis depending on η and uncorrected p_T of the jet. In Figure 6.3 we show kinematic distributions (p_T, η and ϕ) for the six highest p_T jets. We find good agreement between data and QCD simulation for these kinematic variables. Table 6.1 summarizes the basic event selection, as discussed in this section up to now.

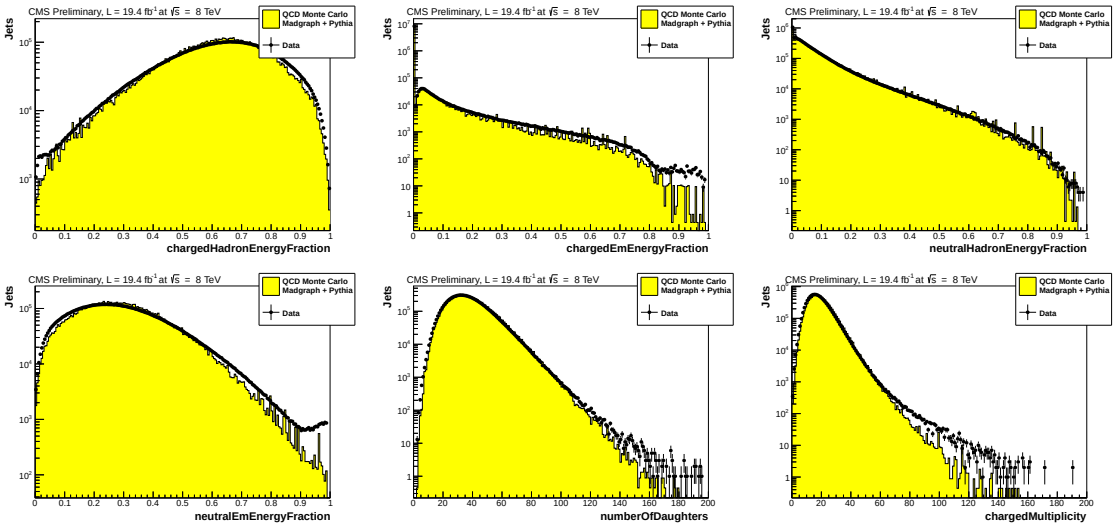

Figure 6.2: Comparison of QCD Monte Carlo (MADGRAPH + PYTHIA) with data for events containing at least six jets with fourth-jet $\geq$ 80 GeV and sixth-jet $\geq$ 60 GeV. Shown are (from left to right and top to bottom) charged hadron fraction, charged electromagnetic energy fraction, neutral hadron fraction, neutral electromagnetic energy fraction, number of constituents, and charged multiplicity.

Basic event selection	
≥ 1 good primary vertex	
$z_{vtx} < 24$ cm	$d_0^{vtx} < 0.2$ cm
≥ 6 particle-flow anti-k_t jets with $R = 0.5$	
$p_T \geq 35$ GeV	$\lvert\eta\rvert < 2.5$
4th-jet $p_T \geq 80$ GeV	6th-jet $p_T \geq 60$ GeV

Table 6.1: Basic event selection criteria.

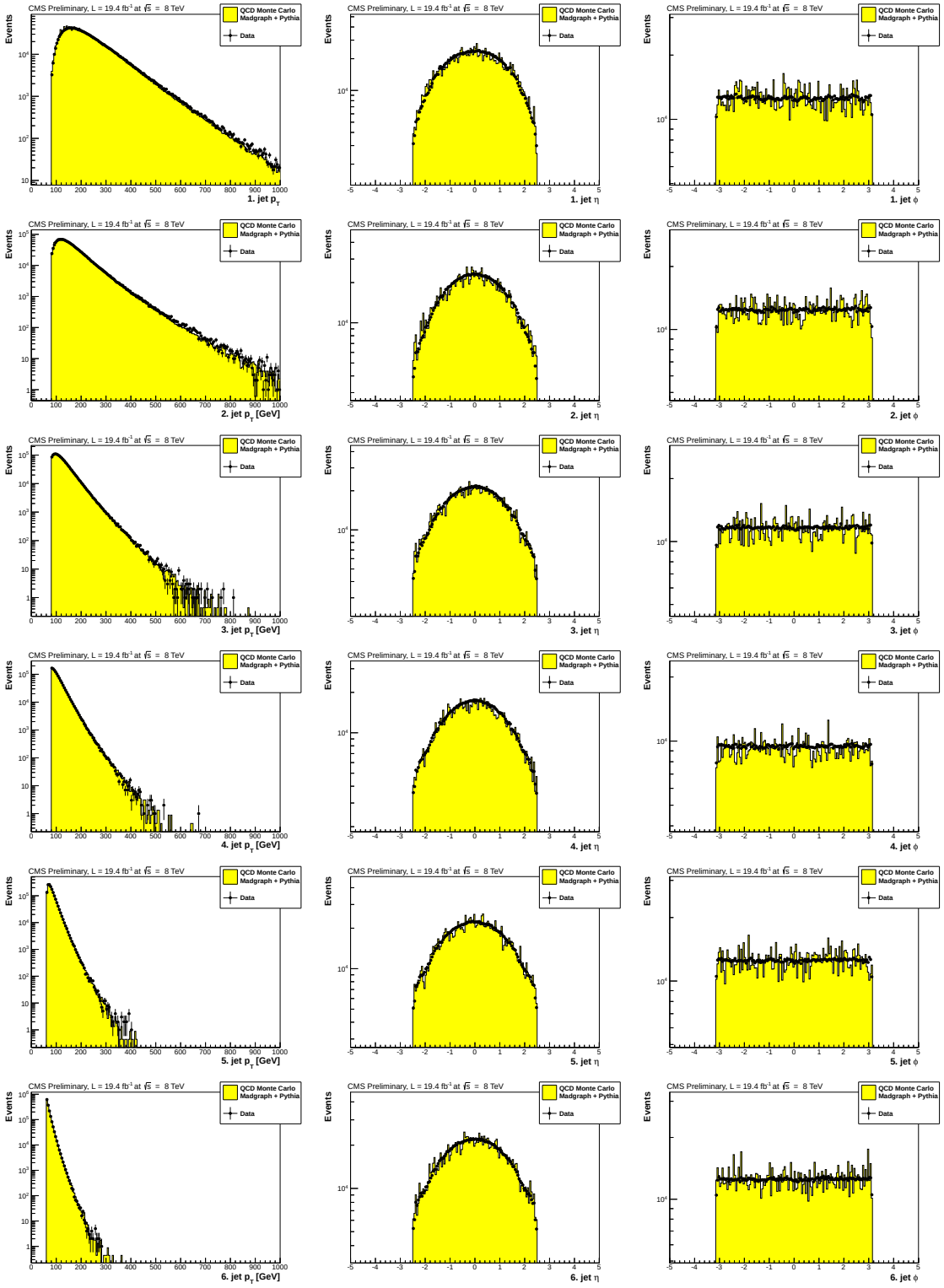

Figure 6.3: Comparison of QCD Monte Carlo (MADGRAPH + PYTHIA) with data for events containing at least six jets with fourth-jet $\geq$ 80 GeV and sixth-jet $\geq$ 60 GeV. Shown are from left to right p_T, η and ϕ of the six highest p_T jets from top to bottom.

6.3 Analysis strategy

In this analysis we search for new pair-produced heavy resonances that decay exclusively into three jets each. Besides the copious amount of QCD multijet background there is another major challenge arising from a combinatorial problem in finding the correct combination of jets representing one of the two new particles. In this section we discuss the so-called jet ensemble technique, which makes use of kinematic features that are present in heavy particle decays but not in QCD multijet events and random combinations of jets.

6.3.1 Jet ensemble technique and simulated signal samples

The jet ensemble technique has been successfully employed previously by the CDF [15] as well as the CMS [16, 17] collaborations. The ensemble here consists of all the possibilities to associate exactly six jets into unique sets of three jets (triplets) which is given mathematically by $\binom{6}{3} = 20$. Once the basic requirements from Table 6.1 for an event are satisfied, the leading six jets in p_T are associated into these 20 triplet combinations such that each unique combination of jets is represented. It is important to note that, for the signal MC events, at best two of the event's 20 jet triplet combinations correspond to the correct daughters of any pair-produced particle decaying into three jets.

In order to reject incorrectly combined triplets, we make use of kinematic features that can distinguish correctly assigned triplets from random combinations. For uncorrelated triplets, the triplet invariant mass (M_{jjj}) scales with the scalar sum of the transverse momenta of its constituents ($\sum_{jjj} |p_T^{jet}|$), whereas for correct combinations of jets in the signal samples the invariant mass is independent of this sum. Figure 6.4 show schematically the final state of six jets on the left and all of the 20 possible combinations in the middle. On the right hand side of this figure we show an illustration of the described correlation between M_{jjj} (plotted on the y-axis) and $\sum_{jjj} |p_T^{jet}|$ (plotted on the x-axis). Correct combinations should pile up along the horizontal dashed line, while incorrect combinations tend to fall along the diagonal line and above shown in the red area. To select predominately correct

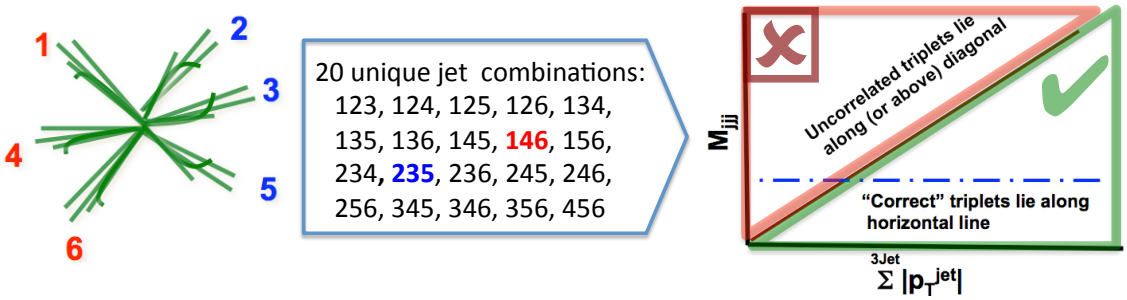

Figure 6.4: Illustration of the jet ensemble technique.

combinations we require each triplet to satisfy the following relation:

$$M_{jjj} < \sum_{jjj} |p_T^{jet}| - \Delta, \tag{6.1}$$

where Δ is an adjustable offset also referred to as the "diagonal" offset. For example, a the black solid line along the diagonal of the rectangle on the right hand side of Figure 6.4 corresponds to $\Delta = 0$. Imposing Equation 6.1 on each triplet selects the ones that fall into the green framed area in Figure 6.4. In order to demonstrate the discussed ideas, we apply the above described method to a hypothetical gluino signal sample as shown in the M_{jjj} vs. $\sum_{jjj} |p_T^{jet}|$ plot on the top in Figure 6.5 for a gluino with light-flavor jets in the final state and a mass of 400 GeV.

In order to visualize the correctly and incorrectly combined triplets we perform the following procedure. Each selected jet is compared to the generated parton content of the event in MC simulation by calculating the spatial distance ΔR between the MC parton and the reconstructed jet. Each pair is ordered by this spatial distance and if the smallest value of ΔR is less then 0.4 we call the jet matched to a generator level parton. Since the full decay chain is known in MC simulation one can now trace back where this parton originated from, finding its mother particle. If all three jets in the triplet are matched to a generator level parton whose mother is identified as the gluino, we call the triplet a "correct" combination. If a triplet consist of jets that are either not matched to a generator level parton (smallest ΔR to large) or if jets originate from both parent gluinos, we call this

triplet an "incorrect" combination.

The right hand side of Figure 6.5 shows again the M_{jjj} vs. $\sum_{jjj} |p_T^{jet}|$ distribution but this time we make use of the aforementioned technique in order to identify correct and incorrect combinations. In this plot the correct combinations are represented by the solid colored region, and as predicted these triplets pile up along a horizontal line around 400 GeV, the mass of the simulated gluino. Incorrect combinations are represented by the grey scatter plot with colored contour lines, and they populate the left top corner of two dimensional space. The red dashed line corresponds to the Δ value of 110 GeV, and we keep all triplets that fall to the right of this line. Figure 6.6 shows the triplet invariant mass distribution for all 20 combinations on the left, where no Δ requirement is imposed. Correct triplets are shown in the gold shaded area and incorrect ones, also referred to as intrinsic combinatorial background, are represented by the white histogram. Without the Δ requirement, correct and incorrect combinations peak at the same position. The shape of the combinatorial background, even for simulated signals containing a three-jet resonance, is not known, making it therefore difficult to pick out a potential signal peak. The right hand side of the same figure shows the triplet invariant mass distribution for triplets that satisfy Equation 6.1 with $\Delta = 110$ GeV. After this selection criterion is imposed the combinatorial background distribution moves further to the left while the correct combinations stay at the resonant mass, allowing a separation of a signal peak from its intrinsic combinatorial background.

The combinatorial background for signal triplets is smoothly falling with increasing triplet mass and is modeled by a functional form that exhibits the same smoothly falling behavior. We define the following four-parameter function, which has been used to model the triplet invariant mass distribution in the previous two CMS analyses [17, 16] and is also widely used in searches for dijet resonances [72]:

$$P^{comb} = P_0 \frac{(1 - \frac{x}{\sqrt{s}})^{P_1}}{(\frac{x}{\sqrt{s}})^{P_2 + P_3 \log \frac{x}{\sqrt{s}}}} \quad . \tag{6.2}$$

The signal peak is modeled with a Gaussian function, which we find suitable to describe the shape. For the gluino model tested here the intrinsic width of the new resonance is negligible with respect to the jet energy resolution, which affects the three-jet invariant

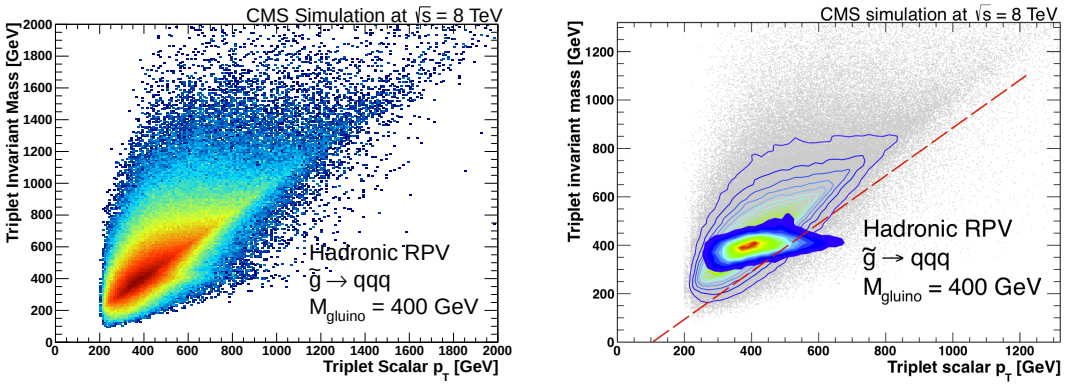

Figure 6.5: The triplet invariant mass M_{jjj} vs. its $\sum_{jjj} |p_T^{jet}|$ is for signal gluinos of mass 400 GeV with light-flavor jets in the decay. The left plot shows all possible 20 combinations, incorrectly and correctly assigned ones on the same scale. On the left, a similar plot is shown as before with the correct gluino combinations represented by the solid colored region and incorrect combinations by the grey area with colored contour lines.

mass resolution [73].

6.3.2 Jet ensemble technique and simulated QCD multijet events

In the previous section we have discussed the jet ensemble technique in the context of a new hypothetical three-jet resonance. This section focuses on the technique tested with simulated QCD multijet events, where no actual resonance is present. The top plot in Figure 6.7 shows the M_{jjj} vs. $\sum_{jjj} |p_T^{jet}|$ distribution for the simulated QCD events. The two plots in the bottom row show the triplet invariant mass distribution on the left when no Δ requirement is imposed (all 20 combinations per event) and on the right when each triplet must satisfy Equation 6.1 with $\Delta = 110$ GeV. The behavior of the triplet invariant mass distribution of jets originating from QCD is very similar to the one described in the previous section for incorrect triplet combinations. The shape is again smoothly falling and the peak position of the distribution changes when the Δ requirement is imposed.

QCD MC is used in the analysis to cross-check results of optimization procedures and for comparisons of kinematic distributions in data with the expectation from simulation. However, it is important to note that whenever the QCD simulation is used we obtain the overall normalization from the data. For the following section, where event selection

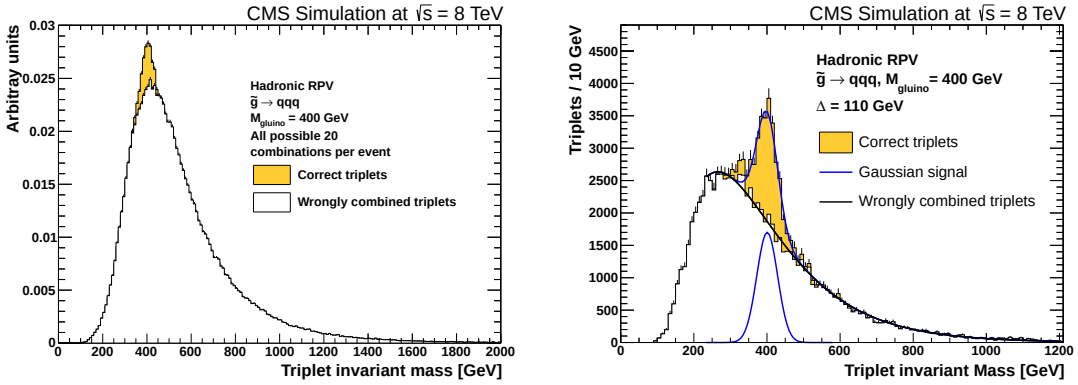

Figure 6.6: M_{jjj} distributions for a 400 GeV gluino with light-flavor jets in the decay. The distribution on the left shows all possible 20 combinations per event. Correct triplets are shown in the gold shaded area and incorrect ones are represented by the white histogram. The plot on the right shows the M_{jjj} distribution for all triplets that satisfy Equation 6.1 with a $\Delta = 110\ GeV$. It can clearly be seen that imposing the Δ requirement separates out the correct combinations in gold from the incorrect ones in white. The distribution is fit with a four-parameter function (black line) and a Gaussian (blue line), where the Gaussian is used to model the expected signal for the search.

optimization is discussed, we estimate the background from the data shape itself using a parametrized function.

6.4 Advanced selection and optimization

One major advantage to this analysis is its simplicity. We probe essentially any hadronic final state where the particle decays to three daughters and do so using only the jet information in the event. To best select signal events, we use two "lever arms" available in the analysis: the p_T threshold of the sixth leading jet in p_T and the offset value Δ for the jet triplet mass requirement in Equation 6.1. Additionally, we study other variables that distinguish between signal and QCD background by comparing event shape variables [38].

6.4.1 Diagonal offset Δ selection

The effect of Δ on the distribution of either incorrectly assigned signal triplets or uncorrelated jet combinations in QCD events is very similar; the peak position moves as a function

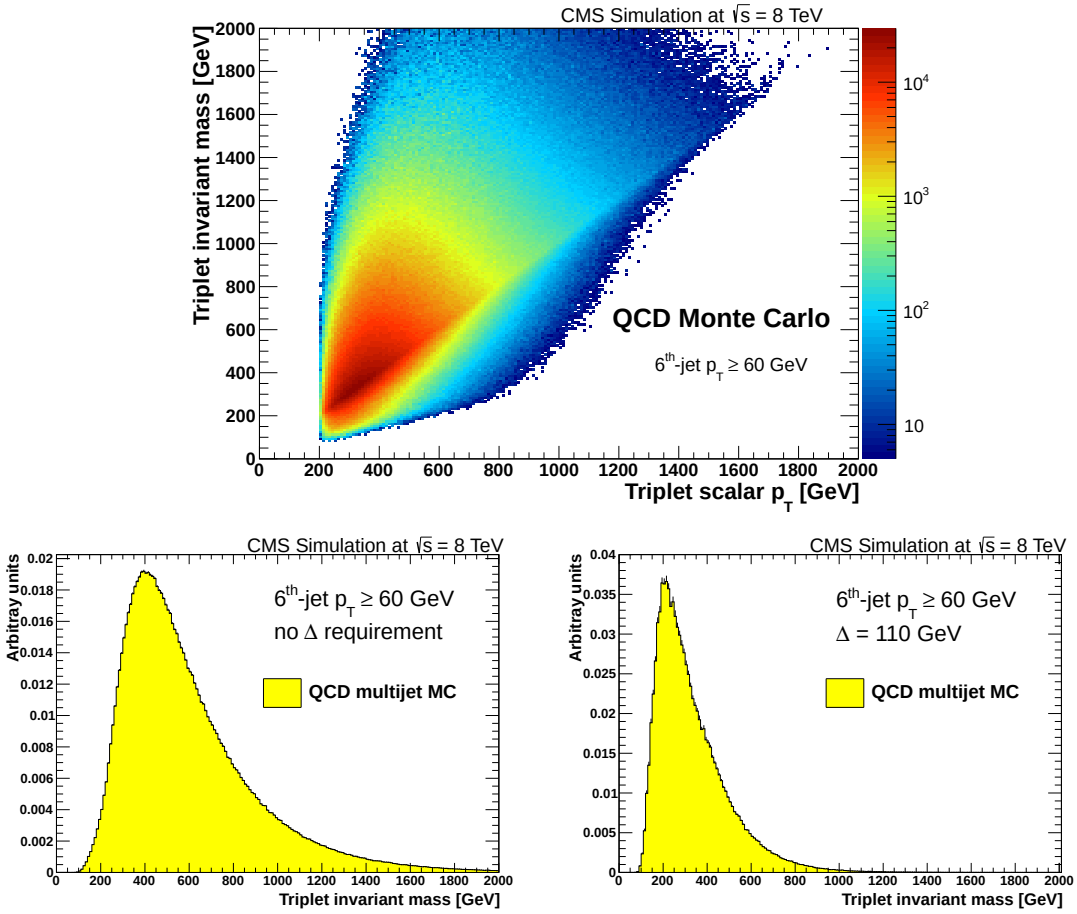

Figure 6.7: M_{jjj} vs. $\sum_{jjj} |p_T^{jet}|$ and M_{jjj} distributions for simulated QCD MC events. The top plot shows the two dimensional distribution for all 20 possible triplet combinations. The bottom row shows the one dimensional M_{jjj} distribution when no Δ requirement is used on the left and for $\Delta = 110$ GeV on the right. All three distributions are shown in arbitrary units.

of Δ. In order to separate the Gaussian signal peak of correct combinations, as shown in Section 6.3.1 for the case of the hypothetical new three-jet resonance with a mass of 400 GeV, from the peak of the intrinsic combinatorial background it is advantageous to chose Δ such that the background peaks to the left of the potential new signal. Changes in the minimum p_T requirements on the jets have the effect of moving the peak position, however not in exactly the same way as Δ. While an increase on the minimum sixth-jet p_T pushes the peak position to higher values, we find that the peak position first decreases and after a minimum increases again as a function of Δ. In order to determine this minimum position we use the triplet invariant mass distributions in data corresponding to 19.4 fb^{-1} for events that pass the selection criteria summarized in Table 6.1 with the additional requirements that events are vetoed that contain b jets. This veto reduces the effect of all-hadronic $t\bar{t}$ events that contribute to the triplet invariant mass distribution around 175 GeV, and could bias the estimate of the peak position in that region. The Δ requirement is varied in steps of 10 GeV, and the triplet invariant mass distributions are shown in Figure 6.8 for three different Δ (70, 110, and 150 GeV) values. In order to determine the peak position, the triplet invariant mass distribution is fit with a Landau function, which describes the area around the peak reasonably well. The Landau function has three parameters: the most probably value (MPV indicates the peak position), the amplitude (indicates the height of the distribution), and the width. We use the MPV as an estimate for the peak position in data and plot this value as a function of Δ as shown in the right bottom plot in Figure 6.8 in red. We find that for a value of $\Delta = 110$ GeV the peak position reaches its minimum.

Since we are considering two different scenarios of RPV gluino decays into either light-flavor jets or one heavy-flavor and two light-flavor jets, we are also testing if the use of b-jet identification has any effect on the peak position of the triplet invariant mass distribution. We repeat the procedure of fitting the M_{jjj} peak position for events where b jets are not vetoed, and find overall the same behavior as a function of Δ. This is shown by the black curve in the bottom right plot of Figure 6.8. The minimum again is reached for $\Delta = 110$ GeV. The difference in the two curves can be explained by the non-negligible contribution of all-hadronic $t\bar{t}$ triplets in the distribution where b jets are not explicitly vetoed, which results in a slightly lower fit value for the peak position.

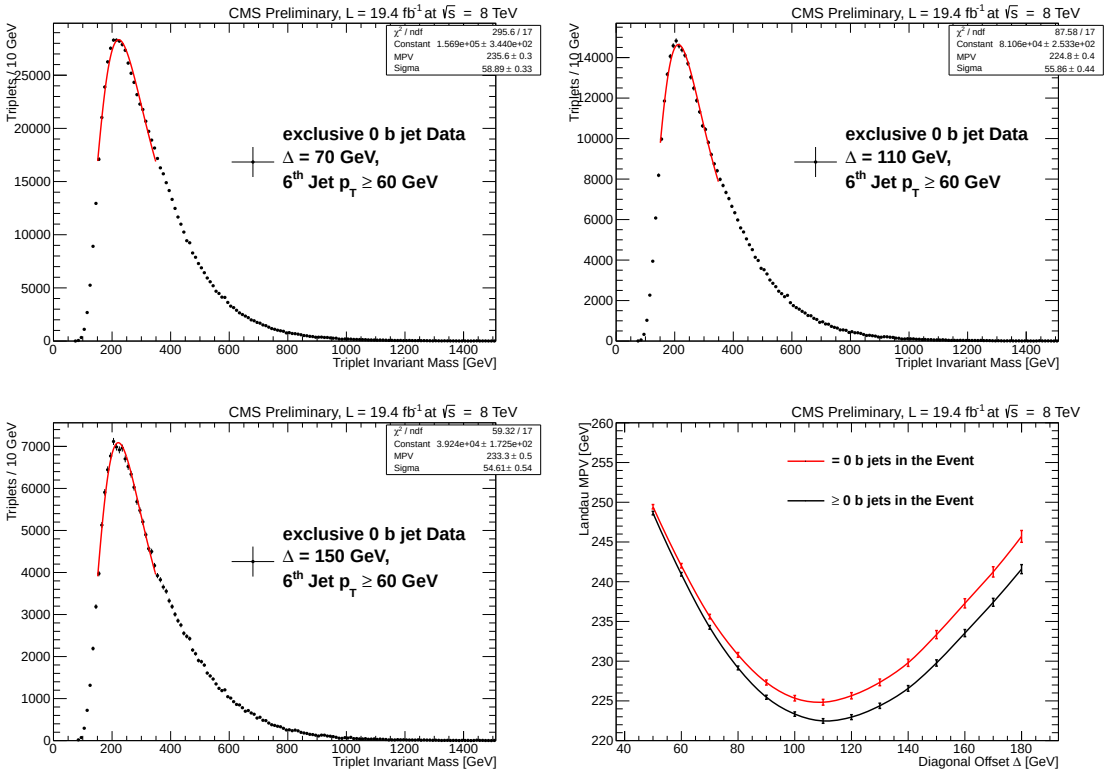

Figure 6.8: The data distributions for the exclusive 0-b jet case for $19.4 fb^{-1}$, with a Landau function to fit the peak position of the data. The bottom right plot shows the MPV of the different Landaus as a function of the diagonal offset. The minimum is reached for Δ of 110 GeV.

We chose a value of $\Delta = 110$ GeV for all subsequent optimization studies.

6.4.2 Optimization of the sixth-jet p_T

After the diagonal offset is chosen, we optimize the sixth-jet p_T for each of the different resonance masses. With increasing particle mass the jets it decays to also increase in p_T, which is the second handle in the analysis that improves the sensitivity to the new physics model under consideration.

The distribution of the p_T of the sixth-jet in data, QCD MC, and three different signal masses is shown in Figure 6.9. Changing the minimum p_T threshold has an effect on triplet invariant mass distribution, pushing the lowest accesible M_{jjj} point to higher values. In order to find the optimal selection criteria we increase the minimum requirement on the

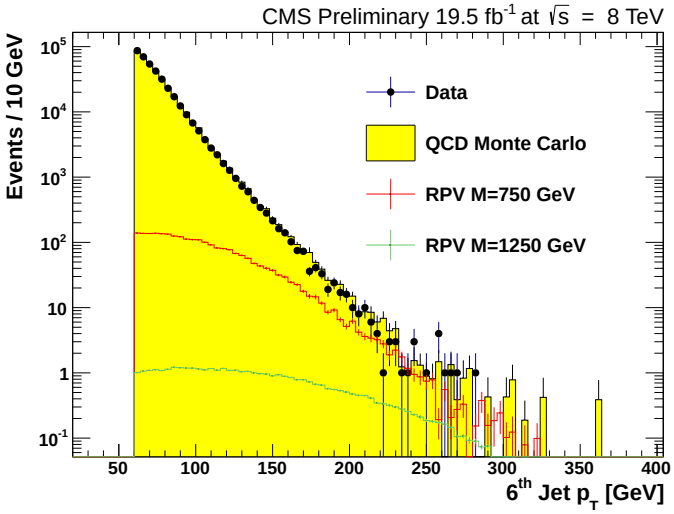

Figure 6.9: Sixth-jet p_T distribution in data, QCD MC, and three different signal masses. Shown are distributions for data, QCD MC and two signal samples with masses of 750 and 1250 GeV.

sixth-jet p_T in 10 GeV steps, for brevity referred to as operating point, from 60 GeV to 130 GeV. While the reconstructed Gaussian peak stays at a fixed mass the shape of the QCD multijet background changes.

