

SEARCH FOR PAIR-PRODUCED DIQUARK
RESONANCES IN PROTON-PROTON COLLISIONS
WITH THE CMS DETECTOR AT $\sqrt{S} = 13$ TeV

By

TIRSO ALEJANDRO GOMEZ ESPINOSA

A dissertation submitted to the

School of Graduate Studies

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

May, 2018



ABSTRACT OF THE DISSERTATION

Search for pair-produced diquark resonances in proton-proton collisions with the CMS detector at

$$\sqrt{s} = 13 \text{ TeV}$$

By **TIRSO ALEJANDRO GOMEZ ESPINOSA**

Dissertation Director:

Prof. Eva Halkiadakis

A search is performed for the pair production of resonances decaying to two light quarks or one light quark and one bottom quark. The search is conducted separately for light resonances between 80 and 400 GeV in mass, when the hadronization of the quarks is collimated enough to be reconstructed as a single jet producing a dijet final state, and for heavy resonances above 400 GeV in mass, when the decay products generate pairs of hadronic jets producing a four-jet final state. The data used were collected with the CMS detector in 2016 at the CERN LHC from proton-proton collisions at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 35.9 fb^{-1} . The mass spectra are analyzed to look for new resonant particles, and are found to be consistent with standard model expectations. These results are interpreted in the framework of R-parity-violating supersymmetry assuming the pair production of scalar top quarks decaying via the λ''_{312} or the λ''_{323} couplings. Upper limits are set at the 95% confidence level on the pair production cross section of top squarks as a function of the top squark mass.

Acknowledgments

It is impossible to list all of the people that contributed and helped me achieved this necessary step in my path to the Nobel prize, however I would like to specially thank:

Eva, thank you for believe in me. Thank you for the long, sometimes frustrating and tedious, hours that you spent with me during these almost six years, teaching me how to become a better scientist and a better person. I will always be grateful to you.

Edgar and Francisco, thank you for helping me start this journey, without your initial support I could not accomplish this.

In general, I would like to acknowledge all the members of the CMS Collaboration and the LHC. Without them, this work and one of the biggest scientific endeavours in human history could not be accomplished. I would like to also thank all the people in the HEX group at Rutgers and the Jets+X and B2G groups at CMS for the insightful review and discussion.

Savvas, thank you for being my partner in crime during these six years together. Thank you for being a real friend, a good (sometimes annoying) roommate, a good physicist, and thank you for all the adventures.

And to my friends (in no particular order) Tote, Aniket, David, Gianna and Ted, Myriam, Thanasis, Kostas, Andrea,... thank you all for all the support. Thank you for being good friends and for making these six years more enjoyable. *Y mis otros amigos*, Vilyn, Karina, Jaredo, Wicho, gracias por estar ahi en las buenas y en las malas.

Y finalmente, *a mi madre*, la verdadera heroína de esta historia. Gracias por todo lo que has hecho por mi. Tratar de ser mejor y alcanzar mis sueños es lo minimo que puedo hacer en agradecimiento.

Dedication

*A mi madre,
la Pachamama y,
¡Salve, oh Patria, mil veces! ¡Oh Patria!*

Table of Contents

Abstract	ii
Acknowledgments	iii
Dedication	iv
List of Tables	ix
List of Figures	xi
1. Introduction	1
2. Theoretical overview	3
2.1. The Standard Model of Particle Physics	3
2.1.1. Electroweak Interactions and Electroweak Symmetry Breaking	5
2.1.2. Quantum Chromodynamics	8
Lattice and perturbative QCD	9
Running constant and asymptotic freedom	10
Quark masses	11
Confinement and proton structure	12
Parton showers and hadronization	14
2.2. Beyond the Standard Model and Supersymmetry	15
2.2.1. R-parity Violating Scenarios	17
2.3. Theoretical benchmark model	18

3. Experimental setup	20
3.1. The Large Hadron Collider	20
3.2. The CMS experiment	22
3.2.1. Superconducting Solenoid Magnet	25
3.2.2. Tracking System	26
3.2.3. Electromagnetic calorimeter	28
3.2.4. Hadronic Calorimeter	30
3.2.5. Muon system	31
3.2.6. Trigger system	32
3.2.7. Luminosity measurements	36
3.2.8. Data Samples	37
4. Event reconstruction and simulation	39
4.1. Object reconstruction	39
4.1.1. Particle-flow algorithm	41
4.1.2. Tracks and vertex reconstruction	41
4.1.3. Jets	43
Jet algorithms	44
Pileup mitigation for jets	46
Jet identification	47
Jet energy corrections	48
Jet energy resolution	51
Jet substructure	52
b-jet identification	55
4.2. Simulated samples	56
4.2.1. Benchmark MC signal samples	57
4.2.2. SM processes	59

5. Boosted search	61
5.1. Trigger Selection	61
5.2. Event Selection	64
5.2.1. Kinematic Variables	64
5.2.2. Inclusive selection optimization	67
5.2.3. b-tagged selection optimization	70
5.3. Signal Acceptance	72
5.3.1. N-subjettiness data/MC scale factors	73
5.3.2. B tagging scale factors	75
5.4. Background Estimate	75
5.4.1. $t\bar{t}$ background	76
5.4.2. Background estimate for inclusive signal region	79
5.4.3. Background estimate for b-tagged signal region	81
5.4.4. Final background estimate	84
5.5. Systematic Uncertainties	84
5.5.1. Signal Systematic Uncertainties	85
5.5.2. Systematic Uncertainty on Background Estimate	87
6. Resolved search	89
6.1. Trigger Selection	89
6.2. Event Selection	91
6.2.1. Dijet Pairing	92
6.2.2. Kinematic Variables	93
6.2.3. Signal Optimization	94
6.2.4. B-tagging optimization	97
6.3. Signal Acceptance	101
6.4. Background Estimate	103

6.4.1. F-test	104
6.4.2. Bias test	107
6.4.3. Results of resolved search	112
6.5. Systematic Uncertainties	113
6.5.1. Signal Systematic Uncertainties	113
6.5.2. Systematic Uncertainty on Background Estimate	115
7. Interpretation of the results	117
7.1. Statistical analysis	117
7.2. Limits on hadronic RPV SUSY	120
8. Conclusion	123
Appendix A. Boosted search	124
A.1. Jet energy scale uncertainties	124
A.2. Jet energy resolution uncertainties	125
A.3. Pileup uncertainties	125
A.4. PDF uncertainties	127
A.5. B tagging uncertainties	129
Appendix B. Resolved search	131
B.1. Alternative dijet pairing algorithms	131
B.2. Jet energy scale uncertainties	135
B.3. Jet energy resolution uncertainties	136
B.4. Pileup uncertainties	137
B.5. PDF uncertainty	137
B.6. B tag uncertainty	139
Bibliography	141

List of Tables

2.1. Summary of previous limits for the benchmark RPV SUSY models.	19
3.1. JetHT primary datasets and corresponding luminosities.	37
4.1. List of variables and selection for different jetID criteria.	48
4.2. Simulated signal samples.	58
4.3. Simulated background samples.	60
5.1. Boosted search: Description of the triggers used.	63
5.2. Boosted search: Preselection criteria.	65
5.3. Boosted search: Threshold values of variables for optimization	67
5.4. Boosted search: Optimized selection criteria.	69
5.5. Boosted search: Significance for different b tagging selections.	71
5.6. Boosted search: Cutflow table for inclusive and b-tagged search.	71
5.7. Boosted search: Optimized top selection criteria.	78
5.8. Boosted search: Definition of the regions for the ABCD method.	79
5.9. Boosted search: Summary of the systematic uncertainties on the signal. . .	87
5.10. Boosted search: Summary of the systematic uncertainties on the background.	88
6.1. Resolved search: Triggers used in the analysis with their L1 and HLT definition	90
6.2. Resolved search: Summary of preselection criteria	92
6.3. Resolved search: Threshold values of variables for optimization.	95
6.4. Resolved search: Optimized selection table.	98
6.5. Resolved search: Cutflow table with sequential selection.	100
6.6. Resolved search: F-test for inclusive search.	108

6.7. Resolved search: F-test for b-tagged search.	108
6.8. Resolved search: Resulting χ^2/ndf and likelihood ratio values for the different functions modeling the QCD multijet background.	110
6.9. Resolved search: Summary of the systematic uncertainties on the signal. . .	116
B.1. Resolved search: Significance and number of events of different pairing methods.	134

List of Figures

2.1. Fundamental particles of the standard model	4
2.2. Examples of Feynman diagrams for some electromagnetic and weak interaction	5
2.3. Examples of Feynman diagrams for some QCD interactions	10
2.4. Example of parton distribution function.	13
2.5. Feynman diagrams for the benchmark models.	18
3.1. CERN accelerator complex	21
3.2. Total integrated luminosity for CMS in 2016.	23
3.3. CMS Detector	23
3.4. Schema of the CMS pixel tracker and silicon sensors	27
3.5. View of the CMS tracker system	27
3.6. View of the CMS electromagnetic calorimeter and $PbWO_4$ crystals.	29
3.7. View of the CMS hadronic calorimeter.	31
3.8. View of the CMS Muon system.	32
3.9. Overview of the CMS L1 trigger system.	33
4.1. Transverse slice of the CMS detector showing the interaction of different particles within the detector	40
4.2. Graphical representation of a jet in the CMS detector	44
4.3. Representation of the different jet iterative algorithms	46
4.4. Schema of jet energy corrections.	49
4.5. CMS jet energy corrections to pileup offset and simulated response in 2016 data.	49

4.6. CMS jet energy corrections to residual jet responses in 2016 data.	51
4.7. CMS jet energy resolution vs p_T of jets for 2016 MC samples.	52
4.8. Graphical diagram of b-jet identification	56
4.9. Missidentification and efficiency of the CSVv2 b tagging algorithm.	57
5.1. Boosted search: Trigger efficiency turn-on curves.	64
5.2. Boosted search: m_{asym} and $\Delta\eta$ distributions after preselection.	66
5.3. Boosted search: jet τ_{21} distributions after preselection.	66
5.4. Boosted search: jet τ_{32} distributions after preselection.	66
5.5. Boosted search: Example of optimization process for $m_{\tilde{t}} = 100$ GeV.	68
5.6. Boosted search: Average pruned mass distribution with MC samples after full inclusive selection.	70
5.7. Boosted search: B tag discriminator for two leading jets.	70
5.8. Boosted search: Average pruned mass distribution with MC samples after full b-tagged selection.	72
5.9. Boosted search: Signal efficiency distribution.	72
5.10. Boosted search: Distributions of the average pruned jet mass for signal MC.	73
5.11. Boosted search: Jet p_T distributions for signal MC samples.	74
5.12. Boosted search: Comparison of the mean of the fitted Gaussian in the signal masses as a function of the true $m_{\tilde{t}}$	75
5.13. Boosted search: Optimization plots for top selection.	77
5.14. Boosted search: Average pruned mass distribution with the top selection.	78
5.15. Boosted search: two disagreement distributions of m_{asym} as a funtion of $\Delta\eta$	80
5.16. Boosted search: transfer factor for ABCD method and closure test for inclu- sive selection.	82
5.17. Boosted search: Comparison between transfer factors with inclusive and b- tagged selection.	83

5.18. Boosted search: Transfer factor and closure test for the b-tagged selection. .	83
5.19. Boosted search: Final background estimate.	84
6.1. Resolved search: Trigger efficiency turn-on curves.	90
6.2. Resolved search: Trigger efficiency turn-on curves for average dijet mass distribution.	91
6.3. Resolved search: Fourth leading jet p_T for QCD multijets MC and some selected signal samples.	91
6.4. Resolved search: Kinematic variables used in the analysis.	94
6.5. Resolved search: Distributions of Δ as a function of the average dijet mass.	94
6.6. Resolved search: Optimization plots.	96
6.7. Resolved search: Optimization plots for delta variable.	97
6.8. Resolved search: CSVv2 distributions for AK4 jets.	98
6.9. Resolved search: Optimization plots for b-tagged selection.	99
6.10. Resolved search: Example of gaussian fits for some inclusive signal samples.	101
6.11. Resolved search: Example of gaussian fits for some b-tagged signal samples.	101
6.12. Resolved search: Signal efficiency.	102
6.13. Resolved search: Gaussian signal shapes for some selected signal samples. .	102
6.14. Resolved search: $R(m)$ as a function of the fit starting point.	104
6.15. Resolved search: Average dijet mass distribution for the inclusive and re- solved search.	105
6.16. Resolved search: Average dijet mass distribution for inclusive selection fitted with different parameter functions.	107
6.17. Resolved search: Average dijet mass distribution for b-tagged selection fitted with different parameter functions.	108
6.18. Resolved search: $\overline{M}$ distribution with different fits for the bias test for the inclusive analysis.	110

6.19. Resolved search: $\overline{M}$ distribution with different fits for the bias test for the b-tagged analysis.	110
6.20. Resolved search: Bias test for inclusive selection.	111
6.21. Resolved search: Bias test for b-tagged selection.	112
6.22. Resolved search: Results of bias test.	112
6.23. Results: resolved search.	113
7.1. Limits on the hadronic RPV stop models	122
A.1. Boosted search: acceptance as a function of the $\overline{m}$ for the JES uncertainty estimate.	124
A.2. Boosted search: $\overline{m}$ distributions with different JES for the uncertainty estimate.	125
A.3. Boosted search: acceptance as a function of the $\overline{m}$ for the JER uncertainty estimate.	126
A.4. Boosted search: $\overline{m}$ distributions with different JER for uncertainty estimate.	126
A.5. Boosted search: acceptance as a function of the $\overline{m}$ for the pileup uncertainty estimate.	127
A.6. Boosted search: $\overline{m}$ distributions with different pileup profiles for the uncer- tainty estimate.	127
A.7. Boosted search: acceptance as a function of the $\overline{m}$ for the PDF uncertainty estimate.	128
A.8. Boosted search: $\overline{m}$ distributions with different PDFs for the uncertainty es- timate.	128
A.9. Boosted search: acceptance as a function of the $\overline{m}$ for the b tagging uncer- tainty estimate.	129
A.10. Boosted search: $\overline{m}$ distributions with different b-tagged scale factors for un- certainty estimate.	130
B.1. Resolved search: Average mass distribution as a function of the mass resolution.	132

B.2. Resolved search: Distributions of the different metrics studied in the dijet pairing method.	132
B.3. Resolved search: M_{asym} distributions for different dijet pairing methods. . .	133
B.4. Resolved search: $\Delta\eta_{\text{dijet}}$ distributions for different dijet pairing methods. . .	133
B.5. Resolved search: Δ distribution as a function of mass for different pairing methods in QCD multijets MC.	133
B.6. Resolved search: Δ distribution as a function of mass for different pairing methods in RPV stop signals at $m_{\tilde{t}} = 700 \text{ GeV}$ MC.	133
B.7. Resolved search: $\overline{M}$ distribution with partial selection for MC samples. . .	134
B.8. Resolved search: $\overline{M}$ distribution comparing different pairing methods in QCD multijets MC.	135
B.9. Resolved search: acceptance as a function of the $\overline{M}$ for the JES uncertainty estimate.	135
B.10. Resolved search: $\overline{M}$ distributions with different JES for the uncertainty estimate.	136
B.11. Resolved search: acceptance as a function of the $\overline{M}$ for the JER uncertainty estimate.	136
B.12. Resolved search: $\overline{M}$ distributions with different JER for the uncertainty estimate.	137
B.13. Resolved search: acceptance as a function of the $\overline{M}$ for the pileup uncertainty estimate.	138
B.14. Resolved search: $\overline{M}$ distributions with different pileup profile for the uncertainty estimate.	138
B.15. Resolved search: acceptance as a function of the $\overline{M}$ for the PDF uncertainty estimate.	139

B.16.Resolved search: $\overline{M}$ distributions with different PDFs for the uncertainty estimate.	139
B.17.Resolved search: acceptance as a function of the $\overline{M}$ for the b tagging uncertainty estimate.	140
B.18.Resolved search: $\overline{M}$ distributions with different b tagging scale factor for the uncertainty estimate.	140

Chapter 1

Introduction

Since the dawn of time, humanity has been trying to accurately described the world we live in. After many centuries of scientific endeavor, a fundamental theory describing how matter interacts in the universe has been formulated. The standard model (SM) of particle physics describes this phenomena through the interaction of elementary particles. The SM is one of the most successful theories ever formulated since many experimental results have been confirmed with astonishing precision. In 2012 a particle with the features of the Higgs boson, the last unseen piece of this theory, was discovered by the two main experiments (ATLAS and CMS) at the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN).

Despite the enormous success of the SM, there are still important unanswered questions. For instance, the different scales between the fundamental particles of matter or the force carrier particles, the value of the Higgs boson mass, the introduction of gravity into particle physics, or the source of astrophysical dark matter in terms of interacting particles. Several theoretical models beyond SM (BSM) have been formulated to solve these questions, supersymmetry (SUSY) being one of the leading extensions of the SM. SUSY is a new symmetry which transforms bosonic states into fermionic states, and vice versa, introducing a supersymmetric partner for each SM particle.

Assuming that some of these BSM particles can interact with SM particles, the particle physics community has performed multiple searches of these new particles in experiments like the Tevatron at Fermilab or the LEP and LHC at CERN. This dissertation describes one of these searches for new physics, studying particles decaying into quarks, using one of

the detectors at the LHC, the Compact Muon Solenoid (CMS).

The content of this dissertation is as follows:

In Chapter 2 the theoretical motivation of this work is presented. There, a brief introduction of the SM is described with special focus on the theory describing the interaction of quarks in the SM. In addition, the BSM and SUSY models are also briefly introduced in this chapter with a more elaborate description of the benchmark theoretical model used in this search.

The LHC and the CMS detector are described in Chapter 3, followed by the description of the physics object reconstruction in Chapter 4. This chapter stresses the concepts and features of how hadronic cascades are detected, reconstructed and further analyzed. Data and simulated samples used in this analysis are also presented in this chapter.

The search is performed in two mass ranges, using different techniques. In Chapter 5 the low mass boosted search is fully described, while Chapter 6 reports on the high mass resolved search. Results of both searches are presented in Chapter 7, with the statistical analysis used to interpret the results in the context of the SUSY benchmark model. The last chapter summarizes the searches and the results obtained.

Chapter 2

Theoretical overview

In this chapter the theoretical background for this analysis is presented. First, a brief introduction of the Standard Model (SM) of particle physics is described, with a special emphasis on hadron interactions. Next, a short description of models involving new physics scenarios is introduced, including the theoretical benchmark model used for this search. The chapter finishes with details about the simulated samples used to model the theory previously described.

2.1 The Standard Model of Particle Physics

The *Standard Model* (SM) of particle physics is our best understanding of how matter interacts in the universe. The SM is a relativistic quantum field theory describing the interactions between matter particles and particles mediating the electromagnetic, weak and strong forces [1]. Only three of the four forces are described by the SM; gravitational interactions are not currently described by the SM. In the context of this theory, elementary particles of matter with half-integer spin obeying the Pauli exclusion principle and Fermi-Dirac statistics are called fermions, while mediator particles with integer spin which are not subject to the Pauli exclusion principle and obey Bose-Einstein statistics are called bosons [1]. For instance, elementary particles of matter (leptons and quarks) are fermions, while the force carrier particles (photons, gluons, W and Z) are bosons.

Figure 2.1 shows all the particles described in the SM. Matter particles are divided into three generations. Particles belonging to the first generation are lighter, more stable

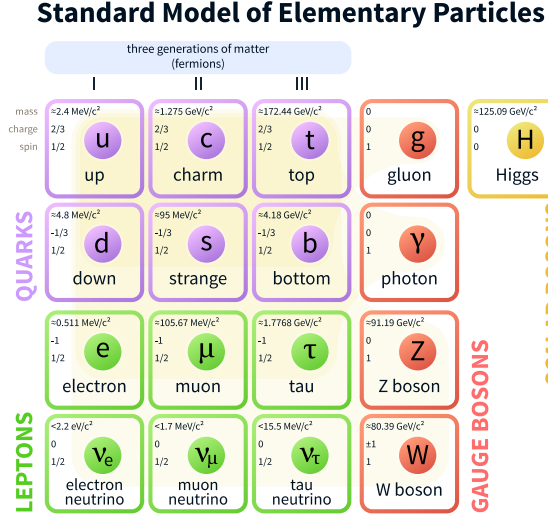


Figure 2.1: Fundamental particles of the SM: fermions, vector bosons and a scalar boson [2].

and they are the constituents of all the observable matter. On the other hand, particles of the second and third generation are heavier, less stable and they are found in highly energetic processes. Electromagnetic interactions between elementary charged particles and photons are described by quantum electrodynamics. Quantum chromodynamics describes the interactions between quarks and gluons, and allow the formation of composite states like hadrons. In the SM, the Higgs mechanism introduces a spontaneous symmetry breaking in the electroweak interaction [1]. This mechanism produces the massless photon, the carrier of the electromagnetic interaction, along with the massive W and Z bosons of the weak interaction. The weak bosons become massive through the interaction with the Higgs field, and the boson mediating the interaction of the Higgs field with itself is the scalar Higgs boson. Thereby, any elementary particle which interacts with the Higgs field becomes massive.

A Higgs-like particle was discovered by the ATLAS and CMS experiments at CERN in 2012 [3, 4]. More recent studies by both collaborations with more data and with more energetic collisions show that this particle has indeed the properties of the Higgs from the SM [5]. The latest measurement of the mass of the Higgs boson is $125.09 \pm$

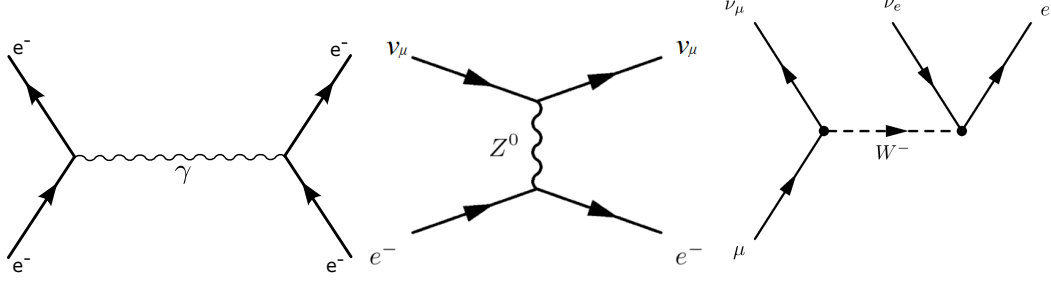


Figure 2.2: Examples of Feynman diagrams for some interactions in the SM. On the left: an electron annihilation and pair production of electrons. On the middle, neutral current exchange of Z boson. On the right, muon decay in a muon neutrino and through an exchange of a W boson an electron and electron neutrino.

$0.21(\text{stat.}) \pm 0.11(\text{syst.}) \text{ GeV}$ [5]. This mass is consistent with the electroweak constraints that the Higgs should be around the weak scale ($\sim 100 \text{ GeV}$) to explain the interactions between elementary particles.

In the group theory formalism, the SM can be described by the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$, where $SU(3)_C$ represents strong interactions and the group $SU(2)_L \times U(1)_Y$ describes the electroweak interactions. In the rest of this chapter, however, a brief description of the quantum field theory of the SM is presented in the Lagrangian formalism. In this formalism, particles and antiparticles are described by fields representing creation and annihilation operators, following this notation: spin-0 particles (bosons) are scalar fields $\phi(x)$, spin-1 particles (bosons) are vector fields $A_\mu(x)$ and spin-1/2 particles (fermions) are spinor fields $\psi(x)$.

2.1.1 Electroweak Interactions and Electroweak Symmetry Breaking

Weak interactions are mediated by the massive vector bosons W and Z. Due to the uncertainty principle, these massive particles must interact in short ranges of less than 10^{-18} m [1]. Leptons and quarks interact with these weak mediator bosons. For instance some electroweak interactions are shown Fig. 2.2: an electron annihilation and subsequently pair production of electrons mediated by a photon (left), the neutral exchange between leptons mediated by the Z boson (middle), and the beta decay or muon decay involving the exchange

of the W bosons (right).

Using the symmetry group $SU(2)_L \times U(1)_Y$, Glashow, Salam and Weinberg unified the electromagnetic and weak interactions in 1967 [1]. Based on this symmetry group, electroweak interactions can be described with gauge bosons W_μ^i , with $i = 1, 2, 3$ for $SU(2)$ and B_μ for $U(1)$ corresponding to gauge coupling constants g and g' [6]. Left-handed fermion fields of each i^{th} fermion family transforms under $SU(2)$ as doublets:

$$\Psi_i = \begin{pmatrix} \nu_i \\ l_i^- \end{pmatrix}, \quad \text{or} \quad \Psi_i = \begin{pmatrix} u_i \\ d_i' \end{pmatrix}, \quad (2.1)$$

where $d_i' = \sum_j V_{ij} d_j$, and V is the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix described later in this subsection. Right-handed fermions are described as $SU(2)$ singlets.

In order to generate mass, a complex scalar Higgs doublet is included in this model through the spontaneous symmetry breaking mechanism. Given the potential:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \frac{\lambda^2}{2} (\phi^\dagger \phi)^2, \quad (2.2)$$

where $\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix}$ is the Higgs doublet and, μ^2 and λ are constant parameters. These constants can be chosen to create ϕ to develop a vacuum expectation value (vev) of $|\phi_{vev}| = \nu = \sqrt{\frac{\mu^2}{2\lambda}}$. After the electroweak gauge symmetry is broken, the only physical remaining particle is the neutral scalar Higgs (H) with mass $M_H = \lambda\nu$.

The Lagrangian for fermion fields after symmetry breaking is given by [6]:

$$\begin{aligned} \mathcal{L}_F = & \sum_i \bar{\psi}_i \left(i\not{\partial} - m_i - \frac{m_i H}{\nu} \right) \psi_i - \frac{g}{2\sqrt{2}} \sum_i \bar{\Psi} \gamma^\mu (1 - \gamma^5) (T^+ W_\mu^+ + T^- W_\mu^-) \Psi_i \\ & - e \sum_i Q_i \bar{\psi} \gamma^\mu \psi_i A_\mu - \frac{g}{2 \cos \theta_W} \sum_i \bar{\psi}_i \gamma^\mu (g_V^i - g_A^i \gamma^5) \psi_i Z_\mu, \end{aligned} \quad (2.3)$$

where the raising and lowering operators T^+ and T^- are related to the weak isospin. In

addition, by defining the weak angle $\theta_W = \tan^{-1}(g'/g)$ and the electric charge $e = g \sin \theta_W$, one gets three boson fields:

- $\gamma = B \cos \theta_W + W^3 \sin \theta_W$ is the photon (γ) field with mass $M_\gamma = 0$,
- $W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2)$ are the charged weak fields (W) with mass $M_W = \frac{e\nu}{2 \sin \theta_W}$,
- $Z = -B \sin \theta_W + W^3 \cos \theta_W$ as the neutral weak field (Z) with mass $M_Z = \frac{M_W}{\cos \theta_W}$.

While the first term in Eq. (2.3) gives rise to fermion masses, the other terms represent the charged-current weak interaction (second term), electromagnetic interactions (third term) and weak neutral-current interactions (fourth term).

Finally, the masses and mixing of quarks arise from the Yukawa interaction with the Higgs field [6]:

$$\mathcal{L}_Y = -Y_{ij}^d \overline{Q}_{Li}^I \phi d_{Rj}^I - Y_{ij}^u \overline{Q}_{Li}^I \epsilon \phi^* u_{Rj}^I + h.c., \quad (2.4)$$

where $Y^{u,d}$ are 3×3 complex matrices, ϕ is the Higgs field, i, j are generation levels, ϵ is the 2×2 antisymmetric tensor, Q_L^I are the left-handed quark doublets, and d_R^I and u_R^I the right-handed down- and up-type quark singlets in the weak-eigenstate. In order to get physical quark states from Eq. (2.4), the complex matrices $Y^{u,d}$ are required to be diagonalized by four unitary matrices, labeled $V_{L,R}^{u,d}$, in the form of:

$$M_{diag}^f = V_L^f Y^f V_R^{f\dagger} \frac{\nu}{\sqrt{2}}, \quad (2.5)$$

where $f = u, d$. This process describes interactions between charged-currents and physical quarks (u_{Lj}, d_{Lk}) given by the couplings:

$$\frac{-g}{\sqrt{2}} (\overline{u}_L, \overline{c}_L, \overline{t}_L) \gamma^\mu W_\mu^+ V_{CKM} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} + h.c., \quad (2.6)$$

where the unitary quark mixing matrix is the Cabibbo-Kobayashi-Maskawa matrix V_{CKM} :

$$V_{CKM} = V_L^u V_L^{d\dagger} \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{ts} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \quad (2.7)$$

2.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the quantum field theory describing the strong interaction in the SM. Since this analysis is focused on final states with quarks, QCD is described in more detail in this section.

The theoretical background of this theory began in 1963 with Gell-Mann and Zweig trying to describe a fundamental principle of the interaction between nucleons and other strongly interacting particles [7]. The generic term for these particles is *hadrons*. They proposed a model where the interactions between hadrons could be described in terms of fundamental particles called quarks. In this theory hadrons with integer spin, called *mesons*, can be described as a composite bound state of a quark-antiquark pair. *Baryons* are hadrons with a half-integer spin and are composed of three quarks. For example, protons and neutrons are baryons while pions are mesons.

In order to describe the electric charge and other quantum numbers of hadrons, QCD assumes six types or *flavors* of quarks divided into two families: up (u), down (d), charm (c) and, strange (s), bottom (b) and top (t). Quarks are assigned fractional electric charge in order to form integer charge hadrons. Therefore, u , c , t have $+2/3$ electric charge while d , s and b have $-1/3$ electric charge, and their anti-particles $-2/3$ and $1/3$ respectively.

Quarks carry an additional quantum number called color. Color was introduced in the theory to keep quark wave-functions antisymmetric, in agreement with the Fermi-Dirac statistics [7]. Quarks are assigned three color indices, named *red*, *blue* and *green*, as a fundamental representation of an internal $SU(3)$ global symmetry. There are eight linearly independent generators of this symmetry group T_a called *gluons*. The index a runs from one to eight which in terms of the 3×3 Gell-Mann matrices (λ_a) can be represented as [8]:

$$T_a = \frac{\lambda_a}{2}. \quad (2.8)$$

The Lagrangian for QCD acting on one of the quark fermion fields $\Psi = (q_1, q_2, q_3)^T$, is given by [8]:

$$\mathcal{L}_{QCD} = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi + g_s\bar{\Psi}\gamma^\mu T_a\Psi G_\mu^a - \frac{1}{4}G_{\mu\nu}^a G^{a,\mu\nu}, \quad (2.9)$$

where the field tensor $G_{\mu\nu}^a$, involving eight gauge fields G_μ^a , are defined as:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f_{abc} G_\mu^b G_\nu^c. \quad (2.10)$$

In Eq. (2.10), the term f_{abc} describes the structure constants of the SU(3) group and g_s describes the strong coupling constant commonly expressed as $g_s = \sqrt{4\pi\alpha_s}$ in terms of the fine structure constant of the strong interaction (α_s).

In QCD, the fundamental parameters of the theory are the coupling constant g_s and the masses of the quarks. The mass of each quark is independent of the theory and has an electroweak origin, as discussed in Section 2.1.1.

Lattice and perturbative QCD

The main two approaches to calculate QCD physics related quantities: the so-called lattice QCD (LQCD) and perturbative QCD (pQCD) [9]. LQCD is the most complete approach, which places each quark on the sides of a hypercubic lattice of lattice spacing a , while gauge fields link each side of this discretized Euclidean space-time [6]. Thus, this quantum field theory is finite by using the lattice spacing as an ultraviolet regulator. LQCD is used, for instance, in the calculation of the matrix elements of the V_{CKM} and it has been successfully used to carry out several hadronic quantities like the quark masses and coupling constant [10].

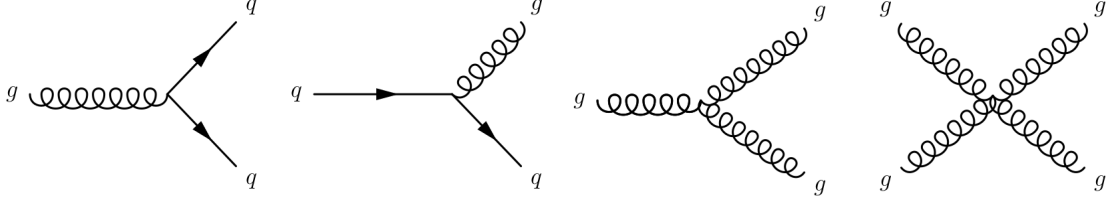


Figure 2.3: Examples of Feynman diagrams for some interactions in QCD. On the left: gluon splitting into two quarks. Middle left: quark radiating a gluon. Middle right: gluon splitting to two gluons. Right: two gluon interaction producing two gluons.

The complexity of the lattice approach is not suitable in all contexts. At hadron colliders for example, the complexity of solving hadronic interactions with lattice calculations would require gigantic lattices and several complex calculations. Fortunately, one can use the perturbative approach used in Quantum Electrodynamics (QED) to calculate observables at lower orders. Using pQCD, one is able to study the interactions of different colored particles using the corresponding Feynman rules. Some of the fundamental interactions allowed by Eq. (2.9) are represented as Feynman diagrams in Fig. 2.3.

Running constant and asymptotic freedom

Calculations of physical quantities, such as cross sections or decay rates, for QCD processes diverge beyond leading order, where all the interesting processes take place. Higher order terms in perturbation calculations are computed by a process called renormalization which allow us to absorb these divergences into a redefinition of fields, masses and coupling constants by expressing the renormalized coupling $\alpha_s(\mu_R^2)$ as a function of an energy scale parameter μ_R [6]. For a given process, the strength of the strong interaction can be inferred by taking μ_R close to the scale of the momentum transfer (Q).

This coupling satisfies:

$$\mu_R^2 \frac{d\alpha_s}{d\mu_R^2} = \beta(\alpha_s) = -(b_0\alpha^2 + b_1\alpha^3 + \dots), \quad (2.11)$$

where each β -function coefficient b_i is referred to as the i -th loop coefficient in the perturbation calculation. One of the main features of QCD is originated from the minus sign of Eq. (2.11). This sign shows that for processes involving large momentum transfer, i.e. at $\sim 100 \text{ GeV}$ scale, the coupling becomes weak, while for scales of the lower than $\mathcal{O}(\text{GeV})$ the coupling is stronger. This feature is called *asymptotic freedom* and it was discovered by Gross, Wilczek and Politzer in 1973 [1].

The b_i coefficients of Eq. (2.11) are given for an effective theory with light quarks, i.e. $m_q \ll \mu_R$, where heavier quarks are decoupled [6]. Considering an energy range where the number of quarks is constant, there is a simple exact analytic solution for Eq. (2.11) if all but the b_0 term are neglected. Then,

$$\alpha_S(\mu_R^2) = (b_0 \ln(\mu_R^2/\Lambda^2))^{-1}, \quad (2.12)$$

here Λ is a constant of integration, usually referred to as Λ_{QCD} . This constant is important in QCD because it corresponds to the scale when non-perturbative calculations diverge. Using Λ_{QCD} one can define light-quarks as quarks with $m_q \ll \Lambda_{QCD}$ and heavy-quarks as $m_q \gg \Lambda_{QCD}$.

Quark masses

After renormalization, the mass terms in the QCD Lagrangian of Eq. (2.9) highly depend on the renormalization scheme and the scale parameter μ_R . This is one of the reasons the quark masses are complex [6]. In QCD, the most widely used renormalization scheme is the so called modified minimal subtraction ($\overline{MS}$) scheme. At high energies and short distances, non-perturbative effects become small enough that quark mass computations can be treated with QCD perturbation theory using the $\overline{MS}$ scheme instead of the more cumbersome LQCD. For scattering process, however, it is common practice to neglect the mass of the quarks which are significantly smaller than the momentum transfer in perturbative QCD

calculations.

Confinement and proton structure

Another property of λ_S is that it becomes larger at large distances, making it impossible for colored charged particles to be isolated. This QCD feature is referred to as quark confinement. Confinement is reflected in the strong potential between color charges, which contains a Coulomb-like potential at short distances and a linearly increasing term at long distances. The potential for color charged particles creates color singlets in the form of mesons and baryons as described above.

This simplified model, where hadrons are described as bound states with two or three non-interacting quarks, has been proven to be too simple to describe the interactions between hadrons. Studies of deep inelastic scattering (DIS) processes of electrons and protons, for instance, have shown that protons can be more accurately represented as a sea of quarks, antiquarks and gluons, constantly interacting with each other [11]. As a result, the three quarks representing the proton are valence quarks embedded in this sea of bounded particles usually referred to as partons.

Additionally, the interactions between hadrons are introduced in the calculation of cross sections with hadrons in the initial-state by considering the most simple DIS between an electron and a proton, $ep \rightarrow e + X$. By denoting the four-momentum of the electron as k , the momentum of the exchanged photon as q and the momentum of the proton as p , one can formulate the differential cross section in terms of the photon is momentum $Q^2 \equiv -q^2$, $x = Q^2/(2p \cdot q)$ and $y = (q \cdot p)/(k \cdot p)$ [6]:

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha}{2xQ^4} \left[(1 + (1 - y)^2) F_2(x, Q^2) - y^2 F_L(x, Q^2) \right], \quad (2.13)$$

where α is the QED coupling and F_2 and F_L are proton structure functions encoding the interaction between the photon and the proton. These functions, or any other function

involving hadrons in the initial-state, are not calculable using the pQCD approach but can be given to zeroth order in α_S in terms of the non-perturbative parton distribution functions (PDF).

PDFs represent the probability density of finding a parton carrying a fraction of the proton momentum (x) at an energy scale Q^2 [11]. These functions are obtained experimentally by fitting the cross section of many experiments in a grid of Q^2 and x [12, 13, 14]. An example of these PDFs at next-to-next-to leading order (NNLO) are shown in Fig. 2.4, where two PDFs are shown for different values of Q^2 . According to the PDF functions, the three valence quarks in the proton dominate at low values of Q^2 while at higher values a sea of low momentum quark-antiquarks pairs are more predominant. Another interesting feature demonstrated by proton PDFs is that quarks and antiquarks on average only carry about half of the momentum of the proton, the other half is carried by the interacting gluons inside the proton.

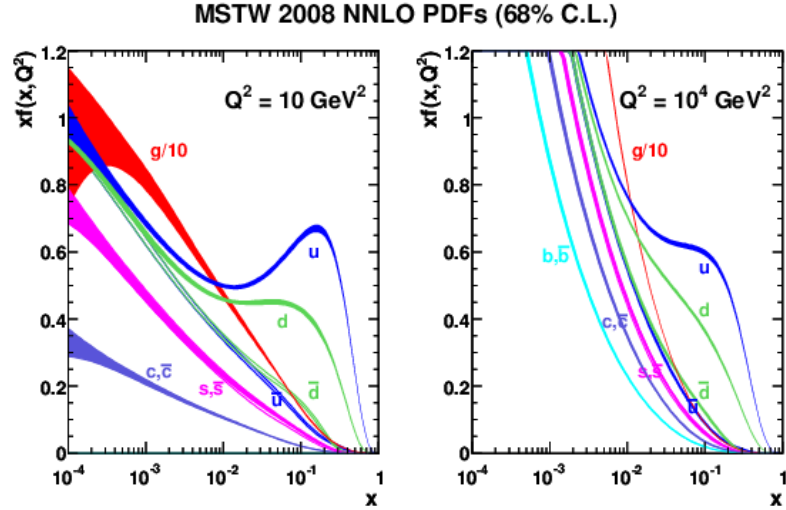


Figure 2.4: Examples of parton distribution functions for the MSTW group at NNLO at $Q^2 = 10 \text{ GeV}^2$ (left) and $Q^2 = 10^4 \text{ GeV}^2$ [13] (right).

Parton showers and hadronization

High energy parton interactions involve large momentum transfers to highly accelerated partons[15]. In the same way that accelerated electrically charged particles emit photons as QED radiation, accelerated colored particles emit gluons as QCD radiation. Since gluons themselves carry colored charge, they can further emit more colored particles until all the energy carried by the partons is just enough to create the lightest colored neutral hadrons. This cascade of colored particles is referred to as *parton showers* and the process when high energy colored particles form hadrons is called *hadronization*. Experimentally, these parton showers are represented as physical objects called *jets*, which are extensively described later in Section 4.1.3.

The hadronization process is a long-distance process which involves low momentum transfers [16] and therefore cannot be treated using perturbation theory. In order to make QCD predictions and simulate QCD behaviors, the following two models are most commonly used:

String model This model is based on LQCD simulations where the potential energy of colored particles increases linearly with the separation at large distances [16]. The string model is used in many of the current Monte Carlo simulators used in hadron colliders, such as the PYTHIA generator described later in Section 4.2.

Cluster model The main assumption of this model is that colored singlet combinations of particles in the parton shower form clusters [16]. These clusters eventually isotropically decay into pairs of hadrons according to the density of states with appropriate quantum numbers. The event generator HERWIG, also described in Section 4.2, uses this hadronization approach.

2.2 Beyond the Standard Model and Supersymmetry

Despite the incredible success of the SM describing most of the particle physics phenomena around us, it cannot explain some fundamental quantities or incorporate other interactions in the model. For instance, the SM does not predict: 1) the mass of any fundamental particle, 2) the reason behind the extreme differences in mass between fermionic generations, 3) the reason why fermions come in exactly three generations. In addition, fundamental particles are treated in an adhoc way in the SM as indivisible based on empirical observations but there is nothing in the theory that predict this behavior. The SM also does not provide any explanation for the observed asymmetry between matter and antimatter, nor any particle candidates to explain dark matter (DM) interactions observed in astrophysics. Additionally, the SM cannot incorporate the description of gravity from general relativity into the model.

