
**A Novel Multidimensional Search for
Diboson Resonances in the Boosted Dijet Final State**

and

Encoding Jet Substructure with a Deep Neural Network

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

Thea Klæboe Årrestad
aus
Norwegen

Promotionskommission:

Prof. Dr. Ben Kilminster (Vorsitz)

Prof. Dr. Florencia Canelli

Prof. Dr. Jesse Thaler

Dr. Andreas Hinzmann

Zürich 2019

Abstract

In this doctoral thesis I will present three different searches for new heavy resonances decaying to pairs of vector bosons in the all-hadronic final state. The analyzed data were collected by the CMS experiment at the LHC during the first three years of data-taking at a collision center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 2.7 (2015), 35.9 (2016) and 77.3 (2016+2017) fb^{-1} , and the searches were the first of their kind to ever be performed at such a high collision energy. The diboson final states under consideration are challenging to resolve due to the bosons being highly energetic (“boosted”), resulting in the two quarks from the decay being collimated and merging into a single jet. This leads to a dijet final state topology where each jet displays some energy substructure. The first search I will present, was one of the two first CMS searches in boosted final states with 13 TeV data to become published, and the first to take advantage of jet substructure at the trigger level. It was a high profile analysis due to a previously observed $3.4\ (1.3)\ \sigma$ excess around 2 TeV in the 8 TeV dataset, as analyzed by ATLAS (CMS), and I brought the search to a published result within six months after 13 TeV data taking began (and within ten months after embarking on my PhD). Following this, in my second analysis I optimized, validated and commissioned the novel PUPPI softdrop jet grooming algorithm for vector-boson tagging and, in addition, developed dedicated mass corrections for the softdrop jet mass. The algorithm and corresponding mass corrections are now the default for vector boson tagging in CMS and used by several analyses. It was the first published result taking advantage of PUPPI softdrop jet grooming. The third and final search I will present introduces a novel multidimensional search framework, which can be used to search for resonances peaking anywhere in the 3D spectrum of the dijet and groomed jet mass spectra. Validated through a search with hadronically decaying vector bosons in the final state, which I will present here, the framework can be used to incorporate all resonance searches with hadronically decaying vector boson or hadronically decaying Higgs boson final states, as well as for generic searches for any boosted object peaking in jet mass. In the context of this search, a simultaneous fit to the $W(q\bar{q})$ and $Z(q\bar{q})$ jet mass peaks from the Standard Model (SM) $V(\bar{q}q)+\text{jets}$ process has been performed, for the first time allowing to constrain the groomed jet mass scale and resolution from this SM process. Finally, I will present a deep neural network for vector-boson tagging, which encodes jet clustering- and substructure-like variables into the neural network layers themselves. This algorithm significantly improves the analysis sensitivity, and can also be used as a stepping stone in the development of a generic anti-QCD tagger. The latter will be of great importance when attempting to use the multidimensional framework for model-independent searches.

Zusammenfassung

In dieser Doktorarbeit werde ich drei verschiedene Analysen über neue schwere Resonanzen vorstellen, die in Paare von Vektorbosonen zerfallen, welche wiederum in Quarks zerfallen. Die analysierten Daten wurden vom CMS-Experiment am LHC während der ersten drei Jahre der Datennahme bei einer Schwerpunktsenergie von 13 TeV gesammelt, was einer integrierten Luminosität von 2.7 (2015), 35.9 (2016) und 77.3 (2016 + 2017) fb^{-1} entspricht. Diese Analysen sind die ersten ihrer Art, die jemals bei einer so hohen Kollisionsenergie durchgeführt wurden. Die Endzustände des Vektorboson-Paare sind schwierig voneinander zu unterscheiden, da die Bosonen hochenergetisch (“boosted”) sind, was dazu führt, dass die beiden Quarks aus dem Zerfall kollimiert werden und zu einem einzigen Teilchenbündel (“jet”) verschmelzen. Dies führt zu einer Dijet-Endzustands-Topologie, bei der jeder Jet eine Energieunterstruktur aufweist. Die erste Analyse, die ich vorstellen werde, war eine der ersten veröffentlichten CMS-Analysen in “boosted” Endzuständen mit 13 TeV-Daten, und die erste, welche die Jet-Substruktur auf Trigger-Ebene nutzte. Es war eine hoch gehandelte Analyse aufgrund eines zuvor beobachteten 3.4 (1.3) σ -Überschusses um 2 TeV im 8 TeV-Datensatz, der von ATLAS (CMS) analysiert wurde. Ich habe die Analyse innerhalb von sechs Monaten nachdem die 13 TeV-Datenaufnahme begonnen hatte zu einem veröffentlichten Ergebnis gebracht. Im Anschluss daran habe ich in meiner zweiten Analyse einen neuartigen Algorithmus (“PUPPI Softdrop”) für das Vektorboson-Tagging optimiert, validiert und in Betrieb genommen und zusätzlich spezielle Massenkorrekturen für die Softdrop Jet-Masse entwickelt. Der Algorithmus und die entsprechenden Massenkorrekturen sind jetzt der Standard für das Vektorboson-Tagging in CMS und werden von mehreren Analysen verwendet. Es war das erste veröffentlichte Ergebnis, bei dem PUPPI-Softdrop genutzt wurde. Die dritte und letzte Analyse führt ein neuartiges multidimensionales Analyse ein, mit dem nach Resonanzen irgendwo im 3D-Spektrum der Dijet- und Jet-Massenspektren gesucht werden kann. Dieses Analyse-Architektur kann für alle Resonanzanalysen über hadronisch zerfallende Vektorbosonen oder hadronisch zerfallende Higgs-Boson-Endzustände sowie generische Analysen über geboostete Objekt, die in der Jet-Masse einen Höchstwert (Peak) haben, verwendet werden. Schließlich werde ich ein Deep Neural Network für das Vektorboson-Tagging vorstellen, welches Jet-Clustering- und substrukturähnliche Variablen in die neuronalen Netzwerkschichten encodiert. Dieser Algorithmus verbessert die Analyseempfindlichkeit erheblich und kann auch als Grundlage für die Entwicklung eines generischen Anti-QCD-Taggers verwendet werden. Letzteres ist von großer Bedeutung, wenn versucht wird, dieses mehrdimensionale Framework für modellunabhängige Analysen zu verwenden.

Contents

I Theoretical background	15
1 The Standard Model	17
1.1 Fundamental particles: quarks and leptons	17
1.2 The Standard Model Lagrangian	19
1.2.1 The quantum chromodynamics sector	21
1.2.2 The electroweak sector	23
1.2.3 The Higgs sector	26
2 Beyond Standard Model Physics	31
2.1 Shortcomings of the Standard Model	31
2.2 Compositeness	33
2.2.1 Heavy Vector Triplet formalism	33
2.3 Warped extra dimensions	36
II Experimental setup	39
3 The Large Hadron Collider	41
4 The CMS detector	45
4.1 Coordinate system	46
4.2 Tracking detectors	47
4.3 Electromagnetic calorimeter	48
4.4 Hadronic calorimeter	51
4.5 Muon chambers	54
4.6 The CMS trigger system	54

5 Event reconstruction	57
5.1 Track and primary vertex reconstruction	57
5.2 The Particle Flow Algorithm	59
5.2.1 Reconstruction of the Particle Flow inputs	60
5.2.2 Particle Flow identification	62
5.3 Pile-up removal	66
5.3.1 Charged Hadron Subtraction	66
5.3.2 Pile up per particle identification (PUPPI)	66
5.4 Jet reconstruction	68
5.4.1 Jet clustering	68
5.4.2 PF jets in CMS	70
5.4.3 Jet energy corrections	71
5.5 Jet substructure reconstruction	73
5.5.1 Grooming	75
5.5.2 N-subjettiness	78
5.5.3 Decorrelated n-subjettiness	80
5.5.4 Vector boson tagging	83
5.6 Monte Carlo Event Generators	86
III Diboson resonance searches in CMS	89
6 Search I: First search for diboson resonances at 13 TeV	91
6.1 A small bump	93
6.2 Analysis strategy	94
6.3 Data and simulated samples	96
6.4 Event selection	97
6.4.1 Triggering	97
6.4.2 Preselection	99
6.4.3 Vector boson tagging	101
6.4.4 Analysis categorization	108
6.5 Background modeling	113
6.6 Signal modeling	117
6.7 V-tagging scale factors	118
6.7.1 Event selection	118
6.7.2 Fitting procedure	122