The optimization procedure is performed as follows:

- Events from simulated MC signal samples and data are selected that pass the minimum event selection criteria given in Table 6.1.
- For each of those events, triplet combinations are formed from the six leading jets. Each jet within a triplet is required to pass the p_T threshold for a given operating point that is being tested.
- Each triplet has to satisfy Equation 6.1 with $\Delta = 110$ GeV.
- The triplet invariant mass distribution is plotted for all triplets surviving the above mentioned criteria. The distribution of signal triplets is divided by the number of events generated and multiplied by the theory cross section from Table 4.2 and by the integrated luminosity to estimate how many triplets one would expect in the given dataset.

- The expected signal distribution is fitted with the sum of the four-parameter function in Equation 6.2 and a Gaussian component, where the Gaussian is defined as the signal in the search. The distribution for a 400 GeV gluino with light-flavor jets in the decay is shown on the left of Figure 6.10 with the aforementioned fits and the estimated Gaussian signal (red). The data distribution is fitted with the same four-parameter function as given in Equation 6.2 and repeated here for convenience:

$$P^{bkg} = P_0 \frac{(1 - \frac{x}{\sqrt{s}})^{P_1}}{(\frac{x}{\sqrt{s}})^{P_2 + P_3 \log \frac{x}{\sqrt{s}}}}, \tag{6.3}$$

  in order to estimate the expected multijet background for a given operating point. This background estimate for a sixth-jet minimum p_T of 60 GeV is shown on the right hand side of Figure 6.10 in red. The expected Gaussian signal from the plot on the left is added to the background as depicted with the blue dashed line.

- For each operating point we define a metric, the signal significance for triplets:

$$S_{Sig}^{trip} = \frac{N_{trip}^{Sig}}{\sqrt{N_{trip}^{Sig} + N_{trip}^{Bkg}}}, \tag{6.4}$$

  where N_{trip}^{Sig} is the number of signal triplets given by the integral of the expected Gaussian signal within a $\pm 2\sigma$ window and N_{trip}^{Bkg} is the number of background triplets estimated from the data background fit, which is integrated in the same window as the Gaussian signal. This is depicted by the black dash-dotted vertical lines in Figure 6.10.

The optimization procedure is conducted for both RPV scenarios considered. Figure 6.11 shows two triplet invariant mass distributions for a gluino with a mass of 400 GeV decaying only into light-flavor jets on the left and a gluino decaying into one heavy-flavor and two light-flavor jets on the right.

For each operating point we repeat the optimization procedure outlined above and obtain a value of the metric S_{Sig}^{trip}. The top two plots in Figure 6.12 show the metric for each operating point on the x-axis and all gluino masses considered on the y-axis. The left plot corresponds to gluinos with light-flavor decays and the right one shows gluinos with heavy-flavor jets in the final state.

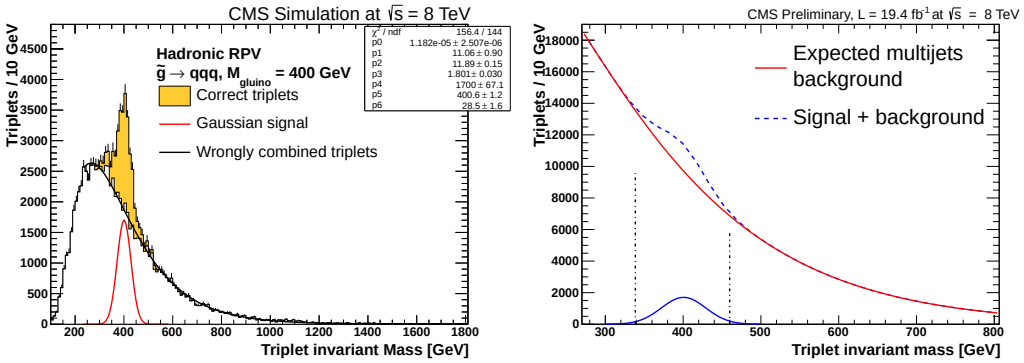

Figure 6.10: M_{jjj} distribution and signal + background distribution for a 400 GeV gluino. The left plot shows the triplet invariant mass distribution for a 400 GeV gluino with light-flavor jets in the decay. Correctly assigned triplets are shown in the gold shaded area while the incorrect triplet combinations are represented by the black lined histogram. The Gaussian component of the signal is shown in red and represents the expected signal in data. The right plot shows the background estimation obtained from a fit to the data triplet invariant mass distribution. The Gaussian signal from the left hand side is added on top of the expected background shown by the blue dashed line. Both distributions are integrated within a $\pm 2\sigma$ window depicted by the vertical black dash-dotted lines.

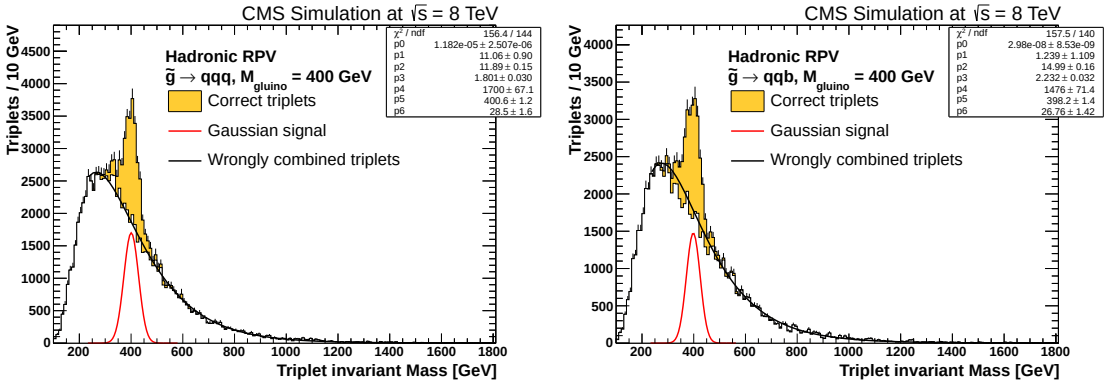

Figure 6.11: M_{jjj} distribution for 400 GeV gluino decaying into only light-flavor jets (left) or one heavy-flavor and two light-flavor jets (right). The distributions are fitted with a Gaussian and the four-parameter function. The gold shaded area in both plots represents triplets matched to MC gluino particles, whereas incorrect combinations are shown by the black lined histogram.

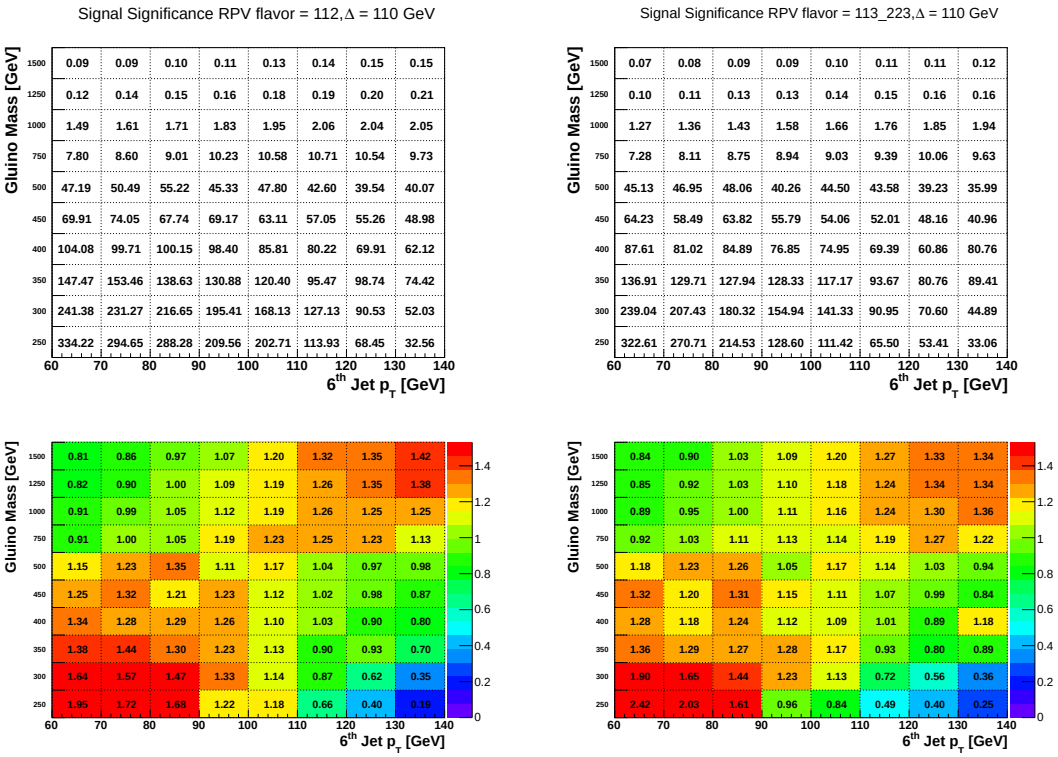

Figure 6.12: The output of the optimization metric of the different signal masses as a function of the sixth-jet p_T with a diagonal offset of 110 GeV. The left plot corresponds to gluinos with light-flavor decays and the right one shows gluinos with heavy-flavor jets in the final state. The bottom row shows the normalized signal significance.

In order to make trends more easily visible we calculated a normalized signal significance, which is defined as the ratio of S_{Sig}^{trip} divided by the average of S_{Sig}^{trip} over all optimization points for a given gluino mass. The bottom two plots in Figure 6.12 show the values of the normalized signal significance for each mass as a function of sixth-jet p_T. The color code represents the highest values of the metric in red and the lowest ones in green and blue. Low-mass three-jet resonances favor lower jet p_T thresholds, which gradually increase towards higher resonance masses. Both flavor scenarios show the same behavior for this optimization. Gluinos with light-flavor decays were already excluded up to a mass of 460 GeV in the 2011 CMS analysis [17], therefore, we are particularly interested in higher masses. We refer to this search as the "inclusive" search since no specific requirements are imposed to identify the flavor content of the final state and we choose one operating point, requiring the sixth-jet p_T to be above 110 GeV. For the gluinos with heavy-flavor jets in the decay we are also interested in the low-mass range, since this is the first search of its kind and the additional handle of b-jet identification is available to further increase sensitivity as discussed in the following Section 6.4.3. We choose two operating points, one for lower resonance masses ($<$ 600 GeV) and one for higher resonance masses ($>$ 600 GeV). The values of 60 GeV and 80 GeV are chosen for the sixth-jet and fourth-jet p_T, respectively, for masses below 600 GeV. Finally, a higher p_T requirement of 110 GeV for gluino masses above 600 GeV is chosen. Table 6.2 summarizes the kinematic selection criteria discussed so far. This full procedure was repeated with QCD MC simulation instead of data for the background estimate which resulted in the same optimization points.

Selection criteria	Inclusive search	Heavy-flavor search	
		low mass	high mass
Mass range	400–1500 GeV	200–600 GeV	600–1500 GeV
Δ	110 GeV	110 GeV	110 GeV
Fourth-jet p_T	110 GeV	80 GeV	110 GeV
Sixth-jet p_T	110 GeV	60 GeV	110 GeV

Table 6.2: Kinematic selection requirements for the three search regions in the analysis.

6.4.3 b-tagging optimization

Using b-jet identification in order to determine the heavy-flavor content of the final state is a natural extension to the previously conducted search for three-jet resonances. In the case of gluinos with RPV decays it is possible to probe a different set of couplings and increase the sensitivity of the analysis with this additional handle. Specifically selecting events, where one or more of the six leading jets is b-tagged, allows us to study the effectiveness of the jet ensemble technique on all-hadronic $t\bar{t}$ events. Several different b-tagging algorithms and working points are available within CMS as discussed in Section 5.4.4. Efficiencies for these taggers are measured in data and compared to simulated MC samples. Scale factors are derived based on the ratio of the efficiency measured in data with respected to the one measured in simulation and they are parametrized as a function of p_T and η of the jet [74]. The scale factors need to be applied to MC events to match the efficiency in data. Instead of defining a weight for each event, we are using a method were the b-tagging status of each jet is either changed from true to false or vice versa, such that on average the measured efficiency in data is reproduced. A simple example is the case where the scale factor is less than one, then a fraction of the b tagged jets corresponding to one minus the scale factor need to be downgraded and regarded as non-tagged. Each jet in MC that can be spatially matched to a generator level particle is assigned a parton flavor (u, d, g, c, or b). Based on this matching, sample dependent b, c, and mis-tag efficiencies are calculated and parametrized as a function of p_T and η of the jets, which are needed to correctly apply the scale factors to the MC samples.

After finding the optimal choice of kinematic variables for both light- and heavy-flavor scenarios, we next focus on finding the best b-tagging selection for the heavy-flavor search, testing the various algorithms and working points, as well as different categories based on the number of tagged jets. The five categories are defined as:

- ≥ 0 b tags in the event
- ≥ 1 b tag in the event
- ≥ 1 b tag in the event and ≥ 1 b tag in the triplet (removes some incorrect combined triplets)

- $\geq$ 2 b tags in the event
- $\geq$ 2 b tags in the event and $\geq$ 1 b tag in the triplet

Based on these categories, events and triplets are selected in data and the gluino samples that included heavy-flavor jets in the final state. We perform a similar optimization procedure as outlined for finding the optimal sixth-jet p_T selection. Adding b tags to the event selection increases the contribution from all-hadronic $t\bar{t}$, and it is therefore necessary to start the fit for estimating the QCD multijet background above the $t\bar{t}$ mass region around 200 GeV. The metric is again the signal significance as defined in Equation 6.4, which is evaluated for the different b-tagging categories for each of the heavy-flavor RPV samples. The top plots in Figure 6.13 show this metric for the Combined Secondary Vertex (CSV) tagger with the medium working point. The bottom plots show each row from the top with the signal significance normalized to its average, to identify trends. Going from no b tags to at least one b tag in the event increases the sensitivity by over 30%. Placing the additional requirement of at least one b tag in the triplet helps to reduce the number of incorrectly combined triplets and increases the sensitivity further.

Considering the different masses, we choose the category of $\geq$ 1 b tag in the event and $\geq$ 1 b tag in the triplet for the heavy-flavor search. After choosing the b-tagging requirements, we compare the signal significance for different working points of the CSV tagger (light, medium, tight) and of the medium working point of the JP tagger. The outcome is shown in Figure 6.14. The CSV medium working point proves to be the most efficient for the heavy-flavor model considered.

6.4.4 Event shape variables

To further improve the sensitivity for higher mass resonances we study the use of different event shape variables. With increasing mass these new particles are produced almost at rest with very little boost, leading to a more isotropic decay in the detector. QCD events, even with high jet multiplicity, should show a more dijet-like structure. One way to examine

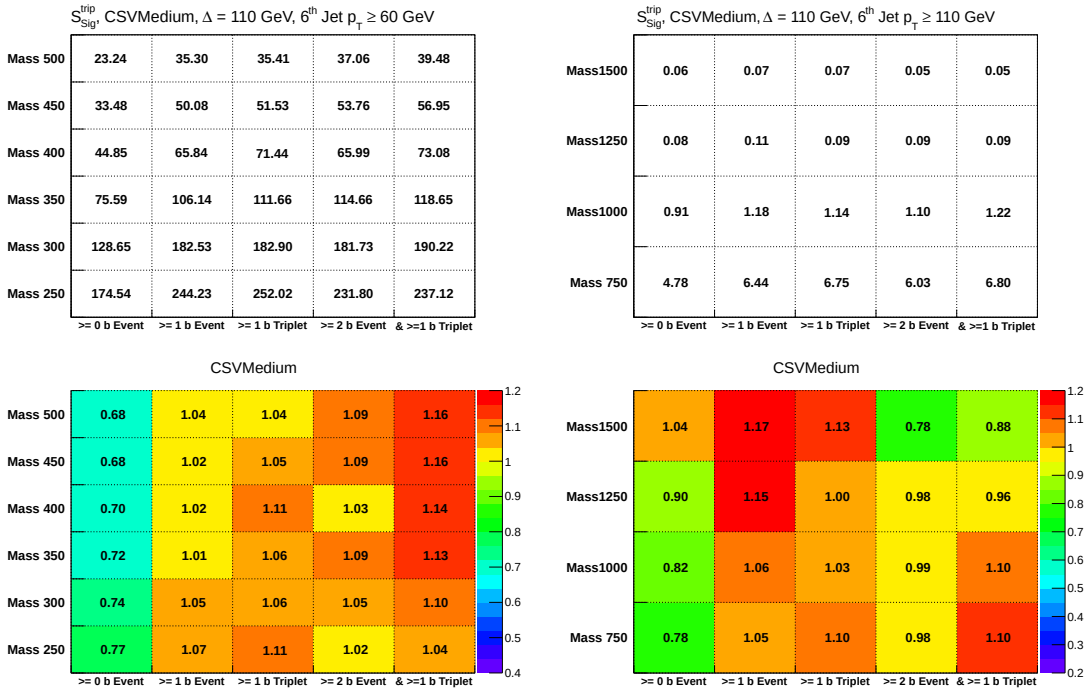

Figure 6.13: The output of the optimization metric over the different signal masses for the five b-tagging categories, for the heavy-flavor RPV samples with masses of 250 – 500 GeV (left) and 750 – 1500 GeV (right), imposing the respective kinematic selection criteria from Table 6.2. The bottom figures show the normalized signal significance to identify trends through the color code, where red indicates improved sensitivity.

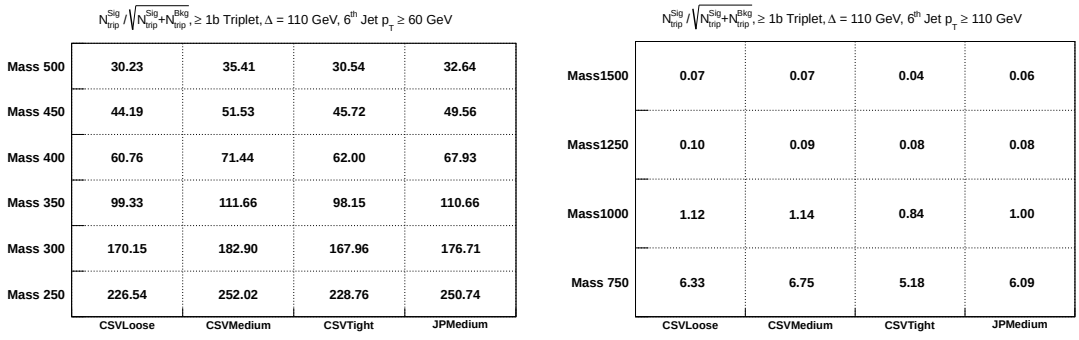

Figure 6.14: Comparison of signal significances for different taggers/operating points.

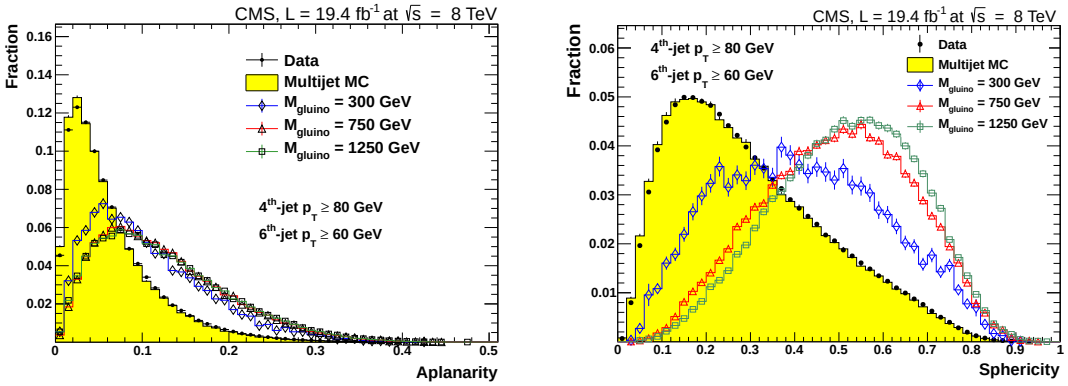

Figure 6.15: The top two plots show the variables aplanarity (left) and sphericity (right) for events passing the full event selection. The data, QCD, and signal shapes are normalized to unity to compare the shapes. The bottom plot shows the signal significance by gluino mass for events exceeding the sphericity value S given on the x-axis.

events with these features is to calculate the sphericity tensor defined as [38]:

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |p_i|^2}, \tag{6.5}$$

where the sum runs over all jets in the event, for events that satisfy the criteria mentioned in Table 6.1. After diagonalizing this tensor, the eigenvalues which satisfy $\lambda_1 > \lambda_2 > \lambda_3$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$ are computed, giving two event shape variables sphericity $S = \frac{3}{2}(\lambda_2 + \lambda_3)$ and aplanarity $A = \frac{3}{2}(\lambda_3)$. Sphericity measures the p_T^2 with respect to the event axis. For dijet events, S will be close to 0, whereas for isotropic events S approaches 1. The variables are calculated on an event-by-event basis using all jets with the basic event selection applied but before triplet combinations are formed. Figure 6.15 shows the variables aplanarity (left) and sphericity (right) for data, QCD MC and three different gluino masses. The QCD MC agrees well with the shapes observed in data and shows that the majority of the events lean toward lower values of S, because of their more dijet-like event structure. Shown in green and red are the distributions for the high-mass gluinos with masses of 750 GeV and 1250 GeV, respectively.

The heavier the resonance, the more spherical an event becomes. For lower masses (shown in blue in Figure 6.15 for a 300 GeV gluino), the shape resembles closer the one

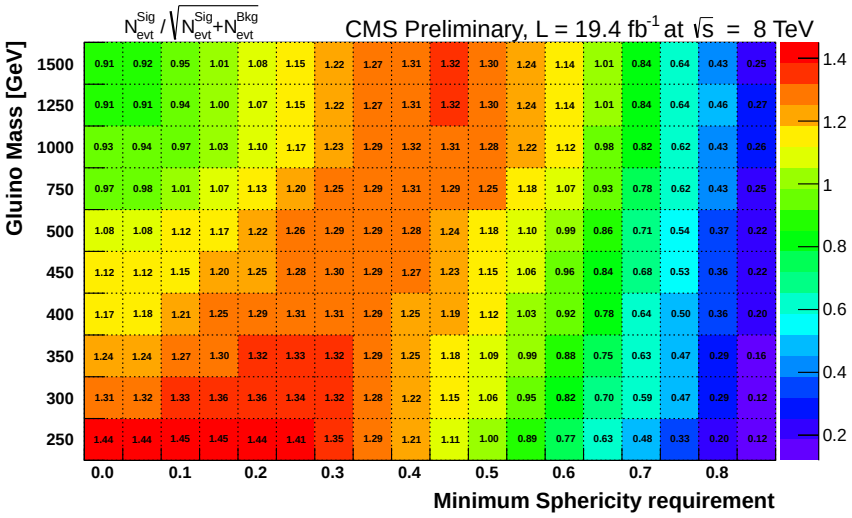

Figure 6.16: Sphericity optimization. Shown is the normalized event signal significance S_{Sig}^{Evt} for all three-jet resonance masses considered (y-axis) and for the different minimum sphericity requirements.

from background. In order to determine the optimal selection criteria we perform another optimization based on the sphericity variable. Unlike changes in the kinematic selection, requirements on the event shape variables have little effect on the shape of the signal and background triplet invariant mass distributions, therefore, the optimization can be performed at the event level. We define the event signal significance as:

$$S_{Sig}^{Evt} = \frac{N_{evt}^{Sig}}{\sqrt{N_{evt}^{Sig} + N_{evt}^{Bkg}}}, \tag{6.6}$$

where N_{evt}^{Sig} denotes all expected events passing the minimum sphericity requirement for a given signal mass, and N_{evt}^{Bkg} represents the number of background events passing the same requirement. Figure 6.16 shows the normalized event signal significance for all three-jet resonance masses considered. Red colors represent minimum requirements on the sphericity variable where the analysis sensitivity improves. Low-mass resonances have a more dijet-like event structure similar to QCD events, therefore, using the sphericity as a discriminating variable is not advantageous. However, for higher masses we find that the sensitivity of the analysis increases when we require events to have $S \geq 0.4$.

6.5 Event selection and optimization summary

We have discussed the basic event selection and analysis strategy for a search for pair-produced three-jet resonances. In order to further optimize the selection criteria we use the model of pair-produced gluinos decaying through RPV into either only light-flavor jets or into one heavy-flavor and two light-flavor jets. Three different search regions are defined: the inclusive search, and the low-mass and high-mass heavy-flavor searches. Table 6.3 summarizes the kinematic and b-tagging selection criteria for all three regions.

Selection criteria	Inclusive search	Heavy-flavor search	
		low mass	high mass
Mass range	400–1500 GeV	200–600 GeV	600–1500 GeV
Jets	at least six jets with $p_T \geq$35 GeV and $\|\eta\| < 2.5$		
Δ	110 GeV	110 GeV	110 GeV
Fourth-jet p_T	110 GeV	80 GeV	110 GeV
Sixth-jet p_T	110 GeV	60 GeV	110 GeV
b-tagging	-	$\geq$ 1 b tags in triplet	
Sphericity	0.4	-	0.4

Table 6.3: Selection requirements for the three search regions in the analysis.

Chapter 7

Statistical methods and background modeling

The main background in this search arises from QCD multijet and $t\bar{t}$ pair production, whereas the latter plays a more important role for the heavy-flavor search. The previous chapter discussed the event and triplet selection criteria as well as optimization strategies based on the benchmark model for three-jet resonances, the gluino with RPV decay. As an introduction to this chapter we start with a discussion of statistical methods for performing fits followed by a more specific description of the background estimates for each of the three search regions as summarized in Table 6.3.

7.1 Statistical analysis

This section describes the statistical methods used in the analysis. We discuss general ideas of fitting data distributions with parametrized functions or binned shapes and estimating their parameters.

7.1.1 Background parameter estimation

Given a set of data points one tries to describe the distribution with a specific model and estimate the best values of the model parameters based on the goodness-of-fit to the data. In this case the data distribution corresponds to the invariant mass of jet triplets that satisfy all selection criteria as outlined in Chapter 6. The expected distribution of M_{jjj} is smoothly falling with increasing triplet invariant mass and we model this background with a four-parameter function:

$$P4(x,\theta) = \frac{dN}{dx} = P_0 \frac{(1-\frac{x}{\sqrt{s}})^{P_1}}{(\frac{x}{\sqrt{s}})^{P_2+P_3\log\frac{x}{\sqrt{s}}}} \quad , \tag{7.1}$$

where N is the number of triplets, x is the triplet invariant mass, $\sqrt{s}$ represents the center-of-mass energy, and $\theta \equiv (P_0, P_1, P_2, P_3)$ are parameters of the function that are estimated by the fit. We discuss three commonly used methods to estimate fit parameters as described in [9]: unbinned and binned maximum likelihood methods, and the method of least squares.

In particle physics we count number of occurrences (events), which are distributed according to a discrete Poisson probability function. In the central limit theorem (i.e. large counts) the Poisson distribution can be approximated by a Gaussian distribution.

7.1.2 Unbinned maximum likelihood method

The unbinned maximum likelihood method for estimating parameters of the model is the default method implemented in the RooFit/RooStats [75] package, and is especially suited for low event counts. We assume there is a set of M measured quantities $x = (x_1, ..., x_m)$, for example, the triplet invariant mass where $x_i \equiv M^i_{jjj}$ of the i'th triplet, that follow a certain probability density function (*pdf*) $f(x, \theta)$, which is normalized to unity (such as a normalized version of Equation 7.1 from above). If the M measurements are independent of each other one can write the combined likelihood for all of the measurements as a product of probabilities :

$$L(\theta) = \prod_i^M f(x_i, \theta). \tag{7.2}$$

Our best estimates of the values of θ is then represented by $\hat{\theta}$, which maximizes the likelihood $L(\theta)$. It is usually more convenient to work with the negative natural logarithm of the likelihood which allows the product to be written as a sum:

$$-\ln L = -\sum_i^M \ln f(x_i, \theta). \tag{7.3}$$

In order to find $\hat{\theta}$, the likelihood equation (equation for maximizing the log-likelihood, or minimizing the negative log-likelihood) needs to be solved for each parameter:

$$\frac{\partial \ln L}{\partial \theta_j} = 0, \qquad j = 0, 1, 2, 3. \tag{7.4}$$

For this computation the overall sample size M is assumed to be fixed. However, it is also possible to treat it as a Poisson-distributed number which is then referred to as extended unbinned maximum likelihood method.