The issue related with the mass difference between fermions is commonly known as the *hierarchy problem*. The mass of the recently discovered Higgs-like particle at the weak scale is consistent with empirical predictions from electroweak measurement constraints to explain the interactions between elementary particles. However, the Higgs mechanism also needs a non-vanishing vacuum expectation value which can only be achieved if the mass of the Higgs boson is of the order of the Planck scale (10^{18} GeV) or if a large quantum corrections from virtual particles coupled to the Higgs field are applied [17]. Calculations of these quantum corrections shows that their energy should be 30 orders of magnitude larger than the mass of the Higgs boson. The process to adjust the theory with these corrections to match experimental observations is referred to as *fine-tuning*.

The most popular and arguably the most elegant method to solve the hierarchy problem is with the introduction of a new set of particles which systematically cancel all contributions to these corrections. Consequently, it strongly suggests a new unknown symmetry related

to fermions and bosons, called *supersymmetry* (SUSY). This model requires a supersymmetric partner, or *sparticle*, for fermion and boson in the SM which neatly cancels any SM contribution to the loop corrections in the Higgs field calculation. Sparticles stabilize the weak scale if they are significantly coupled to the Higgs boson. Supersymmetric partners of SM leptons are called *sleptons*, sparticles of quarks are *squarks*, sparticles of SM bosons are the *photino*, *gluino*, *wino*, *zino* and *higgsino* for the Higgs. Winos and charged higgsinos produce bound states called *charginos* while photinos, zinos and the neutral higgsino produce neutral particles called *neutralinos*.

SUSY sparticles are expected to be heavier than their SM counterparts based on current experimental constraints [6]. In addition, the recently measured mass of the Higgs introduces a natural scale for the SUSY partners of the heaviest SM particles, i.e. particles strongly coupled to the Higgs. *Naturalness*, namely the property suggesting physical quantities not to vary for several orders of magnitude, in SUSY suggest to find the mass of the stops and sbottoms below the TeV scale. Even though naturalness avoids the problem of fine-tuning, it is not a condition required by the theory.

Additionally, SUSY models propose a discrete symmetry called R-parity associated with a Z_2 symmetry group. This symmetry introduces a quantum number defined as $R = (-1)^{3(B-L)+2S}$ where B is the baryon number, L is the lepton number and S is the spin. R-parity is introduced in SUSY models to conserve baryon and lepton numbers. All SM particles have even R-parity while all sparticles have odd R-parity. If R-parity is exactly conserved, odd R-parity sparticles cannot decay into only even R-parity particles, prohibiting the mixing of particles and sparticles. Furthermore, phenomenological studies have shown that R-parity conservation (RPC) demands an even number of sparticles in each interaction vertex as well as a pair production of sparticles in collider experiments [17]. R-parity conservation is also motivated in SUSY models to avoid proton decays.

Besides these features, RPC includes an important extra component to SUSY. If the

lightest particle with even R-parity i.e., the lightest supersymmetric particle (LSP), is electrically neutral and stable, its interaction with matter is weak. Therefore, the LSP may become a good candidate for DM, possibly solving one of the most fundamental problems in modern astrophysics, the source of DM.

RPC incorporates some important features in SUSY and has been widely explored in LHC searches but it is only phenomenologically motivated [18]. Unfortunately, those searches have so far not discovered any RPC SUSY signatures, opening the opportunity to explore scenarios where R-parity is violated (RPV).

2.2.1 R-parity Violating Scenarios

In the absence of R-parity some SUSY models allow the following supersymmetric Yukawa interactions as a superpotential W_{R_p} in the lagrangian [19]:

$$W_{R_p} = \frac{1}{2}\lambda_{ijk}L_iL_jE_k^c + \lambda'_{ijk}L_iQ_jD_k^c + \frac{1}{2}\lambda''_{ijk}U_i^cD_j^cD_k^c + \mu_iH_uL_i, \quad (2.14)$$

where the indices i, j, k are quark and lepton generation indices following the summation convention, while the superindex c denotes the charge conjugation. The trilinear λ_{ijk} couplings permit vertices of sfermions interacting with two fermions and the bilinear μ_i coupling represents a dimensional mass parameter. L_i are the left-handed lepton doublets, E_i the right-handed lepton doublets, Q_i the left-handed quark doublets, U_i and D_j are right-handed quarks, and H_u is the Higgs that gives mass to the up-type quarks. Such a superpotential contains terms which violate lepton number (1st and 2nd term in Eq. (2.14)) or baryon number (4th term), leading to a rapid decay of protons [18]. Assuming RPC the coupling constants of these terms would vanish or they are sufficiently small assuring a lifetime of the proton compatible with the SM.

Under R-parity violation, the coupling of the hadronic term of the potential in Eq. (2.14) (λ''_{ijk}) or the leptonic term (λ_{ijk}) or the mixture of the two (λ'_{ijk}) may be non-zero. For

instance, the hadronic RPV term would produce decays of squarks into multiple quarks in the final state with no invisible neutral particles in the form of missing energy or leptons, while a leptonic RPV term could yield RPV sleptons decaying into a pair of leptons.

Nevertheless, if R-parity is violated, one may still address the issue of naturalness since the spectrum of masses in RPV particles is quite similar to RPC [20]. Moreover, because RPV models violate B or L only, protons cannot decay rapidly unless λ' and λ'' are both greater than zero. On the other hand, these models eliminate the candidate for DM since the LSP may not be stable and would decay to SM particles [20]. Including a DM candidate in SUSY is ideal but is not an essential feature of the model.

2.3 Theoretical benchmark model

The pair production of top squarks (stop, $\tilde{t}$) decaying to light-flavor quarks via λ''_{312} ($\tilde{t} \rightarrow qq'$), and light-flavor quark plus a b -quark via λ''_{323} ($\tilde{t} \rightarrow bq'$), via hadronic RPV is used as the benchmark model for this analysis. The three numerical subscripts of λ'' refer to the quark generations of the corresponding d-s quarks. Figure 2.5 presents the two corresponding Feynman diagrams of the models studied. This analysis is divided into two searches exploiting different decay features.

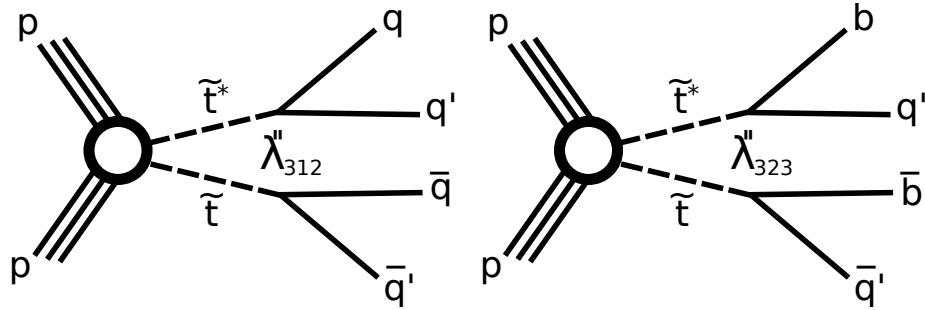


Figure 2.5: Diagram for the benchmark models used in this analysis: pair production of stops decaying via the hadronic RPV coupling λ''_{312} into two light-flavor quarks (left), and to a light-flavor quark and a b -quark via the λ''_{323} coupling (right).

A previous search for $\tilde{t} \rightarrow qq'$ via RPV was performed by CDF at the Fermilab Tevatron [21] and placed 95% confidence level (CL) limits on the production cross section as a function of mass, excluding masses between $50 \leq m_{\tilde{t}} \leq 100$ GeV. Similar searches have been performed at the CERN LHC: CMS [22] excluded masses between $200 \leq m_{\tilde{t}} \leq 350$ GeV at 8 TeV, while the ATLAS exclusion [23] ranges between $100 \leq m_{\tilde{t}} \leq 410$ GeV at 13 TeV. For the $\tilde{t} \rightarrow bq'$ scenario, mass exclusion limits at 8 TeV have been reported by CMS [22] between $200 \leq m_{\tilde{t}} \leq 385$ GeV, and by ATLAS [24] between $100 \leq m_{\tilde{t}} \leq 310$ GeV (boosted topology), and at 13 GeV ATLAS [23] set limits between $100 \leq m_{\tilde{t}} \leq 470$ GeV and $480 \leq m_{\tilde{t}} \leq 610$ GeV.

Table 2.1: Summary of previous limits for the benchmark RPV SUSY models.

Decay	Experiment	$\sqrt{s}$ [TeV]	Search	Mass exclusion
$\tilde{t} \rightarrow qq'$	CDF [21]	1.96	Resolved	$50 \leq m \leq 100$ GeV
	CMS [22]	8	Resolved	$200 \leq m \leq 350$ GeV
	Atlas [23]	13	Resolved	$100 \leq m \leq 410$ GeV
	CMS [25]	13	Boosted	$80 \leq m \leq 240$ GeV
$\tilde{t} \rightarrow bq'$	CMS [22]	8	Resolved	$200 \leq m \leq 400$ GeV
	Atlas [24]	8	Boosted	$100 \leq m \leq 315$ GeV
	Atlas [23]	13	Resolved	$100 \leq m \leq 470$ GeV

Chapter 3

Experimental setup

In this chapter the experimental apparatus used in this thesis is described, as well as the experimental techniques to analyze the collected data. First, a brief description of the Large Hadron Collider is introduced, followed by the Compact Muon Solenoid experiment and its different subdetectors. A summary of the procedure used to collect and measure the data taken is also presented, as well as a quick review of the treatment of the datasets.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the biggest and most energetic particle accelerator ever built. It is part of the CERN accelerator complex located on the Swiss-French border near Geneva, Switzerland. It was built in the tunnel previously constructed for the Large Electron-Positron Collider (LEP), a 27 km underground ring 45 m to 170 m deep [26]. The LHC was designed to collide two proton beams nearly head on in opposite directions at a center-of-mass energy of 14 TeV, as well as proton-lead, lead-lead ions and other ions with lower energy. The current maximum center-of-mass energy reached by the LHC is 13 TeV.

Proton beams at the LHC are not continuous, instead they are made of bunches travelling at time intervals of 25 ns. Each beam contains at most 2808 bunches, where each bunch consists of approximately 115 billion protons. To reach these high energies, the CERN accelerator complex, shown in Fig. 3.1, accelerates the proton beams in different stages. The first stage occurs in a linear accelerator called LINAC 2, where protons previously separated from electrons originating from a bottle of hydrogen, are accelerated to reach 50

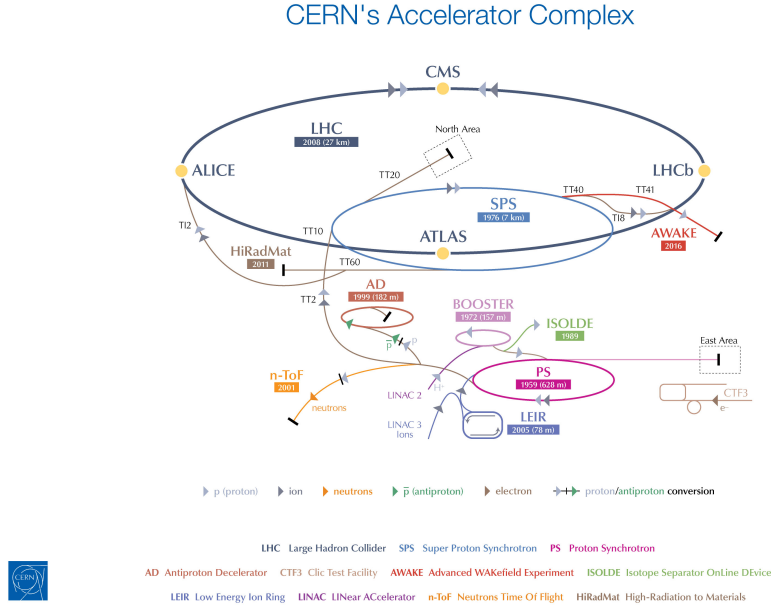


Figure 3.1: The CERN accelerator complex [27].

MeV. The LINAC 2 is the only linear accelerator in the proton's path to the LHC. Protons in the LINAC 2 are connected to a circular accelerator called Proton Synchrotron Booster (PSB, or Booster only) where they are accelerated to energies up to 1.4 GeV. Once they reach that energy, beams are injected into the Proton Synchrotron (PS) and further to the Super Proton Synchrotron (SPS), reaching energies of 26 GeV and 450 GeV, respectively. Finally, proton beams are injected into the LHC ring in two pipes where the beams can be accelerated in opposite directions to a maximum of 7 TeV.

Beams are contained in a single beamline inside the LHC tunnel. The beamline accommodates two separate pipes which are used to accelerate the beams in opposite directions around the ring. In order to keep these two beams running in the same beamline, the LHC uses superconducting dipole magnets for beam circulation and superconducting quadrupole magnets to focus the beams. There are 1200 superconducting magnets in the LHC ring made of superconducting niobium-titanium (NbTi) which can bear magnetic fields of up to 9-10 Tesla. These superconducting magnets require extremely low temperatures to work. Therefore, the LHC cryogenic system uses superfluid helium-4 to cool down the magnets

to a temperature of 1.9 K. The two beams intersect in four different points around the LHC beam, and four different detectors are built at each interaction point around the ring: ATLAS, CMS, LHCb and ALICE. The data used in this thesis were collected with the CMS detector.

Luminosity

The number of collisions in the LHC is measured by the instantaneous luminosity, given by:

$$\mathcal{L} = \frac{kN^2f}{4\pi\sigma_x^*\sigma_y^*}, \quad (3.1)$$

where k is the number of bunches, N is the number of protons per bunch, f is the revolution frequency (approximately 11.24 kHz), and σ^* is the size of the beam at the collision point ($\sigma_x^* = \sigma_y^* \approx 16\mu\text{m}$). The designed luminosity of the LHC is $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, however on June 29, 2016 the LHC reached a peak luminosity exceeding this value [28]. Every bunch crossing is referred to as an *event*. The number of collision events (N_{events}) is related with the integrated luminosity (L) by $N_{\text{events}} = L\sigma_{\text{events}}$, where L is measured in units of inverse barns (b^{-1}) and the cross section σ_{events} is measured in units of barns (b , where $1b = 10^{-24} \text{ cm}^2$). As shown in Fig. 3.2, the LHC delivered a total integrated luminosity of 41.07 fb^{-1} while the CMS Experiment recorded 37.82 fb^{-1} .

3.2 The CMS experiment

The Compact Muon Solenoid (CMS) [29] is one of the two general-purpose detectors built around the LHC ring, the other being ATLAS. It is located on the north side of the ring at one of the interaction points, close to the French city Cessy near the Swiss-French border. CMS is a 14000-ton detector, has a total diameter of 15 m, and an overall length of 28.7 m in a cavern located 100 m underground. The detector is composed of subdetectors built around the proton collision point. A schematic of the CMS detector is shown in Fig. 3.3.

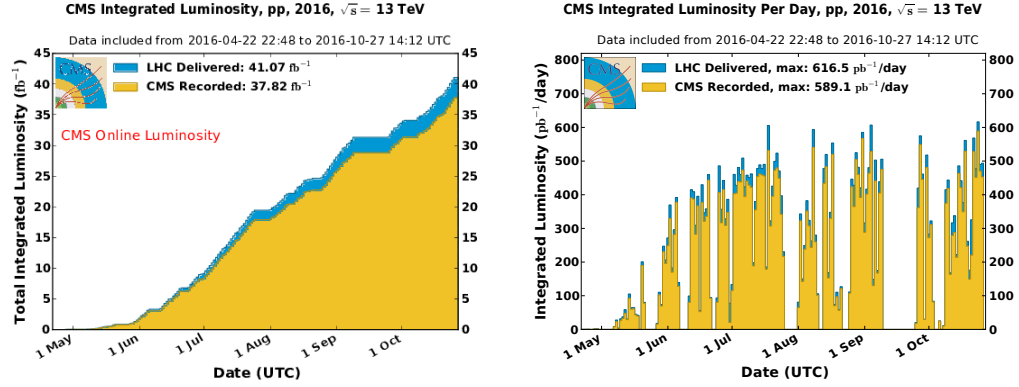


Figure 3.2: Left: Cumulative luminosity measured online versus day delivered to (blue), and recorded by (orange) CMS during stable beams and for p-p collisions at 13 TeV center-of-mass energy in 2016. The delivered luminosity accounts for the luminosity delivered from the start of stable beams until the LHC requests CMS to turn off the sensitive detectors to allow a beam dump or beam studies [28]. Right: Cumulative measured luminosity versus day delivered. This measurement uses the best available offline measurement and calibrations [28].

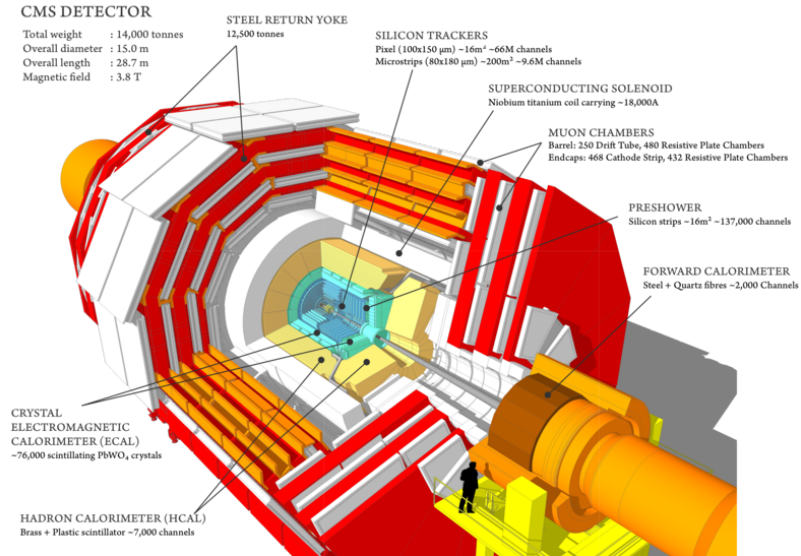


Figure 3.3: Overview of the CMS detector [30].

The central feature of the CMS detector is its superconducting solenoid. It has an internal diameter of 6 m and provides a magnetic field of 3.8 T. Several subdetectors are located within the volume of the solenoid: a silicon pixel and strip tracker, a lead tungstate electromagnetic calorimeter (ECAL), and a hadron calorimeter (HCAL). Outside the solenoid, gas-ionization detectors embedded in the steel return yoke measure the interactions of muons. Extended forward calorimetry complements the coverage provided by the barrel and endcap detectors. Each component of the CMS apparatus is described in the following sections.

Detector coordinates and event variables

The coordinate system used by CMS is centered of the geometrical center of the CMS detector. From this point, the z -axis is located along the direction of the beampipe, the positive x -axis points toward the center of the LHC ring and, the positive y -axis is perpendicular to the x - z plane. Two angles are defined in a cylindrical coordinate system: an azimuthal angle ϕ with respect to the x -axis in the x - y plane and, the polar angle θ with respect to the z -axis.

Due to the high energies of the collisions at the LHC, we consider relativistic invariant coordinates. A Lorentz invariant quantity, *rapidity* (y) given by:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \quad (3.2)$$

where E is the energy and, p_z is the momentum of the particle in the z direction, is defined to take into account Lorentz boosts along the z -axis. In the limit when the mass of a particle goes to zero, y can be approximated as a variable called *pseudorapidity* (η) defined as:

$$y \approx -\ln \left(\tan \frac{\theta}{2} \right) = \eta. \quad (3.3)$$

A value of $\eta = 0$ is perpendicular to the beamline, while $\eta \approx 5$ points along the direction

of the beamline. The direction of a particle in the detector usually is referred to in terms of this angle η .

In addition, it is useful to define variables independent of the z -boost of the colliding particles. Therefore, we define variables in the transverse plane x - y , such as the transverse momentum and transverse energy of a particle:

$$p_T = p \sin \theta, \quad E_T = p \sin \theta. \quad (3.4)$$

Other event variables defined in the transverse plane are the hadronic activity of the event (H_T), the transverse energy (E_T), and the energy imbalance, named missing transverse energy (E_T^{miss}), are given by:

$$H_T = \sum_{j_i} p_T^{j_i}, \quad E_T = \sum_{j_i} E_T^{j_i}, \quad E_T^{\text{miss}} = - \sum_i p_T^i \quad (3.5)$$

where j_i refers to jets described in Section 4.1.3 and i in E_T^{miss} refers to all the particles in the event.

3.2.1 Superconducting Solenoid Magnet

The superconducting solenoid magnet [31] in CMS is a cylindrical coil of superconducting fibers made of niobium-titanium. The solenoid has a 13 m length, an inner diameter of 6 m and it is capable of generating a magnetic field of up to 4 Tesla. The purpose of the strong magnetic field is to bend charged particles coming from the collisions in order to make precise measurements of the momentum of the particles. The inner tracker and the calorimeters are inside the volume of the magnet. The magnetic field outside the solenoid is contained by a 10000-ton iron return yoke of 1.5 m width interlaced with the muon chambers. The superconducting solenoid in CMS was designed to achieve a momentum resolution of 10% for muons with a transverse momentum of 1 TeV.

3.2.2 Tracking System

The tracking system is the closest subdetector to the interaction point, and therefore the most exposed to radiation. It is divided in two parts: the pixel tracker and the strip tracker, both using silicon sensors. The tracking system covers an area up to $|\eta| < 2.5$. This system measures the trajectory of charged particles by reconstructing their positions in different layers. The tracking system allows the reconstruction of muons, electrons and charged hadrons. In particular, the tracking detector is necessary for the identification of long-lived heavy-flavor hadron decays by differentiating particles coming from the secondary vertex, described in detailed in Section 4.1.3. The tracking system was designed to measure a momentum resolution of tracks of 1.5% for $p_T = 100$ GeV charged particles.

Pixel tracker

The first three layers going outwards from the interaction point correspond to the pixel tracker. A graphical representation of this detector is shown in Fig. 3.4 (left). This subdetector has a three circular layer structure in the barrel regions and the endcap disks on each side [32].

The barrel pixel detector is 53.3 cm long and each barrel layer is located at a distance of 4.4 cm, 7.3 cm and 10.2 cm respectively. Each layer is composed of modular detector units named *modules* supported with carbon fibers called *ladders*. There are eight modules in each ladder, and each layer contains 20, 32 and 44 ladders, respectively. The barrel pixel tracker is composed of a total of 1440 silicon pixel modules, and each module contains 66560 pixels. Each pixel module consists of thin, segmented silicon sensors with highly integrated readout chips (ROC). A schematic of the composition of the modules is shown in Fig. 3.4 (right). Each ROC serves a 52×80 array of pixels and each pixel has a length of $150 \mu\text{m} \times 150 \mu\text{m}$. A total of 47M pixels are installed in the CMS barrel pixel detector.

The four endcap disks are arranged in a turbine-like geometry at distances $z = \pm 35.5$ cm and $z = \pm 48.5$ cm. Each disk is split into half-disks to include 12 trapezoidal blazes. There

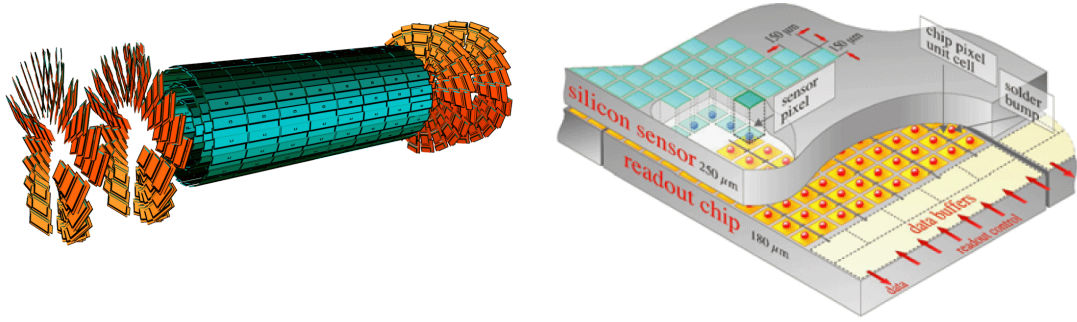


Figure 3.4: Left: sketch of the CMS pixel detector [33]. Right: overview of the different components of the silicon sensors [34].

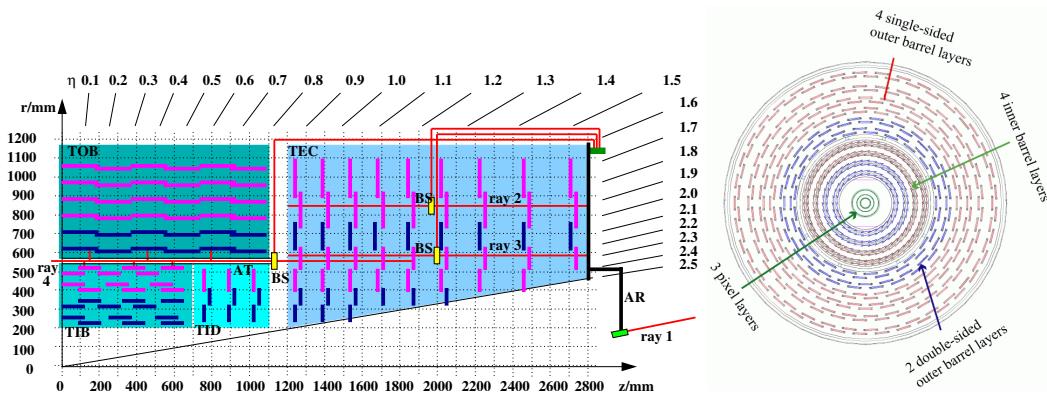


Figure 3.5: Overview of the CMS tracker system. Left: Longitudinal view [33]. Right: perpendicular view [35].

are seven different modules in each barrel and a total of 672 pixel modules in each disk. A total of 17M pixels are included in the endcap pixel detector.

Silicon strip tracker

The next ten layers of the CMS tracker system corresponds to the silicon strip detector. A longitudinal and perpendicular view of the CMS tracker system is shown in Fig. 3.5. The silicon strip tracker (SST) is 5.5 m in length and has a 2.4 m diameter corresponding to an active area of 198 m^2 , the largest silicon detector ever built [36]. The ten layers are structured in two barrels: the tracker inner barrel (TIB) and the tracker outer barrel (TOB). These two barrels are enveloped by a group of wheels per side: the tracker inner disk (TID)

corresponds to the three wheels on each side of the TIB, whereas the tracker endcap (TEC) is composed of 9 wheels on each side of the TOB. The SST has 15148 modules in total and some of them are mounted in a back-to-back configuration called *stereo*. These double-sided modules allow a more momentum resolution of tracks of 1.5% for charged particles with $p_T = 100 \text{ GeV}$.

3.2.3 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is the next detector going outwards from the interaction point after the tracker, and it is designed to accurately measure the energy of electrons and photons. Highly energetic electrons lose energy in matter primarily by emitting photons (bremsstrahlung), while high energy photons convert into electron-positron pairs (e^+e^-) [6]. These two processes (bremsstrahlung and γ conversion) create a subsequent cascade of lower energy electrons and photons which interact with matter mostly by ionization in the case of electrons and the photoelectric effect in the case of photons. To characterize the amount of matter traversed by photons and electrons, we use the quantity called *radiation length* (X_0). It is defined as the mean-distance traversed by a high-energy electron after it loses all but $1/e$ of its energy by bremsstrahlung or, as $7/9$ of the mean-free-path of the electron-positron pairs created by high-energy photons [6]. The *mean-free-path* is defined as the average distance traveled by a particle between collision with another particle. Another quantity used to characterize electromagnetic showers is the Molière radius, defined as the radius of a cylinder containing 90% of the average energy deposition of the shower.

The CMS ECAL is an homogeneous calorimeter made of lead tungstate ($PbWO_4$) crystals [33]. A photo of these crystals before installation is shown in Fig. 3.6 (left). CMS selected $PbWO_4$ crystals for the ECAL due to their short radiation length ($X_0 = 0.89 \text{ cm}$) and small Molière radius (2.2 cm). In addition, lead tungstate crystals emit 80% of light in 25 ns and are radiation hard (up to 10 Mrad). All these properties results in a radiation

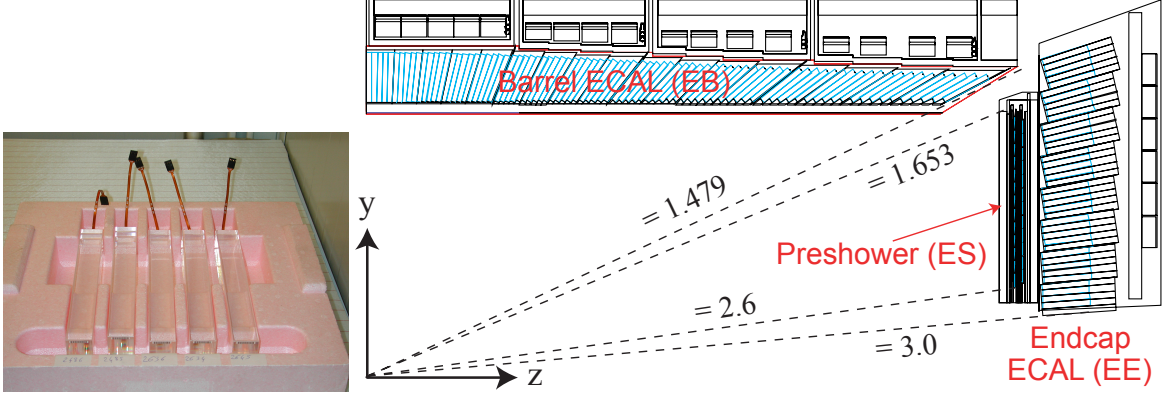


Figure 3.6: Left: Lead tungstate crystals [37]. Right: Overview of the CMS electromagnetic calorimeter [33].

resistant ECAL with fine granularity.

The ECAL is divided in two regions: barrel and endcap, as shown in Fig. 3.6 (left). Covering an $|\eta|$ -region below 1.479 the barrel ECAL (BE) is composed of 61200 crystals, whereas the endcap ECALs (EE) covers a region of $1.479 < |\eta| < 3.0$ and contain 7324 crystals. The BE has an inner radius of 129 m, structured as 36 identical supermodules. Each crystal in the BE is tilted by 3° with respect to the line from the nominal vertex position, covering a region in $\Delta\eta$ - $\Delta\phi$ space of 0.0174 (this region corresponds to an angle of 1°). The length of the crystals in the BE is 230 mm (or $25.8 X_0$) and have a front-face cross section of $22 \times 22 \text{ mm}^2$. The EE are located at a z -distance of $\pm 314 \text{ cm}$, consisting of semi-circular aluminium plates from which a *supercrystal* (a structure of 5×5 crystals) is cantilevered.

The energy resolution of an electromagnetic calorimeter is given by[33]:

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

where σ is the resolution, E is the energy, and S , N , and C are constants. The first term describes statistic-related fluctuations, the second term describes electronic noise and the last term represents a constant describing the non-uniformity of the detector and the uncertainty in the calibration. In CMS, the constants of (3.6) are: $S = 2.8\%$, $N = 12\%$, and

$C = 0.3\%$ [38]. The resolution of the CMS ECAL is optimized for particles with energies above 1 GeV, and corrections made as a function of time for the crystals exposure to high radiation.

3.2.4 Hadronic Calorimeter

The final detector inside the solenoid magnet is the hadronic calorimeter (HCAL), which measures the energy deposits of hadrons and is also crucial for determining transverse missing energy. Heavy particles, which are not stopped by the ECAL, are stopped by the dense material of the HCAL; by losing their energy in inelastic collisions with the electrons of the material. To quantify loss of energy by particles travelling in the HCAL material, we use the quantity called *interaction length* (λ_I). It is defined as the distance after an incident particle has lost $1/e$ of its energy [6].

The HCAL is a sampling calorimeter made of brass, as an absorbing material, which has $\lambda_I = 16.42$ cm, and plastic scintillators as the active material. Its design maximizes the material inside the magnet in terms of interaction length, minimizing the non-Gaussian tails of the energy resolution and providing a good containment and hermeticity for the E_T^{miss} measurement. The CMS HCAL consists of a barrel (HB) covering an $|\eta|$ -region below 1.3, endcaps (HE) in a region $1.3 < |\eta| < 3.0$, an outer barrel (HO) and forward calorimeters (HF) covering a region $2.9 < |\eta| < 5$. The HB consists of 36 identical wedges with a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. Highly energetic hadrons passing the HB and escaping the magnet region are stopped by the HO, which is the only part of the HCAL outside the solenoid. The HO contains 10 mm scintillators covering a region of $|\eta| < 1.26$, increasing the effective thickness of the HCAL to over $\lambda_I = 10$ cm. The different parts of the CMS HCAL are shown in a longitudinal view in Fig. 3.7 (left).

The resolution of hadronic showers is measured as a combination of ECAL and HCAL resolutions:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{100\%}{\sqrt{E}}\right)^2 + (5\%)^2, \quad (3.7)$$

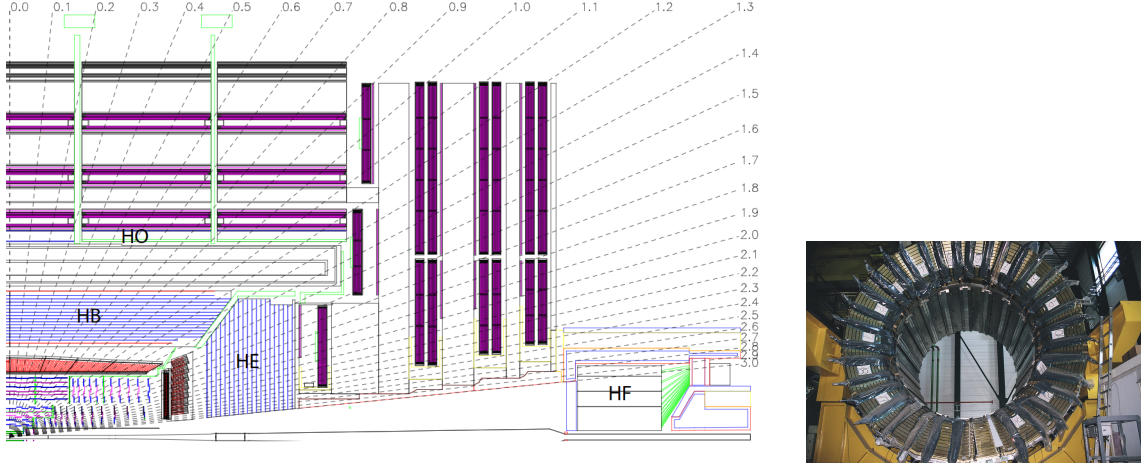


Figure 3.7: Left: Longitudinal view of the CMS HCAL detector [29]. Right: Photo of the CMS HCAL barrel before installation [39].

for the barrel region, where the terms here are similar as to Eq. (3.6).

3.2.5 Muon system

Muons are charged leptons heavier than electrons and are minimum ionizing particles which interact weakly with matter and therefore they interact very little with the calorimeters. The muon system in CMS is a set of gaseous detectors in the outermost part of the whole apparatus after the magnet. Muons travelling through gas ionize atoms which generate an electric field through a voltage difference in cathodes and anodes. The presence of signal is indicated by measuring the negatively charge electrons and the positively charge ions.

Figure 3.8 shows the three different gaseous detectors used to identify and measure muons in CMS. In the barrel region, covering $|\eta| < 1.2$, drift tube chambers (DT) are used, consisting of 250 chambers organized in four concentric cylinders inside the magnet return yoke. The DT are multi-wire chambers where a electric field is configured around a thin wire immersed in a mixed gas of argon and carbon-dioxide. In this region the rate of muons is low due to a small neutron induced background and the magnetic field in the chambers is low [33]. Cathode strip chambers (CSC) cover a region of $0.9 < |\eta| < 2.4$ where the neutron induced background and the magnetic field is high. There are 468 CSC installed

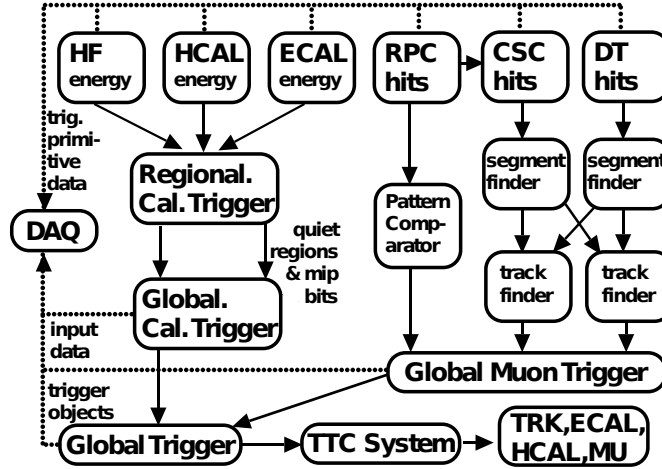


Figure 3.9: Overview of the CMS L1 trigger system [40].

Level-1 trigger

The first level of the CMS trigger system is a hardware-based system with fixed latency [40], referred to as the Level-1 trigger (L1). The total time allocated for the transit of the signals from the front-end electronics to where the L1 trigger logic is located and back to the front-end is $3.2\,\mu\text{s}$. In this time a decision must be made of keeping or discarding an event per bunch crossing. As shown in Fig. 3.9, the Level-1 trigger uses the energy information of the calorimeters and the hit information of the muon systems. The decisions are based on *trigger primitive* (TP) objects, such as photons, electrons, muons or jets with some kinematic thresholds in addition to event variables, like H_T or E_T^{miss} . After the TP are processed in several steps, the combined information is evaluated in a global trigger (GT) where the final decision is taken. The Level-1 trigger is designed to reduce the rate of data to 100 kHz.

The L1 calorimeter system is divided in two: a regional calorimeter trigger (RCT) and a global calorimeter trigger (GCT). The transverse energy information and the quality flags from 8000 ECAL and HCAL towers are processed in parallel for the RCT. The outcome of the RCT is electron/photon (e/γ) candidates, identified by the energy ratio between ECAL and HCAL, and regional E_T sums based on 4×4 calorimeter towers. These outputs are

used by the GCT to sort the e/γ candidates and to form jets using the E_T sums, as well as event quantities such as E_T^{miss} . After these objects are calculated, the GCT sends only four isolated and non-isolated e/γ candidates in three different categories (central, tau and forward jets) to the GT for the final decision.

The L1 muon system uses the information of the three muon detectors to ensure good coverage and redundancy. The first level of reconstruction comes from the muon DT and CSC, where the track of the muons is reconstructed by using the hit information from different layers of the muon chambers, and transverse momentum of the muon candidate is then assigned. These processes are performed in separate track finders which, after sharing some information between the DT and CSC, sends the four highest p_T candidates to the global muon trigger (GMT) to make the trigger decision. The RPC muon system reconstructs the tracks of the muon candidates directly and sends a maximum of four muon candidates in the barrel and four in the endcap to the GMT. Finally, the GMT matches the information from the two sources, with some additional calorimeter information to determine muon isolation, before sending the best four muon candidates to the GT for the final Level-1 decision.

The last Level-1 step is the GT, where a set of selection requirements is applied to the trigger objects passed by the calorimeter and muon systems. In this set of triggers, or *trigger menu*, a maximum of 128 separate selections can be implemented, ranging from simple single-object selections with some kinematic requirements, to multi-object selections with topological conditions.

High Level trigger

The second level of the CMS trigger system is the high level trigger (HLT), which further reduces the rate of data to 100 MHz. The HLT hardware consists of a single computer farm of commercial computers running Scientific Linux with over 13000 CPU cores [40]. After data from the signals coming from the different detectors are processed, the HLT

performs an online reconstruction of physical objects in a similar way to that used in the offline reconstruction as described in Chapter 4. That is, for each event, physical objects are reconstructed by applying some identification criteria in order to select events with potential physics interest. However, rather than reconstructing every single possible object in the event online, the HLT reconstructs objects and regions of the detector whenever needed.

The HLT system is structured around the concept of an HLT *path*, which is a sequence of steps producing and making selections on physical objects. The set of HLT paths is often called the HLT *menu*. A regular HLT path is a sequence of a L1-trigger decisions, or L1-seeds, followed by reconstruction modules and selection filters. If a sequence is not completed, due to not fulfilling a selection filter, the event is not stored for further processing. Physics analyses utilize paths when all the events are recorded after a positive decision of the trigger. The HLT menu contains additional paths where one in every n -th accepted event is recorded. Such paths are referred to as *prescaled paths*, in contrast to paths without a prescale which are referred to as *unprescaled paths*. Prescaled paths are used for calibration of the detectors, monitoring of the online object reconstruction, and to calculate the efficiencies of unprescaled triggers. The selection applied to the prescaled triggers are often looser than the unprescaled ones, and in order to keep the recording rate to manageable levels, the prescaled value is introduced.

Events passing the trigger criteria are sent to archival storage. Those events are stored on disk and then transferred to the CMS Tier-0 computing center for permanent storage and for offline reconstruction. Events are grouped in sets of non-exclusive streams according to the HLT decision. Primary physics streams are not only used, since for instance calibration and monitoring streams are also written for quality purposes. Each stream is divided in different datasets. For example, the physics stream is divided in different primary datasets (PD) according to the physical objects reconstructed by the HLT.

3.2.7 Luminosity measurements

In order to independently measure the amount of collisions delivered by the LHC, and therefore the amount of data, CMS uses the information for five different detectors [41]. Three of them have already been described: the silicon pixel detector, the drift tubes in the barrel and the forward hadronic calorimeter (HF); and two of them are specialized subdetectors designed to make luminosity measurements: the fast beam conditions monitor (BCM1f) and the pixel luminosity telescope (PLT). These detectors are also used to monitor the quality of the beam and the amount of radiation generated by the collisions. The BCM1f, PLT, and HF use separated readouts outside the CMS readout system to guarantee asynchronous luminosity measurements. In addition, these three luminometers are used to provide online measurements of the luminosity during the data-taking, and for redundancy and consistency of the measurements. The silicon tracker and the drift tubes use the standard CMS data acquisition system and provide very low occupancy and good stability over time.