6.7.3	Systematic uncertainties	125
6.7.4	Fit results	127
6.7.5	Impact on search variables	128
6.8	Systematic uncertainties	129
6.9	Results	130
6.10	Limits	130
6.10.1	All-hadronic analysis only	130
6.10.2	Semi-leptonic and all-hadronic combination	131
7	Search II: A new pileup resistant and perturbative safe tagger	139
7.1	Towards robust boosted jet tagging	141
7.2	Analysis strategy	142
7.3	Data and simulated samples	143
7.4	Event selection	144
7.4.1	Triggering	144
7.4.2	Preselection	147
7.5	Developing a new W-tagger	148
7.5.1	Dedicated PUPPI softdrop jet mass corrections	150
7.5.2	W-tagging performance	154
7.5.3	Efficiency scale factors, jet mass scale and resolution	156
7.5.4	W-tagging mistag rate measurement	159
7.6	Mass and purity categorization	162
7.7	Background modeling	163
7.8	Signal modeling	165
7.9	Systematic uncertainties	166
7.10	Results	167
7.10.1	Early analysis	167
7.10.2	Full 2016 dataset	167
8	Search III: A novel framework for multi-dimensional searches	179
8.1	Small bumps and tri-bosons	181
8.2	Analysis strategy	183
8.3	Data and simulated samples	184
8.4	Event selection	184
8.5	Triggering	186
8.5.1	Trigger turn-on modeling	186

8.6	A mass- and p_T -decorrelated tagger	190
8.6.1	Data to simulation scale factors	195
8.6.2	p_T -dependence	198
8.7	The multidimensional fit	201
8.7.1	Modeling of the signal	202
8.7.2	Modeling of the non-resonant background	205
8.7.3	Modeling of the resonant background	209
8.8	Systematic uncertainties	214
8.9	Background model validation	217
8.9.1	Variations of QCD multijet predictions	217
8.9.2	Fit to data control region	217
8.9.3	Bias test in pseudodata	218
8.10	Results	231
8.10.1	Limits	234
8.10.2	Pulls of nuisance parameters	239
9	Summary and outlook	243
IV	Encoding jet substructure in a deep neural network	247
10 A	Lorentz invariance based Deep Neural Network for W-tagging	251
10.1	LoLa	251
10.1.1	Architecture	252
10.1.2	Input	252
10.1.3	The Combination Layer	255
10.1.4	The Lorentz Layer	257
10.2	Performance	259
10.3	Dependence on jet mass and p_T	260
10.3.1	p_T -decorrelation	261
10.3.2	Mass sculpting	263
10.4	Validation	265
10.5	What does LoLa learn?	268
11	Summary	271

V Final conclusions and outlook	273
A V-tagging	279
A.1 mMDT and underlying event	280
A.2 N-subjettiness on groomed jets	281
B Search I: Limits per mass category	285
B.1 Limits per mass category	286
C Search II	295
C.1 W-tagging scale factor: Additional plots	296
C.2 Efficiency scale factors for 2.5 fb^{-1}	298
C.3 Background fit checks	300
C.3.1 Background fit checks in data sideband	301
C.3.2 Background fit checks in QCD MC	308
C.4 F-test in signal region	312
D Search III	317
D.1 Signal fits	318
D.2 2016 kernels	319
D.3 Resonant background	319

Introduction

The Standard Model of particle physics (SM) is one of the greatest accomplishments of fundamental science. The degree to which it can accurately predict observed phenomena is unprecedented, and it has allowed us to incorporate all of particle physics into one single equation that explains what we can see in the world around us. However, it has some shortcomings. One major problem is that the Standard Model and gravity are incompatible at very high energies and that any incorporation of gravity into a quantum field theory framework results in a non-renormalizable theory. The SM also fails to explain why gravity is so much weaker than the electromagnetic and nuclear forces. This, together with a few other problems, has lead scientists to search for extensions to the Standard Model, commonly referred to as *Beyond Standard Model physics* (BSM). These models are usually accompanied by predicted observables not included in the Standard Model, where the observation of these, or the lack thereof, is a way of falsifying or supporting the model.

In this thesis, I look for such observables by searching for new massive particles predicted by SM extensions. These particles have the property that they can decay into vector bosons, W^\pm and Z^0 , and usually have a very small interaction probability. The vector bosons are heavy and unstable and will quickly decay into leptons or quarks. In order to counterbalance the small interaction probability associated with generating such a new heavy particle, a final state with two vector bosons decaying hadronically is required since the branching ratio for a vector boson decaying to hadrons is significantly higher than that into leptons. This final state is complicated by the presence of an overwhelming QCD multijet background and the fact that, due to the high mass of the resonance, the vector bosons are highly energetic and their quark decay products get merged into a single jet due to the small angular opening between them. The latter offers an opportunity to distinguish between vector boson jets and jets coming from a quark or a gluon due to the expected differences in jets mass and geometrical substructure within the jet. Algorithms designed to improve the jet mass resolution and resolve jet substructure are commonly referred to

as *jet substructure methods*. These will be a recurring topic of this thesis due to my own personal contributions to the field.

Three searches for heavy resonances decaying to dibosons in the all-hadronic final state will be presented. The first analysis to be discussed was the first of its kind to be performed at a center-of-mass energy of $\sqrt{s} = 13\text{ TeV}$ and the first published result to take advantage of jet substructure at trigger level. The second led to the development of a novel pileup robust and perturbative safe vector boson tagging algorithm, which afterwards became the default tagging algorithm in CMS. Finally, the third search introduces a completely new way of doing diboson searches in a multi-dimensional space, allowing for the incorporation of all VV, VH and HH searches (where V = W,Z and H = Higgs boson) into one common framework, as well as any generic search for resonances peaking in jet mass and dijet invariant mass.

In addition to the tree searches, I will present a deep neural network for vector boson tagging intended to improve the analysis sensitivity for future searches. In this algorithm, jet clustering and substructure-like variables are embedded into the neural network layers themselves. That makes it a good starting point in the development of a generic anti-QCD tagger capable of distinguishing between the QCD background and any signal with some geometrical substructure peaking in the jet groomed mass spectrum. Such a tagger would, in combination with the multidimensional fit framework, lead to a completely new way of doing model independent searches.

This thesis is organized in five parts. Part I provides the theoretical background and motivation for the searches presented here. First, the Standard Model is introduced together with a discussion of its known shortcomings, followed by a chapter presenting two possible extensions to the Standard Model, both of which are probed in this thesis. Part II consists of a description of the experimental setup used to collect the data which is analyzed here, as well as the different algorithms used in order to reconstruct each event. Part I and Part II mainly consist of work done by others which has been vital for the completion of this work. The remaining three parts are dedicated to my own personal contributions. In Part III, the three searches described above are presented in chronological order, each with a personal introduction motivating the analysis in question. Following this, in Part IV, the deep neural network based vector boson tagger for future analyses will be presented. Both parts end with their own concluding summary and outlook, in Section 9 and Section 11. In Part V, a final summary of the obtained results and a discussion of the future for the analysis is given.