One disadvantage of the unbinned fit is that it does not provide a measure of goodness-of-fit, which quantifies the agreement of the model with the measured data. If the number of measurements is very large, the calculation of the maximum likelihood can be very computing time intensive, therefore it is advisable to bin the data without a loss of precision, and use histograms to perform a binned maximum likelihood fit. This is the method used in the analysis presented.

7.1.3 Binned maximum likelihood method

In case of sufficient statistics one can use binned data to perform a maximum likelihood fit. The method is used in the description of the background as discussed later in Section 7.2. We assume there is a set of M measured quantities, and in the case of the triplet invariant mass distribution this is the bin content at the middle of each bin in the M_{jjj} distribution, assuming a bin width of 10 GeV. If these M measurements are independent of each other, one can write the likelihood of the full histogram as a product of Poisson probabilities for each bin $P(n_i, \mu_i)$, with the expected number of events μ_i given by $P4(x_i, \theta)$ and the observed number of events is n_i. The negative log-likelihood is therefore defined as:

$$NLL = -\ln L(\theta) = -\sum_{i}^{N} \ln P(n_i, \mu_i) = -\sum_{i}^{N} \ln \left(\frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} \right). \tag{7.5}$$

One can add a constant term to Equation 7.5, since the minimization procedure is independent of terms that do not depend on x_i. Therefore, the modified negative log-likelihood can be written as:

$$-\ln LLR = -\sum_{i}^{N} \ln \left(\frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} \right) + \sum_{i}^{N} \ln \left(\frac{n_i^{n_i} e^{-n_i}}{n_i!} \right), \tag{7.6}$$

where the second term represents the probability of observing n_i events given that n_i events are expected [76]. Equation 7.6 is also referred to a the negative log of the likelihood ratio

$(-\ln LLR \equiv -\ln \lambda)$:

$$-\ln \lambda = -\ln \frac{L(n_i, \mu_i)}{L(n_i, n_i)}. \tag{7.7}$$

In the limit of large n_i, when Poisson errors can be approximated by Gaussian errors, one can derive a relation between the LLR and the χ^2 (see Section 7.1.4) distribution:

$$-\ln LLR = -\ln \lambda = \frac{\chi^2}{2}, \tag{7.8}$$

therefore giving a measure of the goodness-of-fit, discussed in more detail in following section. This method is used for the background estimates and all fits that are performed with the RooFit/RooStats packages.

7.1.4 Least square method

If the number of entries in each bin of a histogram is sufficiently large the errors can be approximated by a Gaussian distribution. The χ^2 distribution for the entire histogram is then defined as a sum over all bins:

$$\chi^2 = \sum_i^M \frac{(n_i - \mu_i)}{\sigma_i}, \tag{7.9}$$

where the σ_i is the error on the bin content n_i, which can be approximated by $\sqrt{n_i}$, and μ_i is the expected number of events in each bin. With the minimization of this χ^2 one can determine the best values of the parameters $\hat{\theta}$. Since the χ^2 distribution itself follows a Gaussian function one can define uncertainties on the fit parameters as the values where $\chi^2 \to \chi^2 + 1$. The minimal χ^2_{min} provides an estimate for the goodness-of-fit when divided by the number of degrees of freedom (ndf), which is defined as M (number of bins) - n (number of parameters of the fit). The value of $\chi^2_{min}/ndf \approx 1$ indicates a good agreement between the fit and the observed data within the statistical uncertainties. This method is the standard implementation within the Root package [77], and we use it, for example, in modeling the Gaussian signal parameters as a function of mass, as discussed in Chapter 8.

7.2 Background estimate

In this section we discuss more specifically the background estimates for the three search regions defined in Table 6.3. After applying all event and triplet selection criteria to the data we find that the triplet invariant mass distribution is smoothly falling in the mass range of interest. For the background estimate we do not rely on QCD MC simulation but use a fit directly to the data to extract the expected multijet background. This fit is either based on the function given in Equation 7.1 or estimated from a control region in data. The contribution from $t\bar{t}$ becomes more important when b tags are added, which we will show in Section 7.2.2. Besides being a background for the heavy-flavor RPV search, it also plays an important role to validate the jet ensemble technique on a known SM process.

7.2.1 Inclusive search for three-jet resonances

The event and triplet selection criteria for this part of the search are given in the left most column in Table 6.3 and are summarized below for convenience:

- $\Delta = 110$ GeV
- 6^{th}-jet $p_T \geq 110$ GeV
- Sphericity ≥ 0.4

The data M_{jjj} distribution is smoothly falling and we use the function given in Equation 7.1 in a fit directly to the data distribution to model multijet background. The RooFit/RooStats package is used to perform this fit using the binned maximum likelihood method as discussed in Section 7.1.3. The result of this fit is shown in Figure 7.1 on a normal scale on the left and a log scale on the right in the top row. The data M_{jjj} distribution is represented by the black points and a bin size of 10 GeV is used. The bottom row of the same figure shows the residual (data – fit) and pull ((data – fit)/error) distributions, which show good agreement of the data with fit function. The plot on the right of Figure 7.2 shows a comparison of the fit with the data using a binning which is based on the expected resolution of $\approx 7\%$ for a given triplet mass leading to the following values for

the lower bin edge in GeV:

$$\begin{aligned}\text{bins} = \{&0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 121, 131, 141, 151, 162, 174, 187,\\ &201, 216, 232, 249, 267, 286, 307, 329, 353, 379, 406, 435, 466, 500, 536, 575,\\ &617, 662, 710, 761, 816, 875, 938, 1006, 1079, 1157, 1241, 1331, 1427, 1530,\\ &1641, 1760, 1887, 2024, 2170, 2327, 2495, 2677, 3000\}.\end{aligned} \tag{7.10}$$

Each bin content is divided by the width of the bin to retain the proper shape resulting in a differential triplet invariant mass distribution. Also shown are potential signals of a gluino with masses of 500 GeV in pink (750 GeV in blue).

As a cross-check we also use simulated QCD and $t\bar{t}$ events modeled by the parameters given in Table 4.3. Both shapes of the M_{jjj} distribution are extracted and a combined binned maximum likelihood fit is performed to the data, determining the normalization of the QCD component. The $t\bar{t}$ contribution is scaled to the next-to-next-to-leading-order (NNLO) cross section of $245.8^{+8.7}_{-10.5}$ pb [52]. The outcome of this comparison is shown on the right hand side of Figure 7.2 with QCD in yellow and $t\bar{t}$ in red. For this event selection with the high thresholds on jet p_T the contribution from $t\bar{t}$ is almost negligible and no distinct mass peak is observed. Comparing the shape of the M_{jjj} distribution of QCD MC and $t\bar{t}$ one can see how similar are the shapes of incorrectly combined $t\bar{t}$ triplets and background triplets from QCD multijet events.

We find no significant deviation of the data from the background prediction indicating signs of new physics. We will use the background determined from the four-parameter fit in order to set limits on the production cross section of new heavy three-jet resonances, which is discussed in more detail in Section 9.6.

7.2.2 Search for three-jet resonances with heavy-flavor jets

Here we describe the background modeling for the search for three-jet resonances with heavy-flavor jets in the final state. As determined in Section 6.4.2, we choose the CSV tagging algorithm with the medium working point for this part of the analysis and require each event to have at least one b tag to increase sensitivity to the potential signal. Furthermore,

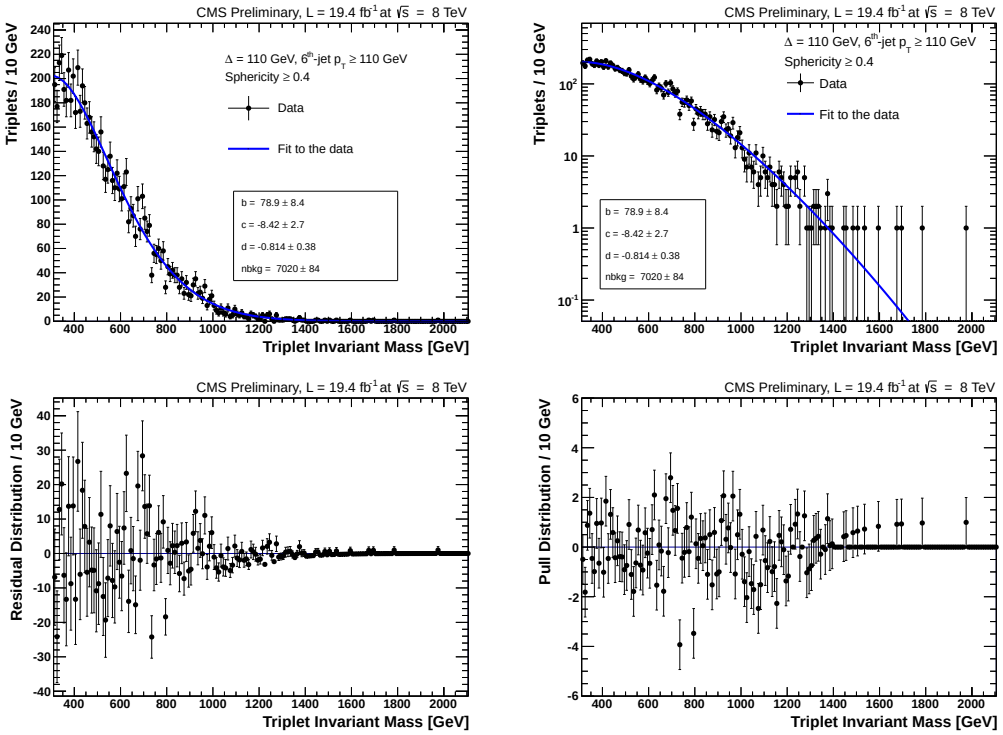

Figure 7.1: Top row: Binned maximum likelihood fit to the data performed with the RooFit package. Normal scale (right), log-scale (left). Bottom row: Residual (data – fit) and pull ((data – fit)/error) distribution.

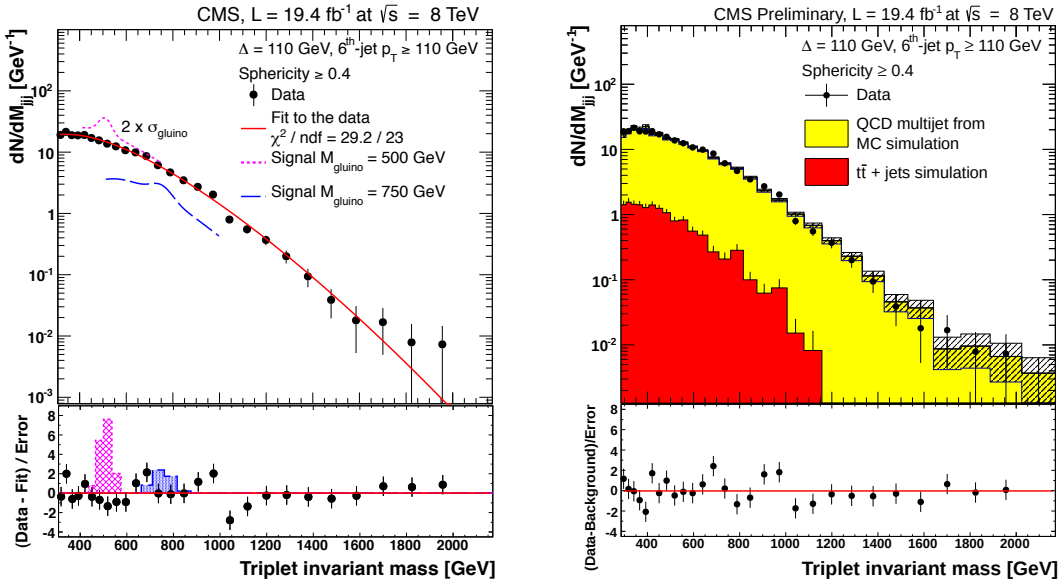

Figure 7.2: Comparison of the differential M_{jjj} distribution in data with the four-parameter fit function (right) and simulation (left). A hypothetical signal is shown on the right plot for a gluino with a mass of 500 GeV (pink) and 750 GeV (blue). The pull distribution are shown for both background estimates underneath the respective plot.

removing triplets without b tags helps to reduce the number of incorrect combinations. The selection requirements for the heavy-flavor search are given in the middle and right column of Table 6.3. The search is divided into two mass regions, which are discussed separately in the following sections.

Low mass region $200 < M_{jjj} < 600$ **GeV**

Events and triplets are selected based on the following criteria:

- $\Delta = 110$ GeV
- 4^{th}-jet $p_T \geq 80$ GeV
- 6^{th}-jet $p_T \geq 60$ GeV
- ≥ 1 b tag in triplet (CSV medium)

Because of the large contribution from $t\bar{t}$ triplets in the low-mass region of the M_{jjj} distribution, we choose to model the background here by a combination of $t\bar{t}$ MC simulation

and a data control region, which models the QCD multijet component. For this control region events are selected where b jets are explicitly vetoed, creating a data sample that is statistically independent from the signal region. The triplet invariant mass distribution from $t\bar{t}$ simulation is scaled to the integrated luminosity of the dataset and the theoretical cross section of 245.8 pb. The sum of this $t\bar{t}$ shape and the expected shape of the multijet background are fitted simultaneously to the triplet invariant mass distribution in data, where the normalization of the QCD background estimate is allowed to float freely. This fit can be seen in Figure 7.3 along with residuals and pulls. The fit range here is choosen to start at 180 GeV, to search for new heavy resonances above 200 GeV. Figure 7.4 shows a comparison of the data and background prediction using resolution based binning. As a cross-check a dedicated study to measure the $t\bar{t}$ cross section is presented in Appendix C and shows good agreement within uncertainties with the theory prediction.

High mass region from $600 < M_{jjj} < 1500$ GeV

Events and triplets are selected based on the following criteria:

- $\Delta = 110$ GeV
- 6^{th}-jet $p_T \geq 110$ GeV
- ≥ 1 b tag in triplet (CSV medium)
- Sphericity ≥ 0.4

Here the same technique for estimating the background is used as discussed previously for the inclusive search in Section 7.2.1. The data is fitted directly with a four-parameter function and the parameters and normalization are determined with the binned maximum likelihood method. Figure 7.5 shows the data distribution with the fit in the top row on a linear (left) and logarithmic (right) scale. Underneath are the residual and pull distributions based on a 10 GeV bin size. Finally, the bottom plot shows a comparison of the fit and the data with resolution based binning.

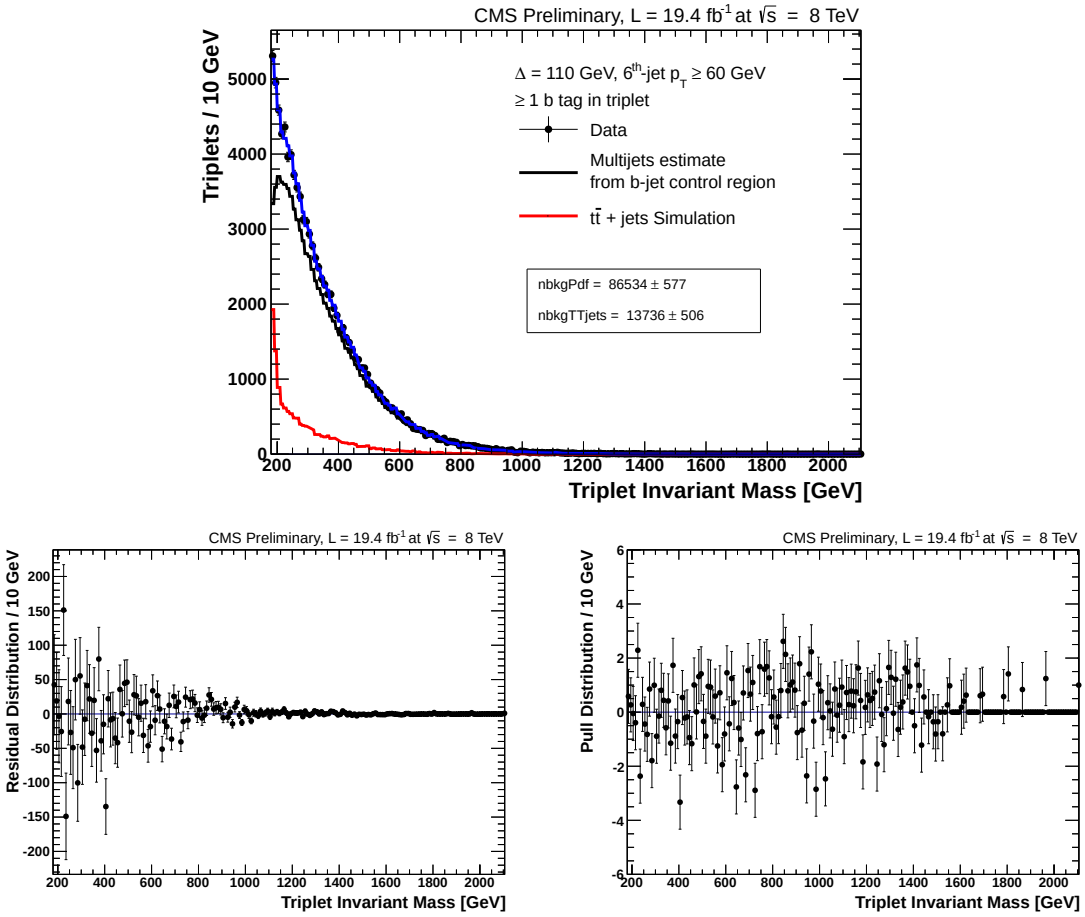

Figure 7.3: Binned maximum likelihood fit to the data distribution in the low-mass region of the heavy-flavor search. The background consists of the triplet invariant mass distribution of simulated $t\bar{t}$ triplets (red) and a QCD prediction from a b-jet control region in data (black). The plots in the bottom row show the residual and pull distributions based on a bin size of 10 GeV.

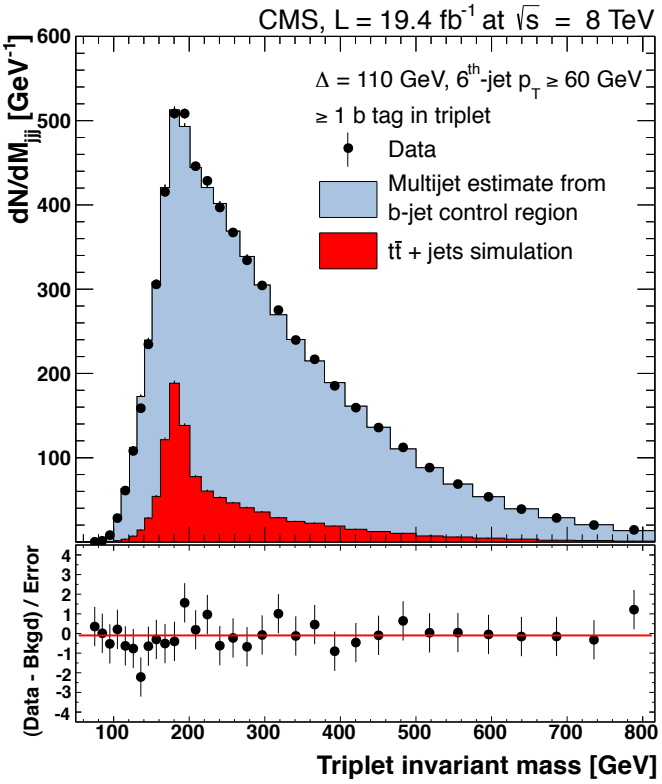

Figure 7.4: Binned maximum likelihood fit to the data distribution in the low-mass region of the heavy-flavor search. The plot shows the data and background fit in resolution based binning. The background from simulated $t\bar{t}$ triplets is shown in red and the QCD multijet prediction is represented by the blue shaded area. The corresponding pull distribution is shown underneath.

Overall we find no significant deviation from the predicted background in neither of the two search regions. Limits on the production cross section of heavy-flavor three-jet resonances will be calculated using the above mentioned methods to estimate the backgrounds for the low- and high-mass heavy-flavor analysis.

Comparison with QCD MC

As an additional cross-check, we use QCD MC simulation as the background template both for the low-mass and high-mass heavy-flavor search. To increase the amount of available statistics the QCD MC sample does not have the b-tagging requirements applied, since the normalization is determined from the fit to the data directly. We find that adding b-tagging requirements to the event selection does not alter the shape of the M_{jjj} distribution for the simulated QCD events as can be seen in Figure 7.6. In Figure 7.7 we show both data distributions in comparison with a background composed of simulated QCD and $t\bar{t}$ triplets. The pull distributions underneath show good agreement between the data and prediction. The shaded error bands represent only the statistical uncertainties on the simulated samples.

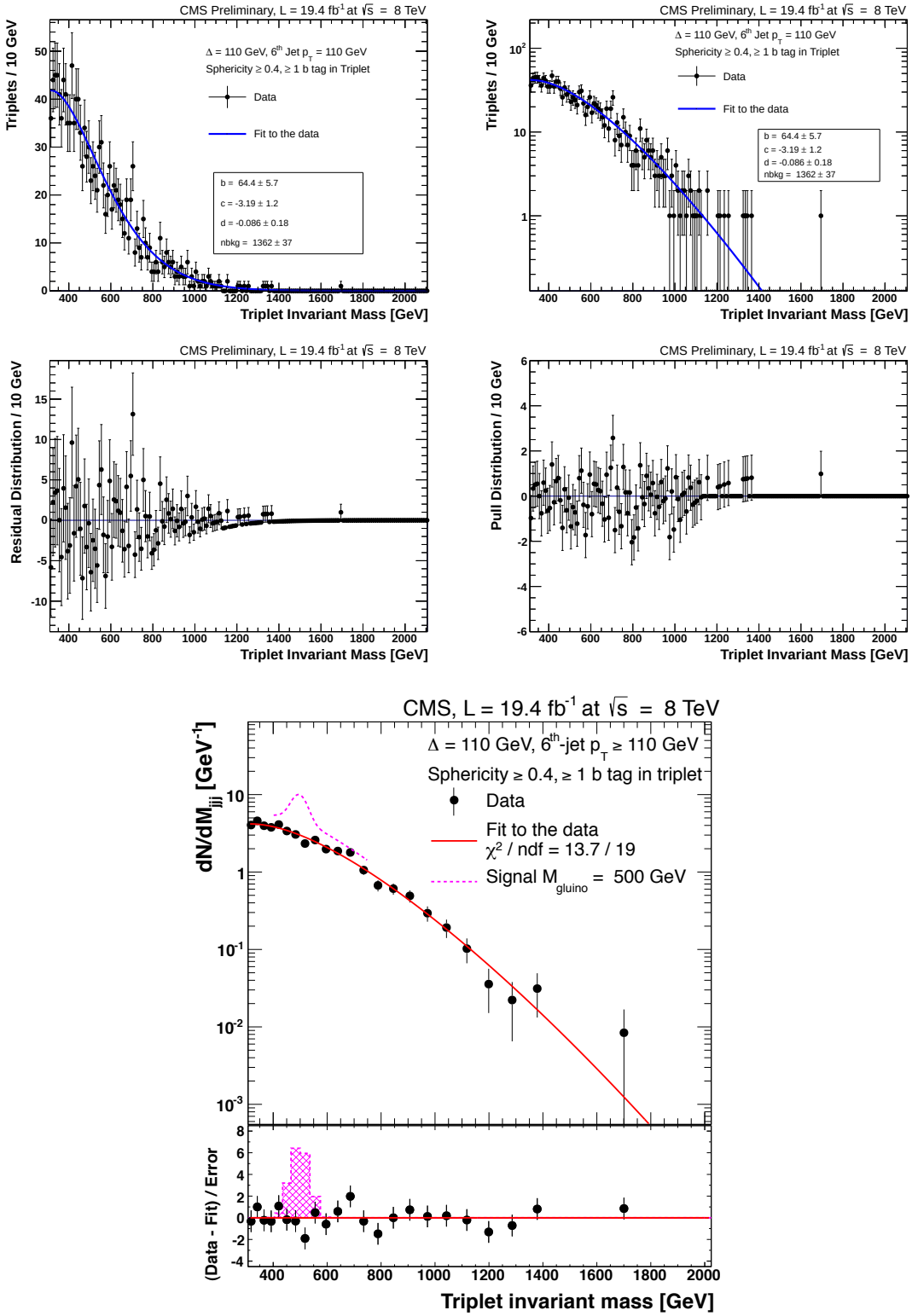

Figure 7.5: Binned maximum likelihood fit to the data distribution in the high-mass region of the heavy-flavor search in linear (left) and log (right) scale. The plots in the middle show the residual and pull distributions. The plot on the bottom shows resolution based binning and a potential gluino signal with a mass of 500 GeV in pink.

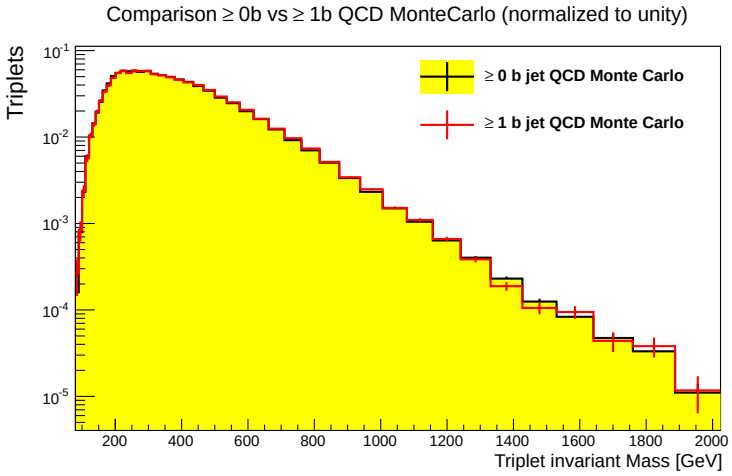

Figure 7.6: Shape comparison of the M_{jjj} distribution in QCD MC when events are selection that contain either $\geq$ 0 b tags (yellow) or $\geq$ 1 b tags (red). The effect on the shape is negligible and only increases the statistical uncertainty on the QCD MC sample. We therefore use the higher statistics inclusive QCD sample to predict the background to model the background for heavy-flavor search.

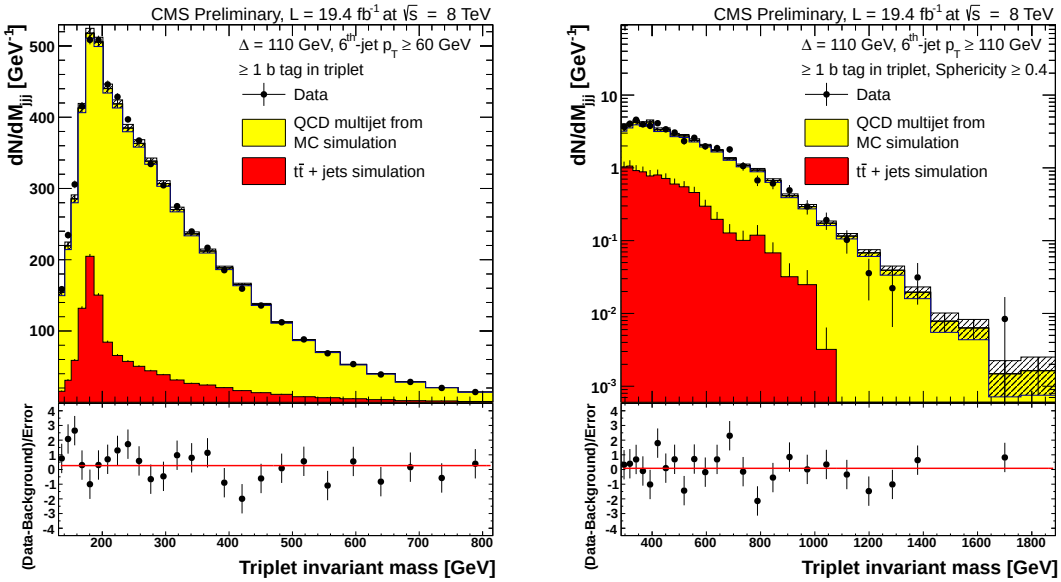

Figure 7.7: Comparison of data M_{jjj} distribution with QCD MC and simulated $t\bar{t}$ events. The $t\bar{t}$ (red) is scaled to the integrated luminosity of the dataset and the theoretical NNLO cross section, while the normalization of the QCD (yellow) component is allowed to float freely when a fit to the data distribution is performed.

Chapter 8

Signal modeling

This search is conducted as a "bump-hunt" looking for a localized Gaussian deviation on top of a smoothly falling spectrum. With the background estimate coming directly from fits to the data distribution we make use of the very similar shapes of incorrectly combined triplets and triplets from QCD multijet background. Both contributions can be described by the same functional form given in Equation 7.1. An example of this can be seen in the two signal fits overlaid on the data distribution on the left in Figure 7.2. Given that the resolution of the reconstructed resonances is solely determined by the jet energy resolution we model the signal using a Gaussian line shape.