To obtain a measurement of the visible cross section of the beams (σ_{vis}) and to establish absolute calibrations of the detectors, the LHC has dedicated collisions between the runs suitable for physics. During these runs known as *Van der Meer scans* (VdM), the LHC scans the two proton beams through one another in the transverse plane. With the machine parameters given by the LHC, VdM scans allow the detectors to measure the size of the beam, i.e. a measurement of σ_{vis} . Using this value, the absolute luminosity can be calculated as:

$$\mathcal{L} = \frac{R}{\sigma_{vis}}, \quad (3.8)$$

where R is the measured rate given by each luminometer. CMS also estimates sources of uncertainty in the luminosity measurement which is used as a systematic uncertainty [41].

3.2.8 Data Samples

In 2016, CMS recorded a total integrated luminosity of 37.9 fb^{-1} [41]. Data were recorded in seven orthogonal run eras in different primary datasets. This thesis uses the dataset which primarily records hadronic events, named *JetHT* dataset. From the list of trigger paths in the JetHT dataset, we use the lowest unscaled multijet triggers available in each PD. A more detailed description of the specific triggers used in the analyses in this thesis is found in Section 5.1 for the boosted search, and in Section 6.1 for the resolved search.

Since not all the data recorded is suitable for physics searches, CMS has a dedicated system to ensure the quality and the high efficiency of the data. The data quality monitoring (DQM) system is in charge of verifying that all the components of the detector were properly calibrated and working during the data-taking. The specific names of the datasets used, along with their corresponding integrated luminosities properly certified by DQM, are listed in Table 3.1.

Dataset Name	Lumi [pb^{-1}]
/JetHT/Run2016B-23Sep2016-v3/MINIAOD	5787.968
/JetHT/Run2016C-23Sep2016-v1/MINIAOD	2573.399
/JetHT/Run2016D-23Sep2016-v1/MINIAOD	4246.701
/JetHT/Run2016E-23Sep2016-v1/MINIAOD	4008.663
/JetHT/Run2016F-23Sep2016-v1/MINIAOD	3101.619
/JetHT/Run2016G-23Sep2016-v1/MINIAOD	7540.302
/JetHT/Run2016H-PromptReco-v2/MINIAOD	8390.540
/JetHT/Run2016H-PromptReco-v3/MINIAOD	215.149
Total for JetHT dataset	35864.34

Table 3.1: JetHT primary datasets and corresponding luminosities.

Once the events are saved on tape and split in datasets, the offline reconstruction starts. This object reconstruction is a more comprehensive and computationally exhaustive process which uses the full recorded information of the event from every detector in the experiment, as well as more sophisticated reconstruction algorithms than the online procedure with additional calibration. The offline reconstruction is described in detailed in Chapter 4. The reconstruction of events performed shortly after they have been recorded is called *prompt*

reconstruction. Often data needs to be reprocessed in order to take into account updated detector alignment conditions or new detector or object calibrations not used in the online, or prompt, procedures. In this case the reconstruction is named *re-reconstruction*. Data in this analysis is re-reconstructed for run eras B-G, with the label 23Sep2016, while it is prompt data for runs labeled H-v2 and H-v3.

Finally, the data is further processed and filtered in order to reduce the event content size and to calculate more advanced object quantities. The next data tier after reconstruction is called *analysis object data* (AOD). The AOD tier reduces the event size to almost 0.05 Mb; for comparison, the size of one event after the trigger decision is about 0.7 Mb while after reconstruction is about 1.4 Mb. An additional data tier called *MINIAOD* further reduces the size of the event to almost 0.005 Mb. All these reduction operations are needed in order to have manageable data files for the physics analyses.

Chapter 4

Event reconstruction and simulation

In this chapter, the event reconstruction and sample simulation performed in the CMS experiment is described. The chapter starts with a brief introduction of how all the physics objects are reconstructed, as well as the particle-flow and vertex reconstruction algorithm used in CMS. Then, a very detailed description about jets is presented, including specific algorithms used in the two searches included in this thesis. The chapter finishes with a summary of the simulated samples used for the benchmark signal and the SM processes investigated.

4.1 Object reconstruction

As described in Section 3.2, CMS is composed of modern detectors built as cylindrical layers around the point where the beams intersect. A simplified description of the interactions of different particles with the different components of the CMS detector is illustrated in Fig. 4.1. Going in an outward direction from the center of the detector, the interaction of charged particles with the sensitive layers of the tracker creates signals (hits) which are used to reconstruct their trajectories (tracks) and their origin (vertices). The magnetic field generated for the CMS solenoid magnet bend charged particles, and by measuring their curvature we subsequently measure their momenta and their charge. Electrons and photons form electromagnetic showers, which are absorbed by the ECAL and detected as clusters of energy in adjacent ECAL towers. Their detection allows us to measure the energy of the electromagnetic particle and its direction. Charged and neutral hadrons start a hadronic shower in the ECAL and are further completely absorbed by the HCAL towers.

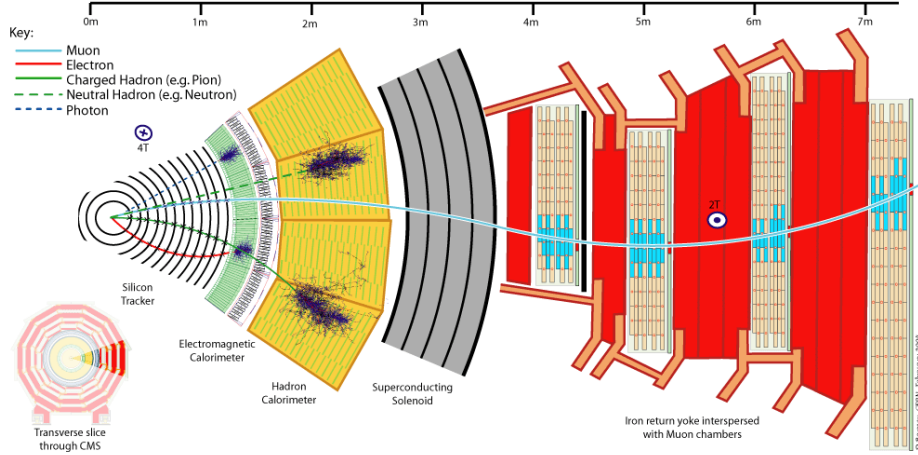


Figure 4.1: Transverse slice of the CMS detector showing the interaction of different particles within the detector [42].

In a similar manner, the clusters of energy around HCAL towers are translated as the energy and location of the hadronic showers. Muons are minimum ionizing particles and thus cross most of the detector leaving few traces of interactions in the calorimeters until they reach the muon system outside the magnet. Hits left by muons in the tracker and the muon systems are used to estimate their momenta. Finally, neutrinos do not interact with the material of the detector and their momenta are inferred by the imbalance momenta in the transverse plane for a given event.

The traditional way of reconstructing objects at hadron colliders has been to link signals in the different sub-detectors [43]. That is, the reconstruction of electrons and photons mainly concerns signals in the ECAL. The identification of muons is based on the information in the tracker and muon systems. Hadronic showers, usually referred as *jets*, are linked to energy deposits in the calorimeters (ECAL and HCAL), while the identification of tau leptons and b-quark jets is based on the properties of their charged particles tracks left in the tracker.

CMS, however, uses a holistic approach based on the correlation of the different detectors to identify each final-state particle. The combination of multiple measurements is used to reconstruct particle properties on the basis of its identification. This approach is called the

particle flow algorithm [43] and it is further explained in the next section.

4.1.1 Particle-flow algorithm

The particle flow (PF) algorithm was developed and used for the first time by the ALEPH collaboration at LEP [44]. This algorithm is well-suited to be implemented in CMS due to its highly-segmented tracker, fine-grained ECAL, hermetic HCAL, strong magnetic field, and excellent muon spectrometer [43]. The result of the PF algorithm is a comprehensive list of final-state particles identified by the combined information of the different layers of subdetectors.

In the PF algorithm used in CMS, charged hadrons are identified by the geometrical connection between one track and one or more calorimeter clusters, with the absence of tracks in the muon system. The combination of the superior measurement of the momentum by the tracker and the energy by the calorimeter provides an overall better determination of the properties of the hadrons. Photons and neutral hadrons are identified by calorimeter clusters with no track geometrically linked in the tracker. In this way, calorimeter clusters can be more accurately calibrated under the photon or neutral hadron hypothesis. Electrons on the other hand are defined as a combination of a track, energy deposition in the ECAL, no energy cluster connected in the HCAL, and a momentum-to-energy ratio compatible with one. Finally, muons are identified by connecting tracks from the inner tracker with the tracks in the muon system.

4.1.2 Tracks and vertex reconstruction

The PF algorithm heavily relies on the reconstruction of tracks from charged particles with high precision. This task, however, is computationally very challenging. The algorithm used in CMS for the reconstruction of track trajectories is called the *combinatorial track finder* (CTF) [45]. In order to create a collection of reconstructed tracks, CMS produces multiple iterations of the CTF. The basic concept behind this iterative process is to start with

an easy to reconstruct track, due to its large p_T or closeness to the interaction point, and remove the hits associated to the trajectory, and repeat the CTF. Each iteration reduces the combinatorial complexity and with it the subsequent steps can look for more complicated tracks.

Once a collection of tracks is created for an event (or a p-p crossing), the next step is to measure the location of the interaction vertices (multiple collisions in one crossing). It is worth noting that for the current conditions of the pp collisions at the LHC, a mean of 30 interactions are expected in each bunch crossing. Therefore, the reconstruction of *primary vertices* (PV), namely vertices close to the beam line, not only includes vertices from a hard interaction but also vertices from soft interactions not coming from the primary vertex. Particles not coming from the primary vertex are collectively referred to as *pileup*. To reconstruct PVs, tracks close to the primary interaction region are selected first. These tracks are grouped based on their z -coordinate at their point closest to the beam line. Finally, a fit of the position of each vertex using associated tracks is performed, resulting in weights associated to each track. The sum of the weights (w_i) roughly corresponds to the number of tracks associated with the vertex, which also defines the number of degrees of freedom (n_{dof}) assigned to the vertex fit; that is $n_{dof} = 2 \sum_i w_i + 3$.

From this list of primary vertices we choose a *good PV*, defined as the vertex which satisfies the following conditions:

- The number of degrees of freedom of all tracks associated to the PV is greater than four.
- The PV position is less than 24.0 cm from the geometric center of the detector in the z -direction.
- The PV transverse distance to the geometric center of the detector is less than 2.0 cm.

The z and transverse distances defining a good PV are based on the measurements of the position of the beams when they intersect in the middle of the detector.

4.1.3 Jets

Jets are widely used in hadron colliders, not only to test and understand high-energy QCD predictions but also to identify the hadronic structure of heavy particles, like top quarks [6]. As described in Section 2.1.2, quarks and gluons cannot be directly observed due to quark confinement. Quarks and gluons go through the hadronization and fragmentation processes almost immediately after having been created and the hadrons subsequently decay. These processes create a collimated spray of energetic charged and neutral hadrons, which are generally called *jets*. The exception to this is the top quarks which is heavy enough and short lived to not hadronize directly, and instead immediately decays.

Jets (representing partons) interacting with the detector create tracks in the tracking system and leave energy deposits in the calorimeters. A graphical representation of a jet travelling through the CMS detector is illustrated in Fig. 4.2. In CMS, information is used from the PF algorithm to provide a set of charged and neutral hadrons which are further used to reconstruct jets. Furthermore, the set of rules to reconstruct and define jets is called the *jet clustering algorithm*. The energy of a clustered jet does not usually correspond to the energy of the particle-level parton due to energy losses in the detector response, as described in Chapter 3. Therefore, a set of corrections are applied to the jets to calibrate them based on their energy scale and resolution. Some features related to jets can be further exploited to differentiate their hadronic origin. For instance, a vertex displaced from the beam line is an indication that a longer-lived hadron has been created in a collision, or if the jet displays internal structure it indicates the signature of a boosted particle where the decay products are reconstructed close together. Since the present work is focused on hadronic decays, and consequently on jets, all these topics are further expanded in detailed in the following sections.

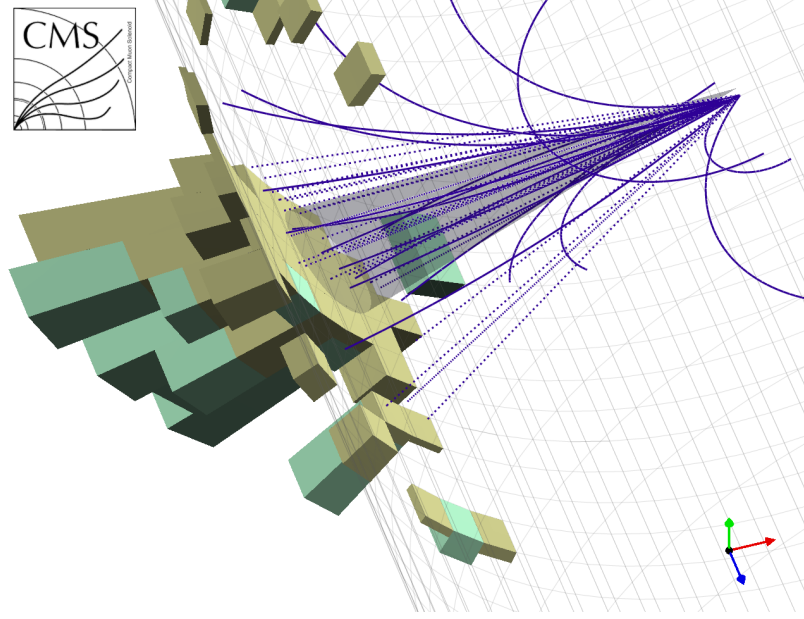


Figure 4.2: Representation of a jet in the CMS detector. Different purple lines represent tracks left by charged hadrons, while the towers represent energy depositions in the ECAL (caqui) and HCAL (green).

Jet algorithms

Jets are not fundamental objects, instead they are our best attempt to reverse-engineer complex processes arising from the hadronization of strongly produced particles. Therefore, there is not a unique jet definition and subsequently there is no unique set of rules to reconstruct them. Different jet algorithms have been used in particle physics experiments to group particles into jets [46], usually invoking some parameters to define a distance between a cluster of particles from the same jet and a recombination scheme to merge the momentum of those particles. A jet is defined by the chosen recombination scheme, or *clustering*, and the distance parameters of the jet algorithm.

These definitions pass some set of requirements, mainly agreed upon by the particle physics community [46]. For instance, a good jet algorithm should give similar results at parton-level and at hadron-level. It is also desired that the reconstructed jets have minimal sensitivity to processes which are difficult to model, such as hadronization or not originating from the primary vertex. In addition, a jet algorithm should not be computationally

expensive, such that the energy corrections of the resulting jets are easy to calculate.

In addition, jets are expected to not drastically change by soft and collinear processes. As part of the fragmentation and some non-perturbative processes, a hadron undergoes many collinear splittings. Furthermore, the emission of soft particles occurs in QCD as part of perturbative and non-perturbative effects. Since these two processes occur randomly and are difficult to model, due to their non-perturbative origin, a good jet algorithm should be very robust against these effects. These two requirements are often called *infrared and collinear* (IRC) safety.

Jet algorithms can be divided into two broader groups: cone and iterative algorithms. Cone algorithms have a "top-down" approach which relies on the unchanged energy flow of the QCD hadronization, especially into a cone. This algorithm draws circles in the $\eta - \phi$ phase space around clusters of energy following certain rules, ending up with a jet with a cone shape. Iterative algorithms incorporate a "bottom-up" approach by using a sequential recombination of jets based on a hierarchical methodology to put together energy deposits or PF particles. Cone algorithms were mostly used at the Tevatron. However, recursive clustering of jets are shown to have better performance and therefore more widely used at the LHC experiments [47].

The recursive algorithm uses the four momenta of the energy deposits or PF particles, which are referred to as *particle candidates* for the rest of this discussion. The algorithm clusters two candidates (i,j) if the distance (d_{ij}) between these two particle candidates is smaller than the distance (d_{iB}) between the i -th candidate and the beam axis. These two distances are defined as [48]:

$$d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \quad (4.1)$$

$$d_{iB} = k_{Ti}^{2p}, \quad (4.2)$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$; k_{Ti} , y_i and ϕ_i are the transverse momentum, rapidity

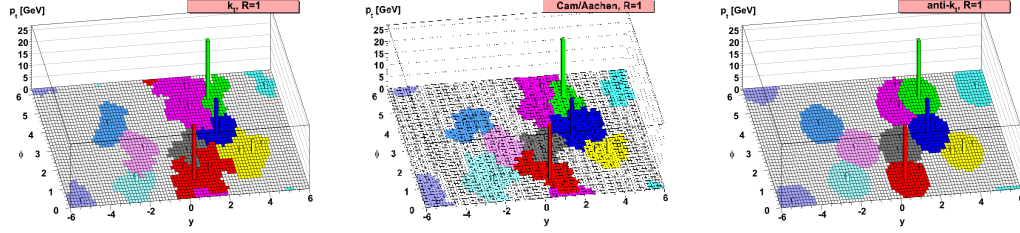


Figure 4.3: Representation of the different jet iterative algorithms in the $\eta - y$ plane with the p_T on the z -axis. Left: k_T algorithm. Middle: CA algorithm. Right: anti- k_T algorithm. Different colors represent different reconstructed jets in the event [48].

and azimuthal angle of the i -th candidate respectively, and R is the cone size of the jet. This process continues recursively with the next closest candidate until $d_{iB} < d_{ij}$, when the i -th candidate is called a jet, or when no other candidates are left in the event. If the parameter p in Eq. (4.2) is $p = 1$ the algorithm is called the k_T algorithm, if $p = 0$ it is called the Cambridge-Aachen algorithm (C/A), and if $p = -1$, it is called the anti- k_T algorithm. For the k_T and C/A algorithms the final jet has an irregular shape while in the case of anti- k_T a jet has a circular shape in $\eta - \phi$ space. A graphical representation of these algorithms in the $\eta - y$ plane is illustrated in Fig. 4.3.

CMS has adopted the anti- k_T algorithm as the preferred technique to reconstruct jets and is what is used in this analysis. If the candidate particles for the jet clustering are calorimeter towers only, these jets are named *calojets*. On the other hand, jets for which their constituent particles are PF particles the jets are called *PFjets*. Moreover, jets reclustered from generator-only particles are referred to *genjets*.

Pileup mitigation for jets

One of the main challenges of pp collisions at the current instantaneous luminosities of the LHC is pileup. When reconstructing around 25 primary vertices on average per bunch crossing, it is highly unlikely that the jet algorithm clusters only particle candidates originating from the same primary vertex. In order to mitigate this effect, CMS mainly considers two approaches: charged hadron subtraction and PUPPI.

In the *charged hadron subtraction* (CHS) approach, the tracks of charged hadrons which are not associated with the primary vertex are removed from the list of PF candidates [47]. Then, the jet algorithm runs on the list of PF candidates after this removal is applied. The benefit of this simple approach is that it reduces the dependence of pileup jet on energy. Additional corrections, described in Section 4.1.3, are applied at the event-level to take into account other remaining hadrons not originating from the PV.

Another technique used in CMS, but not used in this thesis, is called *pileup per particle interaction* (PUPPI) [49]. This technique uses local shape information, event pileup properties and tracking information together to remove particle candidates associated with pileup. The algorithm contrasts the soft radiation from pileup with the collinear structure of QCD events by calculating a local variable, α , which defines the probability of a particle having originated from a pileup vertex. This variable is further used to calculate a per-particle weight to re-scale the four momenta of the particles associated with pileup, eliminating the need of a pileup-based correction.

Jet identification

To further discriminate between real jets coming from a hard parton and jets from noise, *jet identification* (jetID) criteria are applied to each jet used in a CMS analysis. These criteria attempt to reduce the amount of fake hadrons and leptons reconstructed in the PF algorithm which are associated with noise in the calorimeters [47]. The jetID criteria select on quantities which are sensitive to different sources of noise. Table 4.1 summarizes the different jetID working points used in CMS.

The *loose* and *tight* working points target the reduction of calorimeter noise, while the *tight lepton veto* additionally removes potential background coming from mis-reconstructed electrons and muons. The efficiency of these working points is measured in a noise enriched minimum bias sample using a *tag-and-probe* technique. A minimum requirement of at least two back-to-back jets is applied. Once the two highest p_T jets are selected, one jet (tag)

Table 4.1: List of variables and selection for different jetID criteria [47]. We use the tight lepton veto jetID in this thesis.

Jet Variables	$ \eta $ range	Loose	Tight	Tight lepton veto
Charged Hadron Fraction	$ \eta < 2.4$	> 0	> 0	> 0
Charged Multiplicity	$ \eta < 2.4$	> 0	> 0	> 0
Charged EM Fraction	$ \eta < 2.4$	< 0.99	< 0.99	< 0.9
Muon Fraction	$ \eta < 2.4$	—	—	< 0.8
Neutral Hadron Fraction	$ \eta < 2.7$	< 0.99	< 0.99	< 0.9
Neutral EM Fraction	$ \eta < 2.7$	< 0.99	< 0.99	< 0.9
Neutral Multiplicity	$ \eta < 2.7$	> 0	> 0	> 0
	$2.7 < \eta < 3$	> 2	> 2	> 2
	$3 < \eta < 5$	> 10	> 10	> 10

is chosen randomly to satisfy the jetID criteria, while the other jet (probe) is tested to pass the same jetID selection. The efficiency is then defined as the number of probe jets passing the jetID criteria divided by the total number of probe jets. An efficiency of 99% is found for the tight lepton veto jetID in the central detector region ($|\eta| < 2.5$) for jets with $p_T > 30$ GeV [47]. In addition, this working point is found to reject 95% of the background. In this thesis, we use jets selected with the tight lepton veto jetID.

Jet energy corrections

In order to correct the non-linear response in the calorimeters, a set of *energy corrections* (JEC) [50] is applied to jets. The main purpose of the JEC is to relate on average the energy of the reconstructed jet with the energy at particle level. CMS determines the JEC both by using MC simulations and by using data-driven methods. As pictured in Fig.4.4, the JEC are applied in separate steps sequentially and always keeping the same order shown in the figure. Each level of correction takes care of a different effect, applied as a factor (which may have a p_T and/or η dependence) to the four momentum of the jet. In addition, each correction also adds a source of systematic uncertainty to the energy scale of the jets.

Pileup Although electronic noise from the calorimeters is removed during the event reconstruction in the data processing step, an additional correction is needed to remove

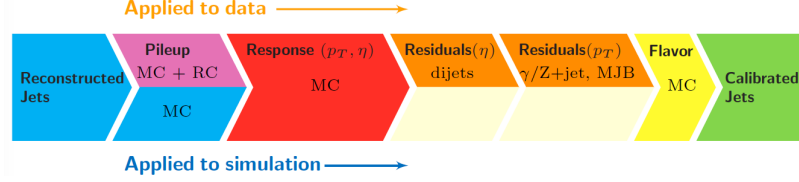


Figure 4.4: Graphical schema of the steps for the jet energy corrections [51].

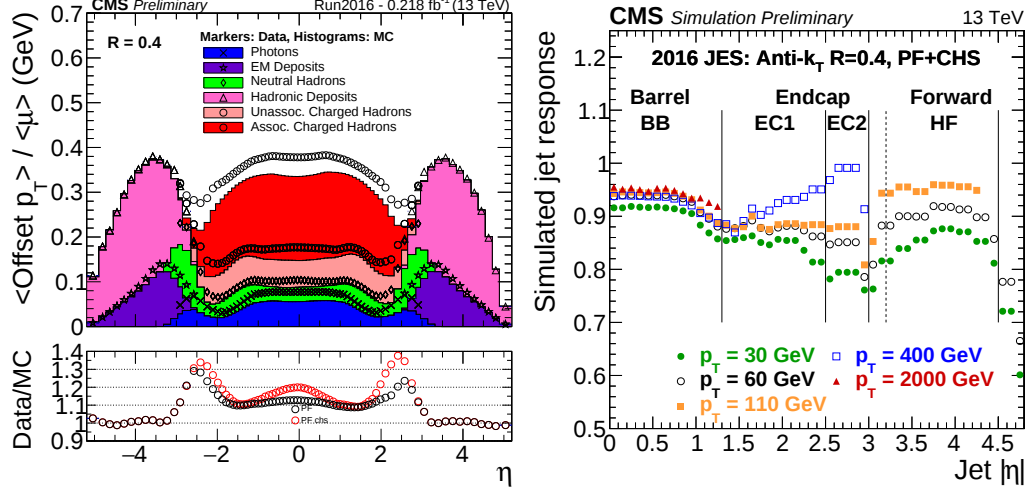


Figure 4.5: Left: Data (markers) vs MC (histograms) normalized by the average number of pileup interactions (μ), and separated by the type of PF candidate [53]. Right: Simulated jet response versus η for AK4 jets [53].

contributions from pileup and extra noisy energy deposits. The *pileup offset* correction is the first in the set of jet energy corrections and its goal is to remove any dataset dependence on luminosity. CMS uses the *hybrid jet area* method [52], which uses the product of the effective area of the jets and the average energy density in the event to estimate the offset energy to be subtracted from the jets. This correction is calculated for MC samples by reconstructing the same events with and without pileup and match the reconstructed jets between these samples. For data, corrections for residual differences with simulations are calculated in a dataset with small noise contribution and no energy deposition from hard interactions, the so called *zero-bias* dataset, using the random cone method [52]. Figure 4.5 (left) shows the offsets in data and MC simulations as a function of η .

Response The second layer of corrections is referred to as MC *response*. The jet calorimeter response in MC simulations does not correspond to the response observed in data and so the MC need to be corrected for this effect. These corrections are calculated in a QCD MC dijet sample by comparing the reconstructed jet after the pileup corrections are applied to the particle-level jets. The simulated jet response is defined [52] as the ratio of arithmetic means of the reconstructed (reco) and particle-level (ptcl) jet p_T :

$$R(p_{T,ptcl}, \eta) = \frac{\langle p_{T,reco} \rangle}{\langle p_{T,ptcl} \rangle} [p_{T,ptcl}, \eta], \quad (4.3)$$

in bins of particle-level p_T and reconstructed η . (Both brackets denote binning variables). The derived corrections are results in a uniformed response of the jets as a function of p_T and η . For 2016 data-taking conditions, CMS calculated the simulated jet response for different p_T jets in different η regions as illustrated in Fig. 4.5 (right).

Residuals The next step is applied only to data, and it is meant to correct the *residual* differences in the jet response as a function of η and p_T . A residual η -dependent correction is calculated by selecting a dijet sample in data using a simple tag-and-probe method. Here, a *tag* jet is located in the central region $|\eta| < 1.3$, and the *probe* jet is unconstrained in η ; in this sense a probe jet is corrected based on the response of the tagged jet. The results of this η -residual correction using the 2016 data is shown in Fig. 4.6 (left). In addition, a p_T -dependent residual is evaluated by selecting $Z(\mu\mu)+\text{jet}$, $Z(ee)+\text{jet}$ and $\gamma+\text{jets}$ events. In this case, we exploit the reconstruction of the Z resonant or the well calibrated γ in the ECAL. We calculate the jet p_T residual correction based on the energy imbalance of the Z or γ and the jet. This response is shown in Fig. 4.6 (right) using the 2016 dataset.

Flavor The final correction is based on the parton *flavor* and is applied to data and MC. We use dijet events to select gluon jets, while the jets in $Z/\gamma+\text{jets}$ events are select mostly quark jets. We match reconstructed jets in MC with parton level information, to

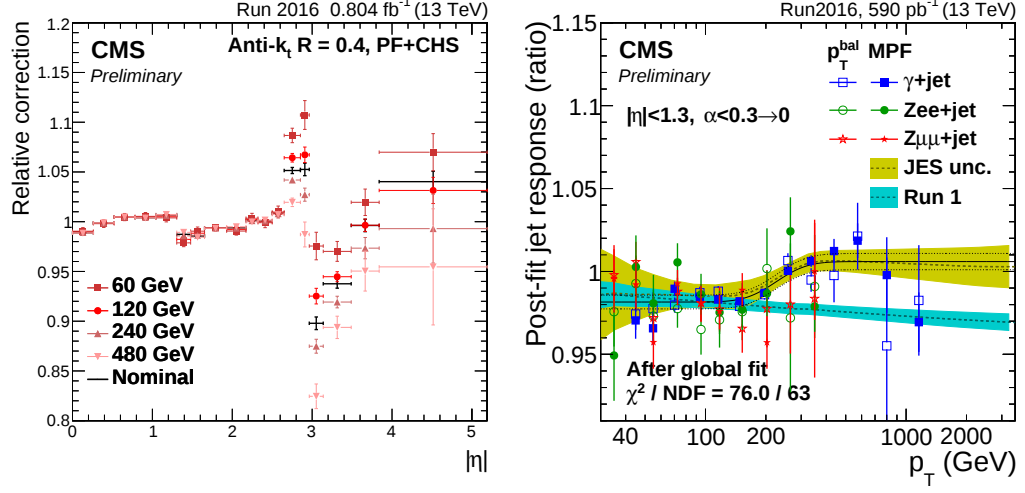


Figure 4.6: Left: Relative energy scale correction for different p_T jets as a function of $|\eta|$ [53]. Right: Data/MC scale factors for the jet response p_T -dependence, evaluated from γ +jet, $Z(\mu\mu)$ +jet and $Z(ee)$ +jet events, combined in a global fit. [53].

calculate a correction based on the difference in response for different jet flavors. These differences arise from the variations in the jet fragmentation energy and the particle compositions of the jets.

Final jet corrections The final JEC corrections are applied to a jet p_T as the product [50]:

$$\mathcal{C} = C_{\text{offset}}(p_T^{\text{raw}}) \cdot C_{\text{response}}(p'_T, \eta) \cdot C_{\text{residual}}(p''_T, \eta) \cdot C_{\text{flavor}}, \quad (4.4)$$

$$\text{Jet}^{\text{corr}} p_T = \mathcal{C} \cdot \text{Jet}^{\text{raw}} p_T, \quad (4.5)$$

where p'_T is the p_T of the jet after the offset correction and p''_T is after the response correction.

Jet energy resolution

The resolution of jets indicate how accurately we measure the p_T of the jets. Jets, being complex objects, regularly have poor resolution compared to other physics objects like electrons, muons or photons [52]. To measure the resolution of jets in MC, generator level jets are spatially matched with reconstructed jets after all the JECs are applied. Then,

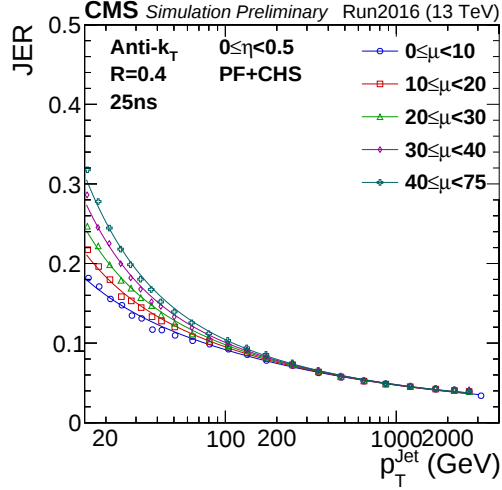


Figure 4.7: Jet energy resolution vs p_T of jets in the central region ($|\eta| < 0.5$) for different levels of pileup (μ) in MC simulations measured by CMS in 2016 [53].

the distribution of the difference in generator level and reco level p_T is fit to a Gaussian function and the width of the fitted Gaussian defines the jet energy resolution (JER).

The JER is also measured in data with similar methods as was previously described in the jet residual section by selecting γ +jets and dijet events. In essence, by tagging a γ or jet we compare the resolution of the *probe* jet to the *tag* object. This method shows a slightly worse resolution in data compared to that measured in MC. Therefore an additional correction factor is derived and applied to MC jets to match the JER in data. In Fig. 4.7, the JER vs p_T of the jets is shown for different pileup scenarios for anti- k_T jets with a 0.4 cone size. The value of the jet resolution for jets with p_T above 100 GeV is about 10%, and is the value used in this analysis.

Jet substructure

One of the main challenges of the LHC is that it is the first hadron collider probing scales significantly higher than the electroweak scale [6]. We are probing the behavior of SM particles when their transverse momenta is at a scale that is considerably higher than their mass. Therefore, due to a relativistic boost at these high momenta, massive particles have very collimated final state products once these decay. If such particles decay to jets,

then they would fail the standard jet clustering methods. To study such highly boosted jet topologies, several tools have been developed in the last couple of years. *Grooming techniques* refer to the methods to remove jet constituents not associated with the hard boosted hadronically decaying particle. Other variables, such as *N-subjettiness*, attempt to quantify how many *prongs* a jet has (based on the collimated decay products) and allows for discrimination of jets originating from a for example W boson or top quark with a jet from QCD jets.

Grooming techniques The mass of the jet is a key kinematic variable of a boosted particle and therefore an improved mass resolution is needed in jet substructure studies. Grooming algorithms seek to eliminate any additional contributions in a jet from pileup. By removing extra constituents the mass resolution of groomed jets is improved. CMS has investigated the performance of several grooming techniques [54]. In this thesis, we use the *trimming* algorithm [55] in the trigger and the *pruning* algorithm [56] in the analysis offline.

The trimming algorithm discriminates jet constituents based on a dynamic p_T threshold. The algorithm reclusters the constituents of the original jet with a smaller cone size R_{sub} , using the k_T algorithm [48]. From the resulting subjets, the trimming algorithm accepts only the ones which satisfy the condition:

$$p_{Tsub} > f_{cut}\lambda_{hard}, \quad (4.6)$$

where f_{cut} is a dimensionless cutoff parameter and λ_{hard} is a hard QCD scale parameter chosen to be equal to the p_T of the original jet. In the trigger used in the boosted search, we use jets reclustered with the AK algorithm with a radius equal to 0.8 (AK8) as the original jet, $R_{sub} = 0.1$ and $f_{cut} = 0.3$.

In the pruning algorithm, the constituents of the original jet are reclustered with the same cone size (in this case 0.8) but using a modified Cambridge-Aachen (CA) algorithm [48]. Constituents i and j are merged in the reclustering algorithm if they satisfy at

least one of the following conditions:

$$z_{ij} = \frac{\min(p_{\text{T}}^i, p_{\text{T}}^j)}{p_{\text{T}}^{i+j}} > z_{\text{cut}}, \quad (4.7)$$

$$\Delta R_{ij} < \frac{2 \times r_{\text{cut}} \times m_J}{p_{\text{T}}} = D_{\text{cut}}, \quad (4.8)$$

where z_{cut} and r_{cut} are parameters of the algorithm, while m_J and p_{T} are the mass and the transverse momentum of the originally-clustered jet, respectively. If both conditions are not met, the softer of the two constituents is removed. For the boosted search analysis we use pruning with $z_{\text{cut}} = 0.1$ and $r_{\text{cut}} = 0.5$. Once the grooming procedure is applied to the raw uncorrected jet, the resulting groomed jet mass needs to be corrected, and therefore we apply the *response* and *residual* jet corrections to the groomed jet mass.

N-subjettiness While the grooming techniques aim to remove additional contributions inside the jet coming from pileup in order to correct its mass [57], jet substructure techniques investigate its internal structure. One of the most studied and widely used substructure variables is n-subjettiness [?]. This method exploits the difference in expected energy flow between the radiation pattern from boosted hadronically decaying particles and gluon or quark jets by counting the number of hard lobes of energy inside a jet.

N-subjettiness uses the k_{T} -algorithm to recluster the constituents of the jet until N subjects are found. These N subjects are used to calculate the quantity:

$$\tau_N = \frac{1}{d_0} \sum_k p_{\text{T},k} \times \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}), \quad (4.9)$$

where $p_{\text{T},k}$ is the transverse momentum of the k -th jet constituent, $\Delta R_{n,k}$ is the distance to the n -th subjet and, the normalization factor d_0 is:

$$d_0 = \sum_k p_{\text{T},k} \times R_0, \quad (4.10)$$

where R_0 is the cone size of the original jet.

It is common to not use the N-subjettiness variables by themselves, but instead use the ratio between them since it provides better discriminating power. For instance, $\tau_{21} = \tau_2/\tau_1$ has been shown have a better separation between jets with two subjets (N=2) in comparison with jets with no internal structure (N=1).

b-jet identification

The longer lifetimes of hadrons originating from b-quarks allow us to differentiate jets coming from light-flavor quarks and those coming from b-quarks. The identification of jets coming from b quarks is called *b tagging* and the basic principle is illustrated in Fig. 4.8. Hadrons originating from b quarks are longer lived and can travel a measurable distance in the detector before they decay. In the detector, b hadrons decay in the tracking system and therefore a vertex displaced from the interaction point can be reconstructed. This displaced vertex is called a *secondary vertex*. The *impact parameter* (d_0 , IP) is the distance between a track and the primary vertex at the point of closest approach in the transverse plane [59]. This parameter is important because it is a Lorentz invariant quantity in the relativistic limit, compared to the decay length. The sign of the IP is determined by the scalar product of the vector pointing from the PV to the point of closest approach with the jet direction. A positive sign determines tracks along the jet axis, which can be used to distinguish prompt tracks and decay products of b hadrons.

CMS developed different algorithms to identify b-jets. In this thesis, we use the *combined secondary vertex version 2* algorithm (CSVv2) [61]. This algorithm uses secondary vertex and track-based lifetime information in a neural network multivariate technique to identify b jets. There are different operating points with different tagging efficiencies for jets using this b tagger [62]. The probability to missidentify b jets for the b tagger used in this thesis is shown in Fig. 4.9 (top). In addition, scale factors are derived and applied to simulation to match the identification efficiencies between data and simulations, as shown in Fig. 4.9

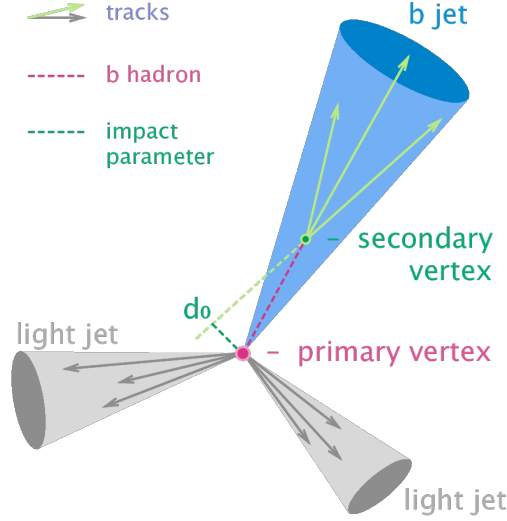


Figure 4.8: Diagram illustrating b-jet identification [60].

(bottom). These scale factors are parameterized as a function of the jet p_T and η and affect individual jets based on the parton flavor of the jet and the efficiency of each jet search. Although the misidentification probability and the scale factors in Fig. 4.9 are shown for AK4 jets the same procedure is performed in AK8 jets with similar results.

4.2 Simulated samples

Modern high energy physics studies heavily rely on simulated samples to understand detector responses to particles, and to compare known SM processes with data to be able to search for new physical phenomena. Despite our good understanding of leading-order processes at hadron colliders, non-perturbative QCD interactions like gluon bremsstrahlung or next-to-leading-order corrections are too complex for us to be calculated with analytical methods. Therefore, simulations are based on Monte Carlo (MC) methods to solve numerically complex matrix element calculations. These methods assume that the processes can be described by PDFs, modeled from data driven or theory methods [6], as described in Section 2.1.2. Two MC generators are mainly used in this thesis: PYTHIA and MADGRAPH. While MADGRAPH is a more specialized matrix element event generator [63], PYTHIA is a multi-purpose generator which includes soft radiation, initial and final-state showers, multiparton

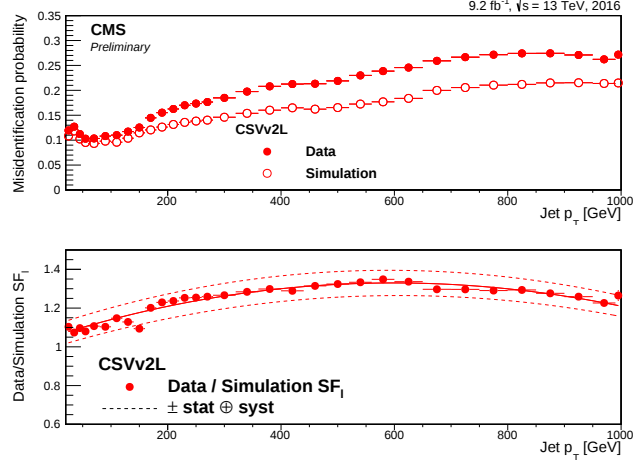


Figure 4.9: Top: misidentification (mistag) probability for light jets in data (solid markers) and simulation (open markers) for the CSVv2 tagger using the loose working point as a function of the jet p_T. Bottom: the data-to-MC ratio or scale factor for the misidentification probability is shown [61].

interactions, fragmentation, and decays [64].

Although MC generators are meant to simulate process from the SM and from hypothetical models, they do not simulate how these particles interact through matter. An additional layer of simulation is used to model how these particles interact with the different components of our detectors. The software used to simulate the passage of energetic particles through detectors such as CMS is called GEANT [65]. Finally, these simulated samples are treated just as one would treat real data in order to reconstruct physical objects.

4.2.1 Benchmark MC signal samples

The benchmark model used in this analysis is the pair production of stops decaying via the hadronic RPV coupling λ''_{312} and λ''_{323} , as described in Section 2.3. We model this signal using a hybrid of MADGRAPH 5_MC@NLO v5.2.2.2 [63] and PYTHIA v8.205 [64]. MADGRAPH 5_MC@NLO at leading order is used for the production of a pair of stops with up to two additional jets from initial state radiation, while PYTHIA is used for the decay of $\tilde{t} \rightarrow qq'$ or $\tilde{t} \rightarrow bq'$ through the λ''_{UDD} hadronic RPV coupling. Two scenarios are considered for this coupling. First, the coupling λ''_{312} is set to a non-zero value such that the branching ratio

Table 4.2: Simulated signal events produced with MADGRAPH and showered with PYTHIA for the λ''_{312} (λ''_{323}).