Part I

Theoretical background

CHAPTER 1

The Standard Model

Everything this thesis is built on has its roots in the Standard Model (SM). The Standard Model of particle physics addresses the question *What is matter made of?* on the smallest possible scale. It links the fundamental constituents of the universe together along with the forces that bind them, in order to describe and predict the laws of nature. The Standard Model is formulated as a quantum field theory, where the fundamental particles are spin-1/2 fermions which interact with one another through the exchange of spin-1 gauge bosons. These interactions come in three forms, mediated by three different types of gauge bosons: The electromagnetic force, mediated through photons; the weak force, mediated through W and Z bosons; and the strong force, mediated by gluons. How the fundamental particles interact also defines which properties they exhibit. In addition, the Standard Model includes a field very different from the others, the Higgs field. The Higgs field interacts with both fermions and bosons and is what gives all particles their mass. One thing the Standard Model fails to incorporate is the force of gravity. This shortcoming is one of the main motivations for looking for alternative models beyond the Standard Model (BSM), which is the main topic of this thesis.

1.1 Fundamental particles: quarks and leptons

It appears that all matter in the universe can be described by a very small collection of fundamental particles, six leptons and six quarks. These are collectively called fermions and are, as far as we can tell, truly elementary (not composed of any other particles). Leptons are particles with integer or zero electric charge, defined in units of electron charge. They come in three flavors, or generations, and their mass increases with generation. Each generation of leptons consists of two particles: one charged lepton and one neutrally charged particle denoted as *neutrino* (ν). The

three generations can be arranged in a doublet structure, and are as follows.

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \quad (1.1)$$

The charged leptons can be positively or negatively charged, defined in units of electron charge e . By convention the leptons of matter are negatively charged, e^- , μ^- , and τ^- , whereas the positively charged leptons, e^+ , μ^+ , and τ^+ are considered their anti-particles. A summary of the lepton properties is listed in Table 1.1. Leptons interact with one another through the

Lepton	Mass	Charge
e^-	0.51 MeV	e
μ^-	105.7 MeV	e
τ^-	1776.9 MeV	e
$\nu_i, i = e, \mu, \tau$	$\sum_i m_i < 0.170$ eV	0

Table 1.1: Lepton Properties [1, 2].

electromagnetic and weak forces, which will be explained in more detail in Section 1.2.2. The other six fundamental particles of matter are the *quarks*. They are distinguished from the leptons in that they, in addition to interacting through the weak and electromagnetic forces, interact with one another through the *strong force*, described in Section 1.2.1. This force binds the quarks together to form baryons (like protons and neutrons) or mesons (like pions), and in addition, keeps the quarks from being observed as free particles such that they are only visible through their baryon or meson bound states (commonly referred to as *hadrons*). Also organized in three generations, the six quarks are called *up*, *down*, *charm*, *strange*, *top* and *bottom*, and are organized in flavor doublets as follow

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix} \quad (1.2)$$

Each quark comes with a fractional charge of $\frac{2}{3}$ (u, c and t) and $-\frac{1}{3}$ (d, s and b) of one electron charge. As with the leptons, there are also distinct particles of opposite charge, anti-quarks. Some of the quark properties are listed in Table 1.2.

These 12 fermions, together with their corresponding anti-particles, represent the fundamental particles of the universe and constitute all matter around us. There are four fundamental forces that we know of: gravity, electromagnetism, the weak force and the strong force. Gravity is extremely weak compared to the other forces and we currently lack a quantum field theory of

Quark	Mass	Charge
u	1 – 5 MeV	$\frac{2}{3}e$
d	3 – 9 MeV	$-\frac{1}{3}e$
c	1.15 – 1.35 GeV	$\frac{2}{3}e$
s	75 – 170 MeV	$-\frac{1}{3}e$
t	172.4 ± 0.1 GeV	$\frac{2}{3}e$
b	4.0 – 4.4 GeV	$-\frac{1}{3}e$

Table 1.2: Quark Properties

its interaction, therefore it is typically ignored in high energy physics experiments. All particles that are electrically charged, the charged leptons (e , μ and τ) and all of the quarks, interact through the electromagnetic force. These interactions are governed by the laws of Quantum Electrodynamics (QED), and are mediated through the massless and electrically neutral spin-1 photon. All of the fermions, including the electrically neutral neutrinos, feel the weak force and undergo weak interactions. The weak force is mediated through the vector bosons W^+ , W^- and Z^0 , which are heavy charged particles with a spin of 1. Finally, there is the strong force, mediated by the massless and electrically neutral spin-1 gluon. Only quarks interact via the strong force, and it is that interaction that makes the quarks so fundamentally different from leptons and neutrinos. Their interaction is governed by the laws of Quantum Chromodynamics (QCD). All of these interactions can be represented in one common gauge theory, the Standard Model.

1.2 The Standard Model Lagrangian

The Standard Model is a quantum gauge field theory in which each particle is described as a dynamical field with a value at each space-time coordinate. These fields are governed by a Lagrangian density function, the Standard Model. For instance, the Lagrangian density of a free fermion, one not interacting with any other fields, is

$$\mathcal{L} = \bar{\Psi}(x^\mu)(i\lambda^\mu \partial_\mu - m)\Psi(x^\mu) \quad (1.3)$$

where $\Psi(x^\mu)$ represents any spin-1/2 fermion field, also called *Dirac field*, as a function of space-time; $\bar{\Psi} = \Psi^\dagger \gamma_0$, where γ^0 is one of the gamma matrices γ^μ and is included in order to make $\Psi^\dagger \Psi$ invariant under Lorentz transformation; and m is the mass of the fermion in question. Any interaction between the fundamental particles due to the fundamental forces, can be described as variations in the Lagrangian of quantum fields and are represented as additional terms in the

equation above.

Being a gauge theory, the Standard Model has the property of gauge invariance, meaning that measurable quantities stay the same despite the fields themselves changing. If observables stay the same after a field transformation, there is a symmetry in the system. The symmetries of the Standard Model arise due to the fact that fermions of a given type are indistinguishable from one another. These symmetries result in the presence of *force mediators*, which arise as a representation of infinitesimal generators of some symmetry group. The fermion fields can be arranged as particle multiplets where one transforms into the other under a symmetry transformation. A symmetry transformation produce rotations between the particles of a given multiplet, but never to a field outside of that group. The symmetry group of the Standard Model is the direct product

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y. \quad (1.4)$$

$SU(3)_C$ is the color (C) symmetry allowing the rotation of quarks, arranged in color multiplets, into one another, corresponding to interactions produced by the strong force. The colors are denoted as red, green and blue. $SU(2)_L \otimes U(1)_Y$ represent the electroweak force with weak left-handed isospin L and weak hypercharge Y symmetries, acting on L and Y multiplets, respectively. The particle multiplets of the Standard Model can then be written as $\mathcal{G}_{SM} \ni x = (C, L)_{(Y)}$ and are (for illustrations only for the the 1st generation):

$$Q = (3, 2)_{(1/3)} = \begin{pmatrix} u_r & u_g & u_b \\ d_r & d_g & d_b \end{pmatrix} \sim \text{quark multiplet} \quad (1.5)$$

$$L = (1, 2)_{(-1)} = \begin{pmatrix} \nu_e & e \end{pmatrix} \sim \text{leptonic doublet} \quad (1.6)$$

$$u^c = (\bar{3}, 1)_{(-4/3)} = \begin{pmatrix} u_r^c & u_g^c & u_b^c \end{pmatrix} \sim \text{anti-up quarks} \quad (1.7)$$

$$d^c = (\bar{3}, 1)_{(2/3)} = \begin{pmatrix} d_r^c & d_g^c & d_b^c \end{pmatrix} \sim \text{anti-down quarks} \quad (1.8)$$

$$e^c = (1, 1)_{(2)} \sim \text{positron} \quad (1.9)$$

$$(1.10)$$

The right-handed neutrino (which in needed to generate a Dirac mass for the neutrinos) is not included here because it is a SM *singlet* and does not transform under the SM group \mathcal{G}_{SM} , but it could be written as $\nu^c = (1, 1)_{(0)}$. The positron e^c , on the other hand, is included as its hypercharge is non-zero and it therefore undergoes $U(1)_Y$ interactions. These five multiplets exist for each of the three generations.