8.1 Acceptance and efficiency

To evaluate the sensitivity to the new physics signal, we define two separate quantities, the signal event acceptance (f_1) and the efficiency to select a correct triplet combination ($f2 \times f3$). The event acceptance f_1 is defined as the fraction of events passing the kinematic and b-tagging requirements as given in Table 6.3, with the additional condition that at least one triplet from the event has to satisfy Equation 6.1 with $\Delta = 110$ GeV. To determine the efficiency of selecting a triplet we start with defining the average number of triplets per event passing the Δ requirement denoted by f_2. This value depends on the resonance mass and b-tagging selection and ranges between 1.6 and 4. The third quantity f_3 is the ratio of triplets within the Gaussian signal peak over all triplets that passed the Δ requirement. The evaluation of f_3 is performed for each gluino mass by a similar method to the optimization technique described in Section 6.4. The M_{jjj} distribution for the signal is fitted with a Gaussian for the correctly combined triplets plus a four-parameter function for the distribution of incorrectly combined triplets, with the integral of the Gaussian divided

by all the triplets in the distribution giving f_3. This value also represents an estimate of how often we find the correct combination of three jets out of all 20 possible ones. In summary, the acceptance and efficiency are factorized into three quantities, f_1, f_2, and f_3, defined as follows:

- f_1: Fraction of events passing the kinematic and b-tagging selection, where at least on triplet passed the diagonal offset requirement of $\Delta = 110$ GeV

$$f_1 = \frac{N_{evt}^{pass}}{N_{evt}^{gen}}, \tag{8.1}$$

- f_2: Average number of triplets per event passing the Δ requirement

$$f_2 = \left\langle \frac{N_{trip}}{event} \right\rangle, \tag{8.2}$$

- f_3: Ratio of triplets in the Gaussian peak over all triplets passing the Δ requirement

$$f_3 = \frac{N_{trip}^{Gauss}}{N_{trip}^{all}}. \tag{8.3}$$

To calculate the expected number of triplets for a sample, we combine acceptance times efficiency ($f_1 \times f_2 \times f_3$) and multiply by the corresponding cross section and the total integrated data luminosity. This calculation is equivalent to the ratio of triplets in the Gaussian signal peak divided by the number of events generated:

$$f_1 \times f_2 \times f_3 = \frac{N_{evt}^{pass}}{N_{evt}^{gen}} \times \left\langle \frac{N_{trip}}{event} \right\rangle \times \frac{N_{trip}^{Gauss}}{N_{trip}^{all}} = \frac{N_{trip}^{Gauss}}{N_{evt}^{gen}}. \tag{8.4}$$

Fits to the M_{jjj} distribution of the signal MC sample are shown in shown in Figure 8.1 for gluinos with light-flavor jets in the decay and a minimum sixth-jet p_T of 110 GeV. Figure 8.2 shows fits to the signal MC samples for gluinos decaying into one heavy-flavor and two light-flavor jets. For masses below 600 GeV the fourth-jet p_T has to be above 80 GeV and the sixth-jet p_T above 60 GeV, whereas for higher resonances masses the sixth-jet p_T requirement is increase to 110 GeV. Figure 8.3 shows the variables f_1, f_2, and f_3 (from left to right) as a function of gluino mass. The top row represents the event selection

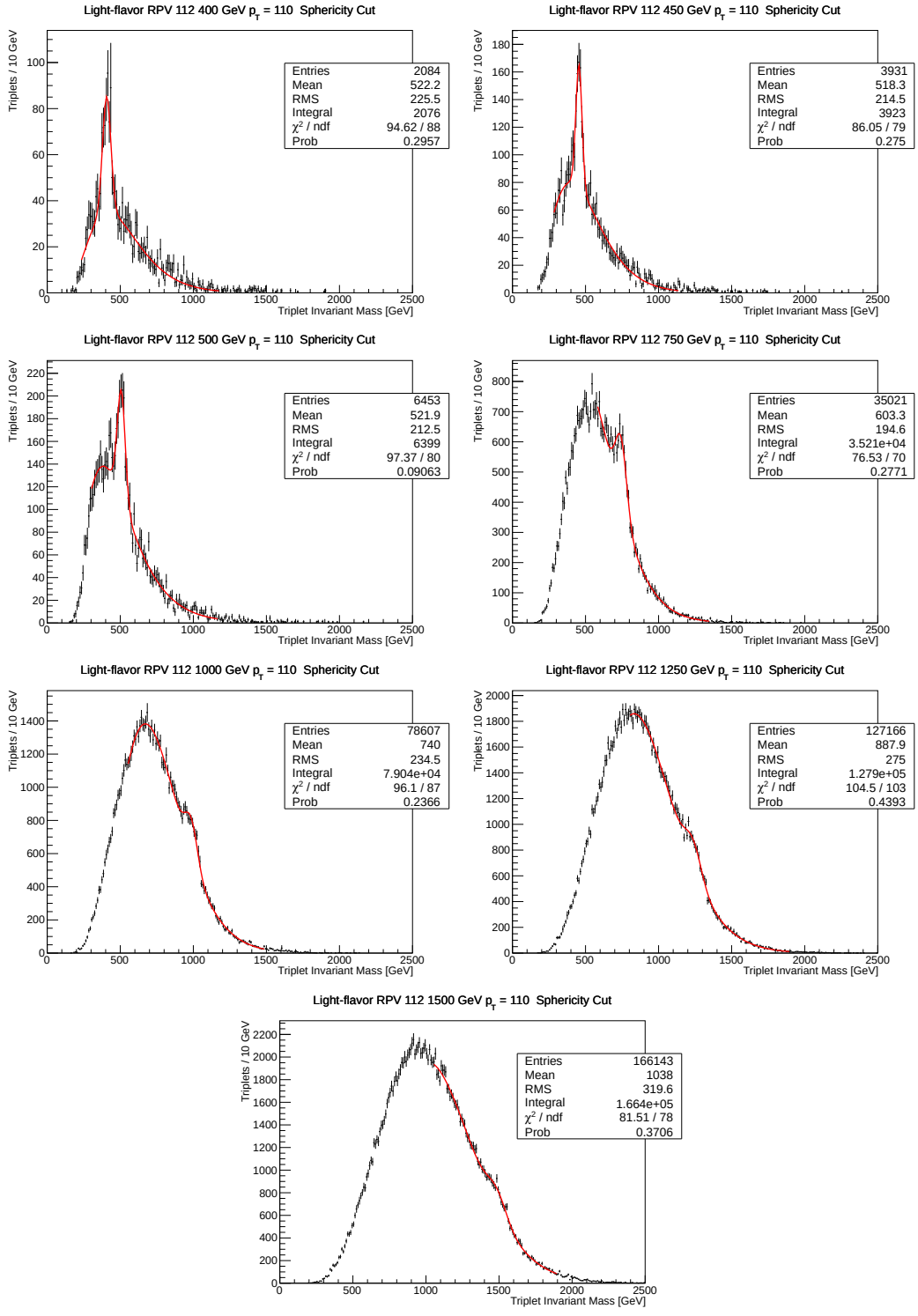

Figure 8.1: Fits used for acceptance parametrization for the inclusive search.

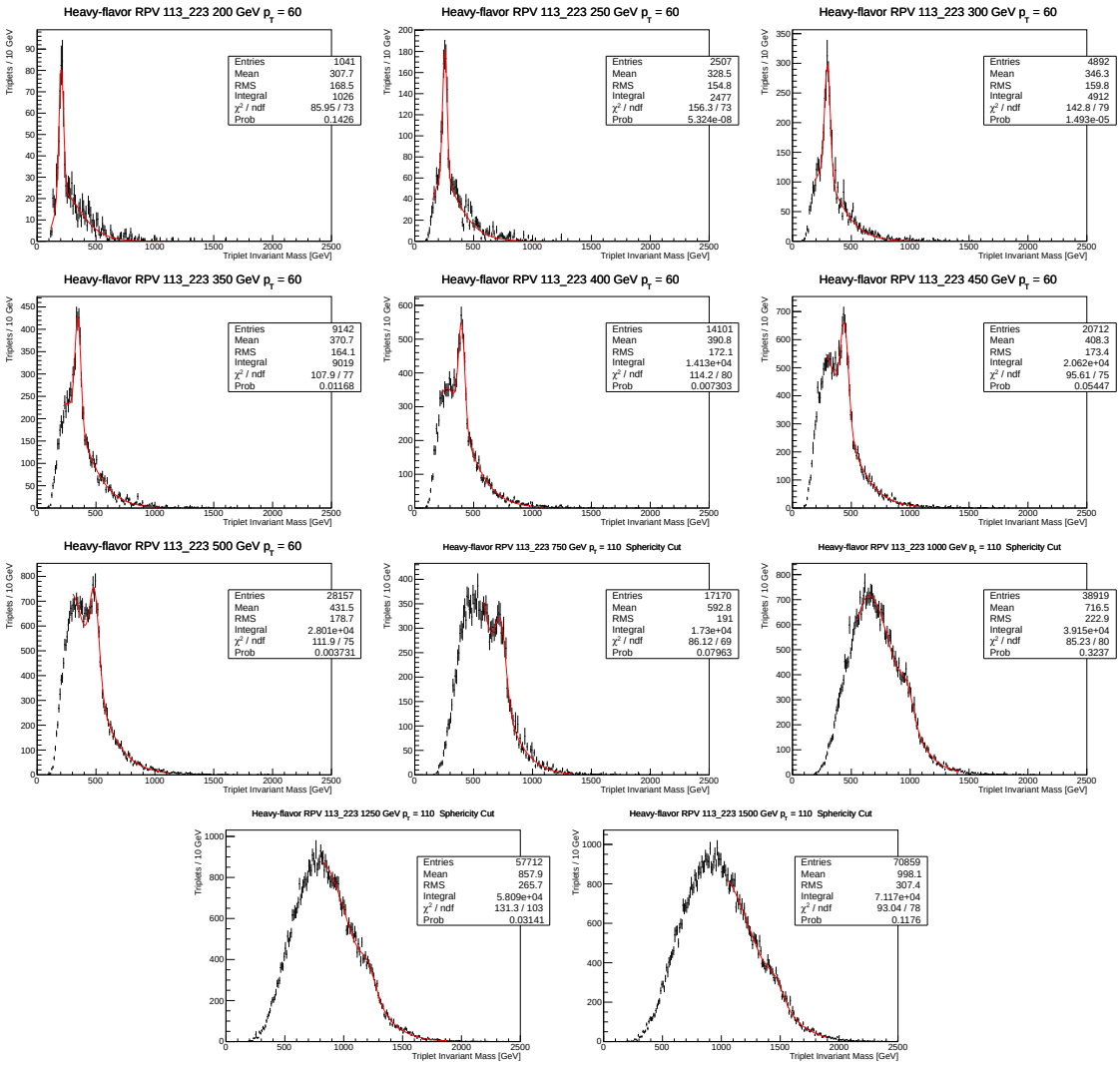

Figure 8.2: Fits used for acceptance parametrization for the heavy-flavor search.

used for the inclusive search, and the two bottom rows show the results for the heavy-flavor search. The top plot in Figure 8.4 shows the event acceptance times triplet efficiency $A \times \epsilon$

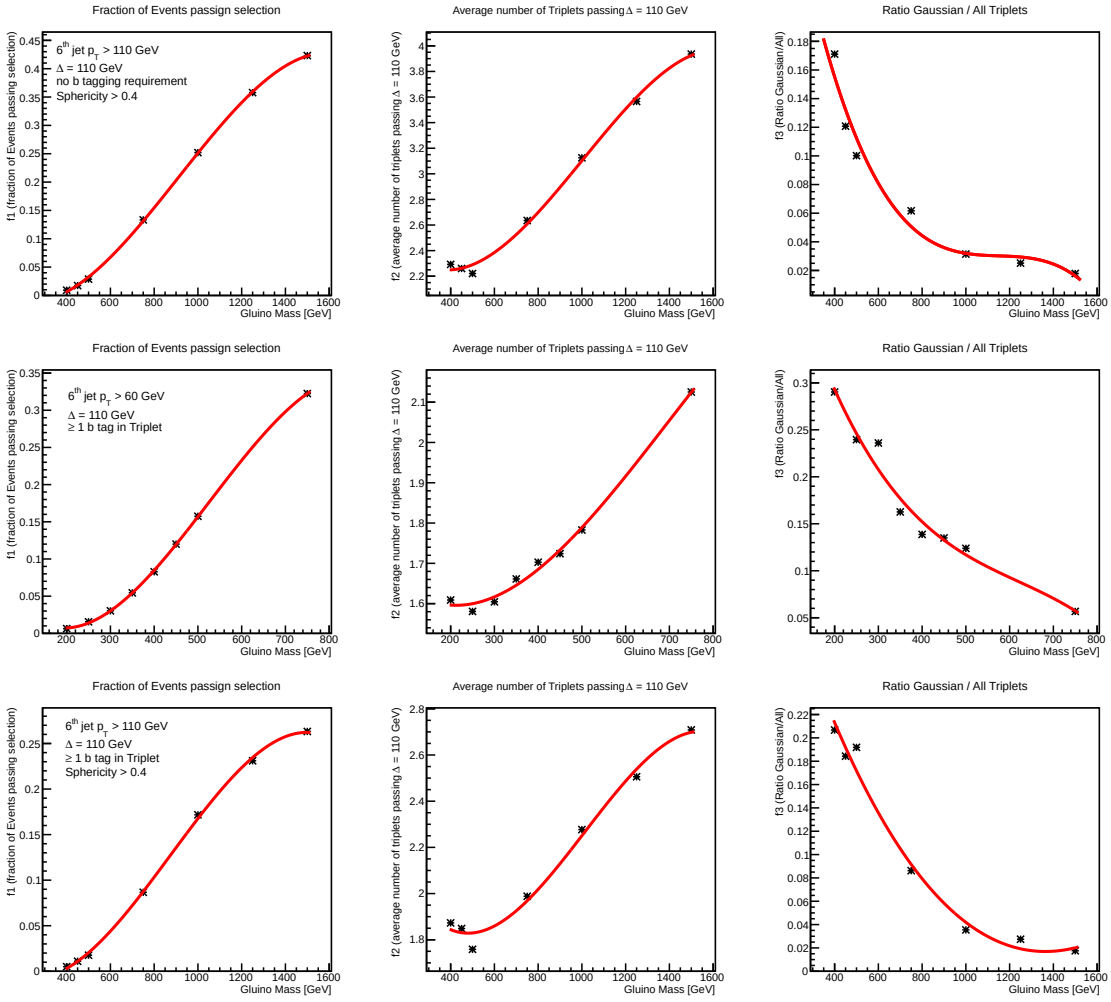

Figure 8.3: From left to right f_1, f_2, and f_3 as a function of gluino mass. The top row shows the values for the inclusive search, and the two bottom rows show the values used in the heavy-flavor search.

for a Gaussian signal for the light-flavor search as a function of gluino mass, with the same kinematic selection criteria used for all masses. The values are fitted with a third-order polynomial, which empirically describes the behavior well. Since two separate kinematic selection criteria are chosen for the heavy-flavor search, the $A \times \epsilon$ changes significantly, and therefore the two respective mass ranges are separated. The bottom left plot in Figure 8.4 shows $A \times \epsilon$ versus gluino mass for the mass range below 600 GeV, and the right plot

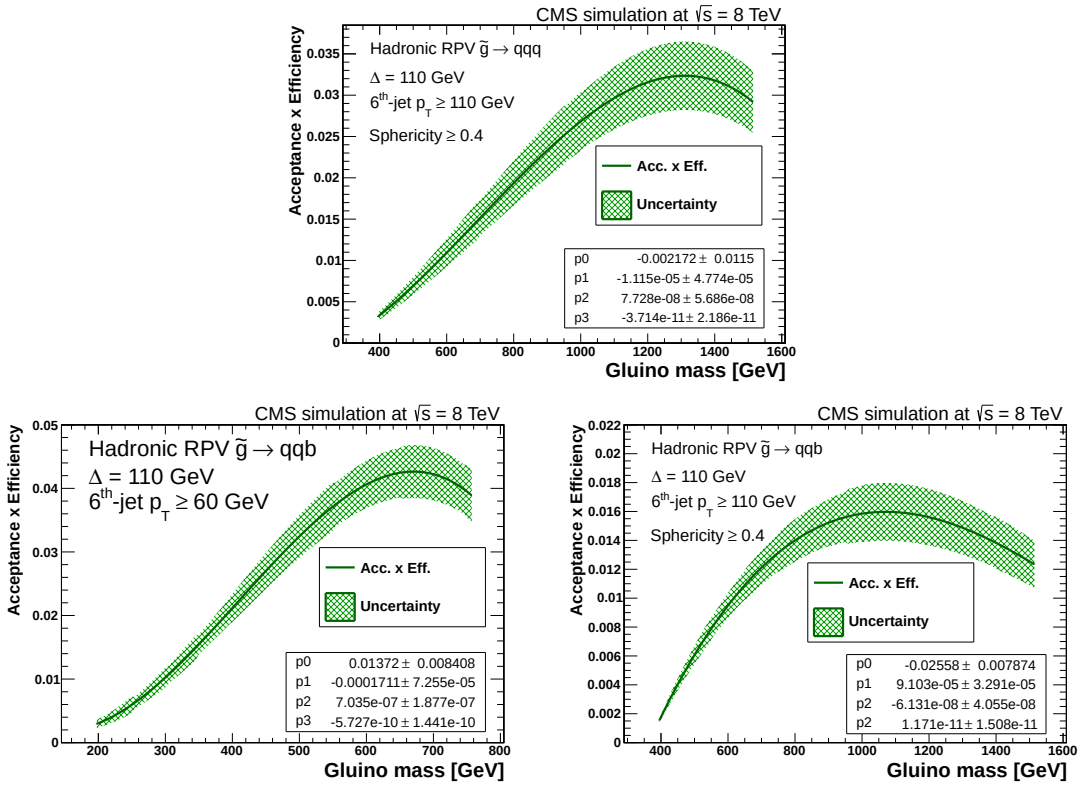

Figure 8.4: Event acceptance times triplet efficiency for a Gaussian signal as function of gluino mass. The values for the inclusive search are shown in the top row. The bottom row corresponds to the $A \times \epsilon$ used in the heavy-flavor search.

shows the higher masses. The points are also fitted with a third-order polynomial, and the parameters based on a χ^2 fit are displayed in the figures. The green shaded areas represent systematic uncertainties on $A \times \epsilon$ and are discussed in more detail in Chapter 9.

8.2 Gaussian signal parameters

In addition to the $A \times \epsilon$ two more parameters are needed to describe the Gaussian line shape, the mean and width. Since we search for a Gaussian signal in a large mass range the mean is set to the point of interest while performing the search. As expected, the Gaussian width increases with the mass of the particle, with the mass resolution ranging from 9% to 5% from lower to higher masses, respectively . Similar values have been obtained for the three-jet mass resolution measurement in Reference [73]. Figure 8.5 shows the signal width

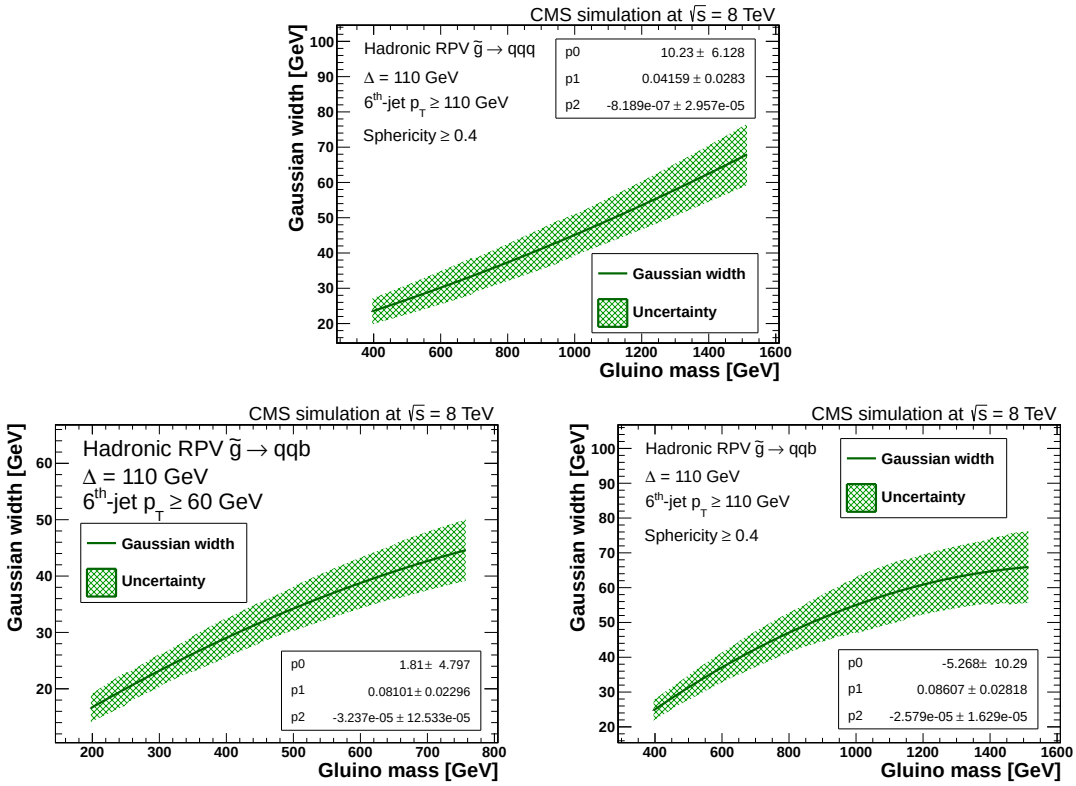

Figure 8.5: Gaussian width as a function of gluino mass. Light-flavor search (top), heavy-flavor search (bottom).

as function of gluino mass. The top plot shows the values of the width used for the inclusive search, while the bottom plots represent the values used in the heavy-flavor search. We find a second-order polynomial to describe the shape, and the estimated parameters based on a χ^2 fit are shown in the plot. The green shaded areas around the values of the Gaussian width correspond to the square sum of the actual uncertainty on the fitted width, as shown in the third column of Tables 9.1 and 9.2, and a 10% uncertainty, which is assigned to account the uncertainty on the three-jet mass resolution. These uncertainties are taken into account when searching for this new physics signal.

8.3 Summary on signal modeling

For this analysis we use a Gaussian line shape to model the potential signal of a pair-produced hadronic resonances decaying into three jets. The expected number of triplets is

defined by the integral of the Gaussian shape and estimated based on the values of $A \times \epsilon$ estimated from simulated gluino signal samples, multiplied by the theory cross section, and the total integrated luminosity of the data sample. The Gaussian width parameter is derived from MC and parametrized as a function of gluino mass.

Chapter 9

Systematic uncertainties and limit setting

As previously noted we find no significant deviation in the data distribution with respect to the background estimates described in Section 7.2. In this chapter we discuss the effect of systematic uncertainties on the analysis followed by a general discussion of setting limits.

9.1 Sources of systematic uncertainties for signal and $t\bar{t}$ background MC

Sources of uncertainties that affect the background from $t\bar{t}$ and signal models derived from MC simulation are considered first and include, for example, effects of jet energy scale uncertainties, b-tagging scale factor uncertainties, pileup uncertainties, and the uncertainty on the measurement of the luminosity.

Uncertainty on jet energy scale correction

Section 5.4.2 describes in detail how jet energy corrections and their uncertainties are derived for CMS. Figure 5.5 in the same section shows from left to right: the uncertainty as a function of p_T for central jets $|\eta| = 0$, the uncertainty as a function of p_T for jets with $|\eta| = 2.7$, and the uncertainty as a function $|\eta|$ for a jet p_T of 100 GeV. To evaluate the effect on the MC signal samples a $\pm 1\sigma$ correction is applied to each jet, based on the aforementioned distributions, and the p_T and $|\eta|$ of the jet. Simultaneously, the E_T^{miss} is recalculated as well. We find the effect on the signal $A \times \epsilon$ to be on the order of 10%, which is evaluated based on the method described in Section 9.2.1. We estimate the effect on the fit position of the Gaussian mean to be $\pm$ 1%. For $t\bar{t}$ the effect on the overall normalization is also on the order of 10% and the effect on the shape of the triplet invariant mass distribution is shown in the top left plot in Figure 9.3.

Pile-up re-weighting uncertainties

Additional interactions that occur in the same bunch crossing are referred to as pile-up interactions. They are not part of the hard interaction and contribute additional low p_T objects to the event. The number of pile-up interactions in data depends on the luminosity in each bunch crossing and on the total inelastic cross section, which is used to determine the distribution of additional interactions for the dataset used in this analysis. This method does not rely on the vertex reconstruction algorithms as described in Section 5.3 to avoid any potential biases. Additional interactions are also modeled in simulated MC samples, where each event is assigned a number corresponding to these pile-up interactions. The distribution in simulation is based on the expected distribution in data, however both do not match exactly. The MC truth distribution is shown in Figure 9.1 (left) in green. Each event in simulation is then assigned a weight to reproduce the distribution measured in data shown in black. Uncertainties associated with this procedure originate from the luminosity and inelastic cross section measurements. For the central value of the inelastic cross section we use the CMS recommendation of 69.4 mb [78] and assign a 5% uncertainty corresponding to $\pm$ 3.5 mb leading to new pile-up histograms for the re-weighting. The nominal distribution (black) and the distributions corresponding to a $\pm 1\sigma$ change in the inelastic cross section (blue and red) can be seen in Figure 9.1 (right). The analysis is repeated using the blue and red pile-up distribution from Figure 9.1 and the effect on the signal $A \times \epsilon$ is on the order of 3%.

Uncertainty on the measurement of the integrated luminosity

The measurement of the luminosity is described in detail in Section 3.2.7. The uncertainty on this measurement is estimated to be 2.6% [36] leading to a value of the total integrated luminosity used in this analysis of 19.4 $\pm$ 0.5 fb^{-1}. This uncertainty effects the yield of all MC samples the same way, for example, scaling the expected number of triplets up or down by 2.6%.

Uncertainties on the b-tagging scale factors

As can be seen from Figure 5.8 b-tagging and mis-tagging efficiencies differ between

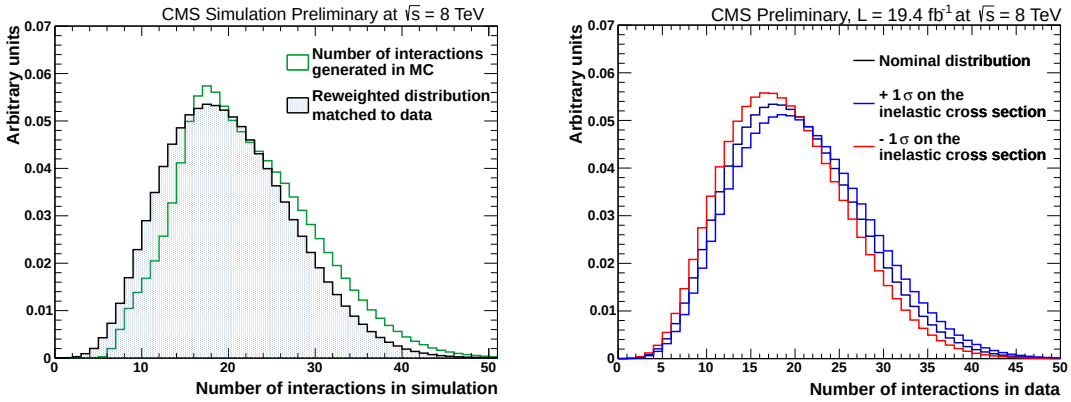

Figure 9.1: Distribution for pile-up re-weighting. The left plot shows the generated MC pile-up distribution in green. The re-weighted distribution is shown in black. On the right hand side the nominal pile-up distribution in data is shown in black and the histograms corresponding to a $\pm 1\sigma$ shift in the inelastic cross section are shown in blue and red.

data and MC. To correct the MC samples to match the efficiency in data, scale factors are applied to the MC samples, and these factors have associated systematic uncertainties. We vary both types of scale factors simultaneously by $\pm 1\sigma$ up and down and generate systematic signal and $t\bar{t}$ histograms. The effect is small fairly small, about 3% on the $A \times \epsilon$ for most masses.

Jet energy resolution uncertainties

Uncertainties on the jet energy resolution will effect the expected Gaussian width of the potential signal. The jet energy resolution is described in more detail in Section 5.4.3. In order to take into account the effect on the three-jet invariant mass due to uncertainties on the jet energy resolution we assign a 10% uncertainty on the Gaussian width parameter. This uncertainty is taken into account when fitting data for a hypothetical signal where the width is allowed to vary within the said 10% of its width predicted by MC simulation.

9.2 Sources of systematic uncertainties for signal samples

The signal is modeled by a Gaussian shape whose parameters are extracted from a fit to the triplet invariant mass distribution of simulated signal MC samples, the uncertainty on

the estimated parameters are discussed in this section. Signal samples are generated with the PYTHIA event generator and the effect of an increased or decreased amount of initial and final state radiation (ISR/FSR) is described as well.

Uncertainty on the fit parameters

Most of the signal fits are performed with the standard implementation in ROOT using the χ^2 method, which is described in Section 7.1.4. The errors on the fit parameters are automatically evaluated during the minimization procedure with the MINUIT package. For each mass point we assign an uncertainty on the $A \times \epsilon$ based on the error on the Gaussian amplitude. Similarly, each width parameter is assigned an uncertainty based on the error estimated by the fit.