Dataset Name	Events	Cross Section [pb]
RPVStopStopToJets.UDD312(UDD323)_M-80	79766521(38857473)	3746.
RPVStopStopToJets.UDD312(UDD323)_M-100	19503575(16404738)	1521.110
RPVStopStopToJets.UDD312(UDD323)_M-120	7595597(7609865)	689.799
RPVStopStopToJets.UDD312(UDD323)_M-140	3266223(1209093)	342.865
RPVStopStopToJets.UDD312(UDD323)_M-160	2009933(150947)	184.623
RPVStopStopToJets.UDD312(UDD323)_M-180	825959(826295)	106.194
RPVStopStopToJets.UDD312(UDD323)_M-200	495711(499827)	64.509
RPVStopStopToJets.UDD312(UDD323)_M-220	370206(368580)	40.594
RPVStopStopToJets.UDD312(UDD323)_M-240	252437(246555)	26.476
RPVStopStopToJets.UDD312(UDD323)_M-260	—(236161)	17.684
RPVStopStopToJets.UDD312(UDD323)_M-280	—(193804)	12.157
RPVStopStopToJets.UDD312(UDD323)_M-300	537728(535481)	8.516
RPVStopStopToJets.UDD312(UDD323)_M-350	236097(234079)	3.786
RPVStopStopToJets.UDD312(UDD323)_M-400	144116(127096)	1.835
RPVStopStopToJets.UDD312(UDD323)_M-450	49390(49693)	0.948
RPVStopStopToJets.UDD312(UDD323)_M-500	51048(51320)	0.518
RPVStopStopToJets.UDD312(UDD323)_M-550	51848(51419)	0.296
RPVStopStopToJets.UDD312(UDD323)_M-600	50578(50172)	0.174
RPVStopStopToJets.UDD312(UDD323)_M-650	24859(24708)	0.1070
RPVStopStopToJets.UDD312(UDD323)_M-700	26074(26005)	0.0670
RPVStopStopToJets.UDD312(UDD323)_M-750	9986(10115)	0.04314
RPVStopStopToJets.UDD312(UDD323)_M-800	9986(9963)	0.0283338
RPVStopStopToJets.UDD312(UDD323)_M-850	107278(10750)	0.0189612
RPVStopStopToJets.UDD312(UDD323)_M-900	11058(9772)	0.0128895
RPVStopStopToJets.UDD312(UDD323)_M-950	10556(10374)	0.00883465
RPVStopStopToJets.UDD312(UDD323)_M-1000	11400(11332)	0.00615134
RPVStopStopToJets.UDD312(UDD323)_M-1100	8658(9080)	0.00307413
RPVStopStopToJets.UDD312(UDD323)_M-1200	9755(9660)	0.00159844
RPVStopStopToJets.UDD312(UDD323)_M-1300	10748(10659)	0.000850345

(B) of $\tilde{t} \rightarrow qq'$ is 100%. The second case instead sets a non-zero value for λ''_{323} , resulting in $\tilde{t} \rightarrow bq'$ with a B of 100%. All superpartners except the stops are taken to be decoupled and no intermediate particles are produced in the stop decay. PYTHIA is also used for the showering of the particles with the tune CUETP8M1 [66]. Samples were produced with stop masses of 80 to 300 GeV in 20 GeV increments, from 300 to 1000 GeV in 50 GeV increments, and from 1100 to 1500 GeV in steps of 100 GeV, as shown in Table 4.2. GEANT4 [65] is used for the simulation of the particles through the detector and the reconstruction was done using the CMS software framework CMSSW_8_0_20. Since the B is assumed to be 100%, this signal is considered a *simplified* topology. The signal cross sections we used are the ones calculated by the LHC SUSY cross section working group at NLO+NLL [67, 68].

4.2.2 SM processes

Although the main QCD multijets background in these analyses are modeled with a data-driven techniques, several MC samples are also used to optimize our selection, to study the comparison between data and simulation, and to model additional sub-leading backgrounds.

Since we are searching for new resonances in all hadronic final states, QCD multijets is the main non-resonant background. Studies using two different sets of samples are performed: one where the QCD multijets events are generated using MADGRAPH and divided in HT bins; and the second is divided in p_T bins and generated using PYTHIA. Although studies have been performed using both samples, we ultimately use the PYTHIA sample for the results presented in this document because of the better agreement with data.

We also consider additional SM backgrounds: the hadronic decay of W bosons with additional initial state radiation jets ($W(q'\bar{q})+jets$) and $Z(q\bar{q})+jets$ samples were generated in MADGRAPH, WZ was generated in PYTHIA, $t\bar{t}$ and WW samples were generated in POWHEG [69, 70, 71], while the hadronic ZZ was generated in MC@NLO. PYTHIA is used for the parton showering of the MADGRAPH, POWHEG and MC@NLO samples. The dataset names with corresponding cross sections and number of events are listed in Table 4.3.

Table 4.3: Simulated background events.

Dataset Name	Events	Cross Section [pb]
QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	17035891	347700
QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	18929951	32100
QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	15629253	6831
QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4767100	1207
QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1203212	119.9
QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1991645	25.24
QCD_Pt_170to300_TuneCUETP8M1_13TeV_pythia8_ext	7838066	117276
QCD_Pt_300to470_TuneCUETP8M1_13TeV_pythia8	4150588	7823
QCD_Pt_300to470_TuneCUETP8M1_13TeV_pythia8_ext	18253032	7823
QCD_Pt_470to600_TuneCUETP8M1_13TeV_pythia8	15458074	648.2
QCD_Pt_470to600_TuneCUETP8M1_13TeV_pythia8_ext	15458074	648.2
QCD_Pt_600to800_TuneCUETP8M1_13TeV_pythia8	9622896	186.9
QCD_Pt_600to800_TuneCUETP8M1_13TeV_pythia8_ext	19324269	186.9
QCD_Pt_800to1000_TuneCUETP8M1_13TeV_pythia8	3992112	32.293
QCD_Pt_800to1000_TuneCUETP8M1_13TeV_pythia8_ext	15704980	32.293
QCD_Pt_1000to1400_TuneCUETP8M1_13TeV_pythia8	2929643	9.4183
QCD_Pt_1000to1400_TuneCUETP8M1_13TeV_pythia8_ext	6982586	9.4183
QCD_Pt_1400to1800_TuneCUETP8M1_13TeV_pythia8	396409	0.84265
QCD_Pt_1400to1800_TuneCUETP8M1_13TeV_pythia8_ext	2477018	0.84265
QCD_Pt_1800to2400_TuneCUETP8M1_13TeV_pythia8	397660	0.114943
QCD_Pt_1800to2400_TuneCUETP8M1_13TeV_pythia8_ext	1552064	0.114943
QCD_Pt_2400to3200_TuneCUETP8M1_13TeV_pythia8	399226	0.00682981
QCD_Pt_2400to3200_TuneCUETP8M1_13TeV_pythia8_ext	596904	0.00682981
QCD_Pt_3200toInf_TuneCUETP8M1_13TeV_pythia8	391735	0.000165445
WWTo4Q_13TeV-powheg	1998400	51.723
WJetsToQQ_HT-600ToInf_13TeV-madgraphMLM-pythia8	1026587	95.14
ZZTo4Q_13TeV_amcatnloFXFX_madspin_pythia8	30454227	6.842
ZJetsToQQ_HT600toInf_13TeV-madgraph	996000	5.67
WZ_TuneCUETP8M1_13TeV-pythia8	2995828	22.82
TT_TuneCUETP8M1_13TeV-powheg-pythia8	77229341	831.76

Chapter 5

Boosted search

This chapter describes the search for boosted paired dijet resonances using jet substructure techniques. This analysis uses the average pruned jet mass spectrum ($\overline{m}$) in dijet events to perform a *bump-hunt* in the mass range of 60-450 GeV. The chapter starts by describing the trigger used in the analysis and it continues with the description of the event selection. Two AK8 jets are selected and these jets are cleaned from pileup contamination using a grooming algorithm. The selection criteria are optimized in order to remove as much of the background events as possible. The selection targeting the $\tilde{t} \rightarrow qq'$ signature is referred to as inclusive while the selection targeting the $\tilde{t} \rightarrow bq'$ signature is referred to as b-tagged. The modeling of the backgrounds in the $\overline{m}$ spectrum are studied and QCD multijets is found to be dominant. Although the background shapes are studied with MC simulations, the final QCD multijets background is estimated primarily from data using a so-called *ABCD method*. The chapter ends with the description the systematic uncertainties. The interpretation of the results of the boosted search are presented together with that of the resolved search in Chapter 7.

5.1 Trigger Selection

To maximize the acceptance of events containing boosted jets with substructure several hadronic triggers are *ORed*, including new HLT trigger paths developed for Run 2 based on AK8 jets and a groomed jet mass. One of these triggers uses the H_T of AK8 jets (H_T^{AK8}) with $p_T = 150$ GeV and $|\eta| < 2.5$, and a requirement of at least one AK8 jet with trimmed jet mass

above 50 GeV (HLT_AK8PFHT700_TrimR0p1PT0p3Mass50). Due to the increase in instantaneous luminosity during the 2016 data-taking, the H_T^{AK8} trigger threshold was increased from 700 GeV to 750 GeV during the last data-taking era (Run2016H). Additional substructure triggers were introduced, such as the AK8 single jet (HLT_AK8PFJet360_TrimMass30) and dijet (HLT_AK8DiPFJet280_200_TrimMass30TrimMass30_BTagCSV_p020) substructure triggers which are based on a p_T requirement of the leading and second leading jet, respectively, and one of the jets with a trimmed jet mass requirement of 30 GeV. The dijet substructure trigger includes a b tagging requirement in a calojet with a loose or medium CSV working points.

In addition to these trigger paths, this analysis also uses the nominal HT trigger to increase the sensitivity of our search. This HLT path (HLT_PFHT800) uses AK4 jets with $p_T > 30$ GeV and $|\eta| < 3.0$ and an $H_T^{\text{AK4}} \geq 800$ GeV for the majority of the 2016 data-taken; the H_T^{AK4} trigger threshold increased to 900 GeV for the Run2016H era. A single jet trigger with p_T above 450 GeV (HLT_PFJet450) is also included to recover inefficiencies found in the hadronic L1 triggers in Run2016H dataset. Table 5.1 summarizes the triggers used in this analysis, indicating their L1 seeds and HLT definitions.

The efficiencies of these triggers are studied and measured using an orthogonal datasets selected with muons (SingleMuon dataset), to avoid any kind of bias. In SingleMuon dataset, an unscaled trigger requiring one muon with $p_T > 50$ GeV (HLT_Mu50) is chosen as the reference trigger for the trigger efficiency measurement. The trigger efficiency of the logical *OR* of all hadronic triggers in Table 5.1 is measured in 2-dimensions using the offline H_T^{AK8} and the offline most energetic pruned jet mass as shown in Fig. 5.1 (left) for all the data-taking eras combined. A loose selection of at least two AK8 jets is applied. The efficiency of the *OR* of the triggers is measured relative to events passing the reference HLT_Mu50 trigger:

$$Eff = \frac{\text{events passing reference plus OR of analysis triggers}}{\text{events passing reference trigger}} \quad (5.1)$$

Table 5.1: Triggers used in boosted search, indicating their L1 seeds and HLT definitions. Values in parenthesis show the thresholds for the Run2016H dataset.

Trigger Name	L1 seeds	Definition
HLT_AK8PFHT700(750)_TrimR0p1PT0p3Mass50	L1_HTT240 OR L1_HTT225 OR L1_HTT270 OR L1_HTT280 OR L1_HTT300 OR L1_HTT320	AK8PFJets with $p_T > 150$ GeV and $ \eta < 2.5$ $AK8HT > 700(800)$ GeV one jet with trimmed mass > 50 GeV trimming parameters: $r = 0.1$ and $p_{T_{frac}} = 0.3$
HLT_AK8PFJet360_TrimMass30	L1_SingleJet180 OR L1_SingleJet200	one AK8PFJet with $p_T > 360$ GeV one jet with trimmed mass > 30 GeV trimming parameters: $r = 0.1$ and $p_{T_{frac}} = 0.3$
HLT_AK8DiPFJet280(300)_200_TrimMass30_BTagCSV_p087(20)	L1_SingleJet180 OR L1_SingleJet200	one AK8PFJet with $p_T > 280(300)$ GeV and another with $p_T > 200$ GeV one jet with trimmed mass > 30 GeV trimming parameters: $r = 0.1$ and $p_{T_{frac}} = 0.3$ and one calojet with loose and medium CSV b tagging
HLT_PFHT800(900)	L1_HTT240 OR L1_HTT225 OR L1_HTT270 OR L1_HTT280 OR L1_HTT300 OR L1_HTT320	AK4PFJets with $p_T > 40$ GeV and $ \eta < 3.0$ $HT > 800$ GeV
HLT_PFJet450	L1_SingleJet180 OR L1_SingleJet200	AK4PFJets with $p_T > 450$ GeV
HLT_Mu50	L1_SingleMu22 OR L1_SingleMu25	Muons with $p_T > 50$ GeV and $ \eta < 2.5$

Figure 5.1 (left) shows that applying an OR of these triggers ensures that the events used in the search are fully efficient in the region above $H_T^{\text{AK8}} > 900$ GeV. Figure 5.1 (right) shows the efficiency as a function of the pruned jet mass after this H_T^{AK8} selection and that we achieve full trigger efficiency above 60 GeV in pruned jet mass. The small inefficiencies measured in the pruned jet mass at very low mass are counted as a systematic uncertainty.

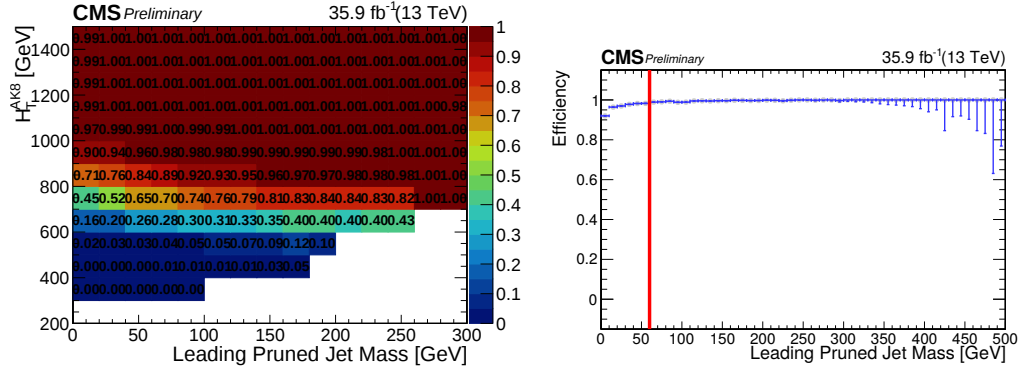


Figure 5.1: On the left, trigger efficiency turn-on curve for the *OR* of the hadronic triggers described in Table 5.1 as a function of offline H_T^{AK8} and offline leading pruned jet mass, measured relative to the SingleMuon dataset. On the right, leading pruned jet mass after selecting $H_T^{\text{AK8}} > 900$ GeV. Red line indicates where the search starts at pruned jet mass > 60 GeV.

5.2 Event Selection

Events are selected if they pass the *OR* of the hadronic triggers described in Section 5.1, and if they contain at least two AK8 jets passing the tight lepton veto jetID (as defined in Section 4.1.3) with jet $p_T > 150$ GeV and a jet $|\eta| < 2.5$. Based on the trigger studies, selecting events with $H_T^{\text{AK8}} \geq 900$ GeV guarantees fully efficient trigger. This preliminary selection (*preselection*) criteria is summarized in Table 5.2.

5.2.1 Kinematic Variables

The pruning algorithm is used to compute the mass of the two most energetic jets. This search assumes that two identical particles are produced and decay to boosted jets, therefore

Table 5.2: Summary of preselection criteria

Variable	Cut Value
Jet p_T	$> 150 \text{ GeV}$
Jet $ \eta $	< 2.5
N jets	$= 2$
H_T^{AK8}	$> 900 \text{ GeV}$

the average mass of the jets is defined as:

$$\overline{m} = \frac{m_{j1} + m_{j2}}{2}, \quad (5.2)$$

where m_{j1} refers to the pruned mass of the leading p_T jet and m_{j2} is the jet pruned mass of the second leading p_T jet. This is the main kinematic variable used in the search.

In order to reduce events coming from known SM processes, we further select on the following variables:

Mass asymmetry: This variable reduces the number of events with large mass imbalance between the two AK8 jets, and it is defined as $m_{\text{asym}} = |m_{j1} - m_{j2}| / (m_{j1} + m_{j2})$.

N-subjettiness: The ratio of the N-subjettiness variables, $\tau_{21} = \tau_2/\tau_1$ and $\tau_{32} = \tau_3/\tau_2$ for both jets, discriminates between signal-like jets with two-prong structure, and background-like jets which have overwhelmingly one-prong structure (or no structure) for QCD multijets and three-prong structure for top quark jets.

$\Delta\eta$: Assuming the signal to be predominantly produced in the central η region, compared to the more wide spread QCD multijets, we use the absolute value of the difference in η between the two jets: $\Delta\eta = |\eta_{j1} - \eta_{j2}|$.

Figures 5.2-5.4 show these kinematic distributions at preselection for data, MC backgrounds and benchmark $\tilde{t}$ signals. The distributions are normalized to show differences in shapes. In next section the procedure used to optimize the event selection considering the kinematic variables above is described.

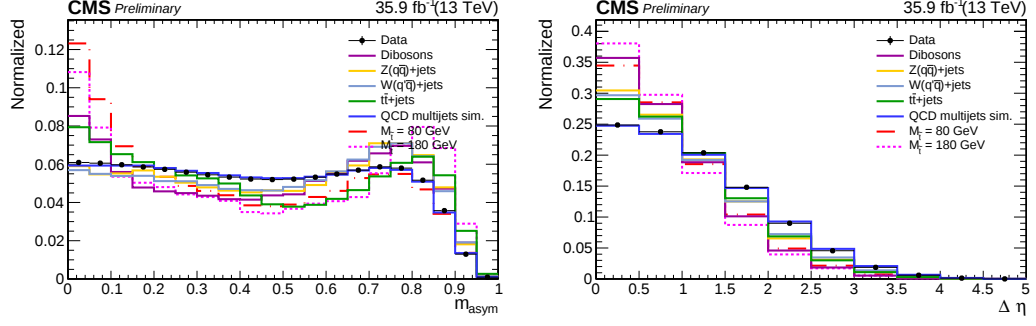


Figure 5.2: Normalized distributions showing the comparison between data, RPV signal $\tilde{t} \rightarrow qq'$ at $m_{\tilde{t}} = 80$ GeV and $m_{\tilde{t}} = 180$ GeV, and MC backgrounds after the preselection shown in Table 5.2. Left: m_{asym} . Right: $\Delta\eta$.

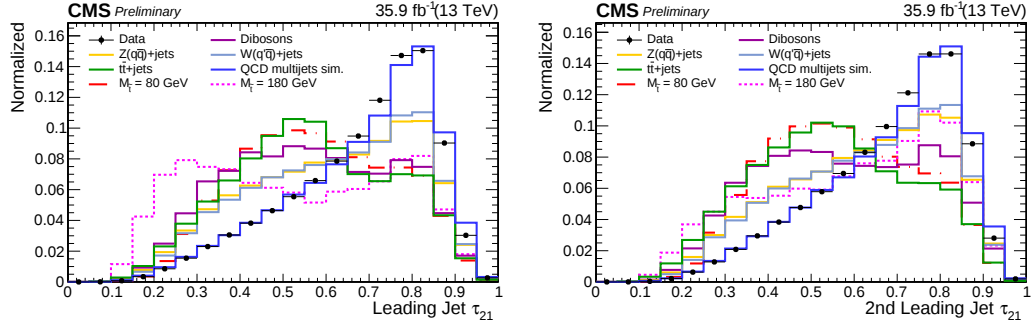


Figure 5.3: Normalized distributions showing the comparison between data, RPV signal $\tilde{t} \rightarrow qq'$ at $m_{\tilde{t}} = 80$ GeV and $m_{\tilde{t}} = 180$ GeV, and MC backgrounds after the preselection shown in Table 5.2. Left: Leading jet τ_{21} . Right: Second leading jet τ_{21} .

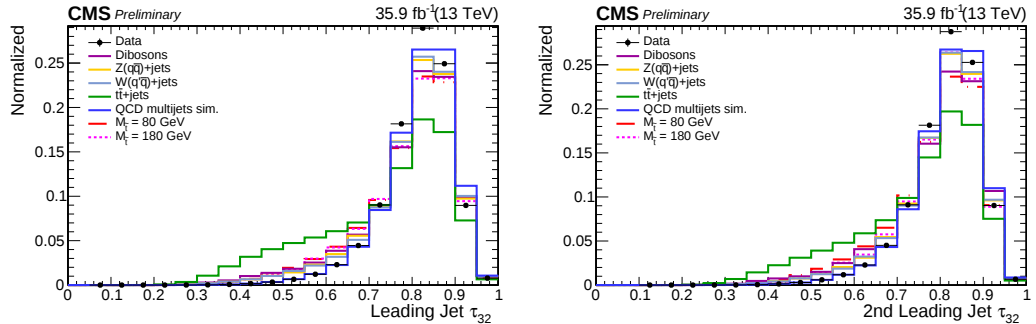


Figure 5.4: Normalized distributions showing the comparison between data, RPV signal $\tilde{t} \rightarrow qq'$ at $m_{\tilde{t}} = 80$ GeV and $m_{\tilde{t}} = 180$ GeV, and MC backgrounds after the preselection shown in Table 5.2. Left: Leading jet τ_{32} . Right: Second leading jet τ_{32} .

5.2.2 Inclusive selection optimization

In this section we describe the selection optimization of the kinematic variables described in Section 5.2.1 for the inclusive analysis. The procedure used is an iterative optimization based on the $S/\sqrt{B}$ metric where S is the number of signal events and B is the number of background events. We use the signal MC samples from Table 4.2 and the different MC background samples described in Table 4.3. All the MC background components are treated together as a whole. The variables described in Section 5.2.1 are used to scan in the ranges summarized in Table 5.3. Each separate selection constitutes an operating point. After a selection for a given operating point is applied, the number of events in S and B are counted in the $\bar{m}$ spectrum in a mass window of 30 GeV around the mass point studied.

Table 5.3: Optimization threshold values for each variable used in the optimization procedure.

Variable	Starting Value	Ending Value	Step Value
m_{asym}	0	1	0.05
1st and 2nd jet τ_{XX}	0	1	0.05
$\Delta\eta$	0	5	0.5

Figure 5.5 illustrates an example of each step in the optimization process assuming $m_{\tilde{\tau}} = 100$ GeV. The goal is to choose a selection on the variable with the highest $S/\sqrt{B}$ ratio, apply that selection and then repeat the process until all optimized selection is achieved. Using this procedure we find the variables (solid colored lines) and the selection (dots) that give the highest $S/\sqrt{B}$. For example, Fig. 5.5 (top left) shows the value of $S/\sqrt{B}$ per variable after preselection. The X-axis shows arbitrary units for simplicity since we are comparing different variables with different ranges in the same plot. Figure 5.5 shows that m_{asym} has the highest $S/\sqrt{B}$ ratio (for $m_{\tilde{\tau}} = 100$ GeV). Therefore, we apply $m_{\text{asym}} < 0.1$ selection and the procedure is repeated. Figure 5.5 (top right) shows $S/\sqrt{B}$ after preselection and $m_{\text{asym}} < 0.1$, and so on. This procedure is repeated until no further gain is observed.

This procedure is performed also for other $m_{\tilde{\tau}}$ signal points and we found a similar optimized selection criteria can be applied for all the masses in the boosted search. This

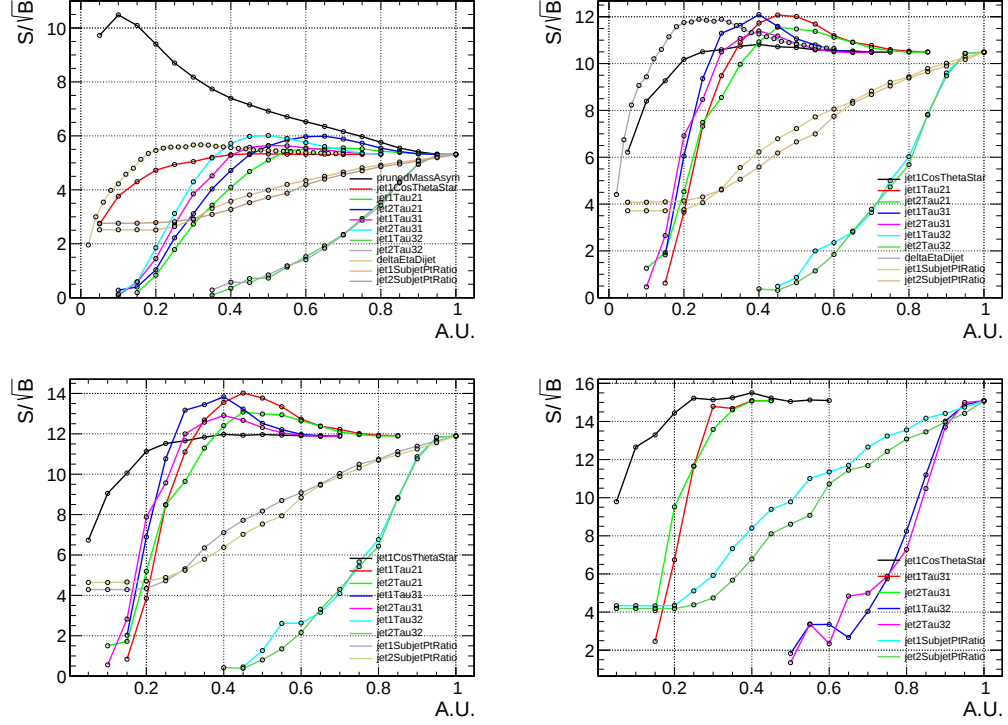


Figure 5.5: The optimization procedure. The $S/\sqrt{B}$ distributions for a given selection as a function of arbitrary units for $\tilde{t} \rightarrow qq'$ at $m_{\tilde{t}} = 100 \text{ GeV}$. Each colored line represents each variable studied, while the dots represent a selection in a given variable. Top Left: shows variables after the preselection. Top Right: after applying the preselection and a cut in $m_{\text{asym}} < 0.1$. Bottom left: after applying the preselection, $m_{\text{asym}} < 0.1$ and $\Delta\eta < 1.5$. Bottom right: after applying the preselection, $\tau_{21} < 0.45$, $m_{\text{asym}} < 0.1$ and $\Delta\eta < 1.5$.

selection is summarized in Table 5.4.

We found that the optimized value for the τ_{21} variable is similar to the one used in CMS to identify boosted hadronic W bosons, referred as *W-tagging* [72]. We subsequently apply the efficiencies and scale factors calculated for this W-tagging working point. In addition, we select on τ_{32} to veto $t\bar{t}$ events; this also allows us to later invert this requirement to select $t\bar{t}$ events in order to validate the $t\bar{t}$ MC background. Again, we find that the *Top Veto* optimization is consistent with the recommended CMS working point for boosted top tagging [73]. In Section 5.4.1 we further discuss the top background validation.

Table 5.6 shows the MC yields for a benchmark RPV stop signal at $m_{\tilde{t}} = 80$ GeV and the SM backgrounds yields expected in 35.9 fb^{-1} of data.

Table 5.4: The values for optimized selection for the inclusive and b-tagged boosted search. τ_{21} and τ_{32} are applied to both jets. These selections include the preselection requirement of Table 5.2.

Inclusive selection	AK8 jets
	jet $p_T > 150 \text{ GeV}$
	jet $ \eta < 2.5$
	$H_T^{\text{AK8}} > 900 \text{ GeV}$
	$m_{\text{asym}} < 0.1$
	$\tau_{21} < 0.45$
	$\tau_{32} > 0.57$
	$\Delta\eta < 1.5$
B-tagged selection	Inclusive selection plus two loose b-tagged jets

After applying the selection in Table 5.4, we show the average pruned mass distribution in data compared to the MC backgrounds in 35.9 fb^{-1} in Fig. 5.6. We also show expected signal distributions for a few benchmark points. This figure is just a demonstration to show the general behavior of the backgrounds using MC samples. From this figure, we categorize two types of background components: a dominant non-resonant background composed mainly of QCD multijets, and a resonant background component, i.e $t\bar{t}$, $W(q'\bar{q}) + jets$, $Z(q\bar{q}) + jets$ and dibosons (WW, ZZ, WZ). In this analysis, these two types of backgrounds are modeled in different ways and are further described in Section 5.4.

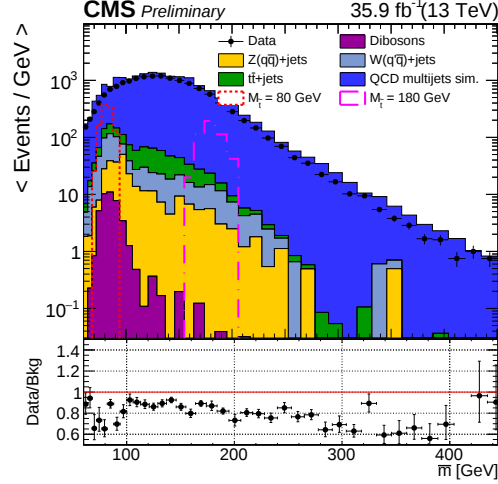


Figure 5.6: Data/MC comparison of the $\bar{m}$ distribution after the optimized inclusive selection from Table 5.4 is applied. Data is shown in dots while the different background MC components are stacked with different colors. Signal shapes at different masses are also shown in dashed lines. The main QCD multijets background is not estimated from MC in this analysis, but ultimately in a data-driven way described in Section 5.4.2.

5.2.3 b-tagged selection optimization

In order to target the heavy flavor resonances, we also apply b tagging. Again, an optimization procedure based on $S/\sqrt{B}$ is used (as in Section 5.2.2) for the different b tagging working points and on the number of b-tags applied to a jet; the b-tagged selection is applied on top of the inclusive selection of Table 5.4. Figure 5.7 shows the data/MC comparison for the CSVv2 b tag discriminator for the two p_T leading jets used in this analysis.

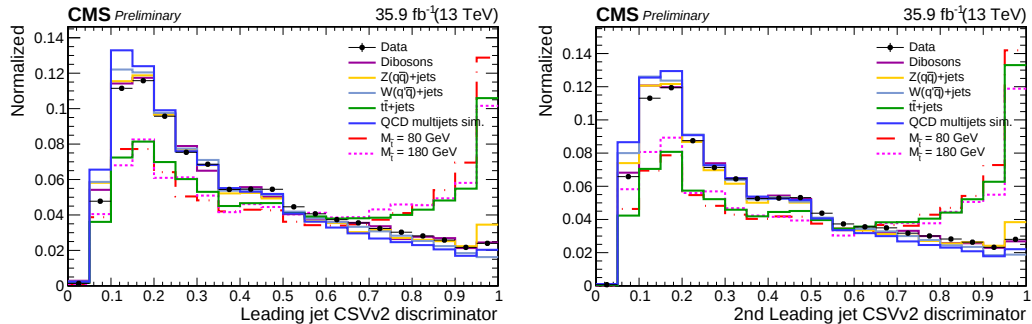


Figure 5.7: Data/MC comparisons of b tag CSVv2 discriminator for the leading (left) and second leading p_T jets (right). All selection listed in Table 5.2 are applied.

Table 5.5: Optimization values of $S/\sqrt{B}$ for different CSVv2 b tag working points for number of b tags and for different signal mass points.

$m_{\tilde{t}}$	80 GeV		120 GeV		180 GeV	
Working points	1 b tag	2 b tag	1 b tag	2 b tag	1 b tag	2 b tag
CSVv2L (0.5426)	7.94	2.03	15.47	5.23	12.84	6.68
CSVv2M (0.8484)	2.32	0.31	4.78	0.43	5.16	0.83

Two CSVv2 b tagging working points are provided by CMS for AK8 jets [74] and we evaluate the $S/\sqrt{B}$ significance for different RPV stop mass points, shown in Table 5.5. From this table, we choose to select two jets with CSVv2 loose working point applied since it gives us the best $S/\sqrt{B}$ for all the $m_{\tilde{t}}$ considered.

Table 5.4 summarizes the final selection for the boosted b-tagged search while the Table 5.6 shows the cutflow table including the b tag requirement. Finally, Fig. 5.8 shows the $\overline{m}$ data to MC distribution for the b-tagged selection using only MC samples and some selected signal mass points.

Table 5.6: Cut flow table with sequential selection. MC samples scaled to 35.9 fb^{-1} . Two signal samples at $m_{\tilde{t}} = 80 \text{ GeV}$ are shown for the two RPV models.

Selection	$\tilde{t} \rightarrow q\bar{q}' \ m_{\tilde{t}} = 80$		$\tilde{t} \rightarrow b\bar{q}' \ m_{\tilde{t}} = 80$		$t\bar{t}$	
	Events	%	Events	%	Events	%
preselection	66475.64 ± 345.70	100	51635.35 ± 447.94	100	84817.93 ± 187.10	100
$\tau_{21} < 0.45$	5359.75 ± 98.74	8.1	3103.08 ± 110.42	6.0	8529.32 ± 59.54	10.1
$\tau_{32} > 0.57$	5083.52 ± 96.15	7.6	2924.66 ± 107.18	5.7	6561.68 ± 52.18	7.8
$m_{\text{asym}} < 0.1$	3158.20 ± 75.70	4.8	1585.40 ± 78.89	3.1	2129.14 ± 29.70	2.5
$\Delta\eta < 1.5$	2971.09 ± 73.44	4.5	1456.96 ± 75.64	2.8	1763.00 ± 27.02	2.1
2 b tag	—	—	871.41 ± 58.40	1.7	703.34 ± 17.04	0.8
Selection	W($q\bar{q}$) + jets		Dibosons		Z($q\bar{q}$) + jets	
	Events	%	Events	%	Events	%
preselection	204125.09 ± 851.70	100	4455.46 ± 57.73	100	14551.26 ± 56.34	100
$\tau_{21} < 0.45$	7657.37 ± 165.94	3.8	409.32 ± 17.11	9.2	651.76 ± 12.00	4.5
$\tau_{32} > 0.57$	6944.84 ± 157.99	3.4	374.25 ± 16.33	8.4	604.87 ± 11.56	4.2
$m_{\text{asym}} < 0.1$	1748.54 ± 79.54	0.9	162.93 ± 10.52	3.7	138.51 ± 5.52	1.0
$\Delta\eta < 1.5$	1358.60 ± 70.12	0.7	131.41 ± 9.45	3.0	112.23 ± 4.97	0.8
2 b tag	153.58 ± 23.37	0.1	8.28 ± 2.07	0.2	14.97 ± 1.79	0.1
Selection	QCD multijets		Total Bkg.		Data	
	Events	%	Events	%	Events	%
preselection	33625161.82 ± 14738.49	100	33933111.56 ± 5825.21	100	28352892 ± 5324.74	100
$\tau_{21} < 0.45$	498488.09 ± 2334.70	1.5	515735.86 ± 718.15	1.5	423578 ± 650.83	1.5
$\tau_{32} > 0.57$	474187.54 ± 2258.89	1.4	488673.19 ± 699.05	1.4	400997 ± 633.24	1.4
$m_{\text{asym}} < 0.1$	138186.62 ± 1294.63	0.4	142365.75 ± 377.31	0.4	118554 ± 344.32	0.4
$\Delta\eta < 1.5$	103598.04 ± 1093.55	0.3	106963.28 ± 327.05	0.3	85921 ± 293.12	0.3
2 b tag	14125.82 ± 590.28	0.1	15005.99 ± 122.50	0.1	11570 ± 107.56	0.1

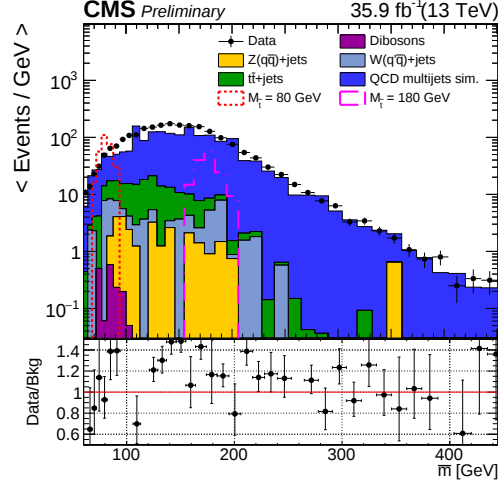


Figure 5.8: Data/MC comparison of the $\bar{m}$ distribution after the optimized b-tagged selection from Table 5.4 is applied. Data is shown in dots while the different background MC components are stacked with different colors. Signal shapes at different masses are also shown in dashed lines. The main QCD multijets background is not estimated from MC in this analysis, but ultimately in a data-driven way described in Section 5.4.2.

5.3 Signal Acceptance

In this section, we describe the signal acceptance for this analysis. Recall, the boosted analysis has two signal regions, an inclusive signal region targeting stops decaying via the coupling λ''_{312} , and a b-tagged signal region targeting stops decaying via the coupling λ''_{323} . Both selections are summarized in Table 5.4.

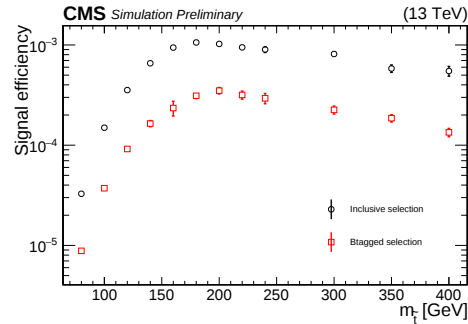


Figure 5.9: Signal efficiency for all the mass points used in the boosted analysis.

Figure 5.9 shows the signal efficiency as a function of $m_{\tilde{t}}$ of all masses for the inclusive selection (black) and the b-tagged selection (red). The efficiency is defined as the number

of signal events surviving all the selection in a mass window within two sigma of the true stop mass divided by the total number of events generated for that mass. Figure 5.10 shows the signal shapes of the $\bar{m}$ distributions after all the selection is applied and as well as the scale factors described below.

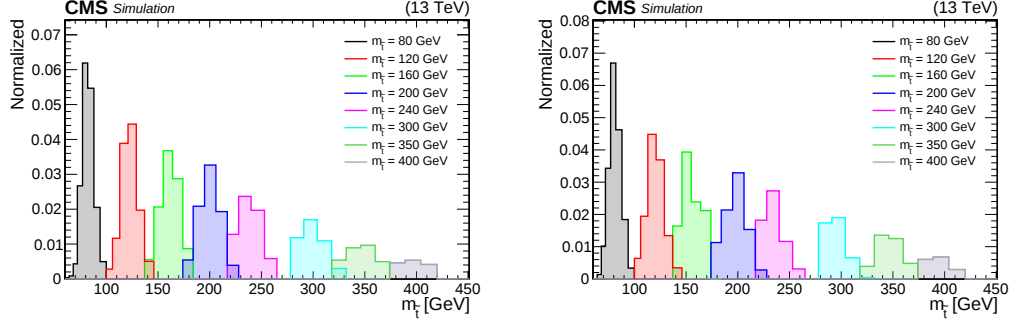


Figure 5.10: Distributions of the $\bar{m}$ for signal MC after the inclusive (left) and b-tagged (right) selections are applied.

5.3.1 N-subjettiness data/MC scale factors

The two-prong tagger (τ_{21}) selection is known to have differing efficiencies in MC and data. The data/MC scale factors are measured by CMS [72] for boosted W bosons in $t\bar{t}$ events using the same τ_{21} selection as is applied in this analysis. From that study, the data/MC scale factor for tagging boosted W bosons with $\tau_{21} < 0.45$ and with CHS applied to the jets is 1.10 ± 0.06 (stat) ± 0.11 (syst). Additionally, the scale factor is determined to have a p_T -dependence and the uncertainty associated with that is reported to be $4.10 \times \ln(p_T/200 \text{ GeV})$, again measured with W bosons from $t\bar{t}$, which have a jet $p_T \sim 200 \text{ GeV}$. Figure 5.11 shows the jet p_T spectra for the RPV stop signal MC used in this analysis. Due to the differences in the jet p_T spectra, an additional systematic is applied to the scale factor of ± 0.037 . Since in this analysis, both jets are required to pass the τ_{21} selection this scale factor is applied two-fold to the signal MC yielding a total scale factor of 1.21 ± 0.29 . This scale factor is included in the $\bar{m}$ spectrum of all the simulated samples used in the boosted analysis and the uncertainty is applied as a systematic uncertainty.