The numbers representing each multiplet correspond to which representation it belongs to. For

instance, we see that the quark multiplet transforms as a triplet under $SU(3)_C$, a doublet under $SU(2)_L$ and has a non-zero hypercharge, corresponding to a non-trivial representation under $U(1)_Y$. That corresponds to saying that quarks interact through all of the three fundamental interactions. From that notation, it is also clear that leptons do not carry color charge, and will only interact via the electroweak interactions.

From the multiplets above, we see that u^c and d^c transform as singlets under $SU(2)_L$, meaning they do not interact. This, however, does not mean they do not feel the electroweak force. The electroweak gauge bosons W and Z are not directly part of $SU(2)_L$, rather, they are a linear combination of $SU(2)_L$ and $U(1)_Y$ and anything with a non-zero weak hypercharge Y will interact with them. This is also true for the photon, which also is a linear combination of $SU(2)_L$ and $U(1)_Y$. These interactions; the strong, weak and electromagnetic, will be explained in more detail in the following sections.

1.2.1 The quantum chromodynamics sector

The group $SU(3)_C$ describes the strong interaction mediated by gluons, and is described by the quantum gauge theory Quantum Chromodynamics (QCD). The group is generated by 8 linearly independent matrices $T^a = \frac{\lambda^a}{2}$, where λ^a are the Gell-Mann matrices [3]. The generator matrices do not commute with one another, but rather satisfy the commutation relation $[\lambda^i/2, \lambda^j/2] = if^{ijk}\lambda_k/2$. This property makes the $SU(3)_C$ group *non-Abelian*, which consequently results in the gluons themselves being charged and displaying self-interactions. Gluons are charged with one unit of color and one unit of anti-color. Quarks, the only fundamental particles interacting with the strong force, are arranged in the simplest representation of $SU(3)$ and come with one unit of color or anti-color.

The generators are collectively referred to as the *group representation* of $SU(3)_C$. Any group element can be written as $e^{-i\theta^a g_a}$, where a runs from 1 to 8, θ^a are real numbers and g_a represent one of the eight linearly independent $\lambda/2$ matrices (generators). Given one representation, one can always find another one through any local gauge transformation that leaves the commutator unchanged. In this case, the group elements are the quark fields, and a local gauge transformation of fields becomes

$$\Psi(x^\mu) \rightarrow e^{-ig_s\theta^a(x^\mu)T^a} \Psi(x^\mu), \quad (1.11)$$

where g_s is the strong coupling, $\theta^a(x^\mu)$ some arbitrary function and a runs over the eight generators of the group. In order to keep the Lagrangian in Equation 1.3 invariant under such a transformation, an additional term must be added, replacing the partial derivative ∂_μ with the

covariant derivative

$$D_\mu = \partial_\mu + ig_s \mathcal{A}_\mu^a T^a, \quad (1.12)$$

introducing a tensor \mathcal{A}_μ^a that represents the 8 gluon fields. The QCD Lagrangian then becomes

$$\mathcal{L}_{QCD} = -\frac{1}{4} \mathcal{F}_a^{\mu\nu} \mathcal{F}_{\mu\nu}^a + \bar{\Psi}(x^\mu) (i\lambda^\mu D_\mu - m) \Psi(x^\mu), \quad (1.13)$$

where $F_{\mu\nu}^a$ is the gauge field of the group, the gluon field tensor

$$F_{\mu\nu}^a = \partial_\mu \mathcal{A}_\nu^a - \partial_\nu \mathcal{A}_\mu^a + g_s f^{abc} \mathcal{A}_\mu^b \mathcal{A}_\nu^c. \quad (1.14)$$

The first term in Equation 1.13 represents the quark-gluon interaction, leading to vertices like the one on the left in Figure 1.1. The second, the gluon field kinetic term, picks up the *structure constant* f^{abc} due to the commutation relation of the λ matrices. This term creates self-interactions between the gluon fields, like the two shown on the right in Figure 1.1. These self-interactions

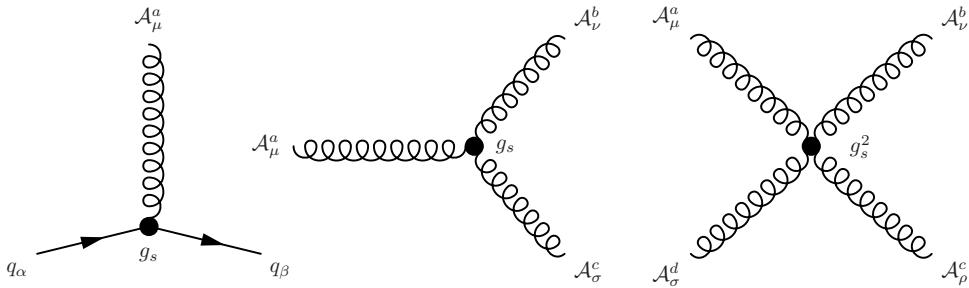


Figure 1.1: The QCD interaction vertices: quark interaction with the gluon field (left), and three-gluon (middle) and four-gluon (right) self-interaction vertices.

have severe consequences: any bare color charge, like a quark, will be surrounded by a sea of virtual gluons and quarks that share the same color. When probing the quark color at higher and higher energies, corresponding to shorter and shorter distances, the color charge decreases until only the bare charge is visible. There, the quarks are essentially free and can be observed as distinguishable particles. This property is referred to as *asymptotic freedom*. For the same reasons, when going further and further away from a bare color charge, the sea of charges surrounding it makes the observed charge increase. That results in a strong attractive force between color charges at large distances, where the potential energy between the two grows linearly with the distance between them as

$$V(r) = -\frac{4\alpha_s}{3r} + kr, \quad (1.15)$$

where r is the distance between the quarks and α_s is the coupling strength of QCD. When the

distance between the quarks grows very large, this potential energy is enough to create real quark-antiquark pairs from the vacuum in order to reduce the potential energy, a process called *fragmentation*. Whenever one tries to separate quarks from one another they will fragment, which consequently means that quarks are never observed on their own. Rather, they form colorless (uncharged under the color charge) bound states of mesons or baryons (collectively called hadrons), a property called *color confinement*. The energy for which the confinement into hadrons occurs, also called *hadronization*, is defined through experimental measurement and found to be $\Lambda_{QCD} = 100 - 500 \text{ MeV}$ (around the mass of the lightest hadrons). The effective charge between the quarks, α_S , changes as a function of energy as

$$\alpha_S(Q) = -\frac{6\pi}{33 - 2n_f} \ln(Q/\Lambda_{QCD}) \quad (1.16)$$

where Q is the energy of the probe used to measure the charge and n_f is the number of quark flavors (u, d, c, s, b, t) at that energy. The value α_S is around 0.1 for energies between 100-1000 GeV.

1.2.2 The electroweak sector

The electromagnetic and weak interactions arise from the breaking of $SU(2)_L \otimes U(1)_Y$ symmetry. While the unification of the electromagnetic and weak force is obtained under the $SU(2)_L \otimes U(1)_Y$ group, the predicted gauge bosons of such a group are not observed in nature: three charged massless vector bosons and one neutral massless boson. Rather, the W^\pm , Z^0 and the photon arise from the spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y$ to $U(1)_{EM}$. This happens due to the *Higgs mechanism*, and exactly how this occurs will be the topic of Section 1.2.3.