Initial- and final-state radiation (ISR/FSR) uncertainties

Initial- and final-state radiation refers to the spontaneous emission of particles (mostly gluons or photons) by incoming or outgoing partons of the hard interaction. The modeling of these effects depends on the generator and as a consequence the energy and number of jets present in each event could be affected. For the MC signal samples we use the PYTHIA event generator and follow a description similar to Reference [79] to estimate the effect of an increased or decreased amount of ISR/FSR. In this reference two parameters are identified that regulate the amount of ISR/FSR in PYTHIA and we use the same settings to evaluate the effect on the gluino model. The PYTHIA parameter k_{ISR} (PARP(67)), when multiplied by the hard scattering scale, gives the maximal p_T in the initial-state shower. The default value is set to 2.5 and we vary this value by ± 0.5. The second parameter in charge of modeling the final-state radiation is PARP(71) which is set to 4.0 as a default value. We set this value to 2.5 or 8 for a decreased or increased amount of FSR. Additional MC samples are generated where the aforementioned parameters are varied up and down. In summary, the following PYTHIA parameters are used for the MC signal generation:

- More ISR/FSR
  - PARP(67)=3.0
  - PARP(71)=8.0

- Less ISR/FSR
  - PARP(67)=2.0
  - PARP(71)=2.5

We generate three mass points with these systematic variations 300, 500 and 750 GeV and pass them through the full CMS detector simulation. The effect on the $A \times \epsilon$ of more or less ISR/FSR is evaluated based on the change of number of triplets matched to parent gluinos, and is on the order of 10%.

9.2.1 Summary of systematic uncertainties on gluino signal samples

As noted earlier, the signal is modeled by a Gaussian function by its width, amplitude, and mean. The integral of a Gaussian is simply related to width and amplitude through the following relation: $integral = \sqrt{2\pi} \cdot amplitude \cdot width$, which is used to calculate the $A \times \epsilon$. For most sources of uncertainties we reevaluate the $A \times \epsilon$, plot these values as a function of gluino mass, and fit the values with a third order polynomial. Then we assign an uncertainty on $A \times \epsilon$ for each fitted mass based on the largest deviation from the nominal parametrization. An example of the procedure can be seen for the case of the jet energy scale uncertainty in Figure 9.2. The black line represents the nominal samples as shown in the previous chapter. The blue and red lines represent systematic curves around the central fit values where the jet energy scale is varied by $\pm 1\sigma$, respectively. Table 9.1 and 9.2 summarizes the effect of all systematic uncertainties discussed previously for the signal samples for gluinos with only light-flavor and light- and heavy-flavor decay products, respectively.

9.3 Sources of systematic uncertainties for $t\bar{t}$ background

In addition to all the uncertainties mentioned in Section 9.1, we also consider uncertainties for the $t\bar{t}$ background. The $t\bar{t}$ MC samples are generated with the MADGRAPH event generator interfaced with PYTHIA for parton the showering and hadronization procedure. Uncertainties associated with the modeling in MADGRAPH and the interface between MADGRAPH and PYTHIA are discussed in more detail here. Parameter settings and their systematic

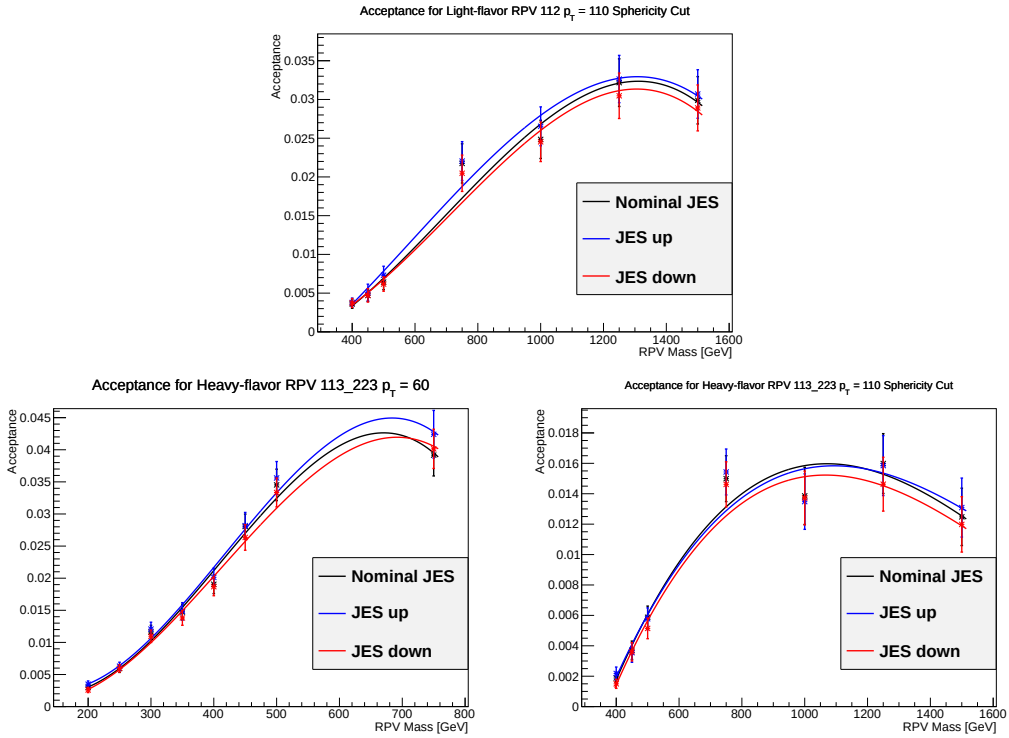

Figure 9.2: Effects of jet energy scale systematic uncertainties on signal $A \times \epsilon$. The top left plot shows the uncertainty on $A \times \epsilon$ for the inclusive search and the bottom two plots show the uncertainties for the heavy-flavor search.

Mass	Gaussian fit parameters		JES	Pile-up	ISR/FSR
[GeV]	Amplitude $\equiv A \times \epsilon$	Width	$A \times \epsilon$	$A \times \epsilon$	$A \times \epsilon$
400	10%	12%	8%	1%	7%
450	10%	12%	12%	4%	6%
500	9%	12%	13%	4%	6%
750	6%	10%	9%	2%	5%
1000	6%	8%	4%	3%	5%
1250	6%	8%	4%	3%	5%
1500	7%	8%	4%	1%	5%

Table 9.1: Systematic uncertainties for the inclusive search. The second and third columns represent uncertainties on the Gaussian fit parameters, where the uncertainty on the amplitude translates into an uncertainty on the $A \times \epsilon$. All other sources of uncertainty, such as on the jet energy scale, pile-up reweighting, and modeling of ISR/FSR, are evaluated as a change on $A \times \epsilon$.

Mass	Gaussian fit parameters		JES	Pile-up	b-tagging SF	ISR/FSR
[GeV]	Amplitude $\equiv A \times \epsilon$	Width	$A \times \epsilon$	$A \times \epsilon$	$A \times \epsilon$	$A \times \epsilon$
200	10%	11%	16%	4%	4%	11%
250	7%	8%	7%	2%	2%	11%
300	6%	6%	3%	2%	1%	11%
350	5%	7%	4%	1%	1%	7%
400	5%	6%	5%	1%	2%	7%
450	4%	5%	5%	1%	3%	7%
500	4%	6%	5%	1%	3%	7%
750	7%	7%	4%	2%	2%	5%
1000	8%	11%	5%	5%	3%	5%
1250	8%	10%	5%	5%	2%	5%
1500	9%	12%	5%	5%	7%	5%

Table 9.2: Systematic uncertainties for the heavy-flavor search. The second and third columns represent uncertainties on the Gaussian fit parameters, where the uncertainty on the amplitude translates into an uncertainty on the $A \times \epsilon$. All other sources of uncertainty, such as on the jet energy scale, pile-up reweighting, b-tagging scale factors, and modeling of ISR/FSR, are evaluated as a change on $A \times \epsilon$.

changes follow one of the CMS $t\bar{t}$ cross section measurements [80].

MADGRAPH/PYTHIA **matching scale uncertainty**

The procedure of interfacing the MADGRAPH and PYTHIA MC generators has been described in more detail in Section 4.2.2. To study the effect of the matching scale parameters on the expected $t\bar{t}$ background we use samples produced with different generator level settings. The values of *xqcut* and Q_{cut} are scaled up and down respectively and the settings can be seen in the list of samples in Table 4.3. Changing these values determines up to what p_T threshold additional jets are produced by PYTHIA, and we study the effect on the shape and normalization of the $t\bar{t}$ M_{jjj} distribution. We do not evaluate this uncertainty for the background QCD multijets sample since these are just used as cross-checks and producing additional samples with these different settings would have been very computing intensive. The $t\bar{t}$ contribution is modeled by the full shape of correctly and incorrectly combined triplets and we compare shapes obtained from the systematic samples to the default one. The effect on the normalization is on the order of 10%, a comparison of the shapes with respect to the nominal triplet invariant mass distribution is shown in bottom plot of Figure 9.3.

Factorization and renormalization scale uncertainty

For $t\bar{t}$ samples produced with MADGRAPH the scale of the hard interaction is set to $Q^2 = m_t^2 + \sum p_T^2$, with m_t being the mass of the top quark mass and the sum including all additional partons from the matrix element calculation. To evaluate the effect of the scale on the analysis the value is multiplied by a factor of 0.5 (scale down) or 2.0 (scale up). A comparison of the two alternative distributions with respect to the nominal sample is shown in the top right of Figure 9.3. The effect on the normalization is on the order of 8% around the $t\bar{t}$ mass peak and on the order of 15% for the distribution of incorrectly combined triplets with masses above 200 GeV.

9.3.1 Summary of uncertainties on the $t\bar{t}$ background

Systematic triplet invariant mass distributions are produced for each source of uncertainty. The full shape information is included when limits are set, as discussed later in Section 9.6, where an extrapolation is performed between the nominal value and the two alternative distributions corresponding to a $\pm 1\sigma$ variation. The three uncertainties with the largest effects are shown in Figure 9.3 for jet energy scale (top left), renormalization and factorization scale (top right), and matching scale (bottom). An additional 5% uncertainty is assigned to account for the uncertainty on the theoretical cross section that is used to determine the expected number of $t\bar{t}$ triplets in the dataset.

9.4 Uncertainties on the QCD multijet background prediction

For all three analysis signal regions the overall normalization of the multijet background component is allowed to float freely and is determined from a fit to the data distribution. In case where the four-parameter function is used to model this background, uncertainties on the fit parameters are taken into account when limits are calculates. More details about the statistical methods for evaluating fit parameters is given in Section 7.1.

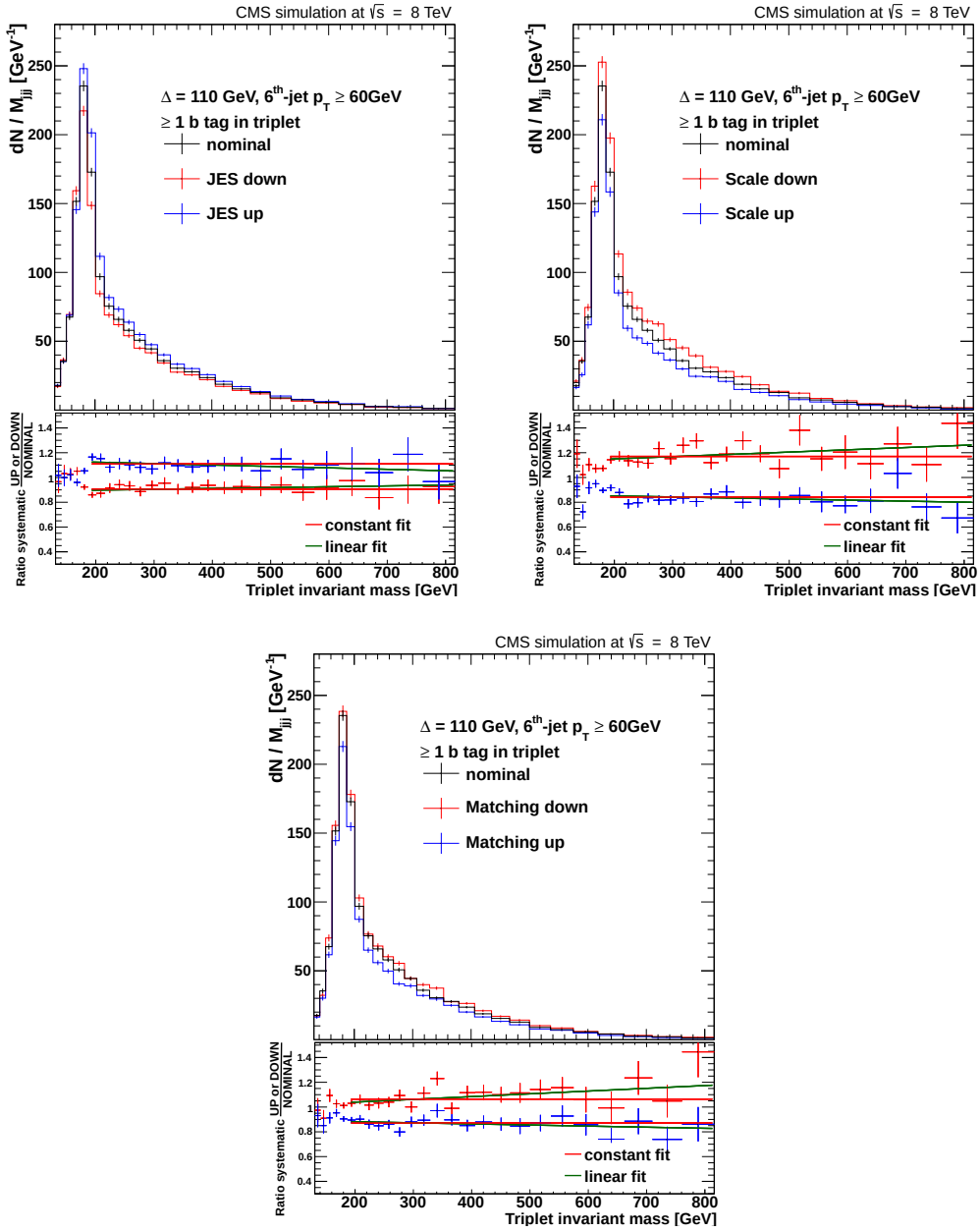

Figure 9.3: Shape uncertainties on the M_{jjj} distribution of simulated $t\bar{t}$ events. Shown are the effects from the uncertainties on the jet energy scale, renormalization scale (top right), and matching scale (bottom). The default distribution is shown in black and the systematic distributions in blue and red.

9.5 Limit setting techniques

In this section we discuss the general theory behind hypothesis testing and limit setting before going into a more detailed description of the method used in the analysis. One needs to compare two different hypotheses with each other and we denote the background-only hypothesis with H_0 and the background + signal hypothesis H_1. When placing a limit on a potential new physics signal one quantifies how incompatible the data is with the H_1 hypothesis. In general there are two philosophies, the Bayesian and Frequentist approaches, when it comes to calculating these limits. In Bayesian statistics, one assigns a "degree of belief" for a parameter of interest by introducing a prior probability, which describes the a-priori knowledge of where that parameter should be. This gives a natural way of introducing systematic uncertainties on the parameters of interest. In the Frequentist approach, probability is interpreted as the frequency of an outcome when the measurement is repeated [9]. We will focus here on a Frequentist approach as discussed in the next section.

9.5.1 Modified Frequentist method - CL_s

In order to ensure a compatible statistical treatment for the CMS and ATLAS Higgs searches and later discovery, a statistical procedure was formulated by the two experiments, and is described in detail in Reference [81]. As mentioned in Section 7.2 we find no significant deviation from the predicted SM background and proceed to set limits. The same approach as used for the Higgs search is adopted for this analysis and we compute exclusion limits based on the modified Frequentist method, also referred to as CL_s [82, 83]. The quantities needed for the calculation are the number of expected signal events (s) multiplied by a signal strength (denoted by r) and the number of expected background events (b), for example in a given bin of the M_{jjj} distribution. Uncertainties are included as so-called nuisance parameters defined by θ, and they affect $s = s(\theta)$ and $b = b(\theta)$, respectively. A *pdf* describing these nuisance parameters $p(\tilde{\theta}, \theta)$ is introduced where $\tilde{\theta}$ reflects what the true value of θ should be. One can then define a likelihood function $L(data|r, \theta)$, where *data*

represents the observation:

$$L(data|r,\theta) = Poisson(data|r \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta). \tag{9.1}$$

For a binned data distribution the *Poisson* function corresponds to a product of Poisson probabilities over all bins M in the histogram with observed events n_i and expected events $r \cdot s_i + b_i$:

$$Poisson = \prod_i^M \frac{(r \cdot s_i + b_i)^{n_i} e^{-(r \cdot s_i + b_i)}}{n_i!}. \tag{9.2}$$

The next step is to define a test statistic $\tilde{q}_r$ in order to compare the data with H_0 and H_1. There is no unique rule on how to define this test statistic, but we follow the "LHC style" prescription. For this prescription profiling is used, which means that the best estimator for the nuisance parameters $\hat{\theta}$ is obtained from the data itself. This leads to the definition of:

$$\tilde{q}_r = -2 \ln \frac{L(data|r,\hat{\theta}_r)}{L(data|\hat{r},\hat{\theta})}, \tag{9.3}$$

where $\hat{r}$ and $\hat{\theta}$ represent the estimators of r and θ that maximize the likelihood in the numerator, and $\hat{\theta}_r$ refers to the estimator of θ for fixed signal strength r given the data. Another condition for the above definition is that $0 < \hat{r} < r$, where the lower bound ensures physical observables (the signal rate is positive) and the upper bound ensures that upward fluctuations $\hat{r} > r$ are not used to rule out the signal hypothesis.

Next we outline the steps to calculate observed and expected limits [81].

Observed limit

1. Calculate $\tilde{q}_r^{obs}$ for the observed data and a fixed signal strength modifier r.

2. Determine values of the nuisance parameters $\hat{\theta}_0^{obs}$ (background-only, H_0 with $r = 0$) and $\hat{\theta}_r^{obs}$ (signal + background H_1) that maximize the likelihood in Equation 9.1.

3. Generate pseudo-data (toys) to construct *pdfs* for the two hypotheses, $f(\tilde{q}_r|r,\hat{\theta}_r^{obs})$ (for fixed r) and $f(\tilde{q}_0|0,\hat{\theta}_0^{obs})$ (for $r = 0$). A generic example for both distributions is shown in Figure 9.4, with $f(\tilde{q}_{r=0}|r = 0,\hat{\theta}_0^{obs})$ in blue and $f(\tilde{q}_{r=1}|r = 1,\hat{\theta}_{r=1}^{obs})$ in red,

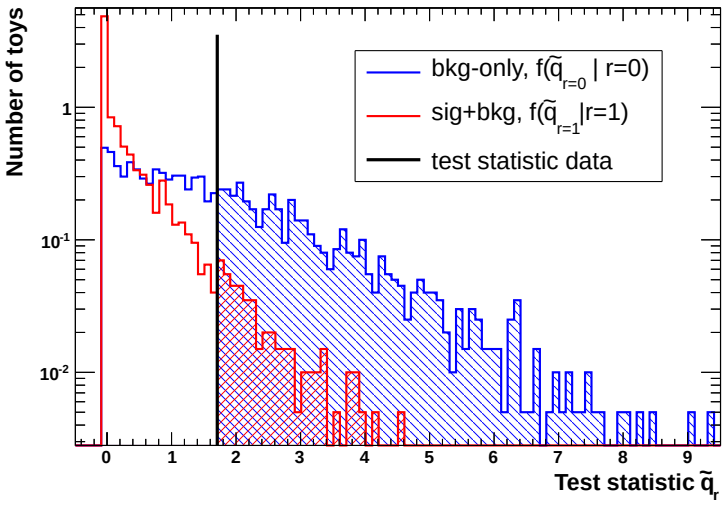

Figure 9.4: Distributions for the test statistic for the background-only hypothesis $f(\tilde{q}_0|0,\hat{\theta}_0^{obs})$ in blue and the signal+background hypothesis $f(\tilde{q}_r|r,\hat{\theta}_r^{obs})$ in red based on 2000 toys each. The value of $\hat{q}_r^{obs}$ for the observed data is shown with a black line. The shaded red and blue areas represent the probabilities p_r and $1-p_b$, respectively.

the value determined in the first step for $\tilde{q}_r^{obs}$ is shown as a black line.

4. Define two p-values as integrals of the aforementioned *pdfs*:

$$\begin{aligned} p_r &= P(\tilde{q}_r \geq \tilde{q}_r^{obs}|H_1) = \int_{\tilde{q}_r^{obs}}^{\inf} f(\tilde{q}_r|r,\hat{\theta}_r^{obs})d\tilde{q}_r, \\ 1-p_b &= P(\tilde{q}_r \geq \tilde{q}_r^{obs}|H_0) = \int_{\tilde{q}_0^{obs}}^{\inf} f(\tilde{q}_0|0,\hat{\theta}_0^{obs})d\tilde{q}_r. \end{aligned} \tag{9.4}$$

5. Calculate CL_s as the ratio for the value of r that is being tested using

$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} = \frac{p_r}{1-p_b}. \tag{9.5}$$

6. If for $r=1$, $\mathrm{CL}_s < \alpha$ the signal model is excluded at the $(1-\alpha)\%$ confidence level (C.L.), where α is conventionally chosen to be 0.05. For this analysis we quote a 95% C.L. limit on the theory cross section times branching ratio ($\sigma\times$ BR), which corresponds to a $r^{95\%}$ multiplied by the respective $\sigma\times$ BR where $\mathrm{CL}_s = 0.05$.

Expected limits

The expected sensitivity to the new physics model is determined from the background-only hypothesis. Pseudo-data is generated based on the expected background, which is then treated as if it had been the observed data. In this context one can simply follow the same steps as outlined in Section 9.5.1 and calculate CL_s values based on the prescription given. The value of $r^{95\%}$ is determined for each of the background-only pseudo-datasets and a cumulative distribution of these $r^{95\%}$ values is generated. The median of the expected limit is then defined where this distribution crosses the 50% quantile. The $\pm 1\sigma \equiv 68\%$ and $\pm 2\sigma \equiv 95\%$ uncertainty bands on the expected limit are determined when the distribution crosses the 16%/84% and the 2.5%/95% quantile, respectively.

Asymptotic description of the test statistic

The method of computing *pdfs* for the test statistic of the two hypotheses can become very computationally intensive with increasing number of toys. The *pdfs* for the profile likelihood test statistic as defined in equation 9.3 can be approximated by an asymptotic formula based on Wilks and Wald theorems as derived in [84]. The *pdfs* of the test statistics for H_1 is given by:

$$f(\tilde{q}_r|r) = \frac{1}{2}\delta(\tilde{q}_r) + \begin{cases} \frac{1}{2\sqrt{2\pi\tilde{q}_r}} \exp\left[-\tilde{q}_r/2\right] & 0 < \tilde{q}_r \leq r^2/\sigma^2 \\ \frac{1}{\sqrt{2\pi}(2r/\sigma)} \exp\left[-\frac{1}{2}\frac{(\tilde{q}_r + r^2/\sigma^2)^2}{(2r/\sigma)^2}\right] & \tilde{q}_r > r^2/\sigma^2 \end{cases}, \qquad (9.6)$$

where $\delta(\tilde{q}_r)$ is a delta function at zero, and $\sigma^2 = \frac{r^2}{\tilde{q}_{r,A}}$ represents the standard deviation on r and is determined from the so-called "Asimov dataset". The Asimov dataset represents a dataset where all "observed" quantities are set to their expectation values and the procedure of calculating the test statistic is repeated as if it was real data. Similar to Equation 9.6 one can define an asymptotic formula for the background-only hypothesis, which can be used to derive the expected limit and the uncertainty bands based on the Asimov dataset. Since this method does not generate toy experiments it is much less CPU intensive than the previous method to extract expected limits. For this analysis we use this asymptotic approximation to set limits. A cross-check performed with the full CL_s method generating

toys gives very similar results.

9.6 Limit setting procedure for a "bump hunt"

In this analysis we are examining the three-jet invariant mass distribution for a Gaussian signal peak on top of a smoothly falling background distribution. As mentioned in Section 7.2, where the background estimates are described, we do not find any significant deviation from the expected background distribution. The three different search regions are summarized in Table 6.3 and the background model is either a parametrized function for the QCD multijet events or a binned background shape based on the QCD estimate from a data control region and simulated $t\bar{t}$ events. We use the RooFit package to perform binned maximum likelihood fits of the expected background to the three-jet invariant mass distribution in data. The normalization component of the multijets background is allowed to float freely and is determined from the fit to the data itself. Uncertainties on the triplet invariant mass distribution of $t\bar{t}$ are determined as described in Section 9.3.1 and incorporated through a quadratical extrapolation of the nominal shape for shifts below 1σ and a linear extrapolation beyond 1σ. Uncertainties effecting the overall normalization of the $t\bar{t}$ background from the systematic shapes are included with a log-normal constraint. The log-normal function is chosen over a Gaussian description of these uncertainties since it avoids the problem of unphysical parameters, for example the cross section should be a non-negative value. When the four-parameter function is used to model the background (high-mass, heavy- and light-flavor search regions) the uncertainties on the original background parameters are included and inflated by a factor of four. These increased uncertainties ensure that fit parameters are not overly constrained, since profiling is used to find the best estimators for the nuisance parameters.

The signal is modeled by a Gaussian with a width shown in Figure 8.5. The expected number of signal triplets (the signal rate s) is given by:

$$s = \sigma^{theory} \times A \times \epsilon \int \mathcal{L}, \tag{9.7}$$

where σ^{theory} is the theory cross section from Table 4.1, $A \times \epsilon$ as defined in Section 8, and

the integrated luminosity is $\int \mathcal{L} = 19.4$ fb^{-1}. Systematic uncertainties on $A \times \epsilon$ as described in Section 9.2.1 are incorporated as nuisance parameters with log-normal constraints. In order to incorporate the effect of the uncertainty on the jet energy resolution we include a 10% uncertainty on the Gaussian width in addition to the uncertainty assigned based on the signal fits, given in the third column of Tables 9.1 and 9.2 . The total uncertainty on the width is shown by the green shaded band in Figure 8.5.

We estimate the upper limit on the gluino pair production cross section by scanning masses between 350–1500 GeV for the light-flavor search and 200-1500 GeV for the heavy-flavor search in 50 GeV steps. The CL$_s$ profile likelihood calculator with the asymptotic approximation evaluates the observed limit as well as the expected limit with a $\pm 1\sigma$ and $\pm 2\sigma$ bands. A tool referred to as "CombinedLimit", which was developed by the CMS collaboration, is used to perform the limit calculations. All previously discussed parameters are organized in a so-called data-card which is then input to the CombinedLimit tool for processing.

9.6.1 Signal injection test

For the limit setting procedure the background estimate is determined from a background-only fit of Equation 7.1 to the data, while when we search for signal a $s + b$ fit is performed described by the sum of Equation 7.1 and a Gaussian component for the signal. In order to show the stability of the statistical method, we perform a so-called signal injection test outlined in the steps below.

1. Pseudo-data is generated with a fixed signal injected. We choose for this test a signal mass of 750 GeV with an expected signal rate multiplied by the input signal strength of $r^{in} = 3$. One of these pseudo-datasets is shown in the left upper plot in Figure 9.5

2. A background-only binned maximum likelihood fit is performed to the pseudo-data generated in step 1 and the corresponding background parameters are extracted (upper left plot in Figure 9.5).

3. Then the same histogram is fit with the $s+b$ function, using the previously determined background parameters as input. These parameters are constrained the same way as

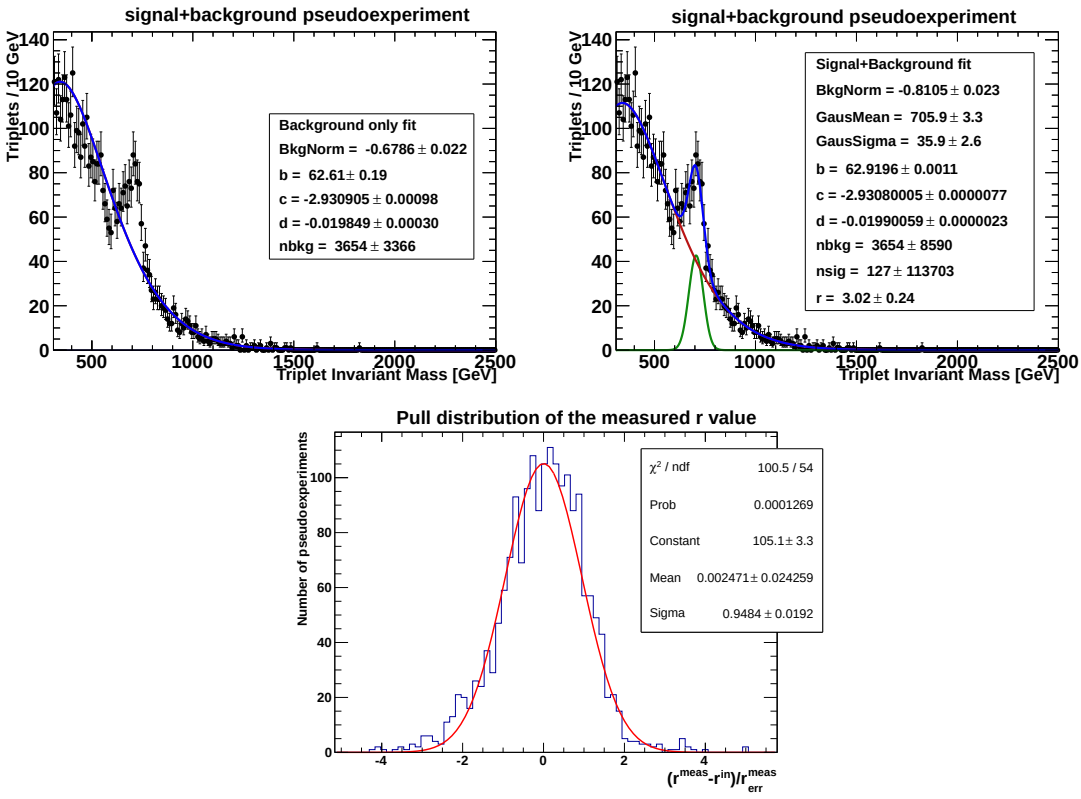

Figure 9.5: Signal injection test. Pseudo-data is generated based on the $s + b$ hypothesis. The first plot on the top left shows the background-only fit in blue to the distribution of one of they pseudo-datasets. On the right hand side the same distribution is fit with the $s + b$ hypothesis, with the background component in blue and the Gaussian signal in green. The bottom plot shows the pull distribution of the measured signal strength evaluated with respect to the injected signal strength divided by the error on the measured value for 2000 pseudo-datasets.

it is done in the limit setting procedure. The fit result is shown in the upper right plot in Figure 9.5.