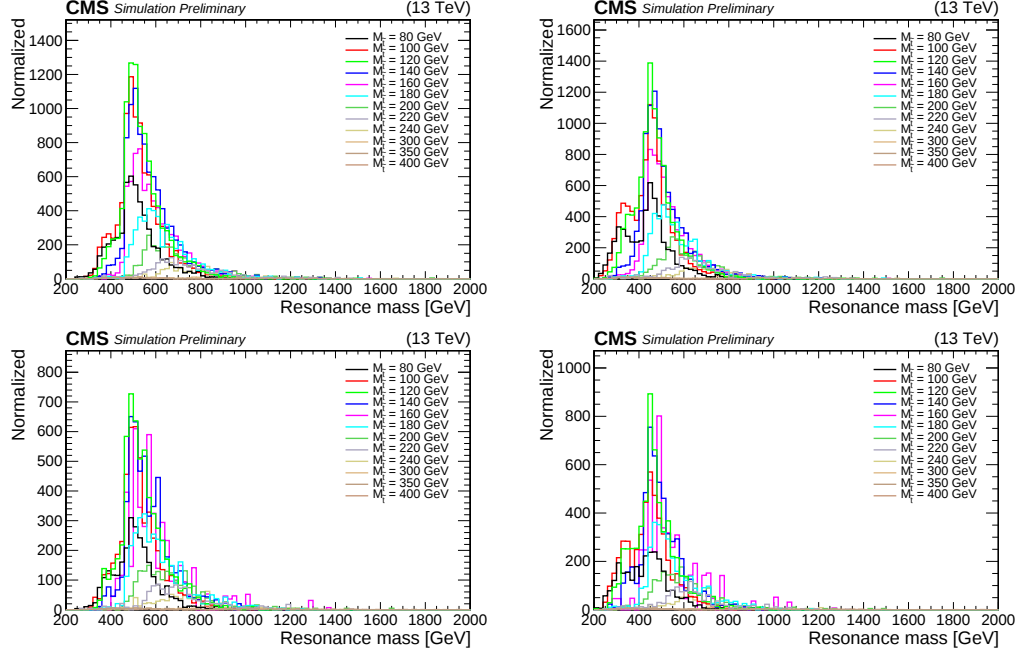


Figure 5.11: Leading jet (left) and second leading jet (right) p_T distributions after all the selection is applied for some selected RPV stop signal mass. Top plots for the inclusive selection, bottom plots for the b-tagged.

In addition, the τ_{21} selection introduces uncertainties on the jet mass scale and resolution. In the case of W-tagging, CMS measures a difference in data compared to MC in the W mass peak and resolution using the pruned jet mass distribution this data/MC ratio of reconstruct peak is measured to be of 1.0 ± 0.023 and for the mass resolution it is 1.23 ± 0.18 . Since we have no other boosted dijet resonance to test these effects, therefore, we apply the same systematic uncertainties on the two-prong tagger in this analysis. We also test for any bias of the $\bar{m}$ spectra using signal MC compared to the MC-truth. Figure 5.12 shows the truth value of the peak of the fitted Gaussian distribution in signal MC compared to the true $m_{\tilde{\tau}}$ value and find at most a 2% difference. This is consistent with the results using W bosons from $t\bar{t}$ [72]. Therefore, we can safely assign a 2% systematic uncertainty on the jet energy scale over the all mass ranges used in the analysis, and 18% for the mass resolution systematic uncertainty.

The τ_{32} selection also contributes to the signal acceptance systematic uncertainty in a

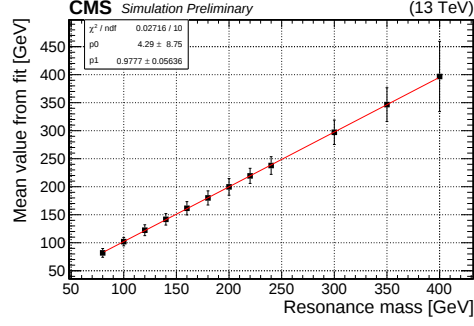


Figure 5.12: Comparison of the mean of the fitted Gaussian in the signal masses as a function of the true $m_{\tilde{t}}$.

similar way. Again, a scale factor is evaluated for the use of the $\tau_{32} < 0.54$ and the observed differences in data and MC in $t\bar{t}$ events [73]. According to this study, the data/MC scale factor for tagging one top quark is 1.07 ± 0.05 . Since we apply a τ_{32} as a veto, the veto scale factor is 0.99 ± 0.01 , which, when applied to two jets is 0.96 ± 0.02 . Similar to the τ_{21} scale factor, the τ_{32} veto scale factor is applied to the $\bar{m}$ distributions of all MC samples and the uncertainty is taken as a systematic uncertainty.

5.3.2 B tagging scale factors

For the b-tagged selection, an additional efficiency scale factor is applied to MC in order to match efficiencies in the identification of b-jets observed in data. The scale factors are evaluated as a function of jet p_T and η and affect individual jets based on the parton flavor of the jet and the efficiency of each jet flavor. These scale factors are applied to the MC samples as event weights following the recommendation of CMS [75].

5.4 Background Estimate

The background estimate in the boosted search is composed by a data-driven QCD multijet component, and a resonant backgrounds modeled with simulated samples which are validated in a $t\bar{t}$ enriched sample.

5.4.1 $t\bar{t}$ background

MC samples are utilized to model the resonant background components, i.e. $t\bar{t}$, $W(q'\bar{q}) + jets$, $Z(q\bar{q}) + jets$ and dibosons. From Figs. 5.6 and 5.8 and Table 5.6, we see that $t\bar{t}$ is the next most significant background after QCD multijets.

In order to validate the $t\bar{t}$ MC sample, we select a $t\bar{t}$ enriched sample by applying a $\tau_{32} < 0.57$ selection on both of the selected jets [73]. The aim is study a $t\bar{t}$ data/MC scale factor and to estimate the uncertainty associated with it.

CMS has developed four different working points corresponding to different selections using the τ_{32} variable to select boosted top jets. Different working points are tested on top of the inclusive selection described in Section 5.2.2 and the value which minimally reduces the signal region while enhancing the $t\bar{t}$ background is chosen for these studies. In addition, the τ_{32} selection is chosen to be orthogonal to the signal selection in order to ensure a signal-free region to validate the $t\bar{t}$ background.

To choose the top selection, two τ_{32} working points are tested, namely $\tau_{32} < 0.57$ and $\tau_{32} < 0.67$. Additionally, the effects of the τ_{21} selection in this region are also tested by comparing results using the following selections:

1. Preselection + $m_{\text{asym}} < 0.1 + \Delta\eta < 1.5 + \tau_{32} < 0.57(0.67) + \tau_{21} < \mathbf{0.45}$
2. Preselection + $m_{\text{asym}} < 0.1 + \Delta\eta < 1.5 + \tau_{32} < 0.57(0.67) + \tau_{21} > \mathbf{0.45}$
3. Preselection + $m_{\text{asym}} < 0.1 + \Delta\eta < 1.5 + \tau_{32} < 0.57(0.67)$

By inverting τ_{32} , the six selections are orthogonal to the inclusive region. The $\bar{m}$ distribution for these six selections are shown in Fig. 5.13, where the top plots correspond to the $\tau_{32} < 0.57$ and the bottom to the $\tau_{32} < 0.67$ working points. The bottom panels show the ratio of data minus all the MC backgrounds except for $t\bar{t}$, divided by the $t\bar{t}$. We see that the different selection have varying contribution from $t\bar{t}$. A similar performance is seen for the selections with inverted τ_{21} and no τ_{21} selection, for both top tagging working points. It

is also noted that the looser τ_{32} selection increases the amount of QCD multijets and other background contributions. Therefore, for the $t\bar{t}$ studies we use the selection summarized in Table 5.7 and shown (again) in Fig. 5.14.

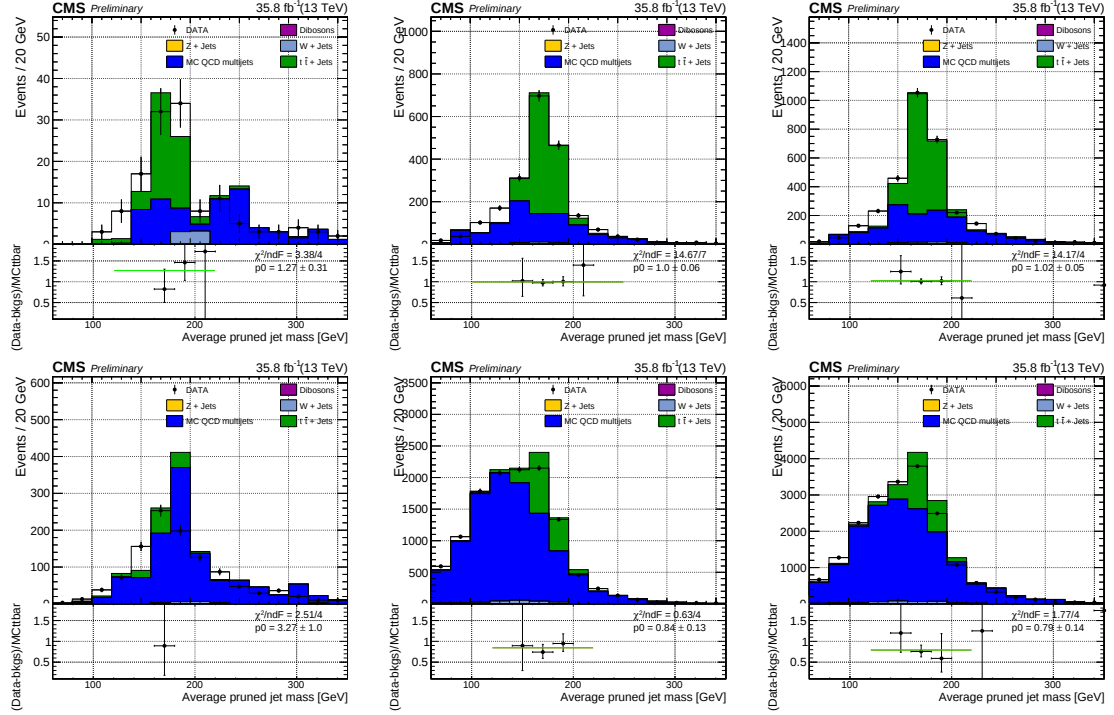


Figure 5.13: Average pruned jet mass distributions for the top region selection. Top row: top tagging working point of $\tau_{32} < 0.57$. Bottom row: $\tau_{32} < 0.67$. On both selections the effect of the τ_{21} selection is tested as described in the text. Left: selection 1 described in the text. Middle: selection 2. Right: selection 3.

The bottom panel of Fig. 5.14 shows the ratio is fit to a constant parameter (p_0), which is found to be $p_0 = 1.03 \pm 0.05$. Therefore the $t\bar{t}$ MC is consistent with that in the data and therefore no scale factor is applied. However, the uncertainty in the scale factor is taken as a systematic uncertainty on the fit. Additionally to this uncertainty, we vary the QCD multijet MC to match the observed discrepancy with data in the tails of Fig. 5.14, and assign the difference as an additional systematic. The total scale factor and uncertainty is $1 \pm 0.05(\text{fit}) \pm 0.09(\text{syst}) = 1 \pm 0.1$.

Finally, since it is difficult to select reasonably pure $W(q'\bar{q})+jets$, $Z(q\bar{q})+jets$ or diboson samples in data using similar control regions, we similarly assign a 10% on these background

Table 5.7: Top selection for $t\bar{t}$ scale factor study. Selection includes the preselection requirement of Table 5.2. Selections in τ_{21} and τ_{32} refers to cuts in both jets.

$\bar{m}$ spectra (GeV)	m_{asym}	$\Delta\eta$	τ_{21}	τ_{32}
60-450	< 0.1	< 1.5	> 0.45	< 0.57

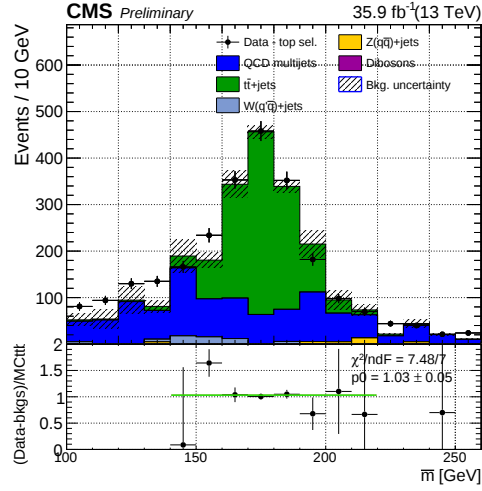


Figure 5.14: Data/MC comparison of the $\bar{m}$ distribution after the top selection of Table 5.7 is applied. Data is shown in dots while the different background MC components are stacked with different colors. Here, the QCD multijets background is estimated from MC and varied to evaluate the systematic effect.

estimates from MC. For $W(q'\bar{q})/Z(q\bar{q}) + jets$ events, higher-order p_T -dependent electroweak NLO corrections are applied to improve the modeling of the kinematic distributions of those events [76, 77, 78, 79, 80]. In addition, it is worth noting that the resonant background components correspond to only roughly 2% of the total background over the entire mass search region after all the selection is applied, as shown in Table 5.6. However, the effect is higher at lower masses where the resonant backgrounds are the largest, as seen in Figs. 5.6 and 5.8.

5.4.2 Background estimate for inclusive signal region

To estimate the QCD multijets background we perform the data-driven so-called ABCD method as done in previous searches [25]. In this method, four regions are defined in a two dimensional kinematic space based on two uncorrelated variables where one region is dominated by signal and the other three regions by background events. Several kinematic variables were explored and we find that the best variables to use for this analysis are m_{asym} and $\Delta\eta$. These two variables are then used to define four regions as shown in Table 5.8. Region A is defined by the nominal selection criteria for the signal region, B and C are regions where the selection of only one of the two variables is applied, and region D when both variables fail the selection.

Table 5.8: Definition of the regions A,B,C,D for the ABCD method to determine the QCD multijet background.

	$m_{\text{asym}} < 0.1$	$m_{\text{asym}} > 0.1$
$\Delta\eta > 1.5$	Region B	Region D
$\Delta\eta < 1.5$	Region A	Region C

Figure 5.15 shows the 2D distribution of m_{asym} as a function of $\Delta\eta$ for the signal MC with $m_{\tilde{t}} = 100 \text{ GeV}$, QCD multijets MC and for data. The correlation factor between these two variables is negligible: for QCD multijets MC it is around -0.01, for signal with $m_{\tilde{t}} = 100 \text{ GeV}$ it is 0.08, while for $m_{\tilde{t}} = 180 \text{ GeV}$ it is 0.03, and for the data it is -0.01. The signal contamination in regions B, C and D, where the background component must

be predominant, is found to be 2.8% in region B, 1.8% in region C and 0.5% in region D for $m_{\tilde{\tau}} = 100$ GeV (1.5%, 1.4% and 0.2% for $m_{\tilde{\tau}} = 180$ GeV, respectively). For other signal mass points, these numbers are found to be similar.

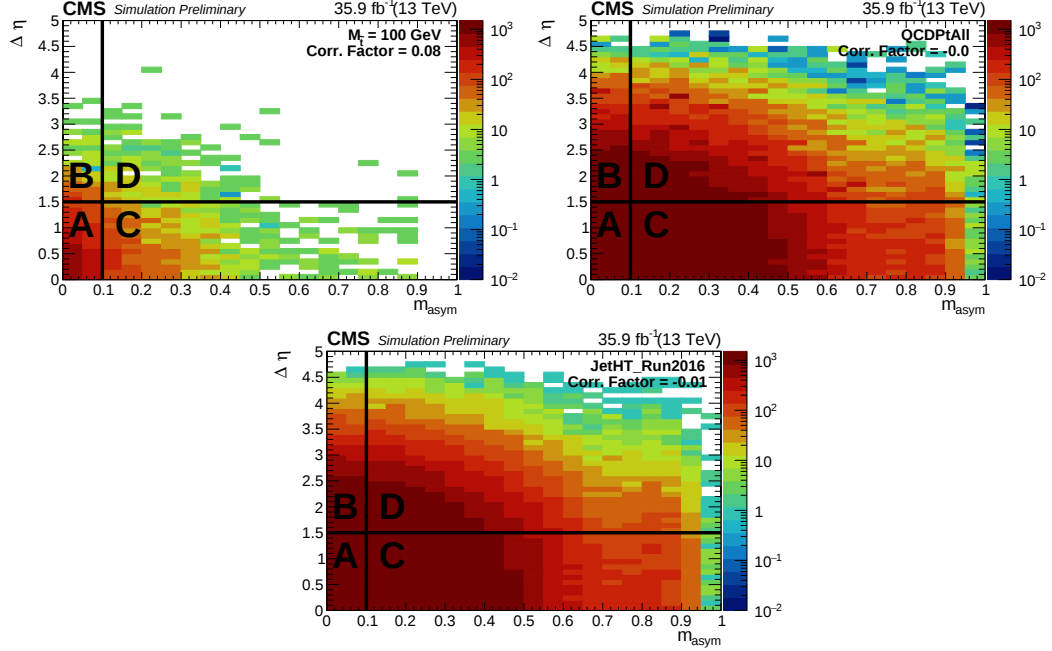


Figure 5.15: Distributions of m_{asym} as a function of $\Delta\eta$ for signal MC with $m_{\tilde{\tau}} = 100$ GeV (top left), QCD multijets MC (top right), and 2016 data (bottom). Regions A (signal region), B (signal+sideband), C (sideband+signal), D (sideband+sideband) are indicated, as described in Section 5.4.2. The correlation factors are also shown.

In the ABCD method, we extrapolate the shape of the $\overline{m}$ distribution of the background in the signal region using the B , C , D sideband regions as follows: Bkg in region $A = (B * C)/D$. We defined the ratio B/D as a transfer factor which is applied to region C to estimate both the shape of the $\overline{m}$ distribution and the scale of the background from data projected into the signal region. Since this method models only the QCD multijets background, the other MC background components ($W(q'\overline{q}) + jets$, $Z(q\overline{q}) + jets$ and dibosons) are subtracted from the data in regions B, C and D, using MC samples. Although the overall effect of subtracting the resonant backgrounds is small it is necessary in order to not double count backgrounds in regions B, C and D.

The transfer factor is parametrized by fitting the ratio of the $\overline{m}$ spectra in region B over

region D using the following sigmoid function:

$$\frac{1}{p_0 + \exp(p_1 + p_2 \bar{m}^2 - p_3 \bar{m}^3)}, \quad (5.3)$$

where p_0 - p_3 are free parameters of the function. This function models the ratio B/D as a function of $\bar{m}$ well in MC and in data, as shown in Fig. 5.16. The transfer factor is shown in Fig. 5.16 and demonstrates the small effect of subtracting the resonant backgrounds. The $\chi^2/ndf = 1.6/5$ in data which indicates that the fit function describes the transfer function well. Although we do not use the transfer factor fit from MC, or the fit to the data without the resonant background subtraction, finding similar results gives confidence that the method is working and the fit is stable. For the final background estimate the fitted transfer factor is multiplied by the $\bar{m}$ distribution of events in region C in data. The uncertainty in the fit transfer factor and the statistical uncertainty in the $\bar{m}$ distribution in region C are assigned as systematic uncertainty on the ABCD method.

To test the stability of the ABCD method, a closure test is performed and is shown in Fig. 5.16 (right). In the case of the resonant backgrounds, the MC samples are used as described in the previous section. We compare the ABCD prediction using MC only and compare this to the predicted events in region A (signal region). The bottom panel of Fig. 5.16 (right) shows the ratio of the MC ABCD prediction to region A and the level of disagreement is found to be within $\pm 10\%$ (blue lines in the ratio plots). This level of agreement is used as a source of systematic uncertainty on the QCD multijet background prediction.

5.4.3 Background estimate for b-tagged signal region

For the b-tagged signal region, the background estimate is performed in a similar way as for the inclusive region. For the ABCD QCD multijets background method unfortunately we find a lack of statistics in the tails of the $\bar{m}$ distributions in the regions B and D when b

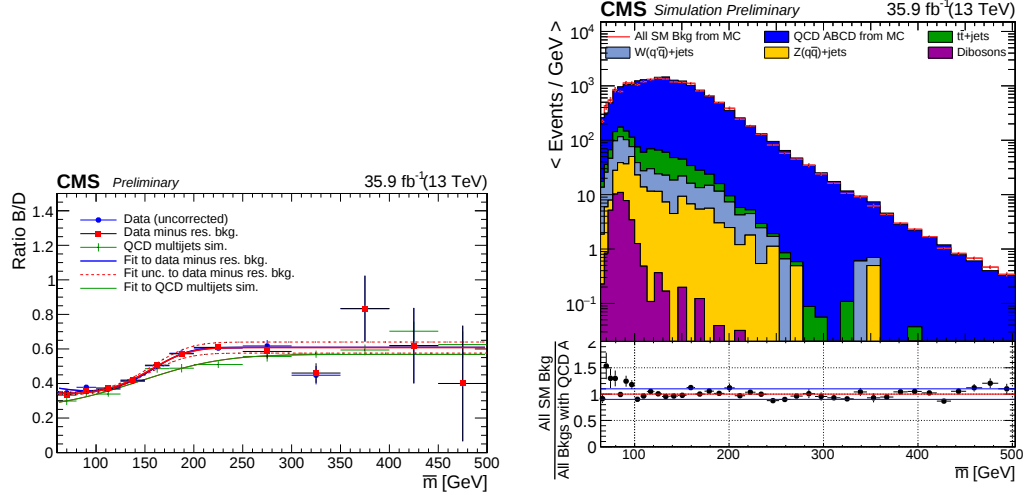


Figure 5.16: Left: Transfer factor ratio B/D as a function of $\bar{m}$ for data (blue dots), data minus resonant MC backgrounds (red dots) and the sum of all the MC backgrounds (green crosses). Fits to these distributions with the sigmoid function are also shown in the colored lines. Right: Closure test. Red line represents the sum of all MC backgrounds in the signal region (region A), the $t\bar{t}$ MC samples is in green, $W(q'\bar{q}) + jets$ in gray, $Z(q\bar{q}) + jets$ in yellow and dibosons in purple. The blue shows the ABCD prediction from MC only. The bottom panel shows the ratio between both distributions. The blue lines indicate the $\pm 10\%$ level of agreement.

tagging is applied and therefore the QCD is not well-modeled. Instead we use the transfer factor from the inclusive selection and apply it to region C with the b-tagged selection. The differences between the fitted sigmoid transfer factors for the inclusive and b-tagged selections from MC (left) and data (right) are shown in Fig. 5.17. Although no significant differences are found in Fig. 5.17, the uncertainty in the transfer factor for the b-tagged analysis is inflated to cover this spread, as shown in Fig. 5.18 (left). Finally, we perform a similar closure test for the b-tagged ABCD background estimate using the inclusive transfer function and is shown in Fig. 5.18 (right). We assign a similar additional 10% systematic due to the MC closure test.

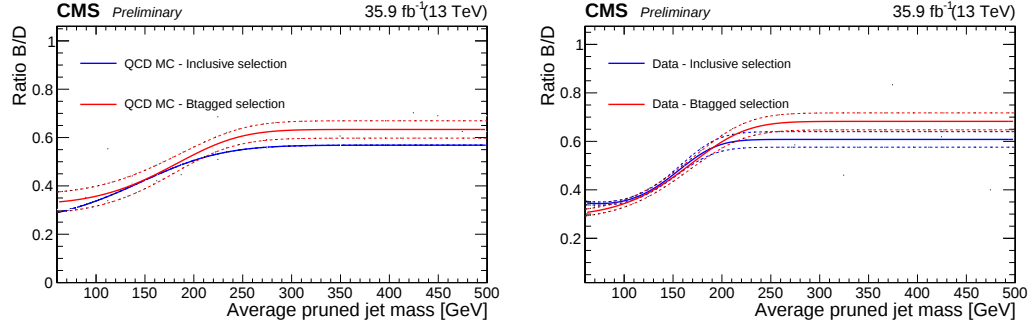


Figure 5.17: Comparison between transfer factors fitted to sigmoid function using the inclusive selection (blue line) and the b-tagged selection (red line). Left: QCD multijet MC. Right: data.

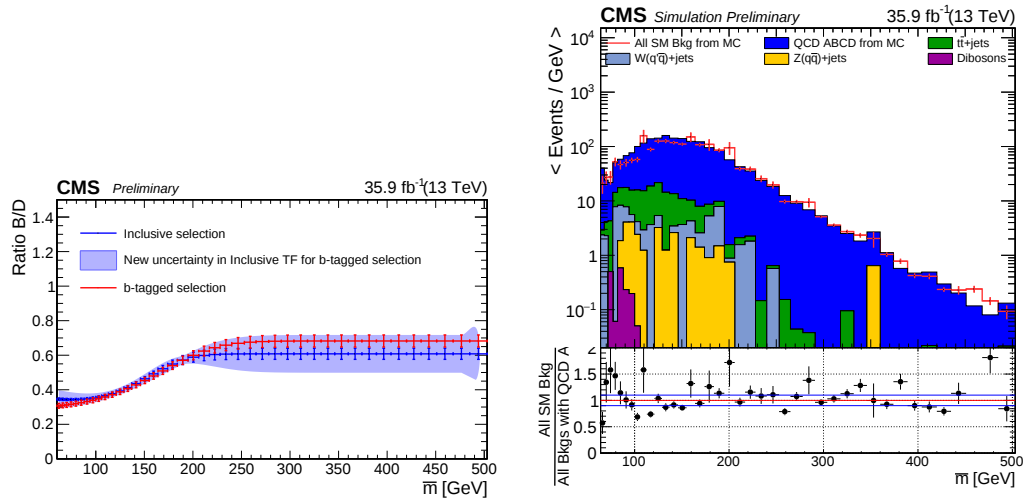


Figure 5.18: Left: inflated uncertainty (blue band) in the transfer factor with inclusive selection (blue crosses) to cover the uncertainties in the transfer factor with the b-tagged selection (red crosses). Right: Closure test. Red line represents the sum of all MC backgrounds in the signal region (region A), the $t\bar{t}$ MC samples is in green, $W(q'\bar{q}) + jets$ in gray, $Z(q\bar{q}) + jets$ in yellow and dibosons in purple. The blue shows the ABCD prediction from MC only. The bottom panel shows the ratio between both distributions. The blue lines indicate the $\pm 10\%$ level of agreement.

5.4.4 Final background estimate

Figure 5.19 shows the $\bar{m}$ distributions for the boosted search. This figure shows a comparison of the data and the final background prediction for the inclusive selection (left) and the b-tagged selection (right). The shaded area around the background estimate corresponds to the total background uncertainty. Since there is no evidence for new physics, we proceed to set 95% confidence level upper limits on the cross section as a function of $m_{\tilde{\tau}}$ in the two RPV SUSY models, as described in Chapter 7.

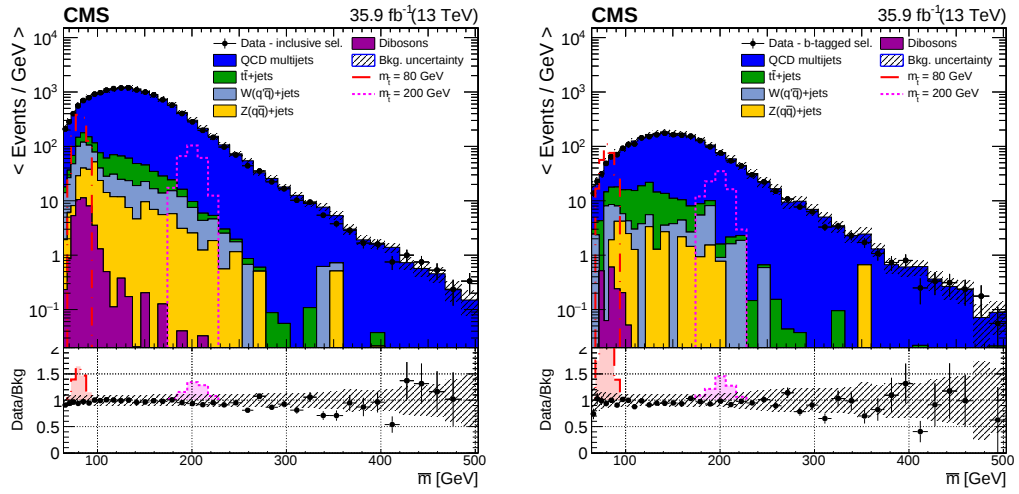


Figure 5.19: Average pruned mass spectrum for the inclusive selection (left) and b-tagged selection (right). We show the data (dots) and the predictions for the data-driven QCD multijets (blue), $t\bar{t}$ (green), $W(q'\bar{q}) + jets$ (gray), $Z(q\bar{q}) + jets$ (yellow), and dibosons (purple). The gray shaded area around the total background prediction shows the total uncertainty on the background estimate. Two signal MC at $m_{\tilde{\tau}} = 80$ GeV and $m_{\tilde{\tau}} = 200$ GeV are also shown for both RPV scenarios. The bottom panels show the ratio of the data to background prediction with statistical errors (black dots) and background systematic uncertainty (shaded area).

5.5 Systematic Uncertainties

In this section we describe the systematic uncertainties on the background estimate and the signal acceptance for the boosted analysis.

5.5.1 Signal Systematic Uncertainties

The sources of systematic uncertainties we consider for the signal are listed and described below. They either affect the yield of the signal acceptance or the mass resolution (shape) of the $\overline{m}$ spectra, and are summarized in Table 5.9.

Integrated Luminosity (Yield): The luminosity measurement is described in detailed in Section 3.2.7. A systematic uncertainty on the integrated luminosity of 2.5% is assigned, according to the most recent measurement using the pixel detector [41].

Trigger (Yield): Based on the observed differences in efficiency of the triggers as a function of the $\overline{m}$ spectrum shown in Fig. 5.1, we assign a 3% uncertainty.

Two-prong tagger scale factor (Yield): As described in Section 5.3.1, the effect of the two-prong τ_{21} selection is also considered. The scale factor for the two-prong tagger is 1.21 ± 0.29 . This scale factor is applied to all MC samples and a total of 23% uncertainty is used.

Anti-three-prong tagger scale factor (Yield): As described in Section 5.3.1, the effect of the anti-three-prong τ_{32} selection is also considered. The scale factor for the anti three-prong tagger is 0.96 ± 0.02 . This scale factor is applied to all MC samples and a 2% uncertainty is used.

Jet Energy Scale (Shape): As as described in Section 5.3, there is a 2% difference in the position of the peak of the W mass distribution in MC compared to data when apply a τ_{21} selection. This is effectively due to the jet energy scale differences affecting the mass. Therefore, a 2% uncertainty is applied as a shape systematic on the $\overline{m}$ distribution.

Jet Energy Resolution (Shape): There is an 18% data and MC difference in the width of W bosons tagged with τ_{21} , as discussed in Section 5.3. Again this is due to energy

scale differences affecting the mass. Therefore, we smear the signal mass spectra in MC by 18% to account for this effect on the signal shape.

Jet Energy Scale (Yield): In addition, jet energy scale corrections applied to the MC samples have associated uncertainties, as described in Section 4.1.3. These uncertainties applied to the jet p_T , are propagated almost linearly to the jet mass. To estimate this effect, the acceptances in the $\bar{m}$ distribution using jets with the nominal JEC correction are compared with the JEC correction ± 1 standard deviation. The differences in acceptance as a function of $m_{\tilde{\tau}}$ are shown in Appendix A.1 From those studies, we found an effect of 1.2% on average for all $m_{\tilde{\tau}}$ mass points.

Jet Energy Resolution (Yield): The jet energy resolution measurement also affects the signal yield as discussed in Section 4.1.3. This uncertainty is estimated by the yield difference entering the $\bar{m}$ distributions using jets with the nominal JER value compared with the distributions using jets with a JER ± 1 standard deviation. The differences as a function of $m_{\tilde{\tau}}$ are shown in Appendix A.2. The effect of the JER on the signal acceptance is found to be on average 1.8% for all $m_{\tilde{\tau}}$ points.

Pileup Re-weighting (Yield): The impact of the pileup in the generation of the signal MC samples is also studied. This uncertainty is estimated by a $\pm 2\%$ standard deviation shift of the minimum bias cross section, as shown in Appendix A.3. From those studies, we find this to be a 1% effect on the signal acceptance.

PDF (Yield): To estimate the uncertainty arising from PDFs, the *PDF4LHC recommendations for LHC Run II* [81] are followed. The root mean square (RMS) of the distribution of the 100 NNPDF MC replicas as the $\pm 1\sigma$ shape variation due to parton distribution functions are evaluated. The effect of the PDF uncertainty is evaluated in Appendix A.4, and it is found to be a 1% effect.

MC statistics (Shape): A bin-by-bin uncertainty due to limited MC statistics in the signal samples is also included as it affects the shapes of the $\bar{m}$ spectra.

Table 5.9: Summary of the systematic uncertainties on the signal. The uncertainty values affecting the resonance shape refer to the value of the uncertainty itself, not the effect on the acceptance.

Source of Systematic	Effect	Value
Luminosity	Yield	2.5%
Trigger	Yield	3.0%
Pileup	Yield	1.0%
PDF	Yield	1.0%
Two-prong Tagger Scale Factor	Yield	23.0%
Three-prong Tagger Scale Factor	Yield	1.0%
Jet Energy Scale	Yield	1.2%
Jet Energy Resolution	Yield	1.8%
B tagging (only for b-tagged selection)	Yield	1.0%
MC Statistics	-	bin-by-bin
Jet Energy Scale	Resonance Shape	2.0%
Jet Energy Resolution	Resonance Shape	18.0%

B tagging (Yield): For the b-tagged selection, we include the uncertainty on the data/MC scale factor of the efficiencies in the identification of b jets, as described in Section 4.1.3. The CMS recommendation is followed for these scale factors using the event weights procedure [75], and the uncertainty is evaluated by varying these scale factors ± 1 standard deviation and studying the difference in signal acceptance as shown in Appendix A.5. The effect of this uncertainty is evaluated to be 1%.

5.5.2 Systematic Uncertainty on Background Estimate

For the data-driven ABCD method used to estimate the QCD multijets background, three possible sources of uncertainties are considered. First, the statistical uncertainty on the number of events in region C is taken into account bin-by-bin. Next there is an uncertainty assigned to the transfer factor shape from the fit parameter errors and are also applied bin-by-bin as a shape systematic. This transfer factor uncertainty is inflated in the case of the b-tagged selection, as discussed in Section 5.4.3. Finally, a systematic uncertainty on the ABCD method derived from the closure test is assigned as described in Section 5.4.2 and found to be 10%; it is taken as correlated among all bins.

Table 5.10: Summary of the systematic uncertainties on the background predictions by source. Uncertainties in the b-tagged selection, which are different from the inclusive selection, are shown parenthesis.

Background	Source of Systematic	Effect	Value
QCD multijets	Closure	Yield	10%
ABCD method	Transfer Factor Fit Uncertainty	Shape	1–8% (5–18%)
	Statistics in Sideband Region (C)	Shape	bin-by-bin
Resonant backgrounds	Systematic in MC Backgrounds	Yield	10%
	MC Statistics	Shape	bin-by-bin

For the case MC resonant backgrounds, two possible sources of uncertainties are applied. First, the statistical uncertainties on the MC samples are considered bin-by-bin. In addition, based on the $t\bar{t}$ scale factor study from Section 5.4.1, we apply a 10% systematic uncertainty on the modeling of each of the MC samples.

The systematic uncertainties on the background predictions are summarized in Table 5.10.

Chapter 6

Resolved search

This chapter describes the search for pair dijet resonances where the decay products are well separated or resolved. This search is performed in events with four high- p_T jets consistent with having been produced by a pair of resonant particles. This chapter starts with the trigger selection and continues with the description of the analysis event selection and the analysis strategy. The average mass spectrum of the two dijet pairs ($\overline{M}$) is used to search for the presence of a localized resonant signal over a large steeply falling background from standard model QCD multijet events. Although we study the background shapes with MC simulations, the final background shape is estimated entirely from data using a functional form. The chapter concludes with the description of the systematic uncertainties.

6.1 Trigger Selection

To maximize the acceptance of events containing four jets we use an *OR* of hadronic triggers selected with AK4 jets. The triggers used are summarized in Table 6.1. Besides the HLT_PFHT800 trigger described in Section 5.1, H_T^{AK4} based triggers with an additional p_T selection in the fourth jet (HLT_PFHT750_4JetPt50) are also included. In addition, the HLT_PFJet450 trigger path is included to mitigate an observed inefficiency at high values of HT caused by the HTT L1 seeds in Run period H.

To measure the efficiency of these triggers an orthogonal dataset **SingleMu** and an orthogonal path HLT_Mu50 is used. This path is also listed in Table 6.1. The efficiencies are measured as a function of H_T^{AK4} after applying a basic kinematic selection on the jets:

$p_T > 80$ GeV and $|\eta| < 2.5$. The trigger efficiency is defined as in Eq. 5.1 for the triggers described in Table 6.1. The efficiency measured as a function of H_T^{AK4} and the p_T of the fourth leading jet is shown in Fig. 6.1 (left) and as a function of H_T^{AK4} (right). This study shows that the triggers are fully efficient for $H_T^{\text{AK4}} > 900$ GeV.

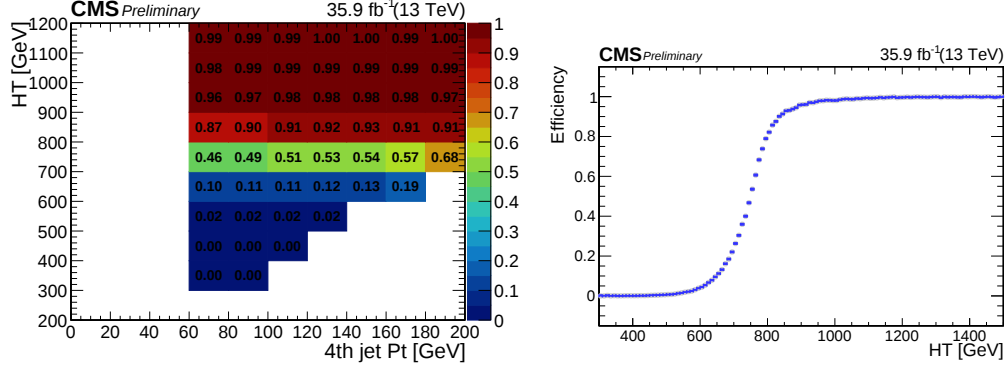


Figure 6.1: Trigger efficiency turn-on curves measured with the SingleMuon dataset. Left: trigger efficiency measured as a function of H_T^{AK4} and fourth leading jet p_T . Right: trigger efficiency as a function of H_T^{AK4} .

Figure 6.2 (left) shows the trigger efficiency as a function of H_T^{AK4} zoomed in the turn-on region. To account for the small inefficiencies observed from 900 to 1000 GeV, we later assign a 3% systematic uncertainty to the overall trigger efficiency. Furthermore, we measure the trigger turn-on as a function of $\overline{M}$ (defined in Section 6.2.3) shown in Fig. 6.2 (right). We conclude that the trigger is fully efficiency for $m_{\tilde{\tau}}$ above 160 GeV.

Table 6.1: Triggers used in the resolved analysis and the L1 and HLT definition are shown. Values in parenthesis indicates the increased thresholds during the data-taking period (Run H).

Trigger Name	L1 seed	HLT Definition
HLT_PFHT750_4JetPt50(70)	L1_HTT225 OR L1_HTT240 OR L1_HTT270 OR L1_HTT280 OR L1_HTT300 OR L1_HTT320	AK4PFJets with $p_T > 50(70)$ GeV and $ \eta < 3.0$ at least 4 jets and $HT > 750$ GeV
HLT_PFHT800_4JetPt50	L1_HTT225 OR L1_HTT240 OR L1_HTT270 OR L1_HTT280 OR L1_HTT300 OR L1_HTT320	AK4PFJets with $p_T > 50$ GeV and $ \eta < 3.0$ at least 4 jets and $HT > 800$ GeV
HLT_PFHT800(900)	L1_HTT225 OR L1_HTT240 OR OR L1_HTT270 OR L1_HTT280 OR L1_HTT300 OR L1_HTT320	AK4PFJets with $p_T > 40$ GeV and $ \eta < 3.0$ $HT > 800$ GeV
HLT_Mu50	L1_SingleMu22 OR L1_SingleMu25	muons with $p_T > 50$ GeV and $ \eta < 2.5$

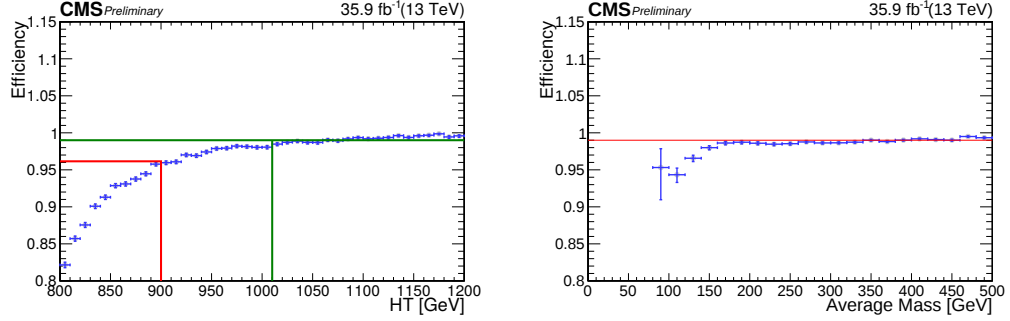


Figure 6.2: Left: trigger efficiency as a function of H_T^{AK4} zoomed in close to the turn-on region. Red line shows where we make the selection in $H_T^{\text{AK4}} \geq 900$ GeV, while the green line represents the 99% efficiency point. Right: Trigger efficiency turn-on curve as a function of the average dijet mass after the H_T cut is applied. Besides small inefficiencies at low masses, the trigger is fully efficient for masses above 150 GeV.

6.2 Event Selection

Events are preselected if they pass the trigger selection described in Section 6.1 and if the AK4 jets pass the TightLepVeto jet ID selection as described in Section 4.1.3. Jets are further required to have $p_T > 80$ GeV in order to reduce the QCD multijets background, as shown in Fig. 6.3. The number of jets required in the analysis was also optimized and we select exactly four jets.

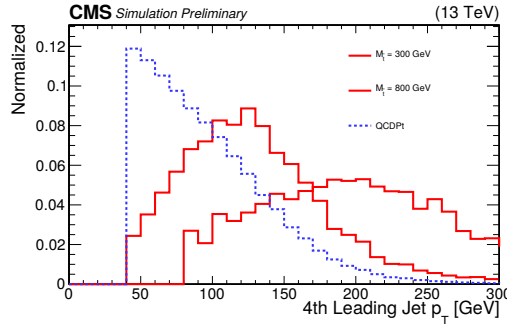


Figure 6.3: Fourth leading jet p_T for QCD multijets MC (blue) and for RPV stop signal MC samples at $m_{\tilde{t}} = 300$ GeV and $m_{\tilde{t}} = 800$ GeV (red). By selecting jets with $p_T > 80$ GeV we can further reject some of the QCD multijets background.