The symmetry under $SU(2)_L$ is called weak left-handed isospin L , and the symmetry under $U(1)_Y$ is the weak hypercharge Y . The name “left-handed” arise from the fact that *parity* is violated in the electroweak interactions. All the fundamental fermions have a *chirality*, defined as the projection of the particles spin along its direction of motion. Charged weak interactions are only observed for fermions of left-handed chirality (vector minus axial coupling, V-A). While the left-handed fermion fields are in the simplest doublet representation of $SU(2)$ with weak isospin $I = 1/2$, the fermions of right-handed chirality are therefore in the singlet representation with weak isospin $I = 0$, meaning they do not interact with the gauge bosons of $SU(2)_L$. The chirality of any fermion Ψ can be defined through the operator γ^5 , the product of the four Dirac matrices [4] $\gamma^5 = i\gamma^1\gamma^2\gamma^3\gamma^4$. Any fermion field can be projected into its chiral components Ψ_L or

Ψ_R through the projection operation

$$\Psi_L = \frac{1 - \gamma^5}{2} \quad \text{and} \quad \Psi_R = \frac{1 + \gamma^5}{2}. \quad (1.17)$$

The gauge field tensor of the group of $SU(2)_L$ symmetry is $W_{\mu\nu}^a$, where a runs over the 3 generators of the group. The conserved charge associated with the group is the *third* component of weak isospin I_3 , and all weak interactions must preserve I_3 . The generators of the group are defined as $T_i = \frac{\sigma_i}{2}$, where σ_i are the Pauli matrices [5]. As for $SU(3)_C$, the group is non-abelian and the generators follow the commutation relation $[\sigma_i/2, \sigma_j/2] = i\epsilon_{ijk}\sigma_k/2$, where ϵ_{ijk} is the Levi-Civita permutation symbol [6]. This in turn implies self-interactions between the gauge bosons of the group. The latter group, $U(1)_Y$ of weak hypercharge Y , is abelian and hence displays no self-interaction. The gauge field tensor $B_{\mu\nu}^a$ interacts with both left- and right-handed fermion fields. Similar to the case for QCD, a local gauge transformation of $SU(2)_L \otimes U(1)_Y$ requires the addition of additional terms in the derivative in order to keep the Lagrangian invariant. The partial derivative ∂_μ is replaced by the covariant derivatives

$$D_\mu \Psi_L = (\partial_\mu + ig_2 T_a W_\mu^a + ig_1 \frac{Y}{2} B_\mu^a) \Psi_L \quad (1.18)$$

$$D_\mu \Psi_R = (\partial_\mu + ig_1 \frac{Y}{2} B_\mu^a) \Psi_R. \quad (1.19)$$

After the substitution, the electroweak Lagrangian can be written as a sum of four terms:

$$\mathcal{L}_{EW} = \mathcal{L}_{gauge} + \mathcal{L}_f + \mathcal{L}_{Yukawa} + \mathcal{L}_\phi. \quad (1.20)$$

The first term, \mathcal{L}_{gauge} , represent the kinetic field tensor and is

$$\mathcal{L}_{gauge} = -\frac{1}{4} W_a^{\mu\nu} W_{\mu\nu}^a - \frac{1}{4} \mathcal{B}^{\mu\nu} \mathcal{B}_{\mu\nu} \quad (1.21)$$

where the gauge field tensors are

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g_2 \epsilon^{ijk} W_\mu^j W_\nu^k, \quad \text{with } i = 1 - 3 \text{ and} \quad (1.22)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (1.23)$$

The non-abelian nature of $SU(2)_L$ leads to trilinear and quadrilinear couplings between the photon and vector bosons as illustrated in Figure 1.2.

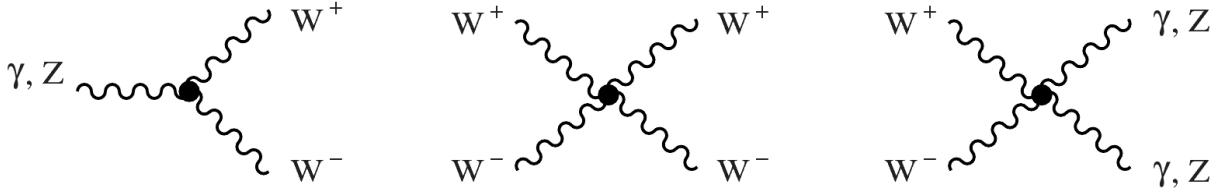


Figure 1.2: The electroweak self-interaction vertices: trilinear (left), and quadri-linear (middle and right) vertices.

The second term describe the boson fields coupling to fermions and is

$$\mathcal{L}_f = \bar{Q}_i i\lambda^\mu D_\mu Q_i + \bar{u}_i^c i\lambda^\mu D_\mu u_i^c + \bar{d}_i^c i\lambda^\mu D_\mu d_i^c + \bar{L}_i i\lambda^\mu D_\mu L_i + \bar{e}_i^c i\lambda^\mu D_\mu e_i^c, \quad (1.24)$$

where we have expressed each term using the particle multiplets in Equation 1.5 and the subscript i runs over the three fermion generations. They all interact differently under $SU(2)_L \times U(1)_Y$ due to their different charges.

Up until now we have considered the Lagrangian before spontaneous symmetry breaking, where we have three charged massless bosons and one massless neutral boson, a constellation we know to be wrong from observations. When I have referred to interaction vertices, I have loosely referred to W , Z and γ vertices without explicitly defining them. We will show in Section 1.2.3 that, due to spontaneous symmetry breaking, the charged gauge boson fields W^\pm in reality are linear combinations of W_μ^1 and W_μ^2 given as

$$W^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \quad (1.25)$$

These are responsible for *charged-current interactions*, which turn up-type fermions into their corresponding down-type fermions within the same generation, and vice-versa. The electrically neutral Z boson and the photon fields are expressed in terms of W_μ^3 and B_μ through the weak mixing angle [7] as

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix} \quad (1.26)$$

where θ_W can be expressed through the $SU(2)_L$ and $U(1)_Y$ couplings g_1 and g_2 as

$$\cos \theta_W = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \quad \text{and} \quad \sin \theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}}. \quad (1.27)$$

The electric charge is defined through weak isospin and hypercharge, and is

$$Q = I_3 + \frac{Y}{2}. \quad (1.28)$$

A few of the fermion-boson vertices are shown in Figure 1.3.

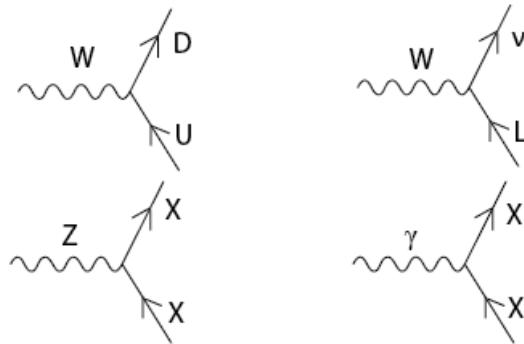


Figure 1.3: The electroweak fermion interaction vertices. Top: charged-current interaction vertex connecting a charged vector boson W^\pm to quarks (left) and leptons (right). Bottom: neutral-current interactions between the neutral Z^0 boson and any fermion (left), and between a γ and electrically charged fermions.

The last two terms, \mathcal{L}_{Yukawa} and \mathcal{L}_ϕ , are related to the Higgs boson and represent couplings to the Higgs field: the Yukawa interaction term which generates masses for the fermions due to the non-zero Higgs vacuum expectation value, and the Higgs bosons interactions with itself and with the gauge bosons. Introduced as an extension to the original Standard Model, the Higgs sector is one of the greatest accomplishments of particle physics in the 20th century and one of the reasons why the LHC was built. How it arises is what we will turn to next.

1.2.3 The Higgs sector

The problem of having massless gauge bosons under $SU(2)_L \times U(1)_Y$, while observing massive W and Z bosons, was independently solved by three different groups and has become known as the Englert–Brout–Higgs–Guralnik–Hagen–Kibble mechanism of spontaneous symmetry breaking [8–10], an accomplishment for which Peter Higgs and Francois Englert shared the 2013 Nobel Prize. It began with the realization that the breaking of a local gauge symmetry could give rise to a final mass for the gauge boson involved. This was first discovered in association with superconductivity, where it was found that when a normal metal becomes superconducting the photon field inside the superconductor would acquire a finite mass [11].