4. For each pseudo-dataset we evaluate the pull on the measured signal strength r^{meas}, where the pull is defined as $(r^{meas} - r^{in})/r^{meas}_{err}$, as can be seen in the bottom plot in Figure 9.5 for 2000 pseudo-datasets. This distribution is a Gaussian centered around zero with a width of one, showing that there is no under-coverage of the extracted signal strength, and we are able to successfully extract the input signal strength.

Chapter 10

Results

The statistical methods used in this analysis have been discussed in detail in Section 9.5.1. In this chapter we discuss the experimental results and the limits placed on the pair production cross section of new hadronic resonances.

10.1 Limit on hadronic RPV

Upper limits at the 95% C.L. are placed on cross section times branching ratio $\sigma \times BR(X \to jjj)$ of a new unknown particle X decaying into three jets. The specific interpretation here assumes gluino pair production, where each gluino decays 100% into three jets as shown in the Feynman diagram in Figure 2.6. Figure 10.1 shows the expected (black dashed line) and observed limits (black solid line) as a function of the triplet invariant mass for the assumed model of gluinos with light-flavor jets in the decay. Figure 10.2 shows these limits for the heavy-flavor search. The green and yellow shaded areas represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands on the expected limit. For both cases the theory cross section line in red includes $\pm 1\sigma$ uncertainties from Table 4.2 as dashed red line. As a conservative estimate, we exclude gluinos with masses below the point where the -1σ theory line intersects with the observed limit. A tabular form of these plots are shown in Table B.1 and B.2 in Appendix B.

We exclude the production of gluinos undergoing RPV decays through the coupling λ''_{112} into light-flavor jets at 95% C.L. for gluino masses below 650 GeV, with an expected exclusion limit below 755 GeV. Previous exclusion limits for this scenario were summarized in Table 2.3, and we are able to extend the last CMS limit by almost 200 GeV. Gluinos that decay through the RPV couplings λ''_{113} or λ''_{223} into one heavy-flavor jet and two light-flavor jets are excluded for masses between 200 and 835 GeV, which is the first mass limit on

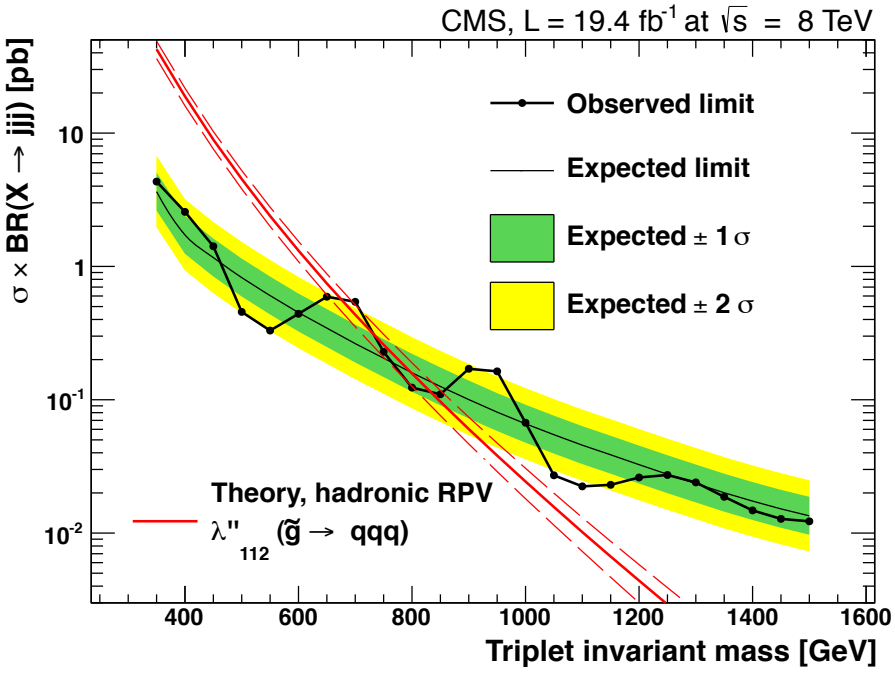

Figure 10.1: Observed and expected frequentist CLs cross section limits for λ''_{112}. Shown are the observed (solid black line) and expected (black dashed line) 95% C.L. cross section limits as a function of the triplet invariant mass for the inclusive search. The $\pm 1\sigma$ and $\pm 2\sigma$ bands around the expected limit are represented by green and yellow shaded areas, respectively. The theory cross section is represented by the solid red line and its uncertainties are shown by the dashed red line [21].

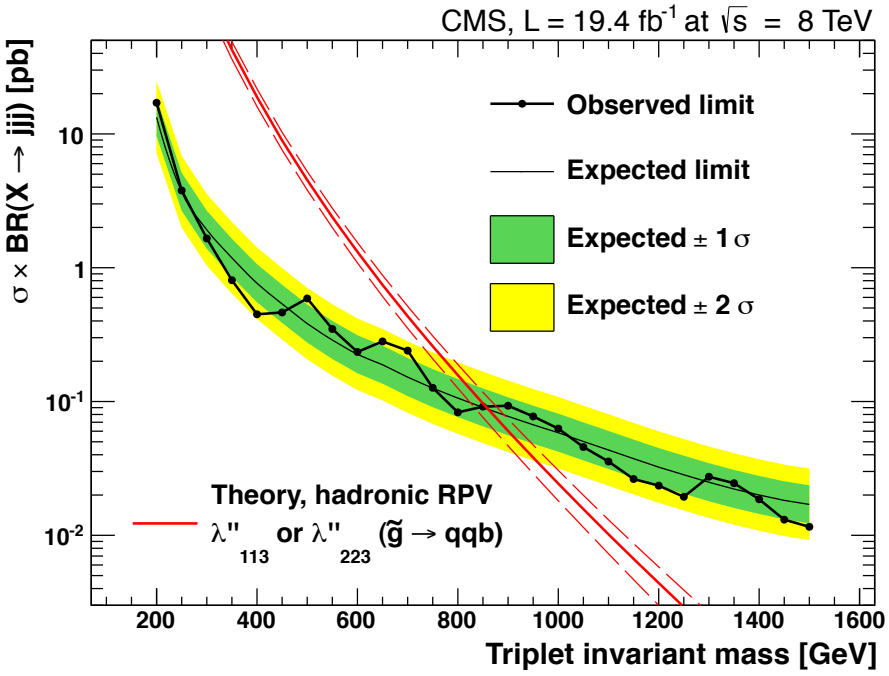

Figure 10.2: Observed and expected frequentist CLs cross section limits for λ''_{113} or λ''_{223} Shown are the observed (solid black line) and expected (black dashed line) 95% C.L. cross section limits as a function of the triplet invariant mass for the heavy-flavor search. The $\pm 1\sigma$ and $\pm 2\sigma$ bands around the expected limit are represented by green and yellow shaded areas, respectively. The theory cross section is represented by the solid red line and its uncertainties are shown by the dashed red line [21].

this model of RPV gluino decays, with the expected exclusion at 825 GeV. The sensitivity increase for the heavy-flavor search is due to the reduction of background with respect to the signal. Therefore, while this search probes a smaller part of phase space it is more sensitive to the RPV couplings allowing heavy-flavor quarks in the final state. A summary of the expected and observed limits is given in Table 10.1.

Mass limit	RPV coupling	
	λ''_{112}	λ''_{113} or λ''_{223}
observed (-1σ theory)	650 GeV	200-835 GeV
observed (central theory)	670 GeV	200-855 GeV
expected (-1σ theory)	755 GeV	200-825 GeV
expected (central theory)	795 GeV	200–860 GeV

Table 10.1: Observed and expected mass limits quoted at the -1σ or central theory cross section intersection point with the expected and observed limits.

10.2 Extra material: Black hole limits applied to three-jet resonance search

With increasing resonance mass, new heavy resonances would mostly be produced at rest and therefore, have little boost. We introduced the use of event shape variables in Section 6.4.4, which helps to distinguish background from high-mass signal events. The sphericity variable as shown in Figure 6.15 provides clear separation between QCD events and gluinos with a mass above 750 GeV. The signal events become more isotropic and therefore also similar to the signatures that is assumed for the CMS analysis searching for black hole production, based on 12 fb^{-1} of proton-proton collision at 8 TeV [85]. This analysis is performed as a search for new heavy objects that decay into final states with high object multiplicity N (jets, electrons, muons, taus, photons, and E_T^{miss}) and a large scalar sum p_T of all these objects S_T. Model-independent limits on cross section times acceptance ($\sigma \times A$) are presented in [85], depending on the object multiplicity and a minimum threshold value of S_T^{min}. The value of S_T is defined as $S_T = \sum_i p_T^i + E_T^{miss}$, where the sum runs over all N reconstructed objects with $p_T \geq 50$ GeV, and E_T^{miss} is only added if its value is above 50 GeV. In order to model these selection criteria we select events in the gluino sample

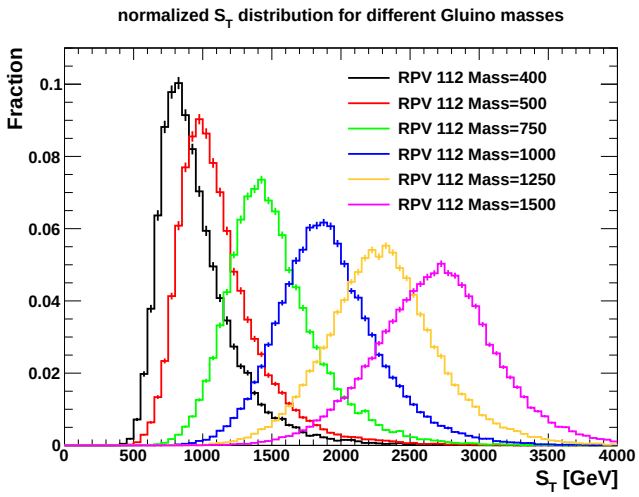

Figure 10.3: S_T distributions for different gluino masses normalized to unity.

with at least six jets with $p_T \geq 50$ GeV and calculate S_T for each event as defined above. The S_T distribution for several different signal masses of gluinos with light-flavor decays is shown in Figure 10.3, where each distribution is normalized to unity to show the trend of an increasing p_T sum with increasing mass. From the bottom right plot of Figure 6 ($N \geq 6$) in [85], we extract the expected limit at the 95% C.L. of $\sigma \times A$ for a given S_T^{min} value, these values are summarized int Table 10.2. In order to compare with the gluino model we

S_T^{min}	$\sigma \times A$
2200 GeV	1.5×10^{-2}
3000 GeV	2×10^{-3}
3600 GeV	7×10^{-4}

Table 10.2: Expected limit on $\sigma \times A$ for a given S_T^{min} value.

evaluate $\sigma \times A$ based on the the number of events passing one of the three minimum S_T requirements from Table 10.2. This value is divided by the number of events generated and multiplied by the theory cross section, which allows a direct comparison with the expected limits of the black hole search. The resulting $\sigma \times A$ as a function of S_T^{min} is shown in Figure 10.4 (left). If a point lies above the gray dashed line (values from Table 10.2), then the black hole limits would have excluded this gluino mass already. Figure 10.4 (right)

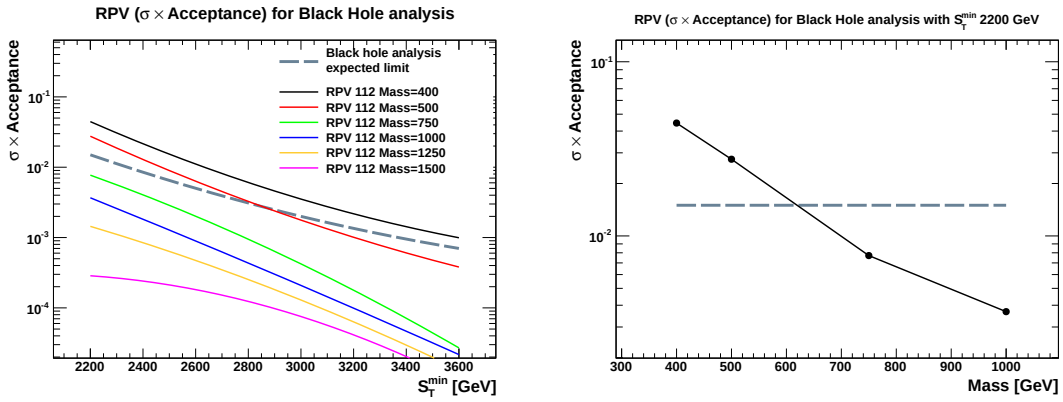

Figure 10.4: The plot on the left shows $\sigma \times A$ for the gluino model with light-flavor jets int he decay as a function of S_T^{min}, if the point lies above the gray dashed line the Black Hole analysis could exclude this gluino mass. The expected exclusion would be around 610 GeV as shown at bottom, where the limit is shown as a function of mass.

shows $\sigma \times A$ for the black hole analysis as a function of mass, so we can deduce an expected mass exclusion limit around 600 GeV. This approach is less sensitive than the expected exclusion obtained from the analysis presented in this thesis. The method of reconstructing the three-jet resonances and looking for a localized deviation in the form of a Gaussian distribution on top of a smoothly falling background, leads to an expected mass exclusion of 795 GeV.

10.3 Conclusion

A search for new hadronic three-jet resonances has been presented using 19.4 fb^{-1} of pp collisions collected with the CMS detector at $\sqrt{s}$ of 8 TeV. The jet ensemble technique was used in order to reconstruct the pair-produced resonances each decaying into three jets. While the technique is applicable to any pair-produced three-jet resonance, the experimental results were interpreted in the context of a hadronic RPV SUSY model, where gluinos are pair-produced and decay 100% of the time into three jets. Two different coupling scenarios were considered allowing gluino decays into only light-flavor jets (λ''_{112}) and gluino decays into one heavy-flavor and two light-flavor jets (λ''_{113} or λ''_{223}).

The inclusive search region covers masses between 400 – 1500 GeV and does not apply selection criteria in order to determine the flavor content of the final state. For this scenario, gluinos decaying through the RPV coupling λ''_{112} were excluded at the 95% C.L. below a mass of 650 GeV.

The heavy-flavor search is the first of its kind. The search is split up into two regions, low-mass (200 – 600 GeV) and high-mass (600 – 1500 GeV), and makes use of b-jet identification techniques in order to probe a different set of RPV couplings. In this search gluinos decaying through the RPV couplings λ''_{113} and λ''_{223} were excluded at the 95% C.L. between 200 – 835 GeV. While b-jet identification reduces the amount of probed phase space, it allows us to validate the analysis technique with the known SM top pair production, where each of the tops decays into three jets. A dedicated study extracting the $t\bar{t}$ cross section is discussed in Appendix C.

Appendix A

PYTHIA cards for gluino pair production with RPV decay

We use the PYTHIA event generator with the RPV violating UUD matrix element turned on via the RVLAMB(1,1,2) parameter to generate the gluinos with only light-flavor quarks in the decay or the RVLAMB(1,1,3) and RVLAMB(2,2,3) parameters to allow heavy-flavor quarks. In this section, we list the PYTHIA card used to generate a sample of pair-produced gluinos with a mass of 200 GeV, where the squark masses are set to 2 TeV. For gluino masses above 750 GeV the squark masses are increased to 7.5 GeV to ensure prompt decays of the gluino.

```
MSEL     = 39          ! turn on SUSY processes
IMSS( 1) = 1           ! generic SUSY scenario
IMSS( 3) = 1           !
IMSS(51) = 0           !  RPV LLE off
IMSS(52) = 0           !  RPV LQD off
IMSS(53) = 3           !  RPV UDD on with user specified couplings

------------------------------ Scenario 1 -----------------------------
RVLAMB(1,1,2) = 0.005  !  light-flavor coupling

------------------------------ Scenario 2 -----------------------------
RVLAMB(1,1,3) = 0.005  !  heavy-flavor coupling
RVLAMB(2,2,3) = 0.005  !  heavy-flavor coupling
-----------------------------------------------------------------------
RMSS( 1) =   10000.0   ! bino
RMSS( 2) =   10000.0   ! wino
RMSS( 4) =   10000.0   ! mu
RMSS( 3) =   200.      ! gluino
RMSS( 5) =   5.00      ! tan beta
RMSS( 8) =   2000.0    ! left squark (1st-2nd generation)
RMSS( 9) =   2000.0    ! right down squark (1st-2nd generation)
RMSS(10) =   2000.0    ! left squark (3rd generation)
RMSS(11) =   2000.0    ! right down squark (3rd generation)
RMSS(12) =   2000.0    ! right up squark (3rd generation)
RMSS( 6) =   10000.0   ! left slepton (1st-2nd generation)
RMSS( 7) =   10000.0   ! right slepton (1st-2nd generation)
RMSS(13) =   10000.0   ! left slepton (3rd generation)
RMSS(14) =   10000.0   ! right slepton (3rd generation)
RMSS(15) =   2.4       ! bottom trilinear
RMSS(16) =  -3.3       ! top trilinear
RMSS(17) =   1.0       ! tau trilinear
RMSS(18) =  -0.1020    ! Higgs mixing angle alpha
RMSS(19) =   10000.0   ! pseudo-scalar Higgs mass
```

Appendix B
Tables for limits on gluino pair production with RPV decay

B.1 Hadronic RPV λ_{112}

Gluino Mass [GeV]	observed limit [pb]	expected limit [pb]	-1σ [pb]	$+1\sigma$ [pb]	-2σ [pb]	$+2\sigma$ [pb]
350	4.3276	3.6373	2.6249	5.0522	1.9734	6.7125
400	2.5654	1.7232	1.2436	2.3935	0.9349	3.1801
450	1.4154	1.1646	0.8404	1.6176	0.6318	2.1491
500	0.4558	0.8241	0.5947	1.1447	0.4471	1.5209
550	0.3303	0.6042	0.4360	0.8392	0.3278	1.1150
600	0.4412	0.4505	0.3251	0.6258	0.2444	0.8314
650	0.5924	0.3446	0.2487	0.4787	0.1870	0.6360
700	0.5426	0.2636	0.1902	0.3661	0.1430	0.4865
750	0.2286	0.2051	0.1480	0.2848	0.1113	0.3784
800	0.1233	0.1592	0.1149	0.2212	0.0864	0.2939
850	0.1099	0.1258	0.0908	0.1747	0.0682	0.2321
900	0.1707	0.1003	0.0724	0.1393	0.0544	0.1851
950	0.1633	0.0807	0.0583	0.1121	0.0438	0.1490
1000	0.0672	0.0660	0.0476	0.0917	0.0358	0.1218
1050	0.0271	0.0547	0.0395	0.0760	0.0297	0.1009
1100	0.0224	0.0456	0.0329	0.0634	0.0248	0.0842
1150	0.0230	0.0386	0.0279	0.0537	0.0210	0.0713
1200	0.0261	0.0326	0.0235	0.0453	0.0177	0.0602
1250	0.0273	0.0276	0.0199	0.0383	0.0150	0.0509
1300	0.0240	0.0235	0.0170	0.0327	0.0128	0.0434
1350	0.0187	0.0201	0.0145	0.0279	0.0109	0.0370
1400	0.0148	0.0174	0.0125	0.0241	0.0094	0.0321
1450	0.0128	0.0152	0.0110	0.0211	0.0082	0.0280
1500	0.0122	0.0135	0.0097	0.0187	0.0073	0.0249

Table B.1: The observed and expected limits including uncertainties on the gluino pair production cross section at the 95% CL for the inclusive search and gluino masses up to 1500 GeV corresponding to Figure 10.1.

B.2 Hadronic RPV λ_{113} or λ_{223}

Gluino Mass	observed limit	expected limit	-1σ	$+1\sigma$	-2σ	$+2\sigma$
[GeV]	[pb]	[pb]	[pb]	[pb]	[pb]	[pb]
200	17.0946	13.3167	9.6104	18.4970	7.2250	24.5757
250	3.7787	3.6902	2.6631	5.1257	2.0021	6.8101
300	1.6560	1.9199	1.3855	2.6667	1.0416	3.5431
350	0.8063	1.1847	0.8550	1.6456	0.6428	2.1863
400	0.4487	0.7648	0.5520	1.0624	0.4150	1.4115
450	0.4630	0.5365	0.3872	0.7452	0.2911	0.9901
500	0.5893	0.3822	0.2758	0.5309	0.2074	0.7054
550	0.3485	0.2881	0.2079	0.4002	0.1563	0.5317
600	0.2346	0.2253	0.1626	0.3129	0.1222	0.4157
650	0.2813	0.1883	0.1359	0.2615	0.1022	0.3475
700	0.2401	0.1521	0.1098	0.2113	0.0825	0.2808
750	0.1266	0.1258	0.0908	0.1747	0.0682	0.2321
800	0.0830	0.1061	0.0765	0.1473	0.0575	0.1957
850	0.0916	0.0903	0.0651	0.1254	0.0490	0.1666
900	0.0930	0.0775	0.0559	0.1076	0.0420	0.1430
950	0.0774	0.0670	0.0484	0.0931	0.0364	0.1237
1000	0.0628	0.0584	0.0422	0.0811	0.0317	0.1078
1050	0.0457	0.0503	0.0363	0.0698	0.0273	0.0928
1100	0.0356	0.0434	0.0313	0.0603	0.0235	0.0801
1150	0.0263	0.0374	0.0270	0.0519	0.0203	0.0690
1200	0.0235	0.0323	0.0233	0.0449	0.0175	0.0596
1250	0.0194	0.0283	0.0205	0.0394	0.0154	0.0523
1300	0.0274	0.0249	0.0180	0.0346	0.0135	0.0460
1350	0.0245	0.0222	0.0160	0.0308	0.0120	0.0409
1400	0.0186	0.0200	0.0144	0.0278	0.0109	0.0369
1450	0.0131	0.0182	0.0132	0.0253	0.0099	0.0337
1500	0.0116	0.0170	0.0123	0.0236	0.0092	0.0314

Table B.2: The observed and expected limits including uncertainties on the gluino pair production cross section at the 95% CL for the heavy-flavor search and gluino masses up to 1500 GeV corresponding to Figure 10.2.

Appendix C
Cross check: $t\bar{t}$ cross section measurement

In this appendix we discuss an additional validation of the analysis technique by extracting the $t\bar{t}$ production cross section, which is a known SM process. For this study we treat $t\bar{t}$ as signal instead of background in the heavy-flavor search. Two different methods are presented using either the complete triplet invariant mass distribution or only the Gaussian approximation to extract the cross section. Both methods yield similar results within uncertainties and are in good agreement with the SM theory cross section prediction. The same event and triplet selection is applied as for the low-mass heavy-flavor gluino search.

C.1 Full $t\bar{t}$ shape for signal extraction

All-hadronic $t\bar{t}$ ($t\bar{t} \to Wb{+}Wb \to jjb{+}jjb$) events account for a significant fraction of the background in the heavy-flavor search. By requiring at least one b-tag in each triplet we reduce the amount of QCD multijet background and enhance the visibility of $t\bar{t}$. Instead of treating these triplets as background we use the data-driven QCD multijet estimate and extract or measure the $t\bar{t}$ cross section. This multijet background is taken from a b-jet control region in the data, where b jets are vetoed. With this event selection there is still a small contribution of $t\bar{t}$ in the data, which is subtracted based on MC estimates and the theoretical cross section of $245.8\ ^{+8.7}_{-10.5}$ pb [52]. We use the M_{jjj} distribution of the resulting QCD multijet and the $t\bar{t}$ MC as templates to fit the M_{jjj} distribution in data with the binned maximum likelihood method. In this fit the background and $t\bar{t}$ normalization are allowed to float freely and the number of signal triplets $N_{t\bar{t}}^{trip}$ its extracted from the fit. We choose the range of the fit to be within 120 – 600 GeV enclosing the most relevant region for the $t\bar{t}$ signal. The result of the fit is shown in Figure C.1 and leads a cross section of:

$$\sigma_{meas}^{t\bar{t}} = \frac{N_{t\bar{t}}^{trip}}{\int L \times A \times f2} = 205 \pm 6 \text{ (stat. only) pb}, \tag{C.1}$$

where the $A \times f2 \equiv f1 \times f2$ is calculated for the full triplet invariant mass distribution from the event acceptance times the average number of triplets passing the Δ requirement, which is found to be $f1 \times f2 = 0.432\%$, $N_{t\bar{t}}^{trip}$ is the number of triplets as measured by the fit, and $\int L$ the total integrated luminosity of the dataset.

To evaluate the effect of systematic uncertainties on the measured $t\bar{t}$ cross section we use dedicated signal shapes for each uncertainty as discussed in Section 9.3.1. The following sources of uncertainty are considered: jet energy scale, matching scale, factorization and renormalization scale, pile-up, b-tagging scale factors, and jet energy resolution. The output fits to the data using these systematic templates are shown in Figure C.2. Table C.1 shows fit output and $A \times f2$ for each systematic sample. The effect on the measured cross section is listed in the last column of the table. Combining all uncertainties we obtain a measured value of the $t\bar{t}$ cross section $\sigma^{t\bar{t}}$, where the uncertainties are symmetrized:

$$\sigma_{full}^{t\bar{t}} = 205 \ \pm \ 6 \text{ (stat.)} \ \pm \ 26 \text{ (syst.)} \ \pm \ 5 \text{ (lumi.) pb}. \tag{C.2}$$

This measured value agrees well within uncertainties with the theory prediction of 245.8 pb.

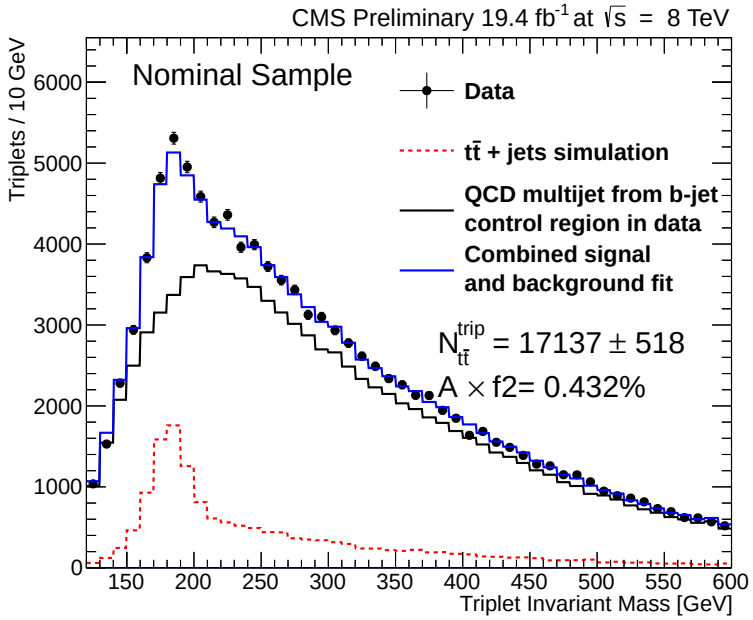

Figure C.1: Result of the maximum binned likelihood fit used to extract the $t\bar{t}$ cross section based on the $t\bar{t}$ template from MC and the QCD multijets template from the b-jet control region in data.