The scalar sum of the transverse momentum of all the jets with the requirements above defines the offline H_T^{AK4} of the event. Based on the trigger studies, a minimum

$H_T^{\text{AK4}} \geq 900 \text{ GeV}$ guarantees events where the trigger is approximately fully efficient. The preselection criteria is summarized in Table 6.2.

Table 6.2: Summary of preselection criteria for resolved analysis.

Variable	Cut Value
AK4 Jets Jet p_T	$> 80 \text{ GeV}$
Jet $ \eta $	< 2.5
N jets	4
H_T^{AK4}	$> 900 \text{ GeV}$

6.2.1 Dijet Pairing

Once an event passes the preselection requirements from Table 6.2, dijet pairs are created from the leading four jets sorted in p_T (j_1, j_2, j_3, j_4). From these jets, there are three unique combinations of dijet pairs: ($j_1 j_2, j_3 j_4$), ($j_1 j_3, j_2 j_4$) or ($j_1 j_4, j_2 j_3$). From these combinations at most one pairing correctly corresponds to the true stop pair decay products in a signal event. To distinguish the dijet pair configuration most likely associated with its parent particle in signal events, a spatial variable is employed that uses the separation ($\Delta R(j_k, j_l)$) of the two jets within each dijet, where $\Delta R(j_k, j_l) = \sqrt{\Delta\eta^2(j_k, j_l) + \Delta\phi^2(j_k, j_l)}$. The dijet pair configuration that results in the minimal value of $dR(j_k j_l, j_m j_n)$ in Eq. (6.1) is selected. We define:

$$dR(j_k j_l, j_m j_n) = |\Delta R(j_k, j_l) - 0.8| + |\Delta R(j_m, j_n) - 0.8|, \quad (6.1)$$

where the 0.8 value guarantees that the two jets with radius 0.4 (AK4) are not on top of each other. As shown in Appendix B.1, two other pair reconstructions were studied. We choose the method that give us better signal significance and a smoother $\overline{M}$ distribution. This pairing was also shown to work best in a previous analysis [22].

6.2.2 Kinematic Variables

This search assumes that two identical particles are produced in pairs, therefore the average mass of the two dijet systems is defined as:

$$\overline{M} = \frac{m_{jj1} + m_{jj2}}{2}, \quad (6.2)$$

where m_{jji} is the mass of i -th dijet (jj) system. This is the main kinematic variable used in the search. In order to reduce background events from QCD multijets processes, the following variables are used for the dijet pair chosen according to the algorithm described in Section 6.2.1:

Mass asymmetry: This variable reduces the number of events with large mass imbalance between the two dijet systems, and it is defined as:

$$M_{\text{asym}} = \frac{|m_{jj1} - m_{jj2}|}{m_{jj1} + m_{jj2}}. \quad (6.3)$$

$\Delta\eta_{\text{dijet}}$: the dijet systems in the signal events are predominantly produced in the central η region, compared to the more wide spread QCD multijet, and therefore we apply a selection in the absolute value of the difference in the η between the dijet system, defined as:

$$\Delta\eta_{\text{dijet}} = |\eta_{jj1} - \eta_{jj2}|. \quad (6.4)$$

Δ : defined for each dijet system as

$$\Delta = pT_{j1} + pT_{j2} - \frac{m_{jj1} + m_{jj2}}{2}, \quad (6.5)$$

where the subindexes j_1 and j_2 referred to jets in a given dijet system (eg. jj_1 or jj_2). Requiring a minimum value of Δ results in a lowering of the peak position value of the $\overline{M}$ distribution from background QCD multijet events. With this selection the

modeling of the background shape can be extended to lower values.

Figures 6.4 and 6.5 show the distributions of these kinematic variables with the preselection applied from Table 6.2, where the discriminating power of each kinematic variable is illustrated by comparing QCD multijets MC and two RPV signal MC. The optimization procedure of these variables is further described in Section 6.2.3.

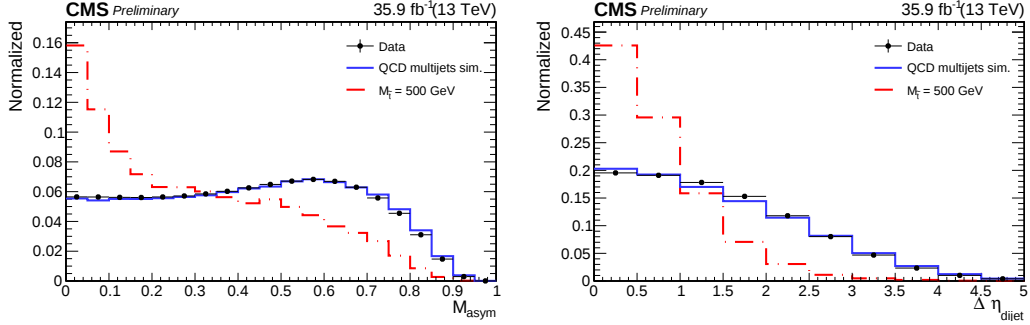


Figure 6.4: Distributions of kinematic variables after preselection from Table 6.2 and after choosing the best dijet pair. Distributions are normalized to unit area. Data corresponding to 35.9 fb⁻¹ is represented as dots, blue lines are QCD multijets MC, while the red lines represent RPV stop signals with masses with $m_{\tilde{t}} = 500$ GeV. Left: M_{asym} . Right: $\Delta\eta_{\text{dijet}}$.

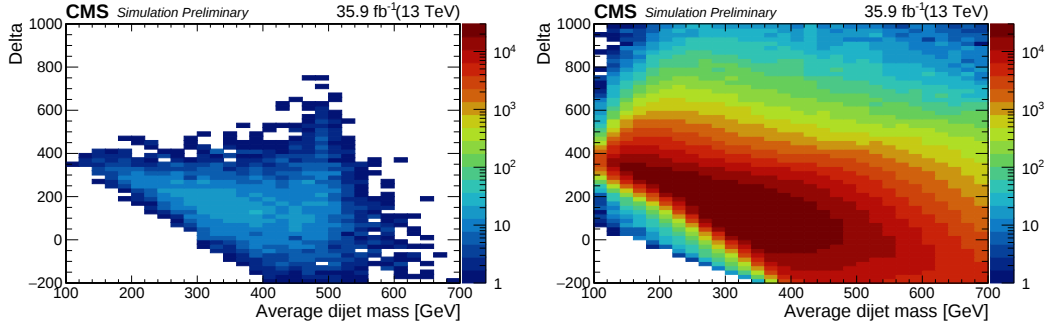


Figure 6.5: Distributions of Δ as a function of the $\overline{M}$ after preselection from Table 6.2 and after choosing the best dijet pair. Left: RPV $\tilde{t} \rightarrow qq'$ signal with $m_{\tilde{t}} = 500$ GeV. Right: data.

6.2.3 Signal Optimization

In this section, the procedure used to optimize the selection criteria using the variables described in Section 6.2.2 is presented. A similar iterative optimization method based on

$S/\sqrt{B}$ is performed as was done in Section 5.2.2, and using the signal MC described in Section 4.2.1 and the QCD multijets MC backgrounds described in Section 4.2.2.

The variables listed in Section 6.2.2 are scanned in the range of values shown in Table 6.3. After a particular selection is applied, we count the number of signal MC events (S) and background MC events (B) in a mass window within 50 GeV around the mass point considered. We repeat this for all values listed in Table 6.3.

Table 6.3: Threshold values for each variable used in the optimization procedure.

Variable	Starting Value	Ending Value	Step Value
M_{asym}	0	1	0.05
$\Delta\eta_{\text{dijet}}$	0	5	0.5
Δ	0 GeV	500 GeV	10 GeV

Figure 6.6 shows the steps in the optimization process for two signal mass points, $m_{\tilde{t}} = 300$ GeV (top) and $m_{\tilde{t}} = 700$ GeV (bottom). The x-axis shows arbitrary units for simplicity since we compare different variables with different ranges in the same figure. For instance, the $\Delta\eta_{\text{dijet}}$ variable, which is scanned from 0 to 5, although it is shown from 0 to 1; while the Δ variable is scanned from 0 to 500 GeV. We first choose the variable with the highest $S/\sqrt{B}$, apply the chosen selection and the process is repeated. For example, Fig. 6.6 (left) shows the $S/\sqrt{B}$ metric per variable after preselection from Table 6.2, whereas Fig. 6.6 (right) it is shown after preselection and a M_{asym} requirement. This procedure is studied for every signal MC mass point generated, and we find the optimized selection criteria is similar for all stop masses.

The Δ variable is further optimized. The angular separation between the dijet pairs creates a double-bump feature in the $\overline{M}$ spectrum [82]. Figure 6.7 (left) shows the $\overline{M}$ spectrum for varying Δ criteria for QCD MC; the first hump corresponds to events where the two dijets are separated a distance of 0.8, and the second corresponds to jets separated between $2\pi - 0.8$ and 2π . From the studies performed in Run I [22] it was shown that cutting on the Δ variable removes the double-bump features present in the average dijet

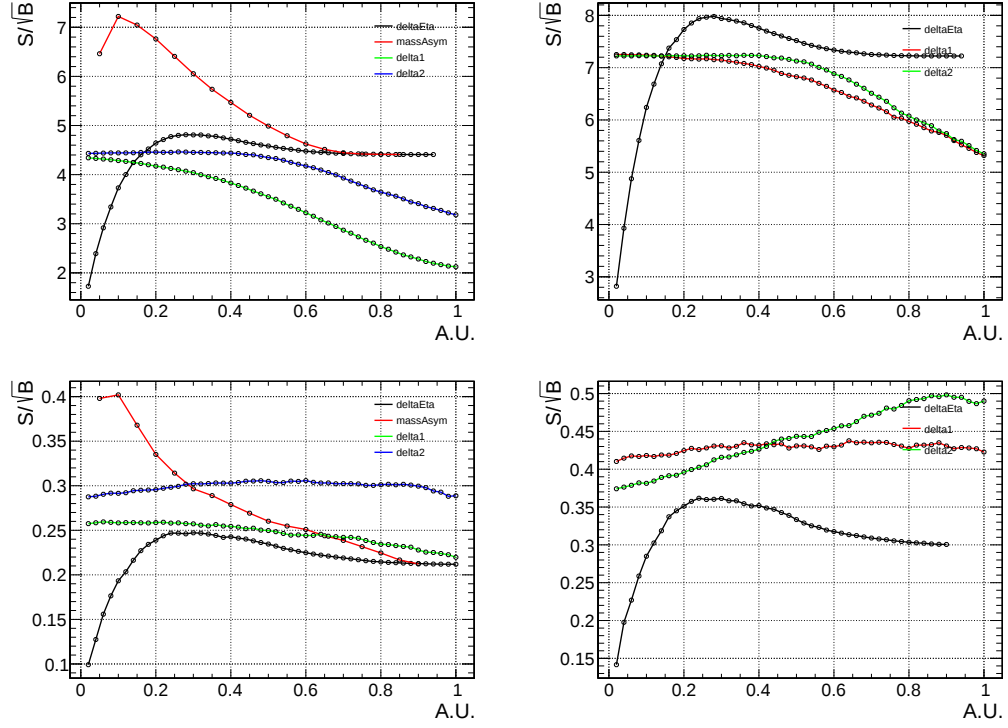


Figure 6.6: The optimization procedure. The $S/\sqrt{B}$ distributions for a given selection as a function of arbitrary units for $\tilde{t} \rightarrow qq'$ at $m_{\tilde{\tau}} = 300 \text{ GeV}$ (top) and at $m_{\tilde{\tau}} = 700 \text{ GeV}$ (bottom). Each colored line represents each variable studied, while the dots represent a selection in a given variable. Left: shows variables after the preselection. Right: after applying the preselection and a cut in $M_{\text{asym}} < 0.1$.

distribution in QCD multijets. The $S/\sqrt{B}$ for different Δ cuts is shown in Fig. 6.7 (right). Optimizing for both $S/\sqrt{B}$ and ensuring a smooth $\overline{M}$ distribution for background, we apply $\Delta > 200$ GeV.

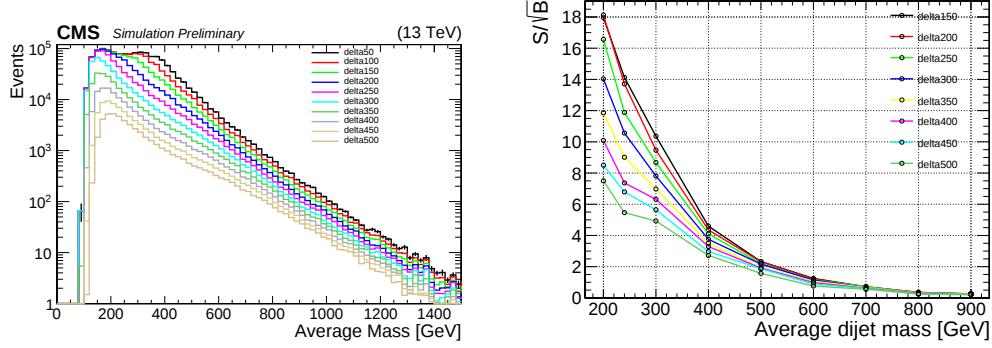


Figure 6.7: Left: The $\overline{M}$ distribution for QCD multijets MC. All selection listed in Table 6.2 and 6.4 are applied except for Δ . Different colors represent different values of the Δ cut. Right: $S/\sqrt{B}$ for different signal mass points and for different Δ selections. We choose the value of Δ with the most smoothly falling distribution and the highest $S/\sqrt{B}$.

Table 6.4 summarizes the selection criteria for the resolved analysis. Table 6.5 shows a cutflow table for QCD multijet MC, signal MC for $m_{\tilde{t}} = 400$ GeV and $m_{\tilde{t}} = 600$ GeV and data.

6.2.4 B-tagging optimization

For the b-tagged analysis, we apply b tagging to target the λ''_{323} coupling, and is applied in a similar way as in the previous analysis [22]. In addition to the inclusive selection, two possible b tag selections are studied:

1. one b tag is identified in any of the four selected jets,
2. one b tag is identified in each dijet pair.

Again, the b tagging algorithm used is the CSVv2 as described in Section 4.1.3. The CSVv2 distribution is shown in Fig. 6.8 with the preselection criteria of Table 6.4. In both cases, three different b tagging working points are tested: loose (CSVv2L), medium

Table 6.4: Optimized selection table for the resolved search, including the preselection shown in Table 6.2.

Inclusive selection	AK4 jets
	jet $p_T > 80 \text{ GeV}$
	jet $ \eta < 2.5$
	$H_T^{\text{AK4}} > 900 \text{ GeV}$
	$M_{\text{asym}} < 0.1$
	$\Delta\eta_{\text{dijet}} < 1.0$
	$\Delta > 200 \text{ GeV}$
B-tagged selection	Inclusive selection plus two loose b-tagged jets

(CSVv2M), and tight (CSVv2T), as recommended by CMS [74].

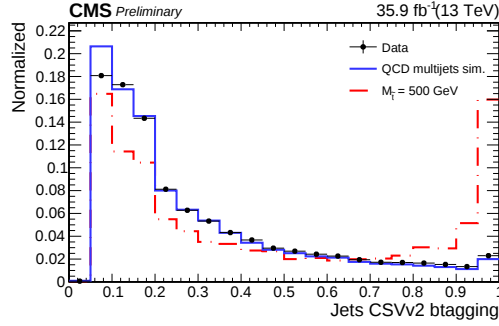


Figure 6.8: The CSVv2 distribution for the four jets used in the analysis after the preselection of Table 6.2.

To optimize the b tagging working point, the shape of the average dijet mass distribution is studied, as well as the $S/\sqrt{B}$ for different $m_{\tilde{t}}$ points, shown in Fig. 6.9. From these figures, the selection with the highest $S/\sqrt{B}$ is chosen which simultaneously minimally affects the shape of the $\overline{M}$ spectrum. We chose that one of the two jets in each dijet pair is b-tagged with the CSVv2 algorithm using the loose working point.

The selection for the b-tagged analysis is summarized in Table 6.4. In addition, a cutflow table is shown in Table 6.5, also indicating the expected signal yields for the signal events with the λ_{312}'' coupling.

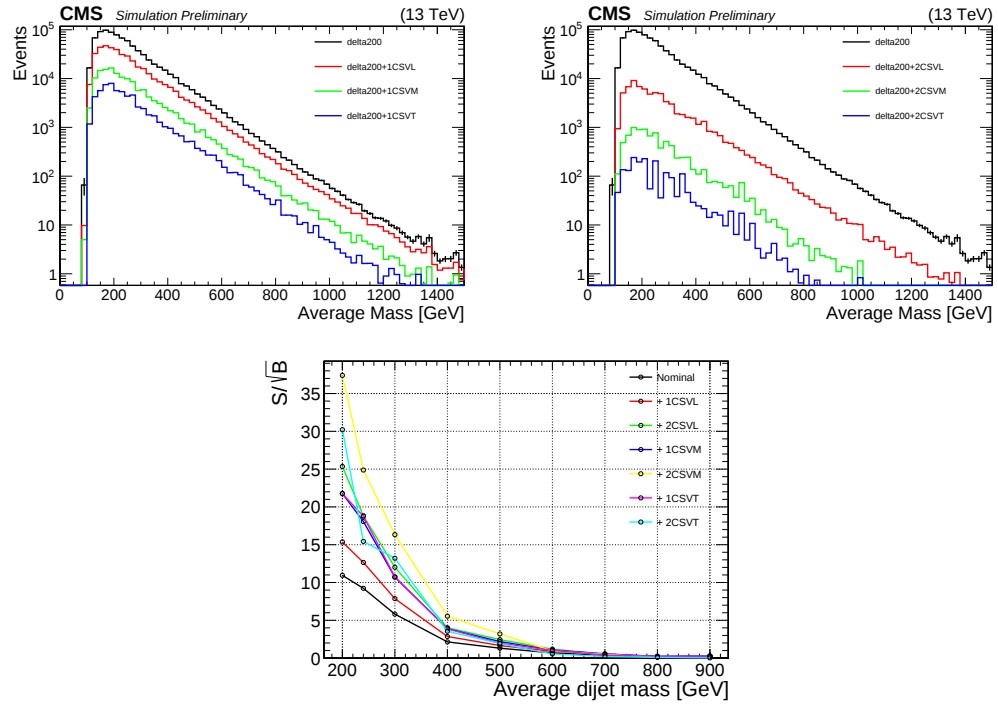


Figure 6.9: Top: average dijet mass distributions for QCD multijets MC after the inclusive selection described in Table 6.4 and for different b tagging criteria. Top left: one b tag is required in any of the four selected jets. Top right: one b tag is required in each dijet pair. Bottom: Distribution of $S/\sqrt{B}$ as a function of mass for different b tagging criteria.

Table 6.5: Cut flow table with sequential selection for the resolved search.

Selection	$\tilde{t} \rightarrow qq' m_{\tilde{t}} = 400 \text{ GeV}$		$\tilde{t} \rightarrow qq' m_{\tilde{t}} = 600 \text{ GeV}$	
	Events	%	Events	%
Processed	144116 ± 379.63	100.0	50578 ± 224.9	100.0
Trigger	36255 ± 190.41	25.0	33863 ± 184.02	67.0
Preselection	14060 ± 118.57	10.0	14523 ± 120.51	29.0
$M_{\text{asym}} < 0.1$	4904 ± 70.03	3.0	3445 ± 58.69	7.0
$\Delta\eta_{\text{dijet}} < 1.0$	3493 ± 59.10	2.0	2458 ± 49.58	5.0
$\Delta > 200 \text{ GeV}$	1478 ± 38.44	1.0	1076 ± 32.8	2.0
Selection	$\tilde{t} \rightarrow bq' m_{\tilde{t}} = 400 \text{ GeV}$		$\tilde{t} \rightarrow bq' m_{\tilde{t}} = 600 \text{ GeV}$	
	Events	%	Events	%
Processed	104822 ± 323.76	100.0	50172 ± 223.99	100.0
Trigger	19349 ± 139.1	18.0	26749 ± 163.55	53.0
Preselection	7857 ± 88.64	7.0	11981 ± 109.46	24.0
$M_{\text{asym}} < 0.1$	2452 ± 49.52	2.0	2654 ± 51.52	5.0
$\Delta\eta_{\text{dijet}} < 1.0$	1716 ± 41.42	2.0	1891 ± 43.49	4.0
$\Delta > 200 \text{ GeV}$	714 ± 26.72	1.0	765 ± 27.66	2.0
2 b tags	385 ± 19.62	0.1	1423 ± 37.72	0.3
Selection	QCD multijets		Data	
	Events	%	Events	%
Processed	214502736 ± 14645.91	100.0	127469181 ± 11290.23	100.0
Trigger	32254186 ± 5679.28	15.0	13987255 ± 3739.95	11.0
Preselection	22096803 ± 4700.72	10.3	9581153 ± 3095.34	7.6
$M_{\text{asym}} < 0.1$	2605567 ± 1614.18	1.2	1081692 ± 1040.04	0.9
$\Delta\eta_{\text{dijet}} < 1.0$	1262116 ± 1123.44	0.6	426695 ± 653.22	0.4
$\Delta > 200 \text{ GeV}$	1011657 ± 1005.81	0.5	225413 ± 474.78	0.2
2 b tags	305 ± 17.46	0.6	191 ± 13.82	0.7

6.3 Signal Acceptance

The selection described in Table 6.4 is applied to the signal MC samples. We fit the $\overline{M}$ spectra to Gaussian functions as shown in Figs. 6.10 and 6.11 for two $m_{\tilde{\tau}}$ points for the $\tilde{t} \rightarrow qq'$ and $\tilde{t} \rightarrow bq'$ with $m_{\tilde{\tau}} = 400$ GeV and $m_{\tilde{\tau}} = 600$ GeV. The signal efficiency is defined as the integral of the Gaussian fit in a 2σ mass window of the true $m_{\tilde{\tau}}$ divided by the total number of events generated. Figure 6.13 shows the shapes of the $\overline{M}$ distributions for signal MC generated with the λ''_{312} coupling (left) and λ''_{323} (right) after all the selection is applied. Figure 6.12 shows the acceptance for the inclusive and b-tagged selections.

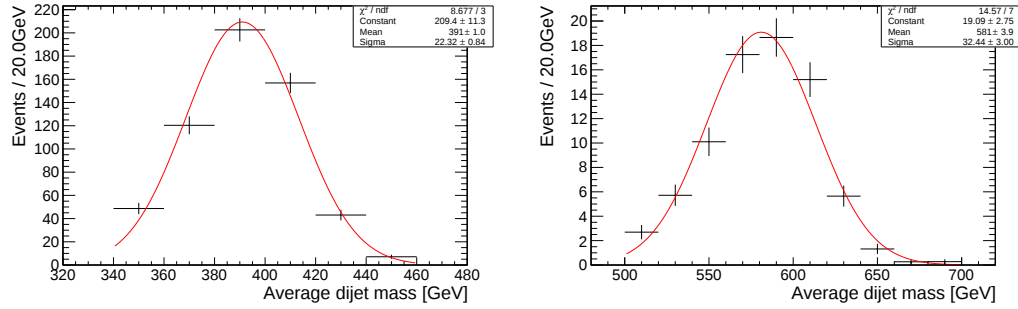


Figure 6.10: Gaussian fits to signal MC samples after applying the inclusive selection shown in Table 6.4. Here we show for the $\tilde{t} \rightarrow qq'$ with $m_{\tilde{\tau}} = 400$ GeV (left) and $m_{\tilde{\tau}} = 600$ GeV (right).

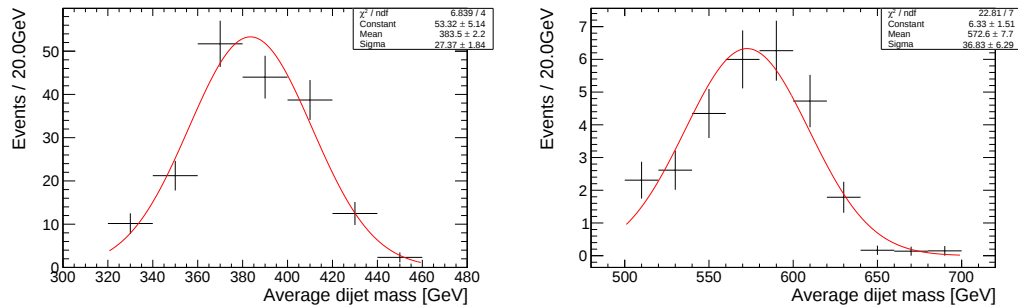


Figure 6.11: Gaussian fits to signal MC samples after applying the b-tagged selection shown in Table 6.4. Here we show for the $\tilde{t} \rightarrow bq'$ with $m_{\tilde{\tau}} = 400$ GeV (left) and $m_{\tilde{\tau}} = 600$ GeV (right).

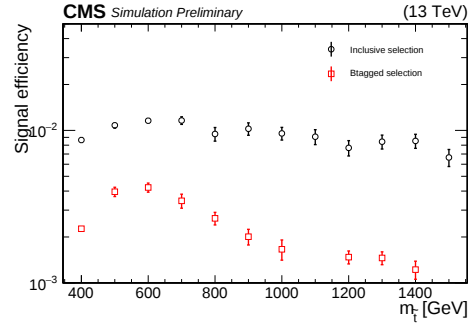


Figure 6.12: Signal efficiency in the resolved inclusive (black) and b-tagged (red) analyses.

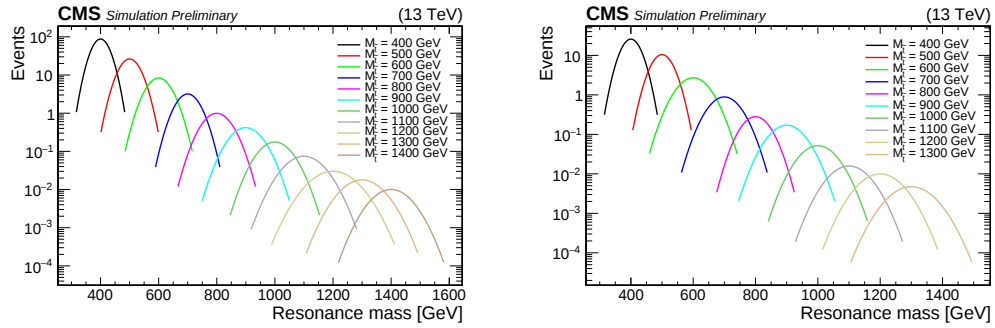


Figure 6.13: Gaussian signal shapes for RPV stop signal MC after all the selection is applied. Left: $\tilde{t} \rightarrow qq'$ signal. Right: $\tilde{t} \rightarrow bq'$ signal.

6.4 Background Estimate

The background composition in this search is dominated by QCD multijet production which its $\overline{M}$ distribution is steeply and monotonically falling. Therefore the $\overline{M}$ spectra of selected events can be modeled simply by using a multiple-parameter fit. This robust technique has been employed in several previously published CMS jet resonance searches [83, 84]. In this analysis, the empirical functional form from the CMS dijet analysis (P3 function) is chosen [83]:

$$\frac{dN}{d\overline{M}} = \frac{p_0(1-x)^{p_1}}{(x)^{p_2}}, \quad (6.6)$$

where $x = \overline{M}/\sqrt{s}$, $\sqrt{s} = 13000 \text{ GeV}$ is the center-of-mass energy of the LHC, and p_0 , p_1 , p_2 and p_3 are fitting parameters.

To determine the optimal point where to start the fit, the following metric is calculated bin-by-bin:

$$R(m) = \frac{1 - eff_{trigger}}{1/\sqrt{N_{sel}}}, \quad (6.7)$$

where $eff_{trigger}$ is the trigger efficiency from Fig. 6.2 (right) and N_{sel} is the number of events in data passing the event selection from Table 6.4, and m in this case is the starting point of the fit. If the value $R(m)$ is lower than 0.5, the inefficiency of the trigger is negligible compared to the relative statistical fluctuations in the data, and so the fit starting point is not affected by the trigger inefficiency. Figure 6.14 shows the value of the $R(m)$ as a function of the starting point of the true fit, the red points shows data and the black points show a pseudoexperiment throw from MC and scale to 35.9 fb^{-1} . From this test, we decide to start the fit at 350 GeV, allowing the search to start at $m_{\tilde{\tau}} = 400 \text{ GeV}$ when taking into account the mass resolution.

Figure 6.15 shows the $\overline{M}$ spectrum for data and the fit with Eq. 6.6 for the inclusive selection (top) and the b-tagged selection (bottom). To quantify the agreement or disagreement of the data and the parametrized fit representing the background, the pull and the residual distributions are shown on the right of both figures. The pull is defined bin-by-bin

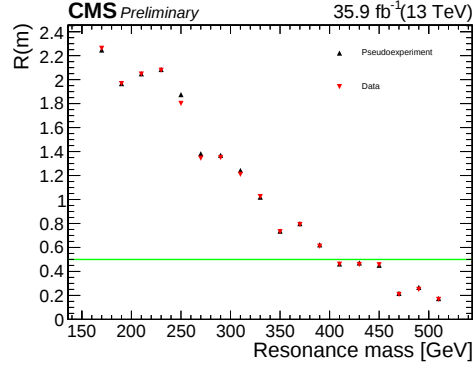


Figure 6.14: $R(m)$ as a function of the fit starting point for data (in red) and for a pseudo-experiment (in blue). The green line represents where the effects of the inefficiencies in the trigger are negligible compared to the statistical fluctuations in data.

as the number of events in data minus the value of the background prediction from the fit divided by the statistical error of the data in that bin. Similarly, the residual is calculated per bin and is defined as the number of events in data minus the value of the background estimate from the fit, divided by the value of the background estimate in that bin. While the pull distribution quantifies the significance of the data compared to the background fit in units of standard deviation, the residual distribution shows the difference in terms of fractions.

In order to ensure that a reasonable number of parameters in this function is chosen, we perform an F-test. Additionally, other functions were studied in order to quantify the bias that we introduce by choosing this particular functional for the fit. Both studies are described below.

6.4.1 F-test

In order to decide on the minimum number of free parameters in the function shown in Eq. (6.6) which best describes the background shape, an F-test is performed [85]. We use the $\overline{M}$ distribution in data and perform different fits using variations of the function shown in

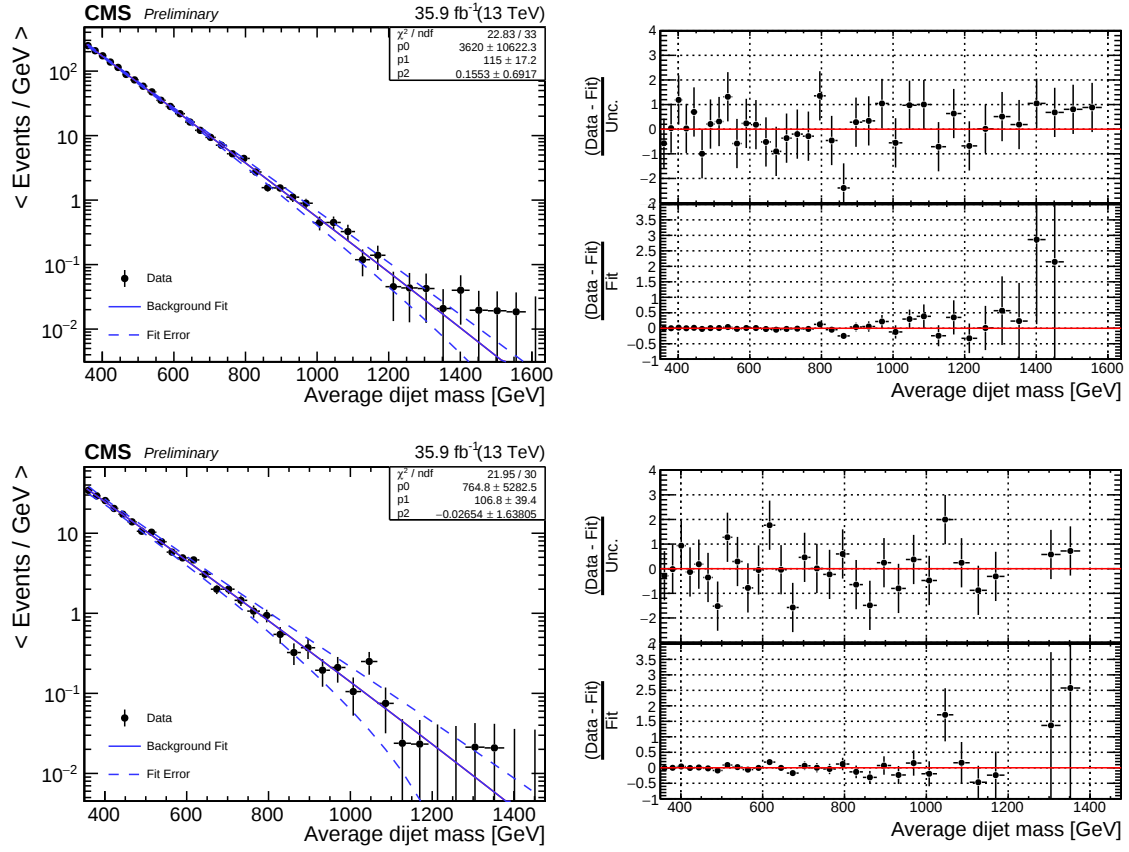


Figure 6.15: Average dijet mass distribution in data corresponding to 35.9 fb⁻¹ for the inclusive selection (top) and the b-tagged selection (bottom) using the functional form in Eq. (6.8). Right: pull (top) and residual (bottom) distributions.

Eq. (6.6), such as:

$$P4 := \frac{p_0(1-x)^{p_1}}{(x)^{p_2+p_3\log(x)}} \quad (6.8)$$

$$P5 := \frac{p_0(1-x)^{p_1}}{(x)^{p_2+p_3\log(x)+p_4(\log(x))^2}} \quad (6.9)$$

The F-test compares 2 functions at a time and requires the following parameters:

- N data points,
- the number of parameters n_j for each model j tested,
- the residual sum of squares for each model, defined as:

$$RSS_j = \sum_{bin} (\text{data} - \text{fit})^2$$

skipping the bins with 0 entries.

Then for two models for which $n_i > n_j$, we calculate the F_{ij} variable:

$$F_{ij} = \frac{\frac{RSS_j - RSS_i}{n_i - n_j}}{\frac{RSS_i}{N - n_i - 1}}, \quad (6.10)$$

under the null hypothesis that model i does not provide a significantly better fit than the model j , the F-distribution will results in $n_i - n_j$ and $N - n_i$ degrees of freedom. If the observed confidence level:

$$CL_{ij} = 1 - \int_{-\infty}^{F_{ij}} \text{F-distribution}(n_i - n_j, N - n_i) \quad (6.11)$$

is smaller than a probability α , which we set to 0.05 (or 5%), then the null hypothesis is rejected (with 95% confidence).

This test is applied recursively to the functions shown in Eqs. (6.6), (6.8), and (6.9),

using a pseudoexperiment thrown from QCD multijet MC scaled to 35.9 fb^{-1} and to data. The fits from each of the functions are shown in Figs. 6.16 and 6.17, for the inclusive and b tagging selections, respectively. The results of these test are shown in Tables 6.6 and 6.7, where the χ^2 and the number of degrees of freedom for each fit is presented, as well as the F-test values. In the tables, for instance F_{34} represents the test of the functions with 3 parameters (Eq. (6.6)) against 4 parameters (Eq. (6.8)). From this test, one can conclude that the P3 function describes the QCD multijet background best, with the minimal number of degrees of freedom. Table 6.7 shows that the test using the b-tagged selection is not conclusive as in the case of the inclusive selection; due to the lack of statistics in the tail of the $\overline{M}$ distribution. The F-test is suggesting to increase the number of parameters in the functional form beyond a P5 function. Therefore, we decide to use a P3 function (Eq. (6.6)) for both the inclusive and b-tagged analyses.

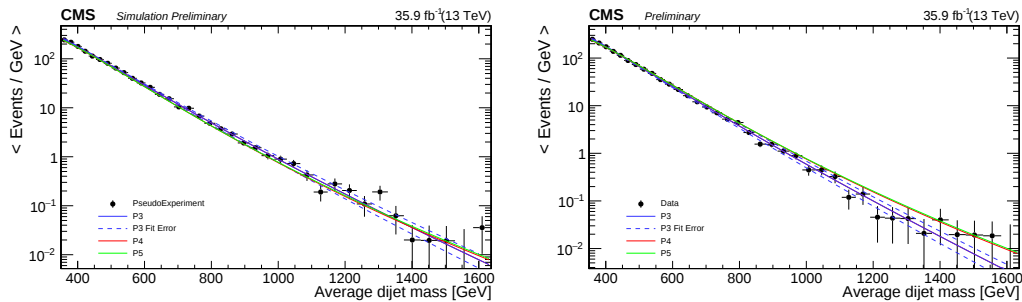


Figure 6.16: Average dijet mass distribution for the inclusive selection. Left: pseudoexperiment based on QCD multijets MC scaled to 35.9 fb^{-1} . Right: data. Different lines represent fits with different number of parameters: P3 (blue, Eq. (6.6)), P4 (red, Eq. (6.8)) and P5 (green, Eq. (6.9)).

6.4.2 Bias test

The parametrization chosen to describe the background shape can introduce a bias in the analysis. To test this, other parameterizations can be used to compare the resulting background estimates. This test helps to cross check how well one can fit for signal in a signal plus background fit. The tests are performed injecting signal, i.e. where the signal strength

Table 6.6: Results of the F-test for inclusive analysis for the different functions described in the text.

Function	Pseudoexperiment			Data		
	Residuals	χ^2	ndof	Residuals	χ^2	ndof
P3	186.5	46.0	33	26.5	22.9	33
P4	157.4	47.8	32	31.7	31.4	32
P5	158.2	40.2	31	19.4	22.2	31
F_{34}	2.0	CL	0.02	5.3	CL	0.03
F_{45}	0.7	CL	0.68	19.5	CL	0.00

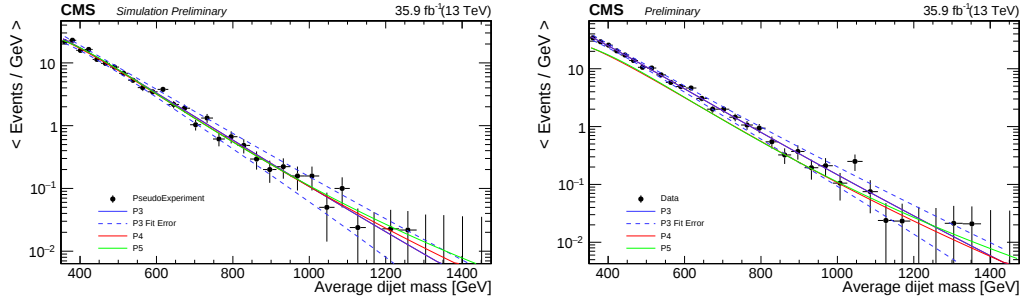


Figure 6.17: Average dijet mass distribution for the b-tagged selection. Left: pseudoexperiment based on QCD multijets MC scaled to 35.9 fb^{-1} . Right: data. Different lines represent fits with different number of parameters: P3 (blue, Eq. (6.6)), P4 (red, Eq. (6.8)) and P5 (green, Eq. (6.9)).

Table 6.7: Results of the F-test for the b-tagged selection for the different functions described in the text.

Function	Pseudoexperiment			Data		
	Residuals	χ^2	ndof	Residuals	χ^2	ndof
P3	18.6	38.8	30	3.9	22.0	30
P4	17.3	38.0	29	4.2	24.8	29
P5	17.8	37.9	28	3.8	22.6	28
F_{34}	2.1	CL	0.16	2.0	CL	0.16
F_{45}	0.8	CL	0.39	3.0	CL	0.09

is 1, in a pseudoexperiment.

In addition to the function in Eq. (6.6), we test the following additional functions:

CDF [86]:

$$\frac{p_0}{m^{p_1}} \left(\sum_{i=0}^N (-1)^i \left(\frac{m}{\sqrt{s}} \right)^i \right)^{p_2},$$

expoPoli [87]:

$$\exp \left(\sum_{i=0}^N p_i (\log(m))^i \right),$$

altPoli [86]:

$$\frac{p_0}{m^{p_1}} \sum_{i=0}^N (-1)^i \left(\frac{m}{\sqrt{s}} \right)^i,$$

Atlas [87]

$$p_0 \frac{(1 - (m/\sqrt{s})^{1/3})^{p_1}}{(m/\sqrt{s})^{\sum_{i=0}^N p_i (\log m)^i}}.$$

First, we need to decide on the number of parameters to use in each class of functions. To do this, we throw a pseudoexperiment from the QCD multijets MC scaled to 35.9 fb^{-1} , fit the mass spectrum to different functions with varying number of parameters and compare the χ^2/ndf and likelihood ratios. This is done for both the inclusive and b-tagged selections and the χ^2/ndf are shown in Table 6.8. We chose the function with the smallest χ^2/ndf of within each class of functions, or CDF3, expoPoli3, altPoli3, Atlas4. Figures 6.18-6.19 show the resulting different fits to the $\overline{M}$ spectra for the inclusive and b-tagged analyses, respectively, for pseudoexperiment from QCD multijets (left) and data (right).

We conclude that these functions are all good ones to used to test for a bias. Then, using these four alternate functions and the default function of Eq. (6.6), we procede with the bias study. We generate pseudoexperiments or toys from the QCD MC fitted to the default function of Eq. (6.6) and inject signal at varying masses with a signal strength of 1. We repeat this 10k times for each signal mass point studied. A maximum likelihood fit of each signal plus background toy is performed using the four alternate functions described

Table 6.8: Resulting χ^2/ndf and likelihood ratio values for the different functions described in the text fitted to a pseudoexperiment thrown from QCD MC scaled to 35.9 fb^{-1} for the inclusive and b-tagged selections. The number in the name of the function represent the number of free parameters in the function.