Spontaneous breaking of a global gauge symmetry

In order to achieve spontaneous symmetry breaking, Jeffrey Goldstone [12] suggested introducing a massive complex scalar field ϕ with quantum numbers of the vacuum and then giving the field a vacuum expectation value. The field

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \quad (1.29)$$

would have a Lagrangian of the form

$$\mathcal{L} = \partial^\mu \bar{\phi} \partial_\mu \phi - \mu_0^2 \bar{\phi} \phi - \frac{\lambda_0}{6} (\bar{\phi} \phi)^2, \quad (1.30)$$

where λ_0 is the coupling constant and μ_0 is the mass. The Lagrangian is invariant under $U(1)$, though in this case under a global symmetry and not a local one. If one takes μ_0^2 to be negative, the potential will get a minima along a circle of radius v such that

$$\phi_1^2 + \phi_2^2 = v^2 \quad \text{and} \quad v^2 = \frac{\mu_0^2}{\lambda_0}, \quad (1.31)$$

and take the form of a “Mexican Hat” as shown in Figure 1.4. The variable v is referred to as the

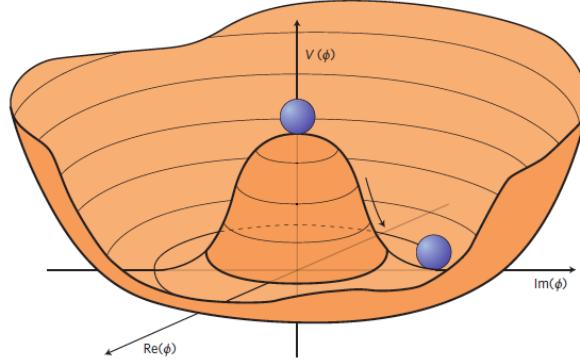


Figure 1.4: The potential $V(\phi)$ for a complex scalar field with $\mu_0^2 < 0$ [13].

vacuum expectation value (VEV). The lowest value of the Hamiltonian is now at $\phi = v$ rather than at $\phi = 0$. Goldstone then translated the field to a minimum energy position $\phi' = \phi + v$ and gave the field a vacuum expectation value, effectively breaking the symmetry between the two field components, but keeping the Lagrangian invariant. This complex field can be expanded around the ground state in terms of two real fields η and ϵ that represent deviations from the

minimum:

$$\phi(x) = \frac{1}{\sqrt{2}} * (v + \eta(x) + i\epsilon(x)) \quad (1.32)$$

The Lagrangian then becomes

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \epsilon)^2 + \frac{1}{2}(\partial_\mu \eta)^2 + \mu_0^2 \eta^2 + \text{additional terms.} \quad (1.33)$$

The third term has the form of a mass term for the scalar η field. However, the ϵ field has no mass term, meaning that the theory contains a massless scalar, referred to as a *Goldstone boson*. This is expressed through *Goldstone's theorem*, which states that whenever a continuous symmetry of a physical system is spontaneously broken, massless scalars will occur. Rather than providing a mass term for the vector bosons, the theory acquired one massive scalar and one massless scalar not observed in nature, requiring the gauge theory of weak interactions to look for solutions elsewhere.

The Higgs mechanism

The solution to the problem came a few years later, in 1964, when spontaneous symmetry breaking of a *local* gauge symmetry was studied rather than a global. For a $U(1)$ symmetry, this requires the Lagrangian to be invariant under $\phi \rightarrow e^{i\theta(x)}\phi$, with the derivative replacement

$$D_\mu = \partial_\mu - ieA_\mu. \quad (1.34)$$

After translating the field ϕ to its true ground state and writing out the Lagrangian,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \epsilon)^2 + \frac{1}{2}(\partial_\mu \eta)^2 - v^2 \lambda \eta^2 + \frac{1}{2}v^2 e^2 A_\mu A^\mu - ev, A_\mu \partial^\mu \epsilon - \frac{1}{4}F_a^{\mu\nu} F_{\mu\nu}^a \quad (1.35)$$

we see that the particle spectrum now contains a massless Goldstone boson ϵ , a massive scalar η and, finally, a massive vector A_μ with $m_A = ev$. Despite the success of having dynamically created the mass for the gauge field, one had to tackle the problem of the massless scalar. The solution was found through the realization that one of the fields was unphysical; by giving mass to the vector A_μ , the polarization degrees of freedom had increased from 2 to 3 through adding a longitudinal polarization. However, this should not be possible when simply translating field variables. It was found that through a simple gauge transformation with a different set of fields

$$A_\mu \rightarrow A_\mu + \frac{1}{ev} \partial_\mu \theta, \quad (1.36)$$

the Goldstone boson would disappear and turn into the longitudinal polarization of the massive gauge boson and that the theory was left with one massive vector gauge boson, A_μ , and another massive scalar, h . This is what is referred to as the *Higgs mechanism*.

The Weinberg-Salam Model

The final step is to formulate the Higgs mechanism such that the vector bosons W^\pm and Z^0 become massive, while the photon remains massless. To do so Sheldon Glashow, Abdus Salam, and Steven Weinberg (all awarded the 1979 Nobel Prize for electroweak unification), added a gauge-invariant term to the electroweak Lagrangian of the following form:

$$\mathcal{L}_{Higgs} = \left| (i\partial_\mu - g_2 T_a W_\mu^a - v g_1 \frac{Y}{2} B_\mu^a) \phi \right|^2 - V(\phi). \quad (1.37)$$

The ϕ_i has to belong to $SU(2) \times U(1)$ multiplets, and the simplest configuration corresponds to four fields in an isospin doublet of weak hypercharge $Y = 1$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \\ \phi^- \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \frac{\phi_1 + i\phi_2}{\sqrt{2}} \\ \frac{\phi_3 + i\phi_4}{\sqrt{2}} \end{pmatrix}. \quad (1.38)$$

To generate the necessary masses, the Higgs potential from the previous section is used, with a vacuum expectation value of

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (1.39)$$

This specific choice of charges, VEV and fields insure that the $U(1)_{em}$ symmetry with $Q = T^3 + \frac{Y}{2}$ remains unbroken, keeping the photon massless. The three others break the symmetry and become massive gauge bosons: the W^+ , W^- and Z^0 . The mass terms for the gauge bosons finally become

$$M_W = \frac{1}{2} v g_1 \quad \text{and} \quad M_Z = \frac{1}{2} v \sqrt{g_1^2 + g_2^2}, \quad (1.40)$$

with the Weinberg angle θ_W defined as

$$\text{with } \frac{M_W}{M_Z} = \cos \theta_W. \quad (1.41)$$

As mentioned in Section 1.2.2, the fermions also get their masses through interacting with the Higgs field, corresponding to the \mathcal{L}_{Yukawa} term in the electroweak Lagrangian in Equation 1.20. This is done in the same way as for the bosons: an additional $SU(2) \times U(1)$ invariant term for

each generation is added, for instance for the electron:

$$-G_e \left[\begin{pmatrix} \bar{\nu}_e & \bar{e} \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e^R + \bar{e}_R \begin{pmatrix} \phi^- & \bar{\phi}^0 \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \right], \quad (1.42)$$

where G_E is the electron coupling. We then spontaneously break the symmetry with

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \quad (1.43)$$

where the neutral Higgs field $h(x)$ is the only remnant of the Higgs doublet after spontaneous symmetry breaking. After substitution, the final Lagrangian for the electron mass becomes

$$\mathcal{L}_{Yukawa}^e = -\frac{G_e}{\sqrt{2}} v (\bar{e}_L e_R + \bar{e}_R e_L) \left(1 + \frac{h}{v}\right). \quad (1.44)$$

We can choose G_e such that

$$m_e = -\frac{G_e v}{\sqrt{2}}, \quad (1.45)$$

and generate the electron mass as

$$\mathcal{L}_{Yukawa}^e = -m_e \bar{e} e - \frac{m_e}{v} \bar{e} e h. \quad (1.46)$$

In summary, all the fermion masses are generated through the interaction of the fermion fields with the Higgs field. From the equation above, we see that the fermion masses are not predicted from the theory as the coupling G is arbitrary. The Standard Model therefore cannot provide an explanation for the difference in hierarchy between the couplings. We also see that the Lagrangian contains an interaction term coupling the Higgs scalar to the fermions and that this term depends on the mass of the fermion. The Higgs boson therefore couples more strongly to heavy fermions than to lighter ones.