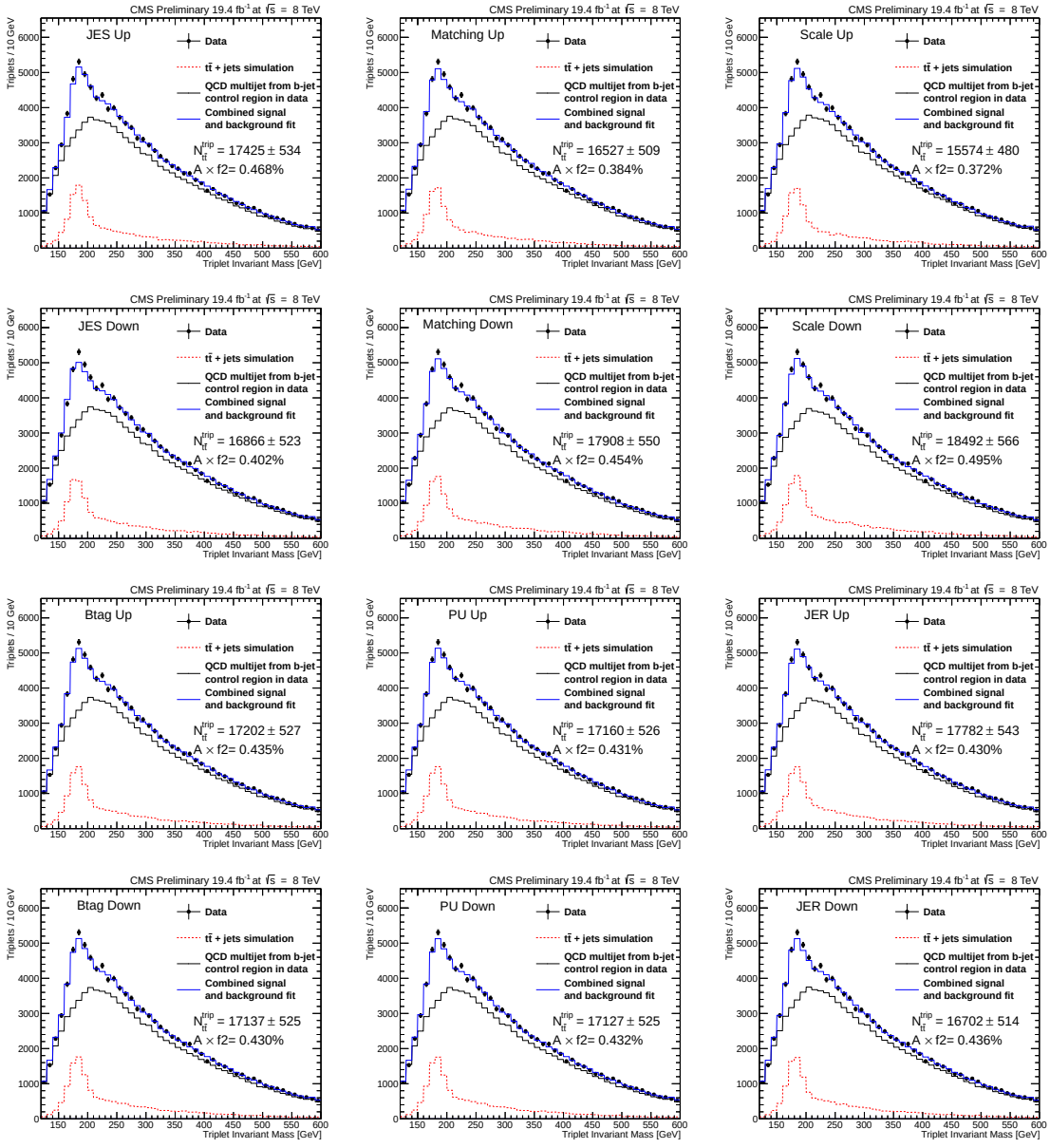

Figure C.2: Results of fits to the data for each systematic uncertainty considered. These fits should be compared to the nominal result from Figure C.1. We considered the following uncertainties: jet energy scale, matching scale, factorization and renormalization scale, pile-up, b-tagging scale factors, and jet energy resolution.

Luminosity 19.4 fb^{-1}		$N_{t\bar{t}}^{trip}$	$A \times f2$	$\sigma_{full}^{t\bar{t}}[pb]$	
Measured values		17137	0.432%	205.2	
Uncertainty		$N_{t\bar{t}}^{trip}$	$A \times f2$	$\sigma_{full}^{t\bar{t}}$ [pb]	$\delta_{t\bar{t}}$ [pb]
JES	Up	17425	0.468%	192.6	-12.7
	Down	16866	0.402%	217.0	11.8
Matching	Up	16528	0.384%	222.3	17.1
	Down	17909	0.454%	203.7	-1.5
Scale	Up	15575	0.372%	216.4	11.2
	Down	18493	0.495%	193.1	-12.2
b-tagging	Up	17203	0.435%	204.4	-0.8
	Down	17138	0.430%	206.1	0.8
pile-up	Up	17161	0.431%	205.6	0.4
	Down	17127	0.432%	205.0	-0.2
JER	Up	17782	0.430%	213.82	8.6
	Down	16702	0.436%	197.93	-7.3
Luminosity 2.6%					± 5.3
Statistical					± 6.5
Total	205 ± 6(stat.) ± 26 (syst.)±5 (lumi.)				

Table C.1: Summary of the effect of different systematic uncertainties on the measured $t\bar{t}$ cross section. The top part of the table shows the central measured cross section value. The values in the fifth column shows the measured cross section when the systematic templates are used. Column six denoted with $\delta_{t\bar{t}}$ represents the difference between the nominal value and the systematic value for a given uncertainty.

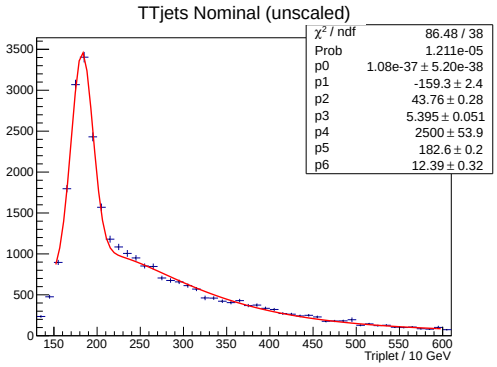

Figure C.3: Result of the fit to the simulated $t\bar{t}$ M_{jjj} distribution to estimate the yield in the Gaussian signal peak. The red line represents the sum of the four-parameter function and a Gaussian.

C.2 Gaussian approximation for signal extraction

For the gluino search the new heavy resonances signal is parametrized as a Gaussian peak. In this section we discuss the effectiveness of the Gaussian approximation by demonstrating its impact on the measurement of the $t\bar{t}$ cross section. First, we evaluate the expected yield with the Gaussian peak from the $t\bar{t}$ M_{jjj} distribution by fitting it with a four-parameter function plus a Gaussian, as is shown in Figure C.3.

Based on this Gaussian fit we extract the width to model the $t\bar{t}$ signal shape and the $A \times \epsilon = f1 \times f2 \times f3$ is determined as previously described in Section 8.1. All systematic uncertainties discussed in the previous section are also taken into account as a change on $A \times \epsilon$. Each distribution is fit with the same function as the nominal one, the results jet energy scale, matching scale, and factorization and renormalization scale are shown in Figure C.4, and for pile-up, b-tagging scale factors, and jet energy resolution in Figure C.5. The jet energy resolution effect is evaluated as a change of the width of the signal Gaussian. The kinematics of the M_{jjj} distribution for the incorrect combination of triplets in the $t\bar{t}$ MC and the QCD multijet background results in them peaking in the same place, as shown in the previous section. Therefore, we use the QCD background estimate from the b-jet control region in data and let it float freely in the signal+background fit, allowing it to cover also the combinatorial background from the $t\bar{t}$ triplets. The result of the combined binned maximum likelihood fit of the multijet background template and the assumed Gaussian signal shape is shown in Figure C.6 (top). The width of the Gaussian is fixed to that observed in the simulation and the normalization and mean position are allowed to float. To account for a change in the width due to jet energy resolution the fit is repeated twice with for the $\pm 1\sigma$ variation on the width, shown in in Figure C.3 (bottom). The effect of all other systematic uncertainties are summarized in Table C.2. Using the Gaussian signal approximation we measure the $t\bar{t}$ cross section as:

$$\sigma^{t\bar{t}}_{Gauss} = 236 \ \pm \ 7 \text{ (stat.)} \ \pm \ 41 \text{ (syst.)} \ \pm \ 6 \text{ (lumi.) pb.} \tag{C.3}$$

The uncertainties for this measurement are larger than when using the full shape in the previous section since we are only evaluating the effect of systematic uncertainties on $A \times \epsilon$ around the mass peak of the triplet invariant mass distribution, instead of including the information on the full spectrum. The measured value agrees well within uncertainties with the theory prediction of 245.8 pb.

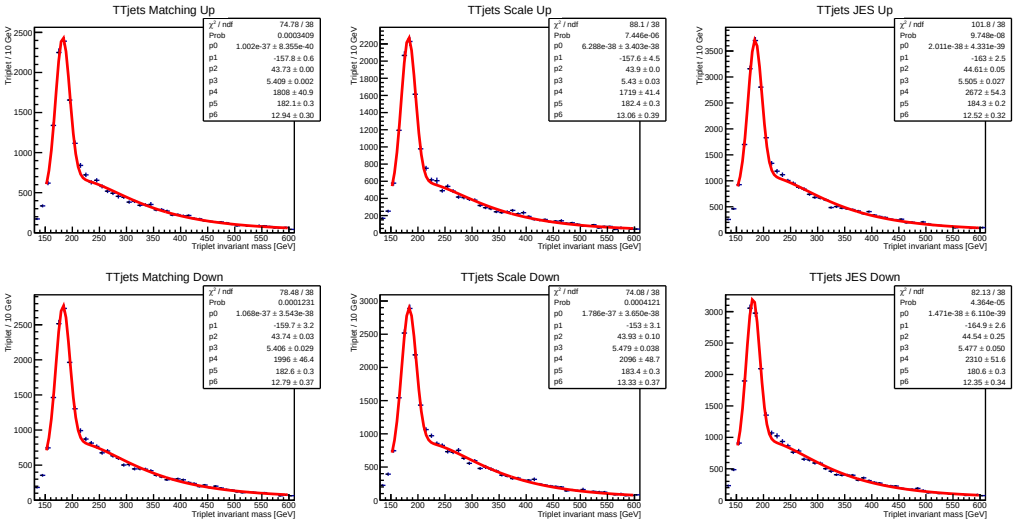

Figure C.4: Results of the fits to the systematic $t\bar{t}$ M_{jjj} distribution to estimate the effect of uncertainties on the signal yield. Shown is the effect of the following uncertainties: jet energy scale, matching scale, and factorization and renormalization scale.

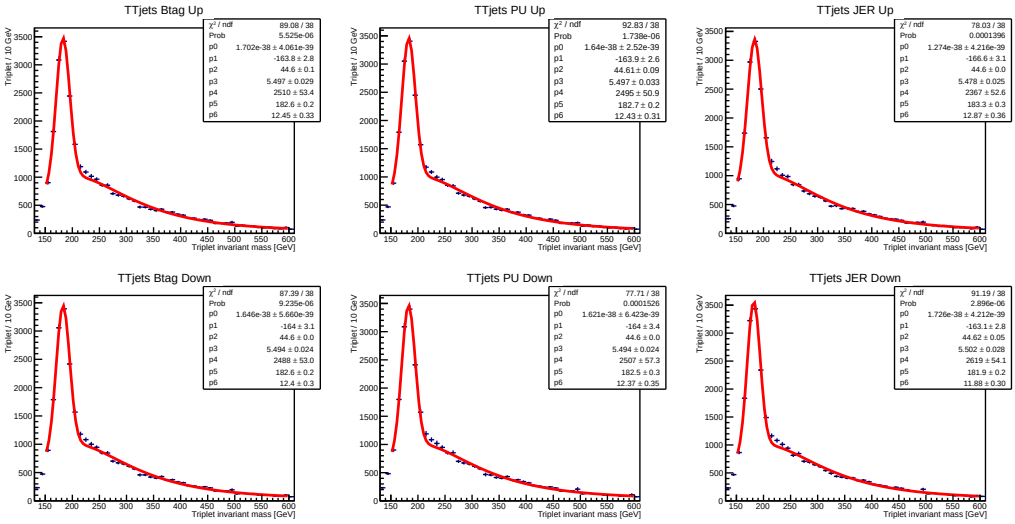

Figure C.5: Results of the fits to the systematic $t\bar{t}$ M_{jjj} distribution to estimate the effect of uncertainties on the signal yield. Shown is the effect of the following uncertainties: pile-up, b-tagging scale factors, and jet energy resolution.

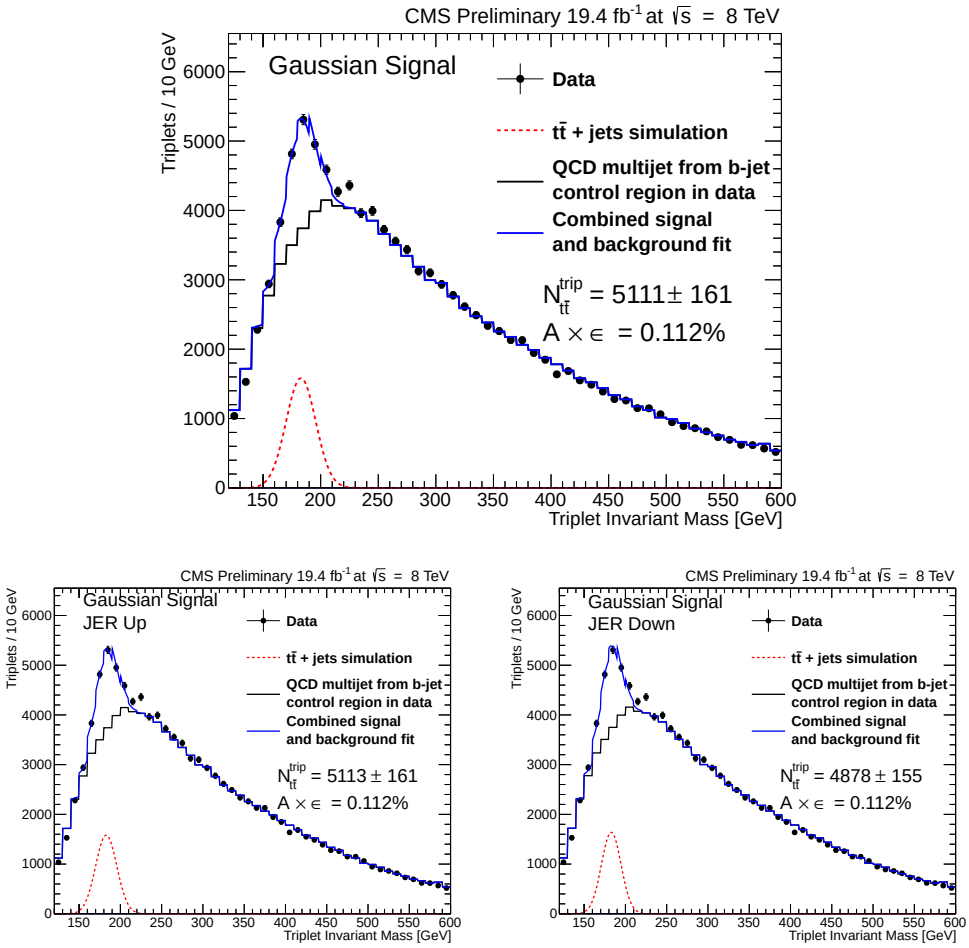

Figure C.6: Result of the fit to of the QCD multijet background template and a Gaussian signal shape to measure number of $t\bar{t}$ triplets. The top plot shows the nominal fit where the width of the Gaussian is fixed to the expected value from MC simulation. The bottom row shows the result of the two systematic fits where the width is increased or decreased to account for uncertainties on the jet energy resolution.

Luminosity 19.4 fb^{-1}		$N_{t\bar{t}}^{trip}$	$A \times \epsilon$	$\sigma_{meas}^{t\bar{t}}$ [pb]	
Nominal Gaussian signal		5111	0.1119%	235.9	
Uncertainty			$A \times \epsilon$	$\sigma_{meas}^{t\bar{t}}$ [pb]	$\delta_{t\bar{t}}$ [pb]
JES	Up		0.1211%	218.1	-17.9
	Down		0.1033%	255.7	19.8
Matching scale	Up		0.1083%	243.9	7.9
	Down		0.1169%	225.9	-10
Scale	Up		0.1123%	235.2	-0.8
	Down		0.1300%	203.2	-32.8
b-tagging	Up		0.1131%	233.4	-2.5
	Down		0.1117%	236.42	0.5
pile-up	Up		0.1123%	235.2	-0.8
	Down		0.1123%	235.2	-0.8
JER Shape	Up	4878	0.1119%	225.23	10.8
	Down	5113	0.1119%	236.06	0.1
Luminosity 2.6%					± 6.1
Statistical					± 7.4
Total	236 ± 7 (stat.) ± 41 (syst.)±6 (lumi.)				

Table C.2: Summary of the effect of different systematic uncertainties on the measured $t\bar{t}$ cross section using the Gaussian signal approximation. The top part of the table shows the central measured cross section value. The values in the fifth column shows the measured cross section when the systematic templates are used. Column six denoted with $\delta_{t\bar{t}}$ represents the difference between the nominal value and the systematic value for a given uncertainty.

C.3 Summary

We have shown in this chapter that we are able to reconstruct all-hadronic $t\bar{t}$ events by using the jet ensemble technique. Additionally, we successfully measured the $t\bar{t}$ production cross section from a fit of the triplet invariant mass distribution in data. Two methods were discussed, using either the full triplet invariant mass distribution or a Gaussian approximation as the signal shape. Both methods agree within uncertainties with each other and the theoretical NNLO prediction. This gives confidence in the search strategy and shows that the technique is designed to discover new three-jet resonances.

Appendix D
Signal injection test with Gaussian approximation and full mass shape

In Section 9.6.1 we discussed a signal injection test showing the stability of the limit setting technique. In this chapter we will show a brief comparison between the Gaussian approximation and the full M_{jjj} distribution for the signal. Therefore, another signal injection test is performed where the complete signal shape of incorrectly and correctly combined triplets, as derived from MC simulation, is injected as the potential signal.

Figure D.1 (left) represents the background four-parameter function for the inclusive search. In the same figure (right) a potential signal M_{jjj} distribution for a gluino with a mass of 750 GeV is shown accompanied by the Gaussian approximation for this mass point. The signal input signal strength parameter r^{in} is set to 1. Based on the background estimate and the full M_{jjj} shape we generate pseudo-data. Figure D.1 shows one example of a signal + background M_{jjj} distribution (the same distribution on left and right). Each of these pseudo-datasets is then fitted with either the four-parameters function + a Gaussian (left, blue) or the four-parameters function + a template of the full M_{jjj} shape (right, red), and the signal strength parameter r^{meas} is determined. In these fits all of the parameters of the background four-parameter function are allowed to float freely.

The bottom plot of Figure D.2 shows distributions of r^{meas} extracted with the Gaussian signal approximation (left, blue) or with the full M_{jjj} shape (right, red). Both fitting methods yield on average $r^{meas} \approx r^{in}$.

We have already presented a similar test in Appendix C for the $t\bar{t}$ cross section measurement, which showed that the cross section measured with the Gaussian signal approximation agrees well within uncertainties with the one measured by using the full M_{jjj} shape. With the signal injection test presented here, we have demonstrated that the Gaussian approximation can also be used to measure the cross section for higher resonances masses, where the QCD multijet background is estimated from a parametrized fit to the data.

While both signal assumptions give a very similar result we use the Gaussian approximation for the analysis since it allows for a more general search for three-jet resonances, with less parameters to model the signal.

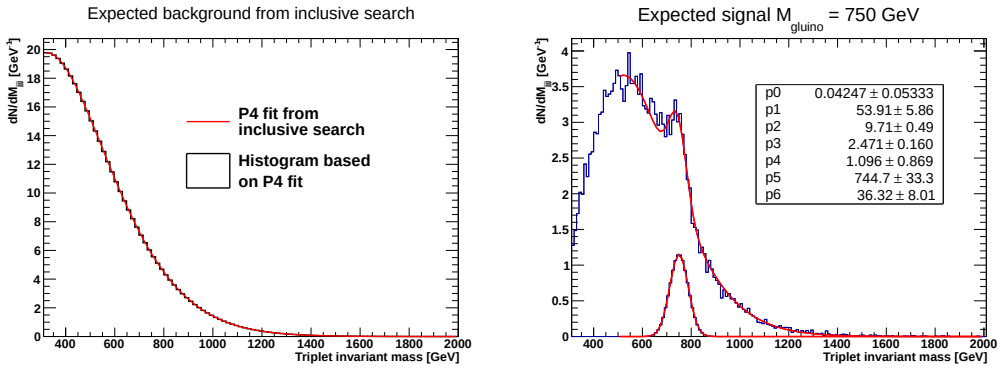

Figure D.1: Background and full signal M_{jjj} distributions for signal injection test. The left plot shows the background estimate for the inclusive search. On the right a potential signal shape for a gluino with a mass of 750 GeV is shown, accompanied by a four-parameter + Gaussian fit in red.

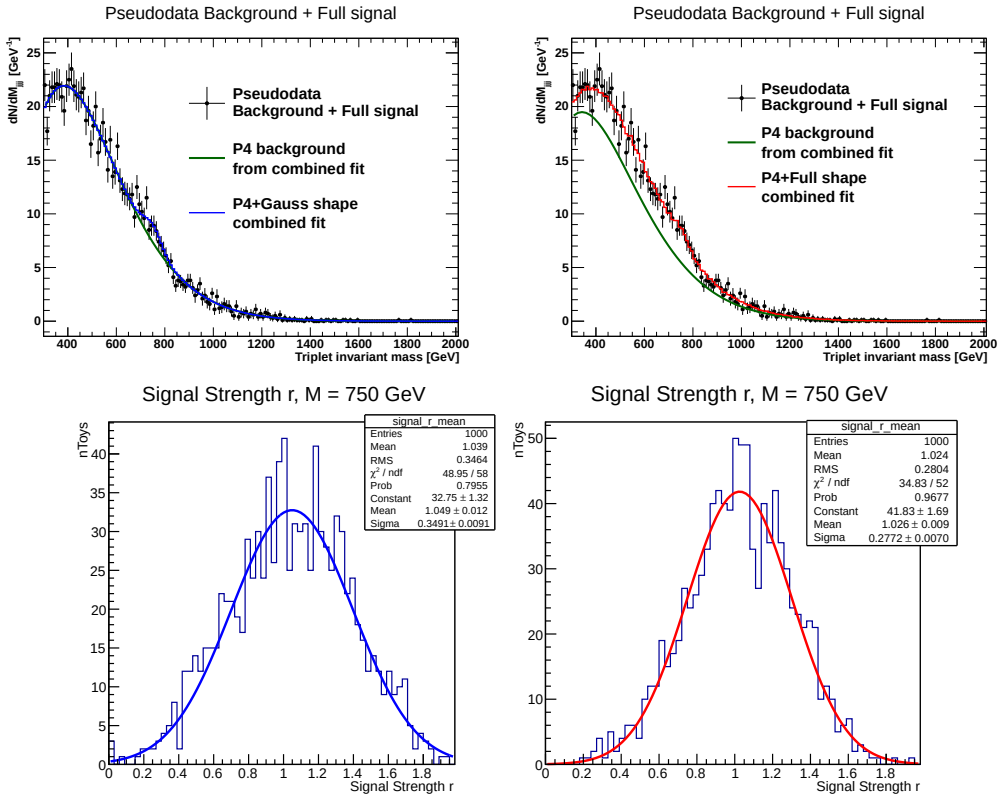

Figure D.2: Pseudo-data for signal injection test and measured signal strength. The pseudo-data is generated from the background + full M_{jjj} signal model and fitted with a four-parameter function + Gaussian (left, blue) or a four-parameter function + the full expected M_{jjj} distribution. For each pseudo-dataset the signal strength r^{meas} is measured and the bottom plot shows the output for 1000 pseudo-datasets.

Appendix E
Signal fits for systematic uncertainty studies

As discussed in Chapter 8, the M_{jjj} distribution is fitted by the sum of the four-parameter function and a Gaussian to extract the $A \times \epsilon$ for each mass point To assess systematic uncertainties, we vary the parameter in question and determine the effect on the signal fit.

- For the jet energy scale uncertainty study we increase and decrease the jet four momentum by a given uncertainty based on jet p_T and η. The fits for the increased value are shown in Figure E.1 for the inclusive search and Figure E.3 for the heavy-flavor search, and for the decreased value in Figure E.2 for the inclusive search and Figure E.4 for the heavy-flavor search.
- For the pile-up re-weighting systematic uncertainty study, we increase and decrease the inelastic cross section used to calculate the re-weighting by 5%. The fits for the increased value are shown in Figure E.5 for the inclusive search and Figure E.7 for the heavy-flavor search, and for the decreased value in Figure E.6 for the inclusive search and Figure E.8 for the heavy-flavor search.
- For the b-tagging systematic uncertainty study we increase and decrease the b-tagging scale factors within their uncertainties. The fits for the increased values are shown in Figure E.9 and for the decreased values in Figure E.10.

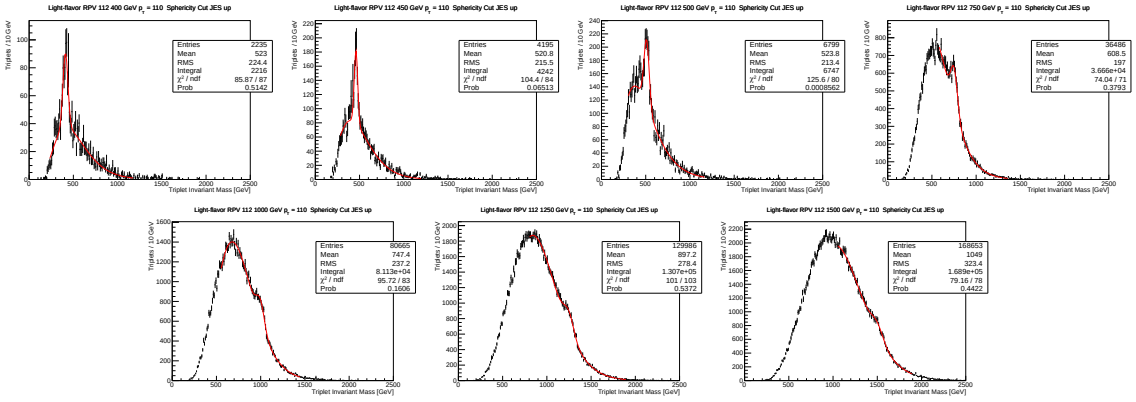

Figure E.1: Fit results to the simluated M_{jjj} distribution from adjusting jet energy scale factors upwards for the inclusive search.

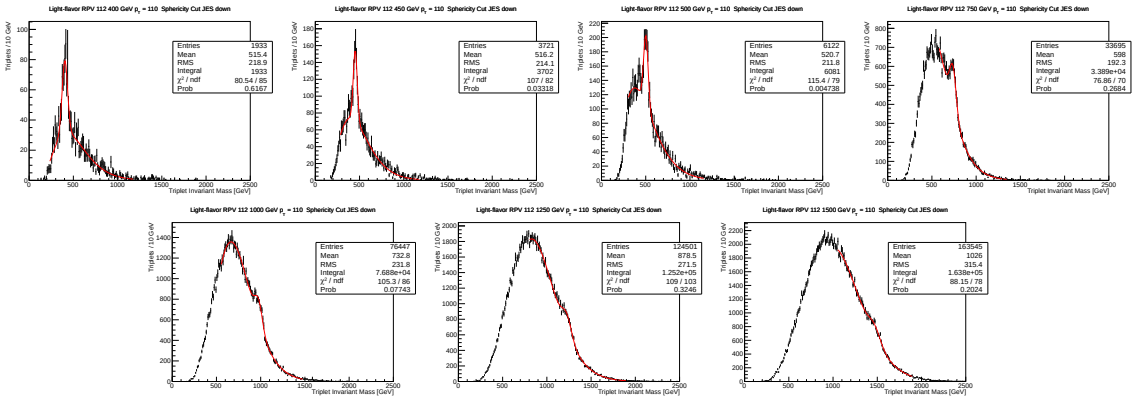

Figure E.2: Fit results to the simluated M_{jjj} distribution from adjusting jet energy scale factors downwards for the inclusive search.

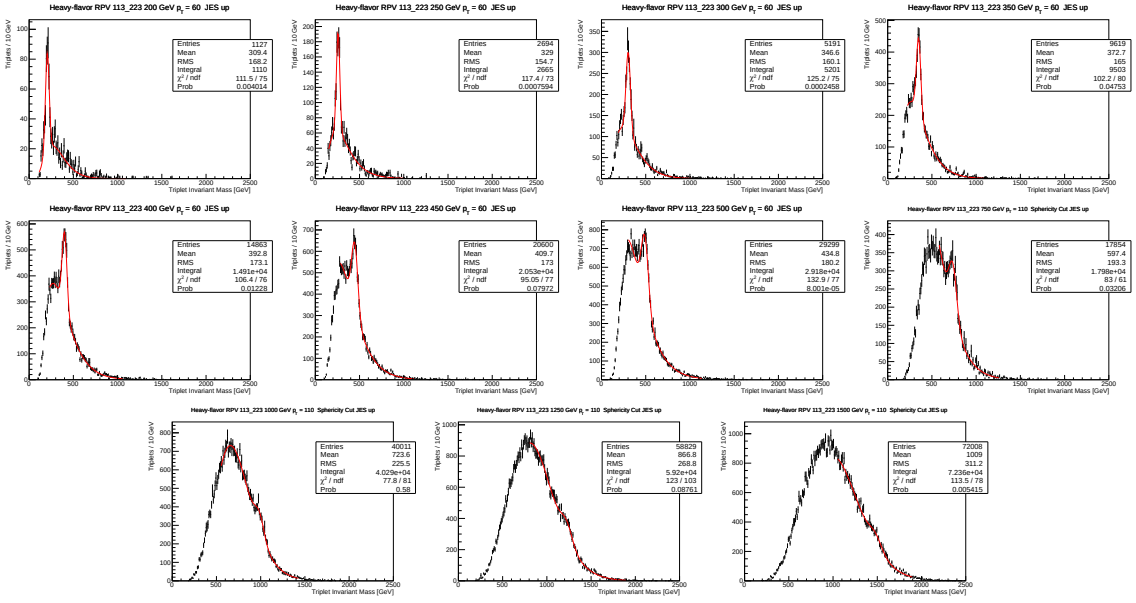

Figure E.3: Fit results to the simluated M_{jjj} distribution from adjusting the jet energy scale factors upwards for the heavy-flavor search.