Function	Inclusive selection		b-tagged selection	
	χ^2/ndf	likelihood ratio	χ^2/ndf	likelihood ratio
CDF3	33.54/35	1.95	25.77/35	1.69
CDF4	29.13/34	1.59	25.57/34	1.68
expoPoli3	31.68/35	1.82	25.91/35	1.70
expoPoli4	29.09/34	1.60	25.74/34	1.70
altPoli3	33.55/35	1.95	25.78/35	1.69
altPoli4	29.14/34	1.59	25.44/34	1.69
Atlas3	29.77/35	1.70	25.51/35	1.68
Atlas4	32.11/34	1.64	25.99/34	1.70

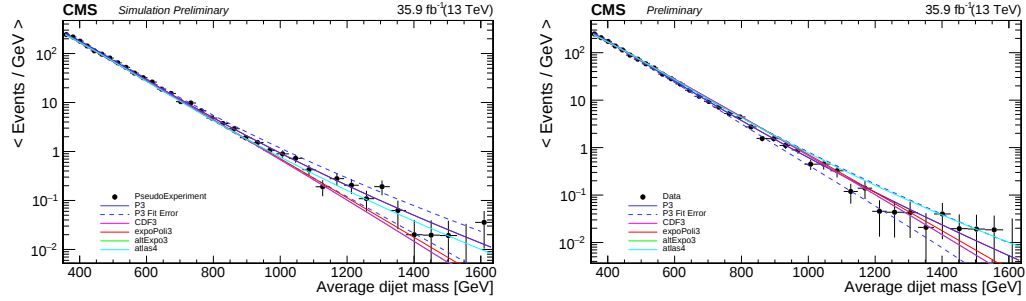


Figure 6.18: Average dijet mass distribution with different fits using a pseudoexperiment based on QCD MC scaled to 35.9 fb^{-1} (left) and data (right) for the inclusive analysis. Different colors represent the different fit functions: CDF3 (pink), expoPoli3 (red), altPoli3 (green), Atlas4 (light blue) and P3 (blue).

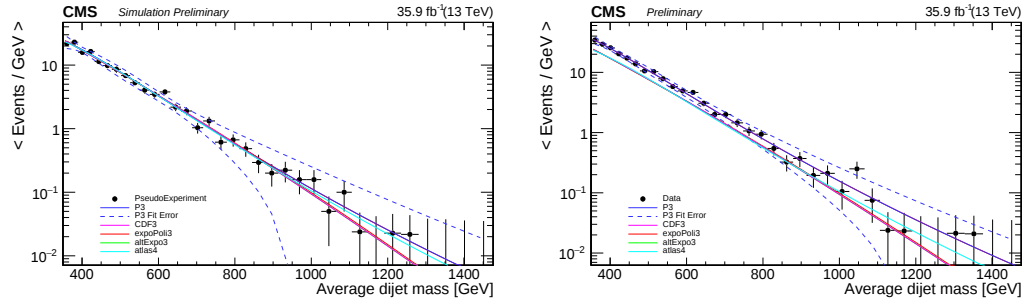


Figure 6.19: Average dijet mass distribution with different fits using a pseudoexperiment based on QCD MC scaled to 35.9 fb^{-1} (left) and data (right) for the b-tagged analysis. Different colors represent the different fit functions: CDF3 (pink), expoPoli3 (red), altPoli3 (green), Atlas4 (light blue) and P3 (blue).

above and the default function of Eq. (6.6). In each fit for each toy, the parameters of each of the four functions are allowed to freely vary. We also allow the signal strength (μ) to vary, which is defined as the number of signal events divided by the number of expected signal events.

The bias introduced by each functional form is measured as:

$$\text{bias} = \frac{\mu_{fit} - \mu_{inj}}{\mu_{err}}, \quad (6.12)$$

where μ_{fit} is the signal strength from the fit, μ_{inj} is the injected signal strength of the tested hypothesis ($\mu_{inj} = 1$ here), and μ_{err} is the error on the fitted signal strength μ_{fit} . Distributions of this bias test from the 10k toys for two different tested signal mass hypothesis are shown in Fig. 6.20 for the inclusive selection and Fig. 6.21 for the b-tagged selection. The different colors represent the results from different background parametrizations tested.

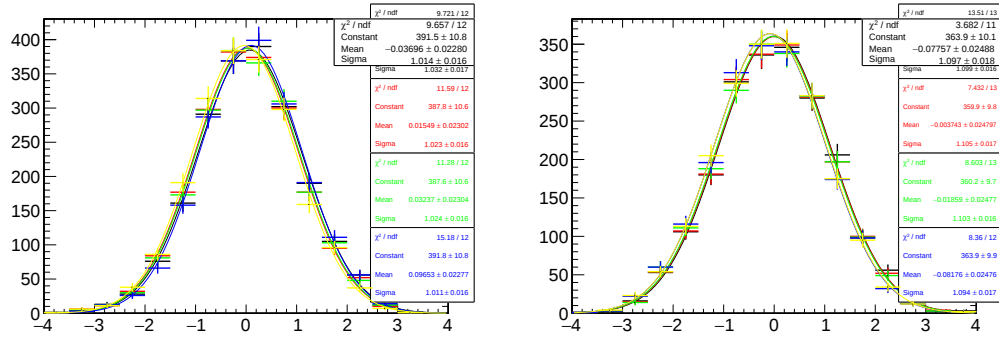


Figure 6.20: Distributions of the bias as defined in Eq. (6.12) for signal plus background toys generated for two different signal mass ($m_{\tilde{\tau}} = 400$ GeV left, $m_{\tilde{\tau}} = 600$ GeV right) for the inclusive selection. The results of fitting the toys to CDF3 (red), expoPoli3 (green), altPoli3 (blue), Atlas4 (yellow) and P3 (black) are shown.

The means of the bias as defined in Eq. (6.12) from the toys of the bias test with all the tested functions as a function of mass are shown in Fig. 6.22 for both the inclusive (left) and b-tagged (right) selections. They are all found to be within 50% of zero. We conclude that all background functions do not introduce a bias in the signal strength. In both plots,

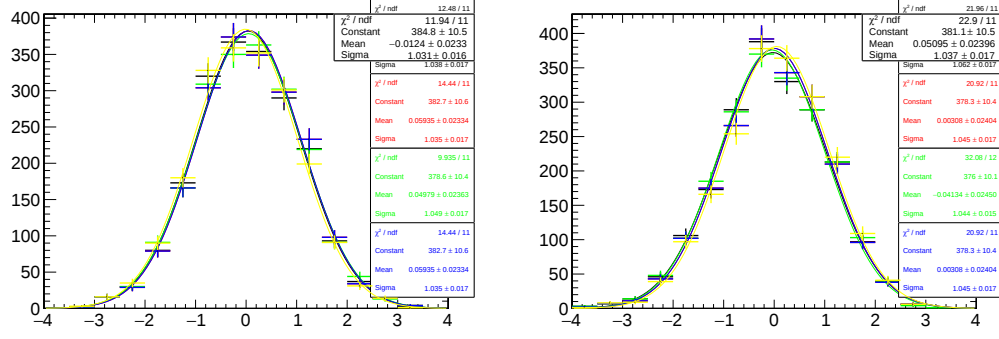


Figure 6.21: Distributions of the bias as defined in Eq. (6.12) for signal plus background toys generated for two different signal mass ($m_\tau = 400$ GeV left, $m_\tau = 600$ GeV right) for the b-tagged selection. The results of fitting the toys to CDF3 (red), expoPoli3 (green), altPoli3 (blue), Atlas4 (yellow) and P3 (black) are shown.

the function (6.6) is compared with the other four parametrizations and it is found a bias smaller than 50% when one uses a different functional form to estimate the background.

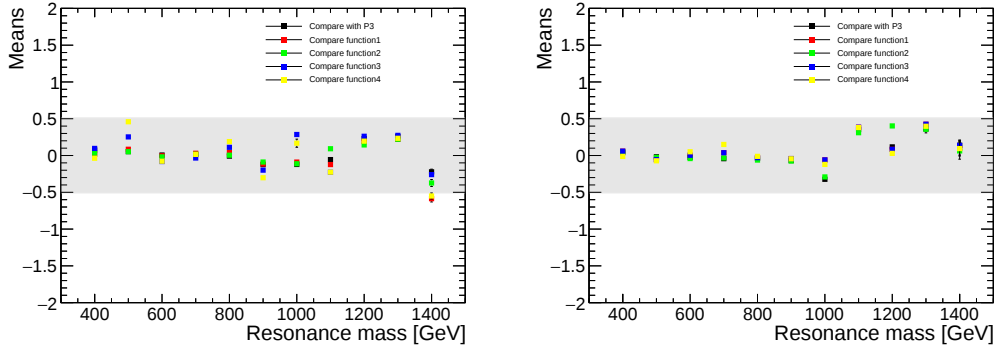


Figure 6.22: Result of the bias study. Bias means as a function of mass for all functions tested for inclusive (left) and b-tagged (right) selections.

6.4.3 Results of resolved search

Fig. 6.23 shows the $\overline{M}$ spectrum for the resolved analysis. This figure shows the results of the inclusive selection (top) and the b-tagged selection (bottom). Since we observe no significant excess of the data over the background prediction we proceed to set 95% C.L. upper limits on the production cross section of the stops as a function of m_τ described in

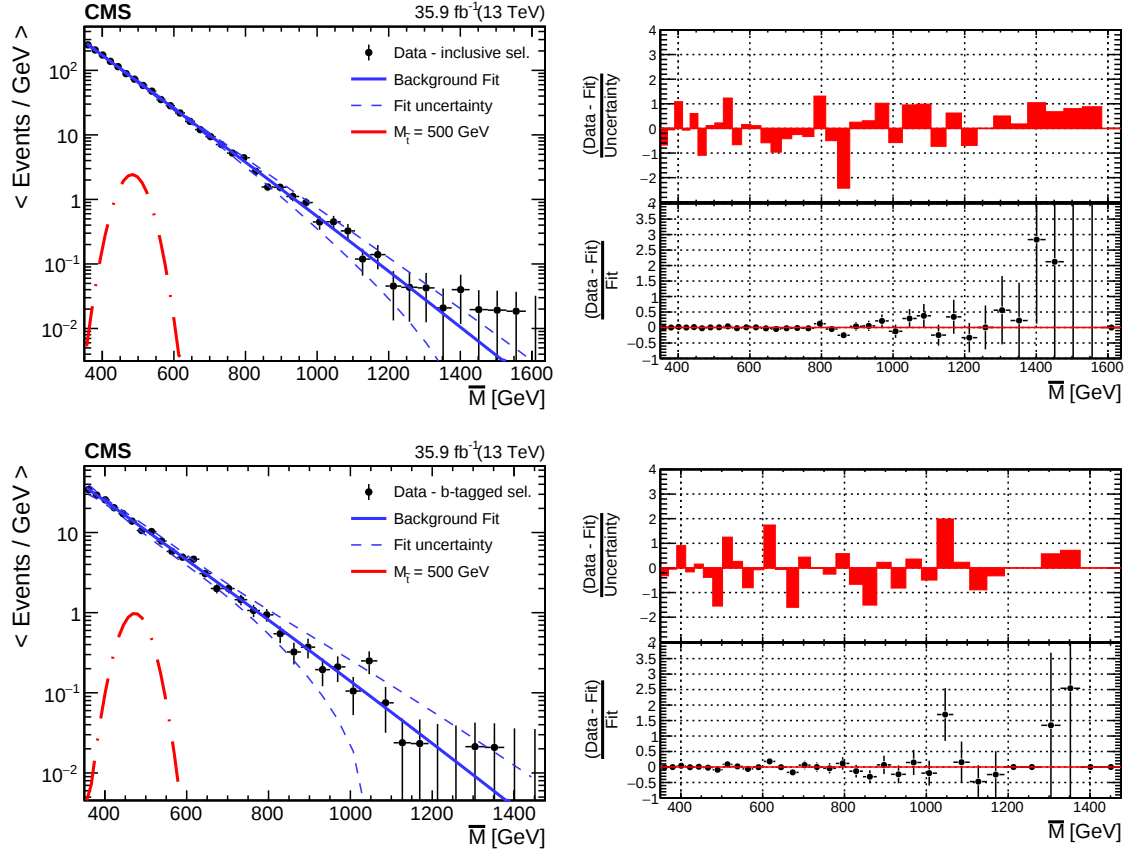


Figure 6.23: $\overline{M}$ spectrum for data (dots) compared to the background fit of the function in Eq. (6.6) (red line). Top: inclusive selection. Bottom: b-tagged selection. Right: the pull and residual distributions as a function of $\overline{M}$. A hypothetical signal at $m_{\tilde{\tau}} = 500$ GeV. in an RPV SUSY model is shown.

Chapter 7.

6.5 Systematic Uncertainties

In this section we describe the systematic uncertainties on the background estimate and the signal acceptance for the boosted analysis.

6.5.1 Signal Systematic Uncertainties

The sources of systematic uncertainties we consider for the signal are listed and described below. They either affect the yield of the signal acceptance or the resolution of the $\overline{M}$ spectra and are summarized in Table 6.9.

Integrated Luminosity (Yield): The luminosity measurement is described in detailed in Section 3.2.7. A systematic uncertainty on the integrated luminosity of 2.5% is assigned, according to the most recent measurement using the pixel detector [41].

Trigger (Yield): Based on the observed differences in efficiency of the triggers as a function of the $\overline{M}$ spectrum shown in Fig. 6.1, we assign a 3% uncertainty.

Jet Energy Scale (Yield): In addition, jet energy scale corrections applied to the MC samples have associated uncertainties, as described in Section 4.1.3. These uncertainties applied to the jet p_T , are propagated almost linearly to the jet mass. To estimate this effect, the acceptances in the $\overline{M}$ distribution using jets with the nominal JEC correction are compared with the JEC correction ± 1 standard deviation. The differences in acceptance as a function of $m_{\tilde{\tau}}$ is shown in Appendix B.2. From those studies, we found an effect of 1.5% on average for all $m_{\tilde{\tau}}$ mass points.

Jet Energy Resolution (Yield): The jet energy resolution measurement also affects the signal yield as discussed in Section 4.1.3. This uncertainty is estimated by the yield difference entering the $\overline{M}$ distributions using jets with the nominal JER value compared with the distributions using jets with a JER ± 1 standard deviation. The differences as a function of $m_{\tilde{\tau}}$ are shown in Appendix B.3. The effect of the JER on the signal acceptance is found to be on average 6% for all $m_{\tilde{\tau}}$ points.

Jet Energy Scale (Shape): Jet energy corrections can also shift the peak of the Gaussian shapes. Therefore, a 2% uncertainty is applied as a shape systematic on the $\overline{M}$ distribution.

Jet Energy Resolution (Shape): The uncertainty in the jet energy resolution can translate into an uncertainty of 10% on the resolution of the $\overline{M}$ distribution [88]. Therefore, we smear the signal mass spectra in MC by 10% to account for this effect on the signal shape. This change in the width stretches or shrinks the shape of the resonance itself.

Pileup Re-weighting (Yield): The impact of the pileup in the generation of the signal

MC samples is also studied. This uncertainty is estimated by a $\pm 2\%$ standard deviation shift of the minimum bias cross section, as shown in Appendix B.4. From those studies, we find this to be a 1% effect on the signal acceptance.

PDF (Yield): To estimate the uncertainty arising from PDFs, the *PDF4LHC recommendations for LHC Run II* [81] are followed. The root mean square (RMS) of the dis-

tribution of the 100 NNPDF MC replicas as the $\pm 1\sigma$ shape variation due to parton distribution functions are evaluated. The effect of the PDF uncertainty is evaluated in Appendix B.5, and it is found to be a 1% effect.

MC statistics (Shape): A bin-by-bin uncertainty due to limited MC statistics of the

signal events is also included as it affects the shape of the $\overline{M}$ spectra.

B-Tagging scale factor (Shape): For the b-tagged selection we include the uncertainty

on the data/MC scale factor of the efficiencies in the identification of b jets, as described in Section 4.1.3. The CMS recommendation is followed for these scale factors using the event weights procedure [75], and the uncertainty is evaluated by varying these scale factors ± 1 standard deviation and studying the difference in signal acceptance as shown in Appendix B.6. The effect of this uncertainty is evaluated to be 1%.

6.5.2 Systematic Uncertainty on Background Estimate

The systematic uncertainty on the background prediction is derived from the uncertainty on the fit parameters of the fit function. The overall normalization of the background parametrization is determined from the data itself and is allowed to float freely in the fit.

Table 6.9: Summary of the systematic uncertainties on the signal. The uncertainty values affecting the resonance shape refer to the value of the uncertainty itself, not the effect on the acceptance.

Source of Systematic	Effect	Value
Luminosity	Yield	2.5%
Trigger	Yield	3.0%
Pileup	Yield	1.0%
Jet Energy Scale	Yield	1.5%
Jet Energy Resolution	Yield	6.0%
PDF	Yield	1.0%
B tagging (only for b-tagged selection)	Yield	1.0%
Jet Energy Scale	Shape	2.0%
Jet Energy Resolution	Shape	10.0%
MC Statistics	Shape	fit parameters

Chapter 7

Interpretation of the results

In this chapter, the results of the search for pair-produced dijet resonances in proton-proton collisions with the CMS detector at $\sqrt{s} = 13$ TeV are presented. As seen in Figures 5.19 and 6.23, we observe no evidence of new physics, therefore, the results of the boosted and resolved analysis are statistically interpreted in the context of RPV SUSY. Here, we present the statistical methods used and evaluate 95% confidence level (C.L.) limits on the cross section as a function of $m_{\tilde{t}}$ in the RPV SUSY models studied.

7.1 Statistical analysis

It is common practice in High Energy Physics searches to statistically test the data against a background-only hypothesis and a background plus signal hypothesis. This process is referred to as limit setting since the output of the hypothesis testing is an upper limit on the production cross section of a new physics signal as a function of a physical observable quantity such as the hypothetical new particle. In this section, a brief description of these techniques are introduced and then the limits on the production cross section as a function of $m_{\tilde{t}}$ for the benchmark model we consider is presented.

There are two main approaches to calculate limits: the so-called frequentist and bayesian approaches [6]. In the frequentist method, the probability of the outcome of a certain experiment is interpreted as how frequent is that outcome can be repeated. In the bayesian approach, the interpretation of the probability requires prior knowledge, or a belief, of where the true value of the parameter of interest should lie. Then, using the Bayesian theorem

this degree of belief is updated by the results from data collected in an experiment. In many cases, these two approaches give similar numerical values [6], even though the formulations are different.

In the LHC experiments, a modified frequentist method was developed for the Higgs discovery and it is the recommended procedure to test the statistical power of the searches [89]. We use this method here. In this method, usually referred to as CL_S [90, 91], we denote the signal events as s , while the background yields are denoted as b . The null results in this method are expressed as a limit on the signal strength (r) that is taken to change the signal cross section by exactly that same scale r .

The uncertainties in the predictions of the signal and background yields are taken as nuisance parameters θ , then the signal and background expectations are given as $s = s(\theta)$ and $b = b(\theta)$, respectively. In this modified frequentist approach, the systematic uncertainties are represented as probability distribution functions (*pdf*) of a real or imaginary measurement $\tilde{\theta}$. These *pdfs* can be normal, log-normal or gamma distributions, and can be defined in the context of Bayes' theorem as [89]:

$$\rho(\theta|\tilde{\theta}) \sim p(\tilde{\theta}|\theta)\pi_{\theta}(\theta), \quad (7.1)$$

where $p(\tilde{\theta}|\theta)$ is the *pdf* of the fictional auxiliary measurement and $\pi_{\theta}(\theta)$ are hyper-prior functions of those measurements. In generic terms, $\rho(\theta|\tilde{\theta})$ reflects the degree of belief on the true value of the nuisance parameter θ , and by re-interpreting its definition as a posterior, as done in Eq. (7.1), it allows us to represent all the systematic uncertainties in the frequentist approach.

Then, one can construct a likelihood function for data or pseudo-data using the signal strength r and the full suite of nuisance parameters θ , as [89]:

$$\mathcal{L}(\text{data}|r, \theta) = \text{Poisson}(\text{data}|r \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta). \quad (7.2)$$

The Poisson function is interpreted as the product of Poisson probabilities to observe n_i events in i bins:

$$\prod_i \frac{(rs_i + b_i)^{n_i}}{n_i!} e^{-(rs_i + b_i)}, \quad (7.3)$$

or for an unbinned likelihood over k events in data:

$$k^{-1} \prod_i (rSf_s(x_i) + Bf_b(x_i)) e^{-(rS_i + B_i)}. \quad (7.4)$$

The functions $f_s(x)$ and $f_b(x)$ are *pdfs* of some observable x , while S and B are total expected number of events for signal and background, respectively.

This likelihood function definition is further used to compare the background-only and signal plus background hypotheses against the data. Recall, the signal component is allowed to be scaled by the factor r . Then, we can build a test statistic $\tilde{q}_r$ based on the profile likelihood ratio:

$$\tilde{q}_r = -2 \ln \frac{\mathcal{L}(\text{data}|r, \hat{\theta}_r)}{\mathcal{L}(\text{data}|\hat{r}, \hat{\theta})}. \quad (7.5)$$

Here, the parameter $\hat{r}$ is constrained between 0 (signal rate positive) and r (guarantees one-sided confidence interval), and $\hat{\theta}_r$ refers to the conditional maximum likelihood estimators of θ . These two estimators, $\hat{r}$ and $\hat{\theta}$, correspond to the global maximum of the likelihood.

Next, for a given signal strength r , the observed value of the test statistics $\tilde{q}_r^{obs}$ is calculated. The values of the nuisance parameters $\hat{\theta}_0^{obs}$ and $\hat{\theta}_r^{obs}$ which best describe the data can be also found by maximising the likelihood function of Eq. (7.2) for the background-only and signal plus background hypotheses.

Then, assuming a signal with strength r (signal plus background hypothesis) or $r = 0$ (background-only hypothesis) one generates toy pseudoexperiment to construct two *pdfs*,

$f(\tilde{q}_r|r, \hat{\theta}_r^{obs})$ and $f(\tilde{q}_r|0, \hat{\theta}_0^{obs})$, respectively. These distributions allow us to define two p-values associated with the two hypotheses:

$$p_r = P(\tilde{q}_r \geq \tilde{q}_r^{obs} | \text{signal plus background}) = \int_{\tilde{q}_r^{obs}}^{\infty} f(\tilde{q}_r|r, \hat{\theta}_r^{obs}) d\tilde{q}_r, \quad (7.6)$$

$$1 - p_b = P(\tilde{q}_r \geq \tilde{q}_r^{obs} | \text{background-only}) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_r|0, \hat{\theta}_0^{obs}) d\tilde{q}_r, \quad (7.7)$$

and to finally compute the confidence level, $CL_S(r)$, as the ratio of these two probabilities:

$$CL_S(r) = \frac{p_r}{1 - p_b}. \quad (7.8)$$

With this technique, a particular signal hypothesis is excluded with a $(1 - \alpha)$ CL_S confidence level if $CL_S \leq \alpha$ for $r = 1$. An exclusion limit at the LHC is usually quoted at 95% confidence level, meaning that the value of r is adjusted until the value of α is 0.95.

This process is performed with real data for the so-called observed limits, and with a large set of background only pseudo-data for the expected limits. In the case of the expected limit, it is customary to also calculate $\pm 1\sigma$ and $\pm 2\sigma$ bands. These correspond to the points at which a cumulative probability distribution, starting from the tails of the pdf , crosses the 16% and 84% quantiles for the $\pm 1\sigma$ band (68%) and at 2.5% and 97.5% for the $\pm 2\sigma$ band (95%).

7.2 Limits on hadronic RPV SUSY

Using the results shown in Figs. 5.19 and 6.23 we set 95% C.L. limits on the pair production cross section of top squarks decaying via the RPV couplings λ''_{312} ($\tilde{t} \rightarrow qq'$) and λ''_{312} ($\tilde{t} \rightarrow bq'$), assuming 100% branching ratios. The exclusion limits are computed using the modified CL_S frequentist approach described in Section 7.1 with a binned profile likelihood as the test statistics in the asymptotic approximation [90, 91].

Results in the boosted search are obtained from combined signal and background binned

likelihood fits to the average pruned jet mass in the data. For each top squark mass scenario, only bins of average pruned jet mass within 2σ of the peak of the mass fitted to a Gaussian in signal are included in the likelihood. For each bin used in the likelihood the individual background components and signal are allowed to float within statistical errors with no correlations between bins. Systematic uncertainties affecting the yield and the shape are summarized in Tables 5.9 and 5.10. All these uncertainties are assumed to be correlated between bins, except for the statistical uncertainty in the MC samples and the uncertainty in the region C in the QCD multijet background estimate. These uncertainties are treated as nuisance parameters, which are profiled, and modeled with log normal priors, except for the uncertainty on the number of events in the sideband region used in the QCD multijet background estimate, which is modeled with a Gamma prior.

In the resolved search, the average dijet mass spectrum in data is compared with the parametrize background fit to look for localized deviations. For each $m_{\tilde{t}}$, shapes for the signal and background as well as data within a mass window of two standard deviations around the true value of $m_{\tilde{t}}$ are used. Signal uncertainties are summarized in Table 6.9. Uncertainties affecting the shape are assumed to be correlated while the ones affecting the yield are uncorrelated. In the case of the background, the uncertainties in the parameter fits are taken as a systematic uncertainties. Here, all systematic uncertainties are modeled with log normal priors.

Figure 7.1 shows the observed and expected 95% CL upper limits on the top squark pair production cross section as a function of $m_{\tilde{t}}$ for the coupling λ''_{312} (left) and λ''_{323} (right), for the boosted and resolved searches. The boosted search covers a range of $m_{\tilde{t}}$ from 80 to 400 GeV, while the resolved search covers $m_{\tilde{t}} \geq 400$ GeV. The dashed pink line indicates the NLO + NLL theoretical predictions for top squark production [67, 68]. Top squark masses are excluded from 80 to 520 GeV assuming the λ''_{312} coupling, and from 80 to 270 GeV, 285 to 340 GeV, and 400 to 525 GeV for the λ''_{323} coupling.

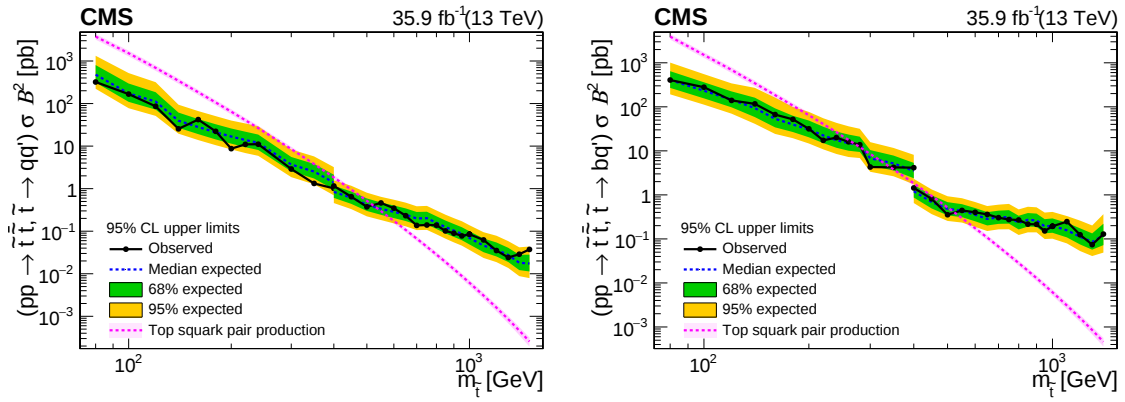


Figure 7.1: Observed and expected 95% CL upper limits on the production cross section as a function of m_t . Left: inclusive search assuming the RPV coupling λ''_{312} . Right: b-tagged search assuming the RPV coupling λ''_{323} . For both scenarios, 100% branching ratios are assumed. The boosted searches cover the region between $80 \leq m_t \leq 400$ GeV, while the resolved searches cover all the masses above 400 GeV. The dashed pink line shows the NLO + NLL theoretical predictions for top squark pair production [67, 68].

Chapter 8

Conclusion

A search has been performed to search for the pair production of resonances decaying to two quarks in two jet events and four jet events from proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS detector using 36.9 fb^{-1} of data collected in 2016. In the boosted search the distribution of the average mass of the selected two jets has been investigated for localized disagreements between data and the SM background estimate. This is the first search of its kind in the CMS experiment and the first covering the region $m_{\tilde{t}} \leq 200 \text{ GeV}$ for the $\tilde{t} \rightarrow qq'$ scenario with boosted techniques. In the resolved search the distribution of the average mass of the selected dijet pairs has been used in four jet events. In both searches, we also select events with b-tagged jets. No significant deviations are found between the data distributions and the expected backgrounds predictions from the SM. Limits at 95% C.L. are set on the top squark pair production cross section as a function of $m_{\tilde{t}}$, when the $\tilde{t}$ decays through the RPV SUSY λ''_{3DD} coupling to final states $\tilde{t} \rightarrow qq'$ or $\tilde{t} \rightarrow bq'$ with a 100% branching fraction. Top squark masses are excluded from 80 to 520 GeV assuming the λ''_{312} coupling, and from 80 to 270 GeV, 285 to 340 GeV, and 400 to 525 GeV for the λ''_{323} coupling.

Appendix A

Boosted search

A.1 Jet energy scale uncertainties

The uncertainties in the jet energy scale are η -dependent in a per-jet basis. We apply ± 1 standard deviation adjustments to the jet mass. The $+1$ standard deviation is called JESup and the -1 is JESDown. Figures A.2 show the distribution of the average mass with the nominal, JESup and JESDown. From the number of events in the mass distribution around 2σ the true mass, we calculate the acceptance for each mass point and we results are shown in Fig. A.1.

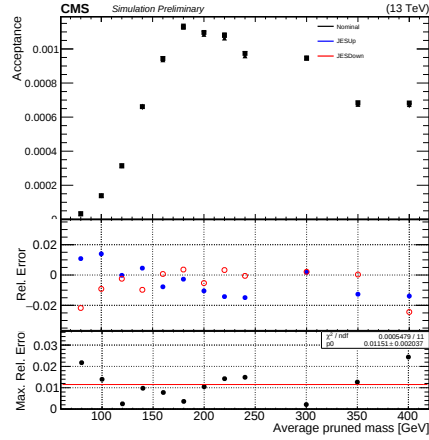


Figure A.1: Acceptance for nominal, JESup and JESDown, for each mass point taken from the distributions shown in Figures A.2.

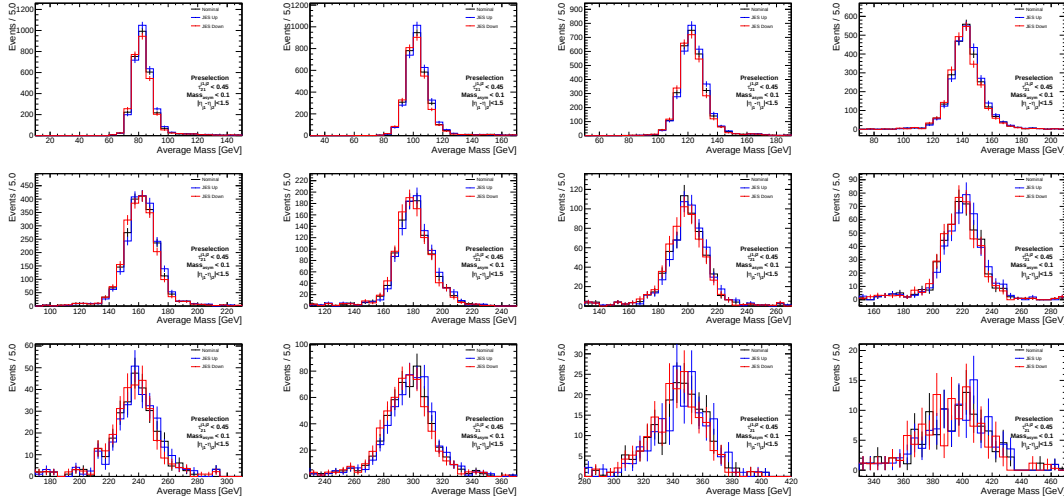


Figure A.2: Average pruned mass distributions for different signals after all the selection is applied. For each mass, the nominal (black), JESUp (blue) and JESDown distribution are shown.

A.2 Jet energy resolution uncertainties

To measure the impact of the jet energy resolution on the average mass, we vary the JER value by $\pm$ standard deviation. The $+1$ standard deviation is called JERup and the -1 is JERDown. Figures A.4 show the distribution of the average mass with the nominal, JERup and JERDown. From the number of events in the mass distribution around 2σ the true mass, we calculate the acceptance for each mass point and we results are shown in Fig. A.3.

A.3 Pileup uncertainties

The MC samples are generated with a pileup distribution that is not identical to the pileup distribution in data. The pileup distribution can be made to match that of the data by re-weighting events based on the relative pileup distributions. However, the modeling and measurement of the pileup distribution in data carries systematic uncertainties, which can be covered by varying the inelastic cross section parameter used to calculate pileup re-weighting by $\pm 5\%$. Figures A.6 show the distribution of the average mass with the nominal, PUUp and PUDown. From the number of events in the mass distribution around 2σ the

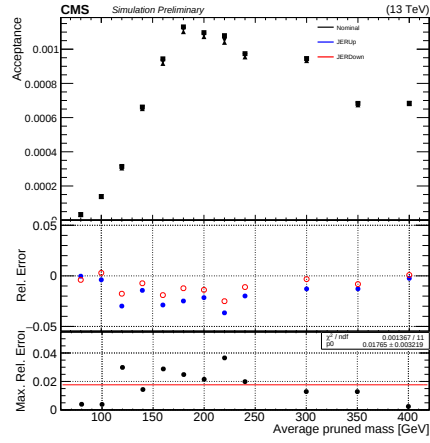


Figure A.3: Acceptance for nominal, JERUp and JERDown, for each mass point taken from the distributions shown in Figures A.4.

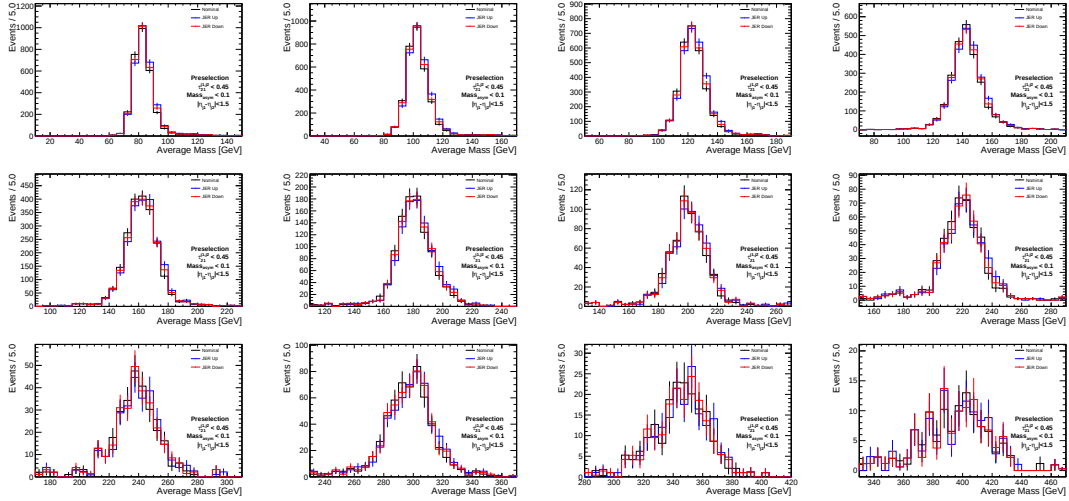


Figure A.4: Average pruned jet mass distributions for different signals after all the selection is applied. For each mass, the nominal (black), JERUp (blue) and JERDown distribution are shown.

true mass, we calculate the acceptance for each mass point and we results are shown in Fig.

A.5.

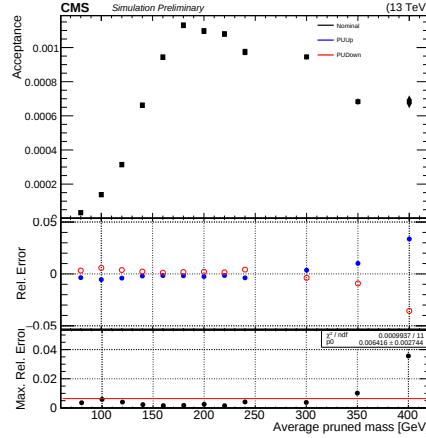


Figure A.5: Acceptance for nominal, PUUp and PUDown, for each mass point taken from the distributions shown in Figures A.6.

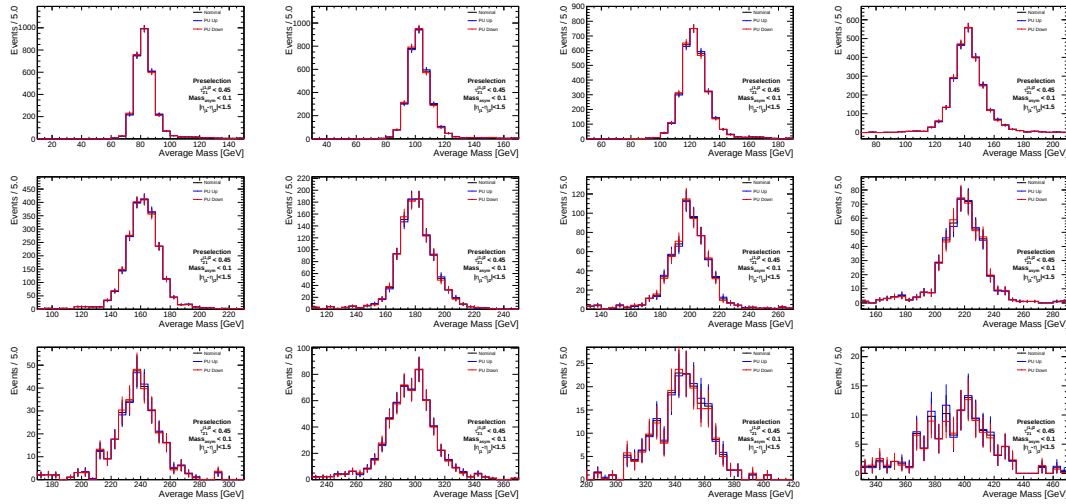


Figure A.6: Average pruned mass distributions for different signals after all the selection is applied. For each mass, the nominal (black), PUUp (blue) and PUDown distribution are shown.

A.4 PDF uncertainties

The impact of the PDFs in the signal MC samples is evaluated with the RMS of the distribution of the 100 NNPDF MC replicas as the 1σ shape variation due to parton distribution

functions as recommended in [81]. The +1 standard deviation is called PDFUp and the -1 is PDFDown. Figures A.8 show the distribution of the average mass with the nominal, PDFUp and PDFDown. From the number of events in the mass distribution around 2σ the true mass, we calculate the acceptance for each mass point and we results are shown in Fig. A.7.

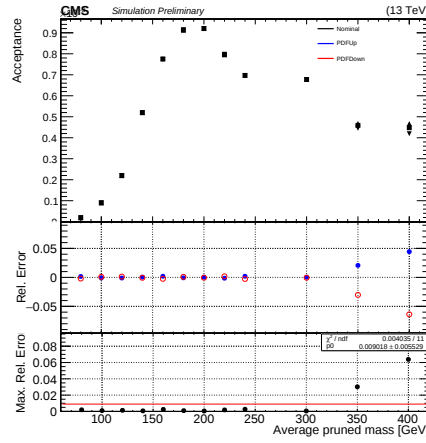


Figure A.7: Acceptance for nominal, PDFUp and PDFDown, for each mass point taken from the distributions shown in Figures A.8.

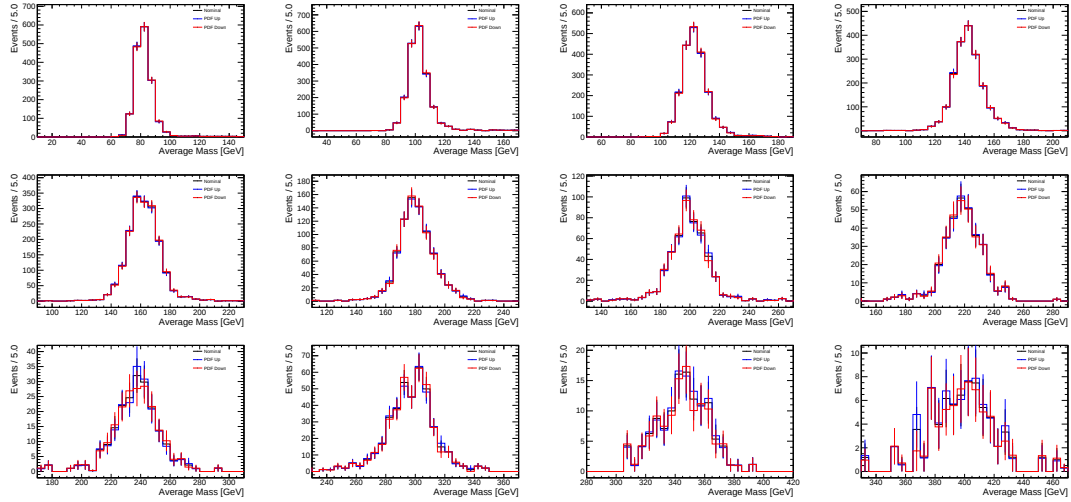


Figure A.8: Average pruned mass distributions for different signals after all the selection is applied. For each mass, the nominal (black), PDFUp (blue) and PDFDown distribution are shown.

A.5 B tagging uncertainties

This uncertainties are p_T , η , and flavour-dependent in a per-jet basis, we apply ± 1 standard deviation adjustments to the jet mass. The $+1$ standard deviation is called btagUp and the -1 is btagDown. Figures A.10 show the distribution of the average mass with the nominal, btagUp and btagDown. From the number of events in the mass distribution around 2σ the true mass, we calculate the acceptance for each mass point and we results are shown in Fig. A.9.