CHAPTER 2

Beyond Standard Model Physics

2.1 Shortcomings of the Standard Model

Despite being an extremely successful and predictive theory, the Standard Model has its shortcomings. The most obvious one is its failure to successfully unite with the gravitational force at very high energies (or, correspondingly, short distances). Gravity is beautifully described in General Relativity (GR) as a classical theory: a force caused by the curvature of space-time in the presence of matter and energy. The theory does not utilize quantum fields and energy is not quantized. The scales between the Standard Model, a quantum field theory (QFT), and General Relativity are completely different: space-time is curved on astronomical scales, where the force of gravity is measurable, while quantum field theories deal with things on the smallest possible scales, where variations in space-time are essentially invisible. Hence, to the Standard Model, space-time is approximately flat and gravity does not exist. In order to have an elegant unified theory of all the forces, attempts have been made to have a quantum field theory of the gravitational force by extending the Standard Model particle family to incorporate a particle to mediate the gravitational force called the *graviton*, a massless gauge boson of spin-2. The problem is that gravity is universally attractive, meaning nothing “cancels” it. That leads to loop divergences that cannot be reabsorbed through renormalization and every effort of integrating gravity in the SM in a renormalizable way has thus far failed. However, it has been shown that the Standard Model and General Relativity are fully compatible in the low-curvature and low-energy regime, and that GR is an inevitable consequence of the quantum mechanics of interacting gravitons. The full non-linear structure of GR can be derived through QFT, yielding graviton couplings to the SM which are uniquely determined. This has led to several proposals for extending the SM in order to incorporate the force mediating gravitons.

In addition to the difficulties of incorporating gravity into a quantum field theory framework, problems occur at small distances at which quantum gravitational effects would become apparent, the Planck scale. This can be represented by the Planck mass, the mass of the smallest possible black hole. When comparing the Planck mass to the masses of the electroweak bosons W and Z, we find that the Planck mass is 10^{16} times heavier than the electroweak bosons, such that there is a *hierarchy* between the mass scales of gravity and the electroweak forces. The reason why this observed hierarchy occurs has to do with the Higgs vacuum expectation value (VEV): the Higgs field has a measured vacuum expectation value of 246 GeV and this is, as discussed in the section above, what gives the W and Z bosons their mass. However, when actually calculating the Higgs VEV and taking all loop corrections into account, it would receive corrections on the order of the Planck energy, yielding a Higgs boson mass 10^{16} times larger than observed. This is called the *hierarchy problem*. Quantum loop corrections of this magnitude only happen for scalar particles such as the Higgs boson. Fermions are protected from such divergences through their chiral structure, and gauge bosons are protected through gauge invariance. The question is then why the Higgs VEV, and consequently the Higgs, W and Z boson masses are so much smaller than the natural mass scale.

Of course, it is possible that the Higgs boson mass just happens to be 125 GeV due to some fine-tuned, large cancellations that keeps the mass from approaching the Planck mass, as is currently held by the Standard Model. However, this is neither very elegant nor very probable without a well-motivated reason why such a cancellation should occur. Rather, in order to resolve the problem of scales, theories Beyond the Standard Model (BSM) have been introduced. The theories that I will probe in this thesis are amongst those.

Two Beyond Standard Model theories will be considered in this thesis: Compositeness and extra dimensional theories. Compositeness attacks the hierarchy problem by assuming that the Standard Model breaks down at an energy between the weak and Planck scales and that, around the TeV scale, the Higgs boson no longer appears to be a single scalar particle but a composite state of two fermions. In the following, I will present the study of composite models in the context of the *Heavy Vector Triplet formalism*, described in Section 2.2.1. Warped extra dimensional theories on the other hand, attempt to solve the hierarchy problem by forcing gravity to be concentrated on another “brane” and letting its strength fall off exponentially through an extra dimension.

2.2 Compositeness

In composite Higgs models, the Higgs boson is assumed to be a bound state of fundamental constituents held together by some new strong force [14, 15]. This removes the hierarchy problem since we no longer have an elementary scalar in the Standard Model, hence no loop corrections going to infinity. The compositeness of the Higgs boson becomes apparent at the energy scale Λ , where Λ has to be at least 10 TeV, since anything below that is ruled out by electroweak precision measurements. The Higgs boson is assumed to be a pseudo-Goldstone boson of some approximate symmetry, where pseudo-Goldstone bosons are bosons with a tiny mass that approach zero in the limit of the symmetry becoming exact. The approximate symmetry is broken at the scale f , where $\Lambda = 4\pi f$. Being a pseudo-Goldstone boson, the Higgs boson mass is protected from divergent quantum loop corrections up to the scale of compositeness and, above that scale, is no longer an elementary scalar. The theory is based on the breaking of two parallel global symmetries $[SU(2)_1 \times U(1)_1] \times [SU(2)_2 \times U(1)_2]$, with Goldstone bosons becoming the longitudinal components of the three predicted gauge bosons of the symmetry group W'^{\pm} and Z' . These have masses of the order of the compositeness scale

$$M(W'^{\pm}) \simeq M(Z') = \frac{g}{\sin 2\theta} f, \quad (2.1)$$

where $\tan \theta = g_1/g_2$, the ratio of couplings of the $SU(2)$ groups. The predicted decay widths are roughly the same for Z' and W' and are as follows:

$$\begin{aligned} \Gamma(W'^{\pm} \rightarrow \ell\nu, Z' \rightarrow \ell\ell) &= \frac{g^2 \cot^2 \theta}{96\pi} M \\ \Gamma(W'^{\pm} \rightarrow q\bar{q}', Z' \rightarrow q\bar{q}) &= \frac{g^2 \cot^2 \theta}{32\pi} M \\ \Gamma(W'^{\pm} \rightarrow WZ, Z' \rightarrow WW) &= \frac{g^2 \cot^2 2\theta}{192\pi} M. \end{aligned} \quad (2.2)$$

Decays into fermions therefore dominate at $\cot \theta \geq 1/2$, whereas decays into bosons are enhanced for very low $\cot \theta$. These generic composite models can be studied with the Heavy Vector Triplet formalism.