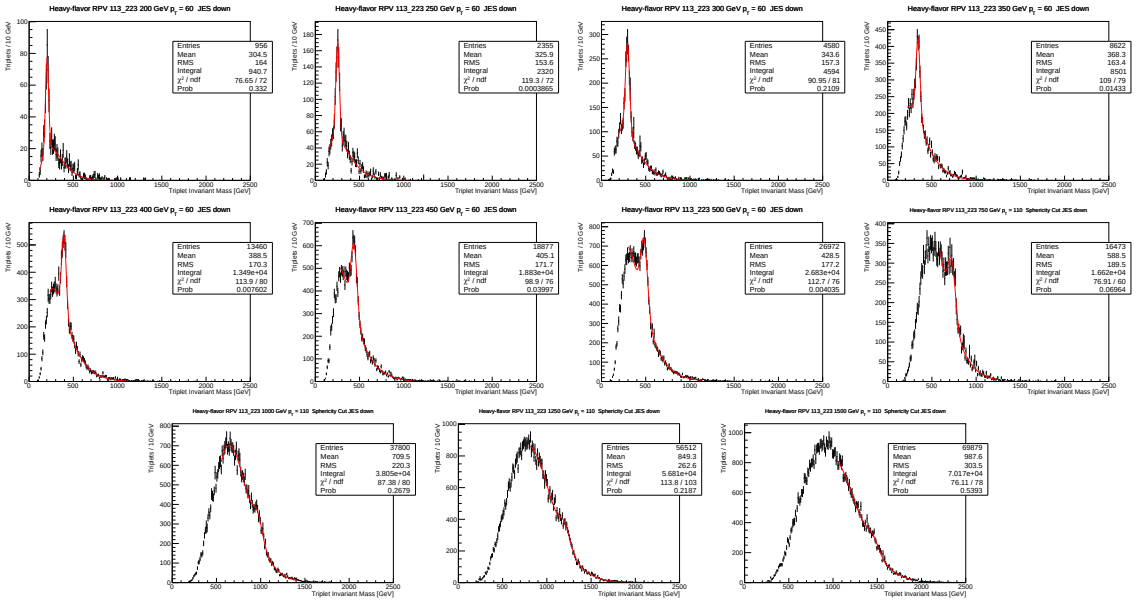

Figure E.4: Fit results to the simluated M_{jjj} distribution from adjusting jet energy scale factors downwards for the heavy-flavor search.

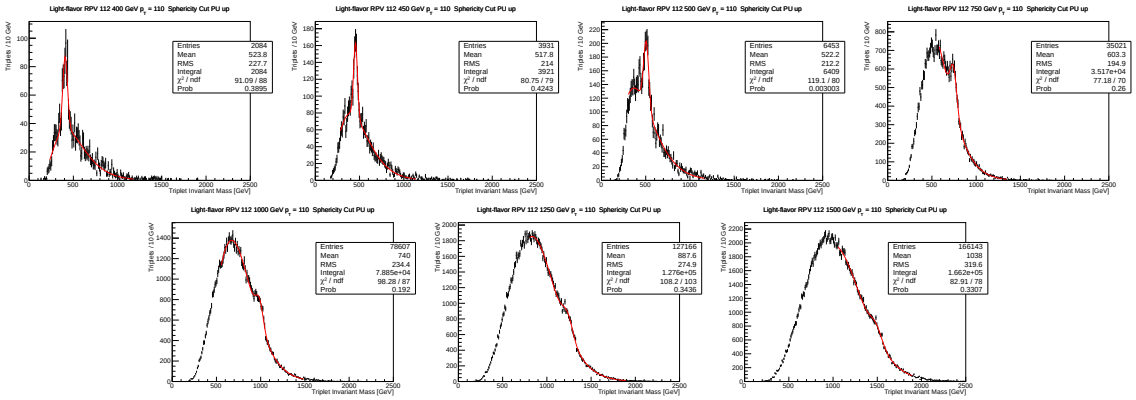

Figure E.5: Fit results to the simluated M_{jjj} distribution from pile-up re-weighting with a increased inelastic cross section for the inclusive search.

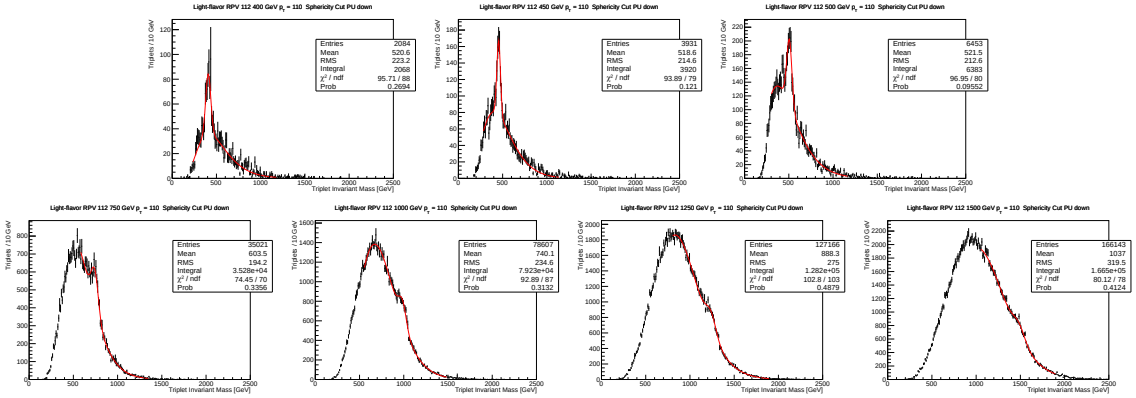

Figure E.6: Fit results to the simluated M_{jjj} distribution from pile-up re-weighting with a decreased inelastic cross section for the inclusive search.

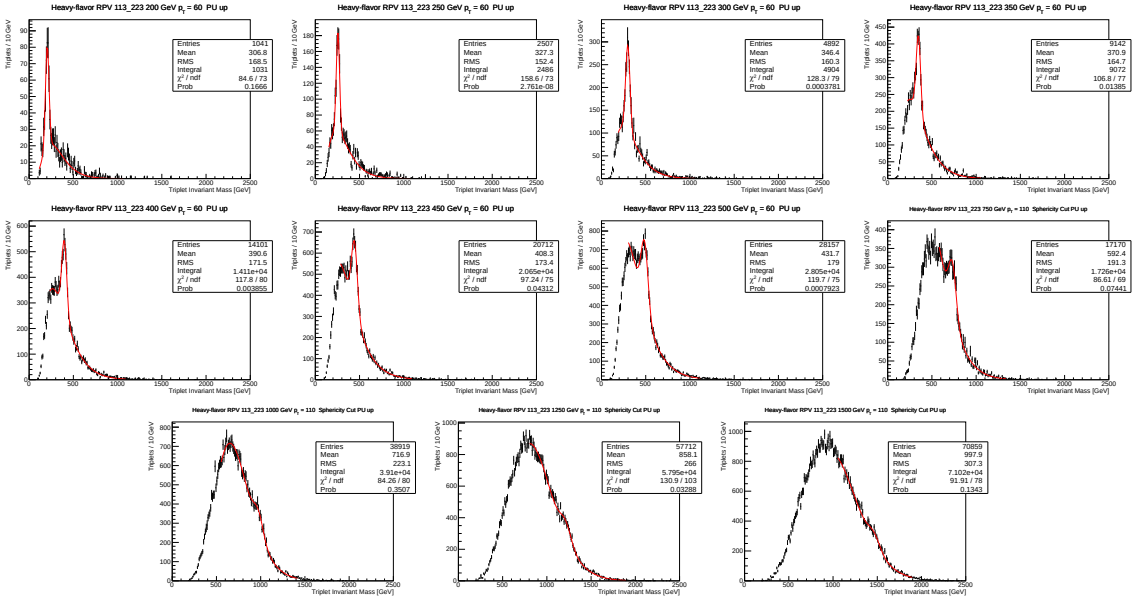

Figure E.7: Fit results to the simluated M_{jjj} distribution from pile-up re-weighting with an increased inelastic cross section for the heavy-flavor search.

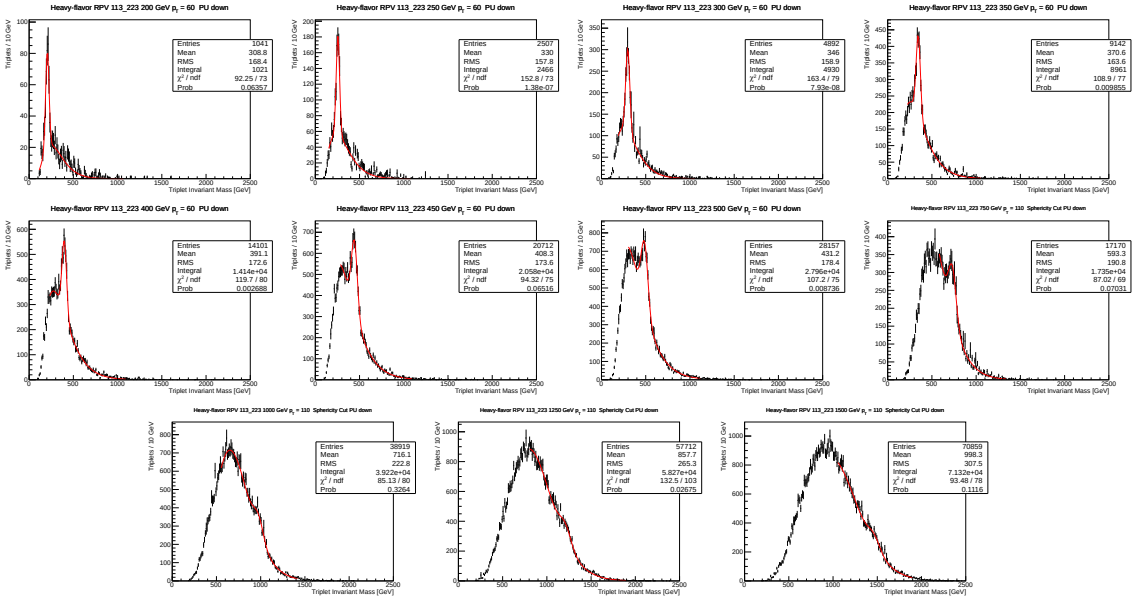

Figure E.8: Fit results to the simluated M_{jjj} distribution from pile-up re-weighting with an decreased inelastic cross section for the heavy-flavor search.

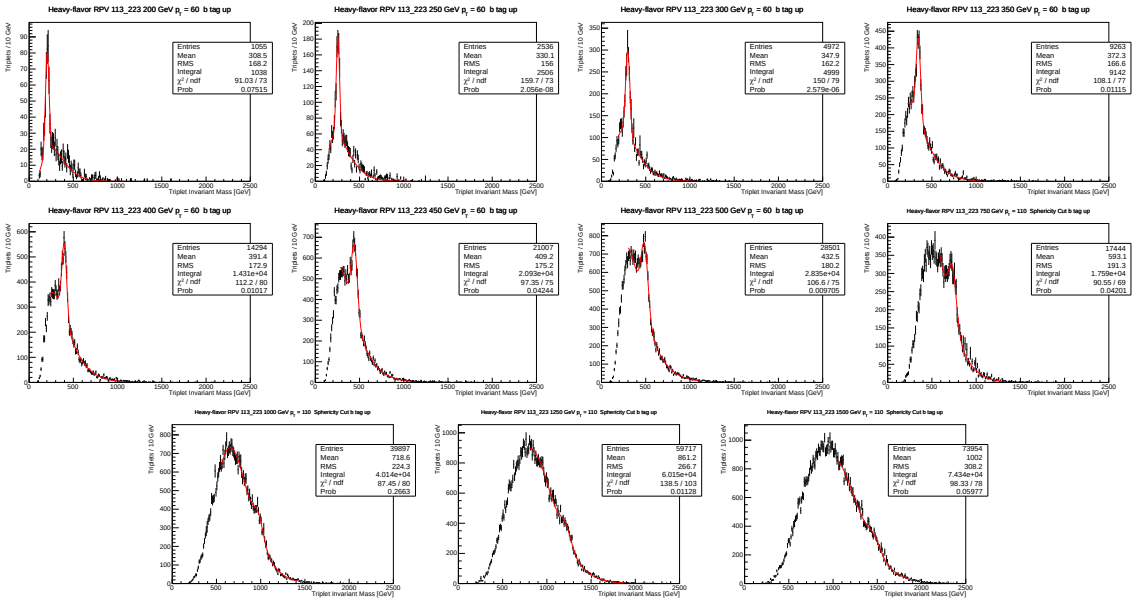

Figure E.9: Fit results to the simluated M_{jjj} distribution from adjusting b-tagging scale factors upwards for the heavy-flavor search.

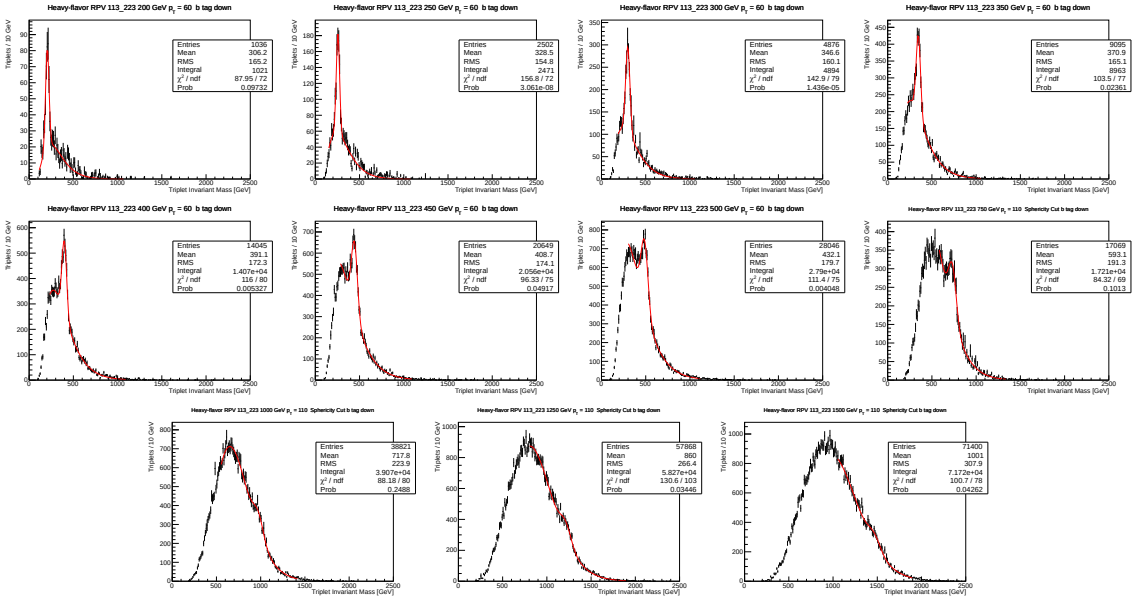

Figure E.10: Fit results to the simluated M_{jjj} distribution from adjusting b-tagging scale factors downwards for the heavy-flavor search.

Bibliography

[1] W. Hollik, "Quantum field theory and the Standard Model," 2010. http://arxiv.org/abs/1012.3883.

[2] Francis Halzen, Allen Martin, *Quarks and Leptons*. John Wiley and Sons, 1984.

[3] Wikipedia, "Particles in the Standard Model." http://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[4] A. Pich, "The Standard Model of Electroweak Interactions," 2012. http://arxiv.org/abs/1201.0537.

[5] David Griffith, *Introduction to Elementary Particles*. WILEY-VCH, 2011.

[6] W. Wagner, "Top quark physics in hadron collisions," *Rep. Prog. Phys*, vol. 68, pp. 2409–2494, 2005. http://arxiv.org/abs/hep-ph/0507207.

[7] [CMS Collaboration], "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 30 – 61, 2012.

[8] [ATLAS Collaboration], "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 1 – 29, 2012.

[9] J. Beringer *et al.*, "Review of Particle Physics (RPP)," *Phys.Rev.*, vol. D86, p. 010001, 2012.

[10] R. Placakyte, "Parton Distribution Functions," 2011.

[11] P. Z. Skands, "QCD for Collider Physics," 2011.

[12] Ellis, Stirling, Webber, *QCD and Collider Physics*. Cambridge University press.

[13] S.P. Martin, "A supersymmetry primer," 2008. http://arxiv.org/abs/hep-ph/9709356v5.

[14] R. Barbier, C. Berat, M. Besancon, M. Chemtob, A. Deandrea, E. Dudas, P. Fayet, S. Lavignac, G. Moreau, E. Perez, and Y. Sirois, "R-Parity-violating supersymmetry," *Physics Reports*, vol. 420, no. 1 - 6, pp. 1 – 195, 2005.

[15] [CDF Collaboration], "First Search for Multijet Resonances in $\sqrt{s} = 1.96$ TeV pp Collisions," *Phys. Rev. Lett.*, vol. 107, p. 042001, 2011.

[16] [CMS Collaboration], "Search for Three-Jet Resonances in pp Collisions at $\sqrt{s} = 7$ TeV," *Phys. Rev. Lett.*, vol. 107, p. 101801, 2011.

[17] [CMS Collaboration], "Search for three-jet resonances in pp collisions at $\sqrt{s} = 7$ TeV," *Phys. Lett. B*, vol. 718, p. 329, 2012.

[18] [ATLAS Collaboration], "Search for pair production of massive particles decaying into three quarks with the ATLAS detector in $\sqrt{s} = 7$ TeV pp collisions at the LHC," *JHEP*, vol. 12, p. 086, 2012.

[19] [ATLAS Collaboration], "Search for massive particles in multijet signatures with the ATLAS detector in $\sqrt{s} = 8$ TeV pp collisions at the LHC," ATLAS Conference Note ATLAS-CONF-2013-091, CERN, Geneva, Aug 2013.

[20] C. Seitz, "Search for hadronic resonances in multijet final states with the CDF detector," Masters thesis, Rutgers University, USA, January 2011.

[21] [CMS Collaboration], "Searches for light- and heavy-flavour three-jet resonances in pp collisions at sqrt(s) = 8 TeV," 2013, arXiv:1311.1799, Submitted to PLB.

[22] [CMS Collaboration], "Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s}$=7 TeV with the CMS detector," *Phys. Rev. D*, vol. 87, p. 112002, Jun 2013.

[23] R. Essig, "Physics Beyond the Standard Model: Supersymmetry, Dark Matter, and LHC Phenomenology," PhD thesis, Rutgers University, USA, 2008.

[24] A. Quadt, "Top quark physics at hadron colliders ," *The European Physical Journal C - Particles and Fields*, vol. 48, no. 3, pp. 835–1000, 2006.

[25] "The CERN accelerator complex", `http://home.web.cern.ch/about/accelerators`,"

[26] "Public CMS Luminosity Information, `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`,"

[27] "The CERN accelerator complex", `http://te-epc-lpc.web.cern.ch/te-epc-lpc/machines/pagesources/Cern-Accelerator-Complex.jpg`,"

[28] [CMS Collaboration], "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, p. S08004, 2008. `http://stacks.iop.org/1748-0221/3/i=08/a=S08004`.

[29] "CMS Detector, `http://bigscience.web.cern.ch/bigscience/Objects/CMS/CMS-eng.jpg`,"

[30] W. Erdmann, "The CMS Pixel Detector," *International Journal of Modern Physics A*, vol. 25, no. 07, pp. 1315–1337, 2010.

[31] [CMS Collaboration], "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV," *JINST*, vol. 8, p. P09009, 2013.

[32] [CMS Collaboration], *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS, Geneva: CERN, 2006.

[33] J. Brooke, "Performance of the CMS Level-1 Trigger," *PoS*, vol. ICHEP2012, p. 508, 2013.

[34] G. Iles, J. Brooke, C. Foudas, R. Frazier, G. Hall, M. Hansen, G. Heath, J. Jones, J. Nash, A. Rose, M. Stettler, and A. Tapper, "Revised CMS Global Calorimeter Trigger Functionality & Algorithms," p. 5 p, 2007.

[35] S. Beauceron, "The CMS High Level Trigger," Tech. Rep. CMS-CR-2012-355. CERN-CMS-CR-2012-355, CERN, Geneva, Nov 2012.

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

[37] S. Agostinelli *et al.*, "GEANT4 – a Simulation Toolkit," *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003.

[38] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[39] J. Alwall, P. Demin, S. de Visscher, R. Frederix, M. Herquet, *et al.*, "MadGraph/MadEvent v4: The new web generation," *JHEP*, vol. 09, p. 028, 2007.

[40] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, "FeynRules 2.0 - A complete toolbox for tree-level phenomenology," 2013.

[41] Michelangelo L. Mangano and Mauro Moretti and Fulvio Piccinini and Michele Treccani, "Matching matrix elements and shower evolution for top-pair production in hadronic collisions," *Journal of High Energy Physics*, vol. 2007, no. 01, p. 013, 2007.

[42] "Introduction to jet-parton matching in MG/ME, `https://cp3.irmp.ucl.ac.be/projects/madgraph/wiki/IntroMatching`,"

[43] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung, "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis," *Journal of High Energy Physics*, vol. 2002, no. 07, p. 012, 2002.

[44] D. Alves *et al.*, "Simplified Models for LHC New Physics Searches," *J.Phys.*, vol. G39, p. 105005, 2012.

[45] W. Beenakker, R. Hoepker, M. Spira, and P. Zerwas, "Squark and gluino production at hadron colliders," *Nucl. Phys. B*, vol. 492, p. 51, 1997.

[46] A. Kulesza and L. Motyka, "Threshold Resummation for Squark-Antisquark and Gluino-Pair Production at the LHC," *Phys. Rev. Lett.*, vol. 102, p. 111802, 2009.

[47] A. Kulesza and L. Motyka, "Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC," *Phys. Rev. D*, vol. 80, p. 095004, 2009.

[48] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, E. Laenen, *et al.*, "Soft-gluon resummation for squark and gluino hadroproduction," *JHEP*, vol. 12, p. 041, 2009.

[49] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, E. Laenen, *et al.*, "Squark and Gluino Hadroproduction," *Int. J. Mod. Phys. A*, vol. 26, p. 2637, 2011.

[50] M. Kramer, A. Kulesza, R. van der Leeuw, M. Mangano, S. Padhi, *et al.*, "Supersymmetry production cross sections in pp collisions at $\sqrt{s} = 7$ TeV," 2012.

[51] "Theory cross section plots for selected SUSY processes, `https://twiki.cern.ch/twiki/pub/CMSPublic/SUSYSMSSummaryPlots8TeV/xsections_strong.pdf`,"

[52] M. Czakon, P. Fiedler, and A. Mitov, "The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$," *Phys. Rev. Lett.*, vol. 110, p. 252004, 2013.

[53] [CMS Collaboration], "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector," CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010.

[54] [CMS Collaboration], "CMS tracking performance results from early LHC operation," *The European Physical Journal C*, vol. 70, no. 4, pp. 1165–1192, 2010.

[55] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2-3, pp. 444 – 450, 1987.

[56] P. S. Maybeck, *Stochastic models, estimation, and control*, vol. 141 of *Mathematics in Science and Engineering*. 1979.

[57] [CMS Collaboration], "Description and performance of the CMS track and primary vertex reconstruction," *CMS Analysis Note*, no. AN-2011/172, 2011.

[58] [CMS Collaboration], "Offline Primary Vertex Reconstruction with Deterministic Annealing Clustering," *CMS Analysis Note*, no. AN-2011/014, 2011.

[59] [CMS Collaboration], "Adaptive Vertex Fitting," *CMS Analysis Note*, no. AN-2007/008, 2007.

[60] "Jets at CMS and the determination of their energy scale." `http://cms.web.cern.ch/news/jets-cms-and-determination-their-energy-scale`.

[61] G. P. Salam, "Towards Jetography," *Eur.Phys.J.*, vol. C67, pp. 637–686, 2010. `http://arxiv.org/abs/0906.1833`.

[62] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008. `http://stacks.iop.org/1126-6708/2008/i=04/a=063`.

[63] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys. Rev. D*, vol. 48, pp. 3160–3166, Oct 1993.

[64] Y. Dokshitzer, G. Leder, S. Moretti, and B. Webber, "Better jet clustering algorithms," *Journal of High Energy Physics*, vol. 1997, no. 08, p. 001, 1997.

[65] [CMS Collaboration], "Search for Anomalous Top Quark Pair Production in the Boosted All-Hadronic Final State using pp Collisions at sqrt(s) = 8 TeV," CMS Physics Analysis Summary CMS-PAS-B2G-12-005, CERN, Geneva, 2013.

[66] [CMS Collaboration], "Determination of jet energy calibration and transverse momentum resolution in CMS," *Journal of Instrumentation*, vol. 6, no. 11, p. P11002, 2011.

[67] "JEC performance plots for 2013." `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsJME2013JEC`.

[68] [CMS Collaboration], "Identification of b-quark jets with the CMS experiment," *JINST*, vol. 8, p. P04013, 2013.

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

[70] [CMS Collaboration], "Photon reconstruction and identification at sqrt(s) = 7 TeV," CMS Physics Analysis Summary CMS-PAS-EGM-10-005, CERN, 2010. Geneva, 2010.

[71] [CMS Collaboration], "Jets in 0.9 and 2.36 TeV pp Collisions," no. CMS-PAS-JME-10-001, 2010.

[72] [CMS Collaboration], "Search for narrow resonances using the dijet mass spectrum in *pp* collisions at $\sqrt{s}$=8 TeV," *Phys. Rev. D*, vol. 87, p. 114015, Jun 2013.

[73] [CMS Collaboration], "Measurement of the 3-jet mass cross section at 7 TeV and determination of alphaS," *CMS Analysis Note*, no. AN-2013/046, 2013.

[74] "CMS Internal twiki: Apply SF for b/c/light tags to directly tagged jets in MC, `https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFUtil`,"

[75] L. Moneta, K. Belasco, K. S. Cranmer, A. Lazzaro, D. Piparo, G. Schott, . W. Verkerke, and M. Wolf, "The RooStats Project," in *13^{th} International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2010)*, SISSA, 2010.

[76] Mark Thomson, ""graduate lectures on statistics 2009, `http://www.hep.phy.cam.ac.uk/~thomson/lectures/statistics/FittingHandout.pdf`","

[77] I. Antcheva, M. Ballintijn, B. Bellenot, M. Biskup, R. Brun, N. Buncic, P. Canal, D. Casadei, O. Couet, V. Fine, L. Franco, G. Ganis, A. Gheata, D. G. Maline, M. Goto, J. Iwaszkiewicz, A. Kreshuk, D. M. Segura, R. Maunder, L. Moneta, A. Naumann, E. Offermann, V. Onuchin, S. Panacek, F. Rademakers, P. Russo, and M. Tadel, "Root - A C++ framework for petabyte data storage, statistical analysis and visualization," *Computer Physics Communications*, vol. 180, no. 12, pp. 2499 – 2512, 2009.

[78] [TOTEM Collaboration], "First measurement of the total proton-proton cross-section at the LHC energy of $\sqrt{s} = 7$TeV," *"EPL (Europhysics Letters)"*, vol. 96, no. 2, p. 21002, 2011.

[79] [CMS Collaboration], "Dijet Azimuthal Decorrelations in *pp* Collisions at $\sqrt{s}$ =7 TeV," *Phys. Rev. Lett.*, vol. 106, p. 122003, Mar 2011.

[80] [CMS Collaboration], "Measurement of the $\mathrm{t\bar{t}}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s}$ =7 TeV," *Journal of High Energy Physics*, vol. 2012, no. 11, pp. 1–40, 2012.

[81] [ATLAS Collaboration], "Procedure for the LHC Higgs boson search combination in summer 2011," Tech. Rep. ATL-PHYS-PUB-2011-011, CERN, Geneva, Aug 2011.

[82] T. Junk, "Confidence level computation for combining searches with small statistics," *Nucl. Instrum. Meth. A*, vol. 434, p. 435, 1999.

[83] A. L. Read, "Presentation of search results: the CL_S technique," *J. Phys. G*, vol. 28, p. 2693, 2002.

[84] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C*, vol. 71, no. 2, pp. 1–19, 2011.

[85] [CMS Collaboration], "Search for microscopic black holes in pp collisions at $\sqrt{s}$=8 TeV," *Journal of High Energy Physics*, vol. 2013, no. 7, pp. 1–34, 2013.