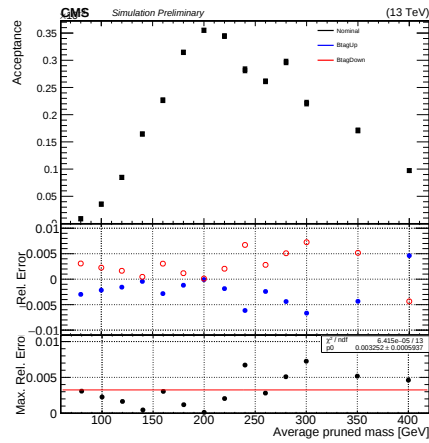


Figure A.9: Acceptance for nominal, btagUp and btagDown for all the signal mass points.

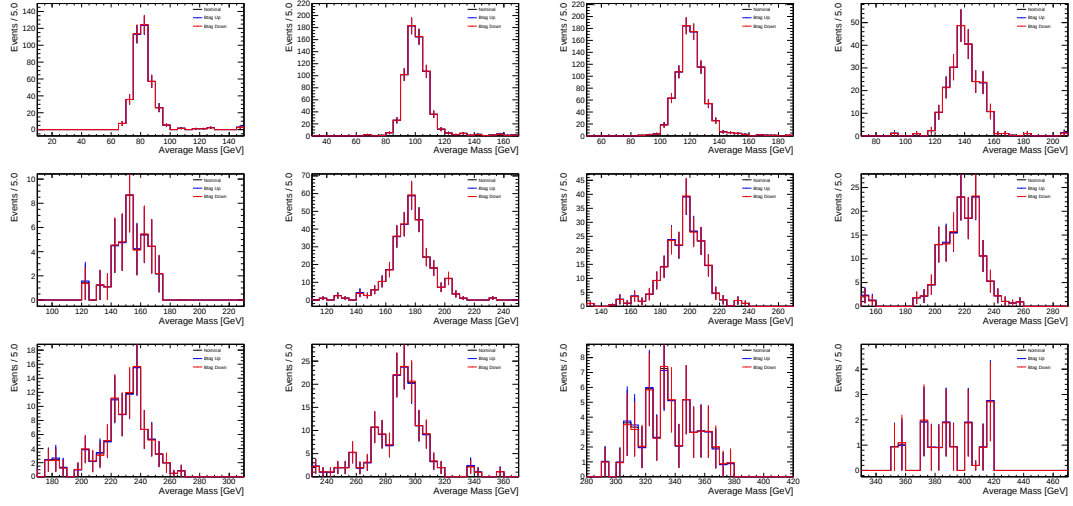


Figure A.10: Average dijet mass distributions for different UDD323 signals after all the btagged selection is applied. For each mass, the nominal (black), btagUp (blue) and btagDown distribution are shown.

Appendix B

Resolved search

B.1 Alternative dijet pairing algorithms

We tested different dijet pairing procedures. The goal is to simultaneously maximize the signal acceptance as well as ensuring a smooth background distribution for the estimation for as low dijet masses as possible. The three different algorithms we studied are:

DeltaR pairing : This is the default procedure described in Section 6.2.1, which was also used in Ref. [22].

Kinematic fit pairing : Minimize a mass χ^2 including mass resolution to choose the pairing. The procedure is as follows:

1. Using signal MC, we calculate the mass resolution of the paired dijet stop signal. Figure B.1 shows the average mass distribution vs the mass resolution (σ_{mass}).
2. Construct the dijet pair combinations in an event.
3. Calculate per pair: $\chi_{klmn}^2 = (M(j_k j_l) - M(j_m j_n))^2 / \sigma^2$, where σ is the signal mass resolution which is a parameter per event depending on the average dijet mass. It is calculated using the value of the fit function evaluated at the average dijet mass of the event.
4. Chose pair with the smallest χ^2 as best dijet pair.

Mass pairing : This is similar to the pairing chosen in the Ref. [92]. We choose the pair with the lowest mass asymmetry variable, as follows:

1. Construct the dijet pair combinations in event.
2. Calculate per pair: $\Delta M_{klmn} = |M(j_k j_l) - M(j_m j_n)| / (M(j_k j_l) + M(j_m j_n))$
3. Chose pair with the smallest ΔM as best dijet pair.

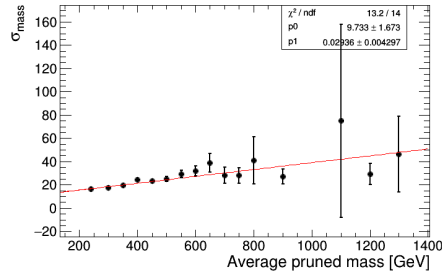


Figure B.1: Average mass as a function of the mass resolution (σ_{mass}) for the RPV stop signal. We match the AK4 jets to the MC truth stop daughters and we calculate the mass of the dijet system. We fit that distribution using a Gaussian function and the width of that Gaussian is the mass resolution.

The distributions corresponding to the minimum deltaR, minimum mass and minimum χ^2 are shown in Fig. B.2.

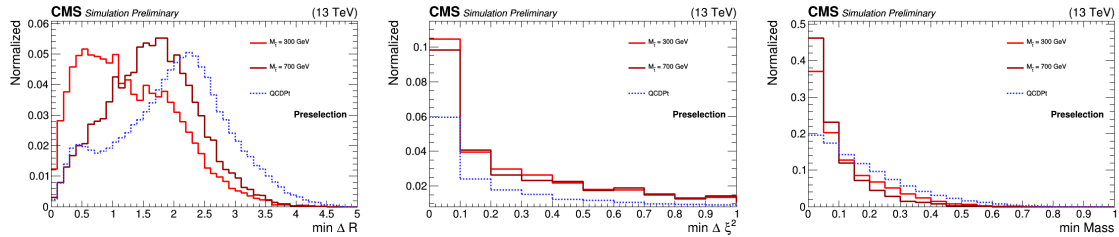


Figure B.2: Distributions of the minimum ΔR (left), minimum χ^2 (middle) and minimum mass (right) from the pairing methods described comparing two selected signal MC at $m_{\tilde{t}} = 300$ GeV and $m_{\tilde{t}} = 700$ GeV, and QCD multijets MC.

We note that the pairing method changes the behavior of the mass asymmetry variable, as shown in Fig. B.3; modifies the behavior of the Delta variable, as shown in Fig. B.5-B.6; and does not modify $\Delta\eta_{dijet}$ shown in Fig. B.4.

Therefore, to choose the best pairing method we compare the performance of the algorithms after preselection plus the $\Delta\eta_{dijet} < 1$ and $\Delta > 200$ GeV selections. These distributions are shown in Fig. B.7, where the distributions are scaled to 36 fb^{-1} . In these figures

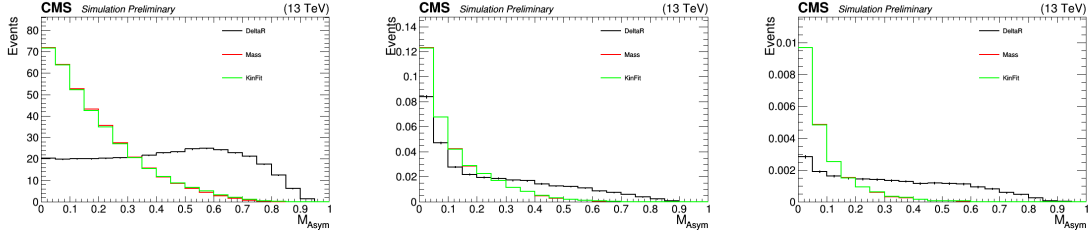


Figure B.3: Mass asymmetry distribution after preselection for QCD multijets MC (left) and RPV stop MC at $m_{\tilde{\tau}} = 300$ GeV (middle), and $m_{\tilde{\tau}} = 700$ GeV (right). Each line represent different pairing method: deltaR (black line), mass pairing (red) and kinematic fit (green).

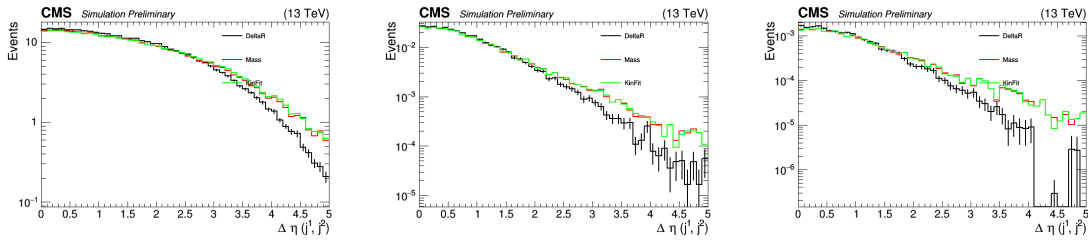


Figure B.4: $\Delta\eta_{\text{dijet}}$ distribution after preselection for QCD multijets MC (left) and RPV stop MC at $m_{\tilde{\tau}} = 300$ GeV (middle), and $m_{\tilde{\tau}} = 700$ GeV (right). Each line represent different pairing method: deltaR (black line), mass pairing (red) and kinematic fit (green).

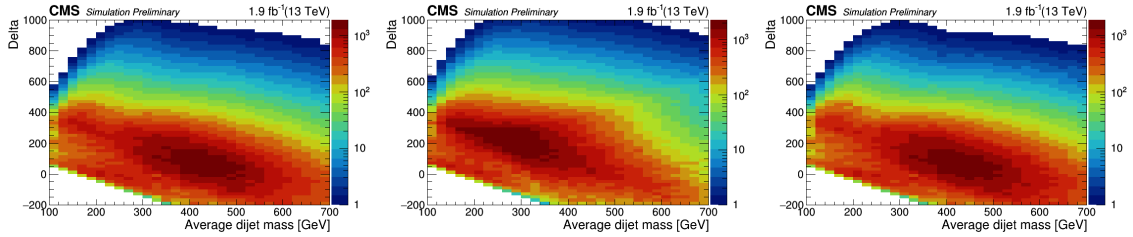


Figure B.5: Δ as a function of average dijet mass distribution after preselection for QCD. Top left: using the deltaR pairing. Top right: using kinematic fit. Bottom: using mass pairing.

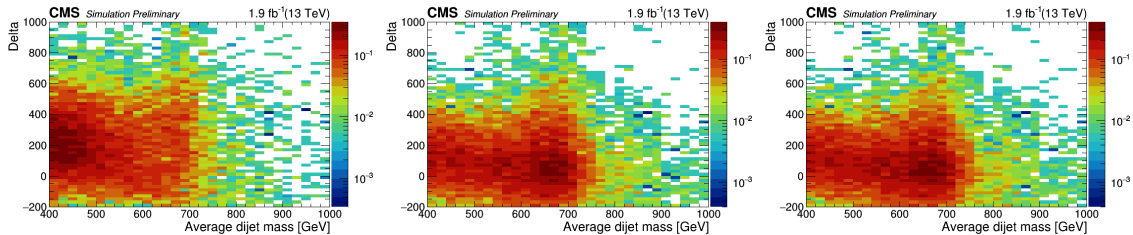


Figure B.6: Δ as a function of average dijet mass distribution after preselection for RPV stop signal at $m_{\tilde{\tau}} = 700$ GeV. Top left: using the deltaR pairing. Top right: using kinematic fit. Bottom: using mass pairing.

we include the distribution with a complete set of cuts as described in Section 6.2.3 for comparison.

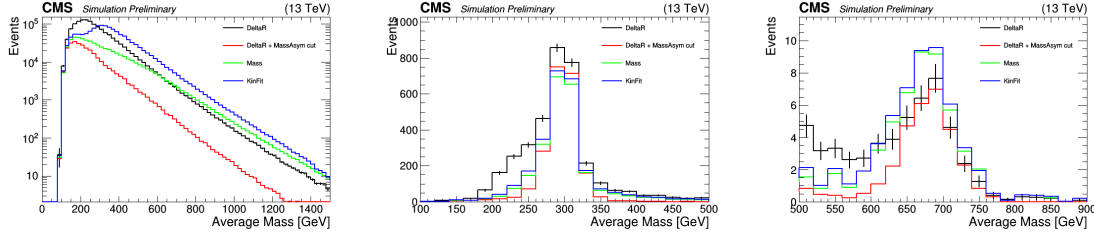


Figure B.7: Average mass distribution after preselection and $\Delta\eta_{\text{dijet}}$ plus Δ selections for QCD multijets (left) and RPV stop signals MC at $m_{\tilde{t}} = 300$ GeV (middle) and $m_{\tilde{t}} = 700$ GeV (right). Distributions are scaled to 36 fb^{-1} . Each line represent different pairing method: deltaR (black line), mass pairing (red) and kinematic fit (green).

Finally, we test the optimization of the Δ cut. We use the minimum mass method and we choose different values of Δ cut. Then we calculate the number of events in two signal mass points, as well as the significance $S/\sqrt{B}$. The values are shown in Table B.1. In addition, Fig. B.8 shows the average mass distribution in QCD multijets MC for these set of selections.

Table B.1: Significance and number of events of different selections using the minimum mass pairing method and different Δ cuts, comparing with the default set of cuts using the $\tilde{t} \rightarrow qq'$ samples.

	Pairing Method	$m_{\tilde{t}} = 300 \text{ GeV}$		$m_{\tilde{t}} = 700 \text{ GeV}$	
		$S/\sqrt{B}$	Nevents	$S/\sqrt{B}$	Nevents
$M_{\text{asym}} + \Delta\eta_{\text{dijet}} + \Delta > 200 \text{ GeV}$	Delta R	9.87	1946	0.74	22
$\Delta\eta_{\text{dijet}} + \Delta > 200 \text{ GeV}$	Mass	6.61	1913	0.39	31
$\Delta\eta_{\text{dijet}} + \Delta > 150 \text{ GeV}$	Mass	6.94	2684	0.37	38
$\Delta\eta_{\text{dijet}} + \Delta > 100 \text{ GeV}$	Mass	6.61	3161	0.38	46

From these studies we conclude that there is no significant gain by using the kinematic fit or the mass pairing methods in terms of signal efficiency. In terms of background shape and rejection, the three methods have similar performance. The main caveat of using the kinematic fit or the mass pairing algorithm is that the distribution of the mass asymmetry looks very similar between background and signal, loosing all its discrimination power. For these reasons we chose the delta R pairing procedure for this analysis, as done in Run I [22].

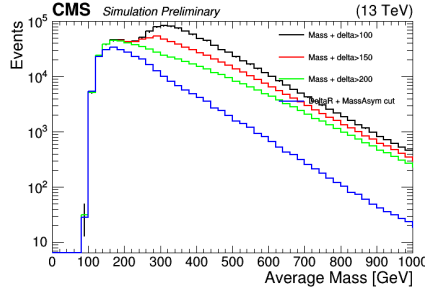


Figure B.8: Average mass distribution comparing different Δ cuts within the minimum mass optimization and the nominal selection for QCD multijets MC. Distributions are scaled to 36 fb^{-1} .

B.2 Jet energy scale uncertainties

These uncertainties are η -dependent in a per-jet basis, we apply ± 1 standard deviation adjustments to the jet mass. The $+1$ standard deviation is called JESUp and the -1 is JESDown. Figures B.10 show the distribution of the average mass with the nominal, JESUp and JESDown. From the number of events in the mass distribution around 2σ the true mass, we calculate the acceptance for each mass point and our results are shown in Fig. B.9.

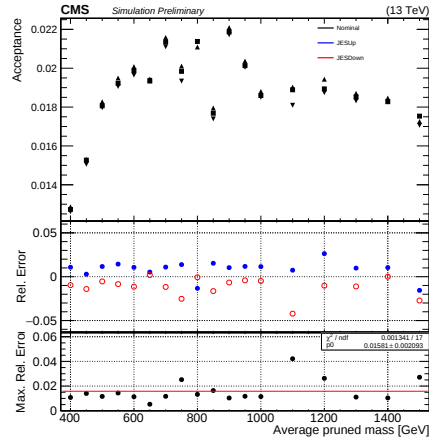


Figure B.9: Acceptance for nominal, JESUp and JESDown for all the signal mass points.

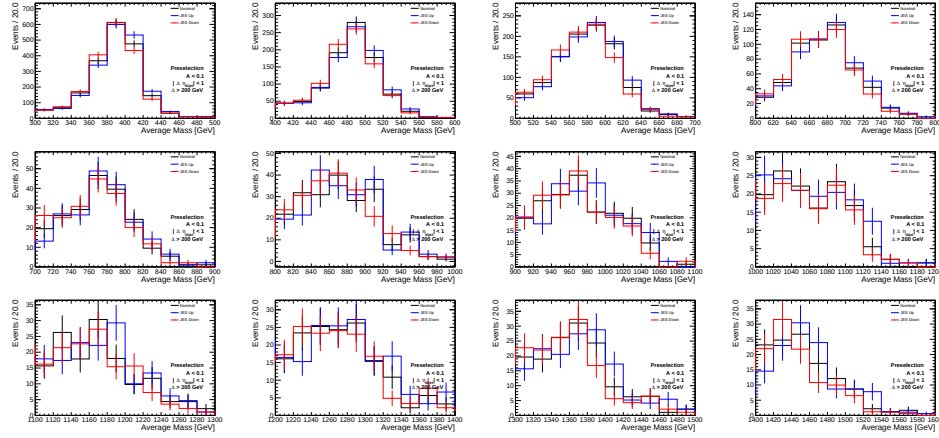


Figure B.10: Average dijet mass distributions for different $\tilde{t} \rightarrow qq'$ signals after all the inclusive selection is applied. For each mass, the nominal (black), JESUp (blue) and JESDown distribution are shown.

B.3 Jet energy resolution uncertainties

To measure the impact of the jet energy resolution on the $\overline{M}$ spectrum, we vary the JER value by ± 1 standard deviation. The $+1$ standard deviation is called JERup and the -1 is JERDown. Figures B.12 show the distribution of the average mass with the nominal, JERup and JERDown. From the number of events in the mass distribution around 2σ the true mass, we calculate the acceptance for each mass point and we results are shown in Fig. B.11.

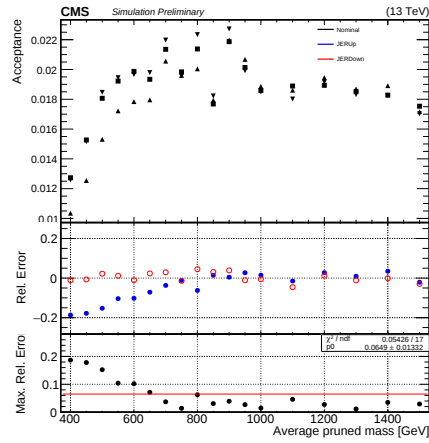


Figure B.11: Acceptance for nominal, JERup and JERDown for all the signal mass points.

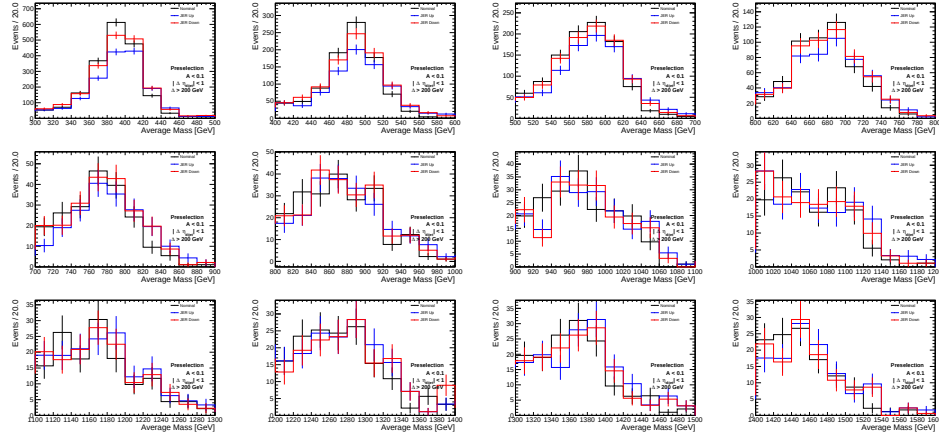


Figure B.12: Average dijet mass distributions for different $\tilde{t} \rightarrow qq'$ signals after all the inclusive selection is applied. For each mass, the nominal (black), JERUp (blue) and JERDown distribution are shown.

B.4 Pileup uncertainties

The MC samples are generated with a pileup distribution that is close to the real pileup distribution in data, but not identical. Thus, by re-weighting events based on the relative pileup distributions, the pileup distribution can be made to match that of the data. However, the modeling and measurement of the pileup distribution in data carries systematic uncertainties, which can be covered by varying the inelastic cross section parameter used to calculate pileup re-weighting by $\pm 5\%$. Figures B.14 show the distribution of the average mass with the nominal, PUUp and PUDown. We calculate the acceptance for each mass point and we results are shown in Fig. B.13.

B.5 PDF uncertainty

To measure the impact of the parton distribution functions in the signal samples, we evaluate the (RMS) of the distribution of the 100 NNPDF MC replicas as the $\pm 1\sigma$ shape variation due to parton distribution functions as recommended in [81]. The +1 standard deviation is called PDFUp and the -1 is PDFDown. Figures B.16 show the distribution of the average mass with the nominal, PDFUp and PDFDown. Based on the bottom plots from each

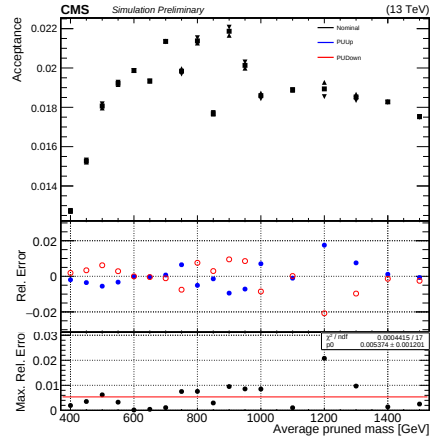


Figure B.13: Acceptance for nominal, PUUp and PUDown for all the signal mass points.

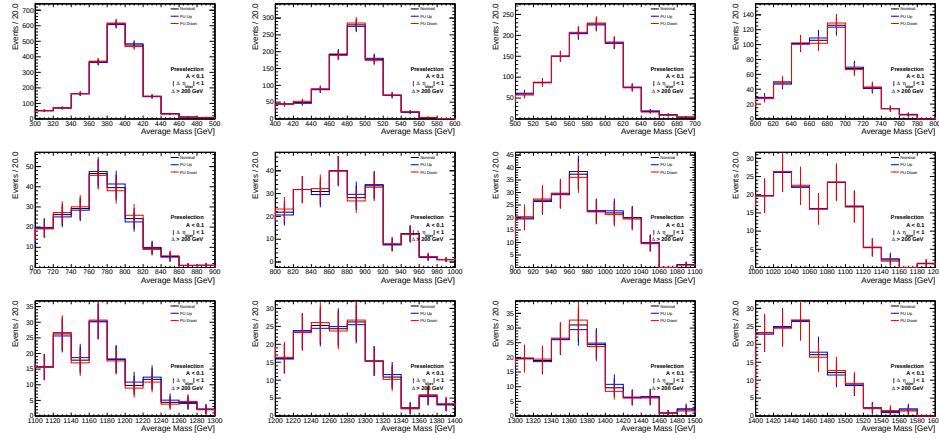


Figure B.14: Average dijet mass distributions for different $\tilde{t} \rightarrow qq'$ signals after all the inclusive selection is applied. For each mass, the nominal (black), PUUp (blue) and PUDown distribution are shown.

those Figures, we applied a flat 12% uncertainty to all the masses which covered all the differences in shape introduced by PDFs. We calculate the acceptance for each mass point and we results are shown in Fig. B.15.

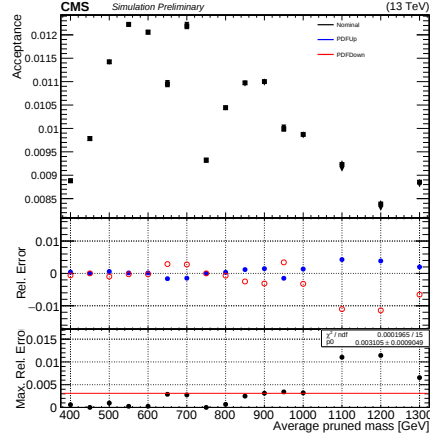


Figure B.15: Acceptance for nominal, pdfUp and pdfDown for all the signal mass points.

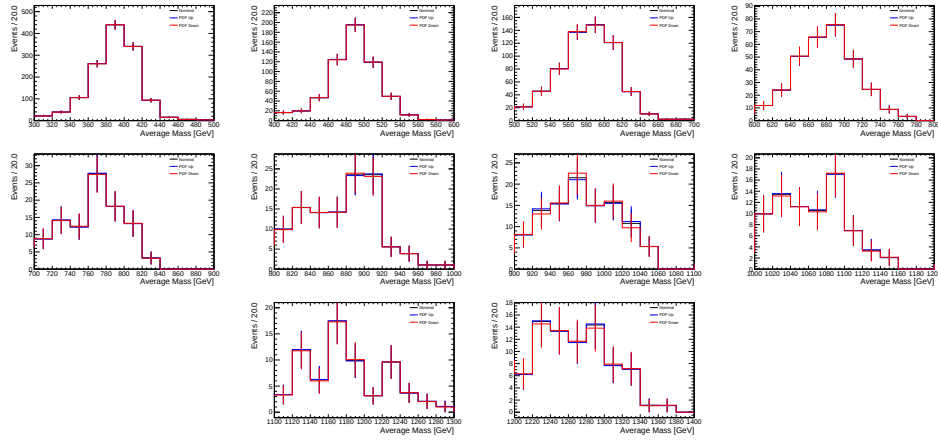


Figure B.16: Average dijet mass distributions for different $\tilde{t} \rightarrow qq'$ signals after all the inclusive selection is applied. For each mass, the nominal (black), pdfUp (blue) and pdfDown distribution are shown.

B.6 B tag uncertainty

To measure the impact of the b tagging working points in the $\overline{M}$ spectrum, we apply ± 1 standard deviation of the scale factor, adjust to the jet mass. This uncertainties are p_T ,

η , and flavour-dependent in a per-jet basis. The +1 standard deviation is called btagUp and the -1 is btagDown. Figures B.18 show the distribution of the average mass with the nominal, btagUp and btagDown. We calculate the acceptance for each mass point and the results are shown in Fig. B.17.

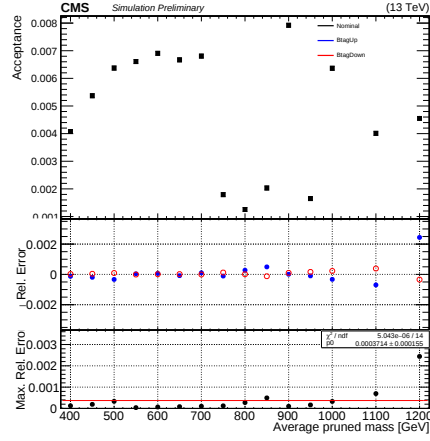


Figure B.17: Acceptance for nominal, btagUp and btagDown for all the signal mass points.

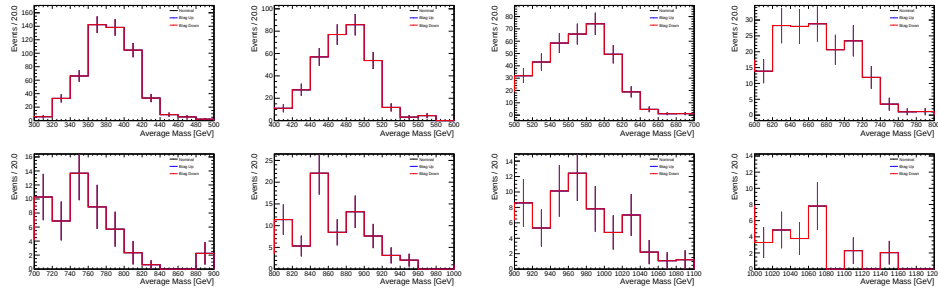


Figure B.18: Average dijet mass distributions for different $\tilde{t} \rightarrow bq'$ signals after all the b-tagged selection is applied. For each mass, the nominal (black), btagUp (blue) and btagDown distribution are shown.

Bibliography

- [1] D. Griffiths, *Introduction to Elementary Particles*. Wiley-VCH, 2008.
- [2] Wikipedia, “Standard model of elementary particles.” https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- [3] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.*, vol. B716, pp. 1–29, 2012.
- [4] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett.*, vol. B716, pp. 30–61, 2012.
- [5] ATLAS and CMS Collaboration, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments,” *Phys. Rev. Lett.*, vol. 114, p. 191803, 2015.
- [6] C. Patrignani and P. D. Group, “Review of particle physics,” *Chinese Physics C*, vol. 40, no. 10, p. 100001, 2016.
- [7] M. E. Peskin and D. V. Schroeder, *An Introduction To Quantum Field Theory, Student Economy Edition (Frontiers in Physics)*. Westview Press, 2015.
- [8] W. Hollik, “Quantum field theory and the Standard Model,” in *High-energy physics. Proceedings, 17th European School, ESHEP 2009, Bautzen, Germany, June 14–27, 2009*, 2010.
- [9] G. P. Salam, “Elements of QCD for hadron colliders,” in *High-energy physics. Proceedings, 17th European School, ESHEP 2009, Bautzen, Germany, June 14–27, 2009*, 2010.
- [10] C. Davies, “Lattice QCD: A Guide for people who want results,” in *Proceedings, 58th Scottish Universities Summer School in physics, SUSSP58, St. Andrews, UK, August 22–28, 2004*, pp. 233–272, 2005.
- [11] J. Feltesse, “Introduction to Parton Distribution Functions,” *Scholarpedia*, vol. 5, no. 11, p. 10160, 2010. revision #91386.
- [12] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, “New parton distribution functions from a global analysis of quantum chromodynamics,” *Phys. Rev.*, vol. D93, no. 3, p. 033006, 2016.
- [13] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC,” *Eur. Phys. J.*, vol. C63, pp. 189–285, 2009.

- [14] ZEUS H1 Collaboration, “Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data,” *Eur. Phys. J.*, vol. C75, no. 12, p. 580, 2015.
- [15] B. Webber, “Parton shower Monte Carlo event generators,” *Scholarpedia*, vol. 6, no. 12, p. 10662, 2011. revision #128236.
- [16] B. R. Webber, “Fragmentation and hadronization,” *Int. J. Mod. Phys.*, vol. A15S1, pp. 577–606, 2000. [eConfC990809,577(2000)].
- [17] S. P. Martin, “A Supersymmetry primer,” 1997. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [18] J. A. Evans and Y. Kats, “LHC Coverage of RPV MSSM with Light Stops,” *JHEP*, vol. 1304, p. 028, 2013.
- [19] R. Barbier *et al.*, “R-parity violating supersymmetry,” *Phys. Rept.*, vol. 420, pp. 1–202, 2005.
- [20] C. Brust, A. Katz, and R. Sundrum, “SUSY Stops at a Bump,” *JHEP*, vol. 08, p. 059, 2012.
- [21] CDF Collaboration, “Search for Pair Production of Strongly Interacting Particles Decaying to Pairs of Jets in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96\text{TeV}$,” *Phys. Rev. Lett.*, vol. 111, no. 3, p. 031802, 2013.
- [22] CMS Collaboration, “Search for pair-produced resonances decaying to jet pairs in proton-proton collisions at $\sqrt{s}=8\text{ TeV}$,” *Phys. Lett.*, vol. B747, pp. 98–119, 2015.
- [23] ATLAS Collaboration, “A search for pair produced resonances in four jets final states in proton-proton collisions at $\sqrt{s}=13\text{ TeV}$ with the ATLAS experiment,” Tech. Rep. ATLAS-CONF-2016-084, CERN, Geneva, Aug 2016.
- [24] ATLAS Collaboration, “A search for R-parity violating scalar top decays in all-hadronic final states with the ATLAS detector in $\sqrt{s} = 8\text{ TeV}$ pp collisions,” Tech. Rep. ATLAS-CONF-2015-026, CERN, Geneva, 2015.
- [25] CMS Collaboration, “Search for low-mass pair-produced dijet resonances using jet substructure techniques in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\text{ TeV}$,” Tech. Rep. CMS-PAS-EXO-16-029, CERN, Geneva, 2016.
- [26] L. Evans and P. Bryant, “LHC Machine,” *JINST*, vol. 3, p. S08001, 2008.
- [27] F. Marcastel, “CERN’s Accelerator Complex,” Oct 2013. General Photo.
- [28] CMS Collaboration, “Public CMS Luminosity Information.” https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016_Proton_Proton_13_TeV_Collis.
- [29] CMS Collaboration, “The CMS experiment at the CERN LHC,” *JINST*, vol. 3, p. S08004, 2008.
- [30] CMS Collaboration, “About CMS.” <https://cms.cern/detector>.

- [31] CMS Collaboration, “The CMS detector magnet,” *IEEE Trans. Appl. Supercond.*, vol. 10, no. 1, pp. 389–94, 2000.
- [32] CMS Collaboration, “Commissioning and Performance of the CMS Pixel Tracker with Cosmic Ray Muons,” *JINST*, vol. 5, p. T03007, 2010.
- [33] CMS Collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS, Geneva: CERN, 2006.
- [34] CMS Collaboration, “Silicon pixels.” <https://cms.cern/detector/identifying-tracks/silicon-pixels>.
- [35] CMS Collaboration, “Silicon strips.” <https://cms.cern/detector/identifying-tracks/silicon-strips>.
- [36] CMS Collaboration, “CMS Silicon Strip Tracker Performance,” *Phys. Procedia*, vol. 37, pp. 844–850, 2012.
- [37] CMS Collaboration, “CMS ECAL Endcap (EE) SC04 Assembly Photos.” CMS Collection., Jun 2004.
- [38] 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. [JINST8,9009(2013)].
- [39] CMS Collaboration, “Hadronic Calorimeter (HCAL) Calibration with Cosmics and Wire Source has started. La calibration du HCAL avec des cosmiques et une source radioactive a commenc  ,” Oct 2005. CMS Collection.
- [40] CMS Collaboration, “The CMS trigger system,” *JINST*, vol. 12, p. P01020. 122 p, Sep 2016.
- [41] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period,” tech. rep., 2017.
- [42] CMS Collaboration, “Interactive Slice of the CMS detector,” Aug 2016.
- [43] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector,” 2017.
- [44] ALEPH Collaboration, “Performance of the ALEPH detector at LEP,” *Nucl. Instrum. Meth.*, vol. A360, pp. 481–506, 1995.
- [45] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” *JINST*, vol. 9, no. 10, p. P10009, 2014.
- [46] G. P. Salam, “Towards Jetography,” *Eur. Phys. J.*, vol. C67, pp. 637–686, 2010.
- [47] CMS Collaboration, “Jet algorithms performance in 13 TeV data,” 2017.
- [48] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,” *JHEP*, vol. 04, p. 063, 2008.
- [49] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification,” *JHEP*, vol. 10, p. 059, 2014.

- [50] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS,” *JINST*, vol. 6, p. P11002, 2011.
- [51] CMS Collaboration, “Jet energy resolution and corrections (jerc) subgroup.” https://twiki.cern.ch/twiki/pub/CMS/JetEnergyScale/JEC_levels.png.
- [52] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV,” *JINST*, vol. 12, no. 02, p. P02014, 2017.
- [53] CMS Collaboration, “Jet energy scale and resolution performances with 13TeV data,” Jun 2016.
- [54] CMS Collaboration, “Pileup Removal Algorithms,” Tech. Rep. CMS-PAS-JME-14-001, CERN, Geneva, 2014.
- [55] D. Krohn, J. Thaler, and L.-T. Wang, “Jet Trimming,” *JHEP*, vol. 02, p. 084, 2010.
- [56] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches,” *Phys. Rev.*, vol. D81, p. 094023, 2010.
- [57] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet substructure,” *JHEP*, vol. 09, p. 029, 2013.
- [58] J. Thaler and K. Van Tilburg, “Maximizing Boosted Top Identification by Minimizing N-subjettiness,” *JHEP*, vol. 02, p. 093, 2012.
- [59] CMS Collaboration, “Identification of b-quark jets with the CMS experiment,” *JINST*, vol. 8, p. P04013, 2013.
- [60] Wikipedia, “b-tagging.” https://en.wikipedia.org/wiki/B-tagging#/media/File:B-tagging_diagram.png.
- [61] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2,” Tech. Rep. CMS-PAS-BTV-15-001, CERN, Geneva, 2016.
- [62] CMS Collaboration, “Performance of b-Tagging Algorithms in Proton Collisions at 13 TeV using the 2016 Data,” Jul 2016.
- [63] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, “MadGraph 5 : Going Beyond,” *JHEP*, vol. 1106, p. 128, 2011.
- [64] T. Sjostrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.
- [65] S. Agostinelli *et al.*, “Geant4 – a simulation toolkit,” *Nucl. Instr. and Methods A*, vol. 506, pp. 250–303, 2003.
- [66] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements,” *Eur. Phys. J.*, vol. C76, no. 3, p. 155, 2016.
- [67] Borschensky, Christoph and Krämer, Michael and Kulesza, Anna and Mangano, Michelangelo and Padhi, Sanjay and Plehn, Tilman and Portell, Xavier, “Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33$ and 100 TeV,” *Eur. Phys. J.*, vol. C74, no. 12, p. 3174, 2014.

- [68] Beenakker, Wim and Borschensky, Christoph and Krämer, Michael and Kulesza, Anna and Laenen, Eric and Marzani, Simone and Rojo, Juan, “NLO+NLL squark and gluino production cross-sections with threshold-improved parton distributions,” *Eur. Phys. J.*, vol. C76, no. 2, p. 53, 2016.
- [69] P. Nason, “A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms,” *J. High Energy Phys.*, vol. 11, p. 040. 29 p, Sep 2004.
- [70] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,” *J. High Energy Phys.*, vol. 11, p. 070. 91 p, Sep 2007. Comments: 91 pages, 2 figures.
- [71] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” Feb 2010. Comments: 57 pages, 4 figures.
- [72] CMS Collaboration, “W/z-tagging of jets.” "<https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetWtagging>".
- [73] CMS Collaboration, “Boosted top jet tagging.” "<https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetTopTagging>".
- [74] CMS Collaboration, “Usage of b/c tag objects for 13 tev data in 2016 and 80x mc.” <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco>.
- [75] CMS Collaboration, “Methods to apply b-tagging efficiency scale factors.” "<https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods>".
- [76] S. Kallweit, J. M. Lindert, P. Maierhofer, S. Pozzorini, and M. Schönherr, “NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging,” *JHEP*, vol. 04, p. 021, 2016.
- [77] S. Kallweit, J. M. Lindert, S. Pozzorini, M. Schönherr, and P. Maierhöfer, “NLO QCD+EW automation and precise predictions for V+multijet production,” in *Proceedings, 50th Rencontres de Moriond, QCD and high energy interactions: La Thuile, Italy, March 21-28, 2015*, pp. 121–124, 2015.
- [78] J. M. Lindert *et al.*, “Precise predictions for V+ jets dark matter backgrounds,” *Eur. Phys. J.*, vol. C77, no. 12, p. 829, 2017.
- [79] CMS Collaboration, “Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at $\sqrt{s} = 13$ TeV,” *JHEP*, vol. 07, p. 014, 2017.
- [80] S. Kallweit, J. M. Lindert, P. Maierhöfer, S. Pozzorini, and M. Schönherr, “NLO electroweak automation and precise predictions for W+multijet production at the LHC,” *JHEP*, vol. 04, p. 012, 2015.
- [81] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. D. Roeck, J. Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, R. McNulty, A. Morsch, P. Nadolsky, V. Radescu, J. Rojo, and R. Thorne, “Pdf4lhc recommendations for lhc run ii,” *Journal of Physics G: Nuclear and Particle Physics*, vol. 43, no. 2, p. 023001, 2016.

- [82] R. Chivukula, M. Golden, and E. H. Simmons, “Multi-jet physics at hadron colliders,” *Nuclear Physics B*, vol. 363, no. 1, pp. 83 – 96, 1991.
- [83] CMS Collaboration, “Search for narrow resonances decaying to dijets in proton-proton collisions at $\sqrt{s} = 13$ TeV,” *Phys. Rev. Lett.*, vol. 116, no. 7, p. 071801, 2016.
- [84] CMS Collaboration, “Search for massive resonances decaying into pairs of boosted W and Z bosons at $\sqrt{s} = 13$ TeV,” Tech. Rep. CMS-PAS-EXO-15-002, CERN, Geneva, 2015.
- [85] CMS Collaboration, “Cms statistics committee recommendations - parametric density estimation.” "<https://twiki.cern.ch/twiki/bin/view/CMS/DensityEstimation>".
- [86] R. M. Harris and K. Kousouris, “Searches for Dijet Resonances at Hadron Colliders,” *Int. J. Mod. Phys.*, vol. A26, pp. 5005–5055, 2011.
- [87] CMS Collaboration, “Search for high-mass $Z\gamma$ resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV,” Tech. Rep. CMS-PAS-EXO-17-005, CERN, Geneva, 2017.
- [88] CMS Collaboration, “Jet energy resolution.” "<https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>".
- [89] CMS Collaboration, “Procedure for the LHC Higgs boson search combination in Summer 2011,” Tech. Rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.
- [90] A. L. Read, “Presentation of search results: the cl_s technique,” *Journal of Physics G: Nuclear and Particle Physics*, vol. 28, no. 10, p. 2693, 2002.
- [91] T. Junk, “Confidence level computation for combining searches with small statistics,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 434, no. 2, pp. 435 – 443, 1999.
- [92] CMS Collaboration, “Search for pair-produced dijet resonances in four-jet final states in pp collisions at $\sqrt{s} = 7$ TeV,” *Phys. Rev. Lett.*, vol. 110, no. 14, p. 141802, 2013.