2.2.1 Heavy Vector Triplet formalism

There are many BSM theories that predict the presence of spin-1 particles with masses at the TeV scale, each with their own list of model parameters. In most cases, however, when looking for such new particles, experiments are not sensitive to the specifics of the model but only the masses

and couplings of the resonances. We can therefore start from a *simplified model* that describes the dynamics of the new spin-1 vector through a simple phenomenological Lagrangian that only retains couplings and mass. In the Heavy Vector Triplet formalism [16], a real vector V_μ^a , where r runs from 1 to 3, is introduced in the adjoint representation of $SU(2)L$ and represents one charged and one neutral heavy spin-one particle with charge eigenstates

$$V_\mu^\pm = \frac{V_\mu^1 \mp i V_\mu^2}{\sqrt{2}} \quad \text{and} \quad V_\mu^0 = V_\mu^3. \quad (2.3)$$

The simplified Lagrangian governing the dynamics is given as

$$\begin{aligned} \mathcal{L}_V = & -\frac{1}{4} \mathcal{D}_{[\mu} V_{\nu]}^a \mathcal{D}^{[\mu} V^{\nu]a} + \frac{m_V^2}{2} V_\mu^a V^{\mu a} \\ & + ig_V c_H V_\mu^a H^\dagger \tau^a \overleftrightarrow{\mathcal{D}}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a J_F^{\mu a} \\ & + \text{additional terms.} \end{aligned} \quad (2.4)$$

The first line describes the kinematic and mass terms of the vector V , and the second line, which is of most interest to us, describes the coupling to the Higgs boson current and the left-handed fermionic currents. In the coupling to the Higgs current, the coefficient c_H leads to vertices involving the Higgs field and the Goldstone bosons, representing the longitudinally polarized SM vector bosons, W and Z. This term therefore governs the decay modes of the vector V into electroweak bosons, the decay mode of interest for this thesis. The second coupling term describes the interaction with leptons and quarks and is governed by the parameter c_F . A formalism is adopted where the interactions are weighted with a coupling g_V and g^2/g_V , where g is the gauge coupling of the group and g_V represents the “typical strength” of the vector interactions. Another interesting feature of the theory is that, after electroweak symmetry breaking provides the heavy vector with its mass, the charged and neutral vectors are found to be mass degenerate and expected to have similar production and decay rates.

After having defined the generic framework, explicit models with fixed values of c_H and c_f can be studied, where only the resonance mass m_V and coupling g_V are left as free parameters. In this thesis, we probe two benchmark models called HVT model A and HVT model B, as introduced in [16]. The reason why these two models are interesting is that the first probes rather weakly coupled extensions of the SM, and the latter, strongly coupled scenarios. That implies very different sizes of g_V , where values of $g_V = 1$ for model A and $g_V = 3$ for model B are used in [16]. For these values of g_V , model A predicts a comparable branching fraction for decays into bosons and fermions, the decay into fermions only enhanced by a factor of 2, while for model B, the

dominant branching fraction is to dibosons with decays into fermions severely suppressed. The branching fraction for the different decay modes of HVT model A and B, are shown in Figure 2.1. For obvious reasons, model B is of most interest for the searches presented here.

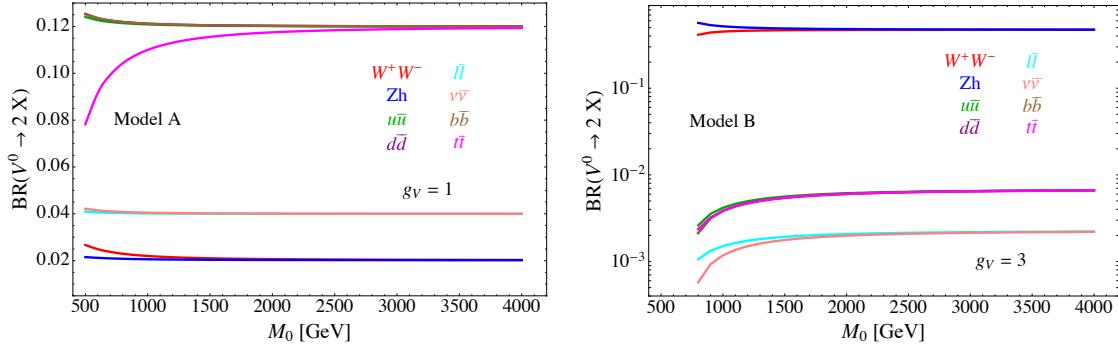


Figure 2.1: Predicted branching fractions of a Z' for two explicit HVT models: Model $A_{g_V=1}$ (left) and model $B_{g_V=3}$ (right) [16].

2.3 Warped extra dimensions

Extra dimensional theories also offer solutions to the hierarchy problem. This thesis focuses on Randall-Sundrum (RS) warped extra dimensional scenarios [17]. In RS models, a new curved spatial dimension y is proposed, leading to a 5-dimensional space-time bounded by two (3+1)-dimensional planes, or *branes*: the UV/Planck and the IR/TeV brane. The new metric now depends on the radius r and the curvature factor k of the new extra dimension, with

$$ds^2 = e^{-2ky} \eta_{\mu\nu} dx^\mu dx^\nu + dy^2; 0 < y < \pi R. \quad (2.5)$$

Gravity is concentrated and relatively strong at the Planck brane at $y = 0$, which is separated from us by the fifth dimension. Our observed four-dimensional reality and the Standard Model particles reside at the TeV brane, at $y = \pi R$. Only gravity, transmitted through gravitons, is allowed to propagate through the warped 5D space-time (the “bulk”) and is not confined to either brane. Figure 2.2 illustrates how the branes and the bulk are connected. Due to the

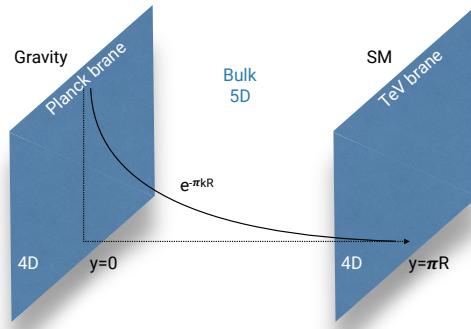


Figure 2.2: The RS model predicts an extra dimension where a 5D space-time stretches between two 4D branes: the Planck brane where gravity is concentrated, and the TeV brane where the SM particles are confined.

warping, the Planck mass on the Planck brane gets reduced by a factor of $e^{-\pi k R}$ at the TeV brane, thereby solving the hierarchy problem. The Planck mass on the TeV brane, which depends on the geometry of the extra dimension, becomes

$$\bar{M}_{Pl}^2 = V_1 M_*^3, \quad (2.6)$$

where V_1 is the volume of the 1 dimensional added warped dimension and M_*^3 is the 5D Planck mass. One distinct prediction of the model, and a way in which we can test its validity, is the prediction of a tower of TeV-scale excitations with spin-2, so called Kaluza-Klein states, that could be observed in high energy experiments.

In this thesis, we are more interested in an alternative to the original RS model called the “bulk” scenario [18, 19]. In this case, the Standard Model particles, besides the Higgs boson, are also allowed to propagate in the bulk. The light 1st and 2nd generation fermions are localized near the Planck brane, yielding small couplings to the Higgs boson that still resides at the TeV brane, explaining their small masses. Similarly, the top quark is now located near the TeV brane, resulting in a stronger Yukawa coupling to the Higgs boson. In addition, with the gravitons located near the TeV brane and the fermions now residing near the Planck brane, the graviton coupling to fermions is strongly suppressed. SM gluons have a flat distribution throughout the bulk, making gluon-gluon production the dominant production channel of gravitons. Due to the weak vector bosons absorbing the Higgs degree of freedom in spontaneous symmetry breaking, their wave-functions fall off steeply near the TeV brane, resulting in a coupling to the gravitons similar to that of the Higgs and the top. The only free parameters of the theory is the mass of the lightest KK graviton and the ratio $\tilde{k} = \frac{k}{M_{Pl}}$, which controls the widths and cross sections of the new resonances. The branching ratios of the Bulk Graviton is shown in Figure 2.3.

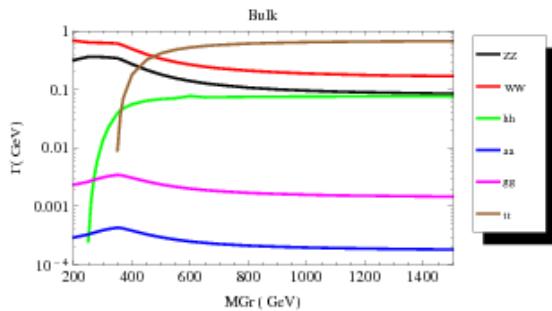


Figure 2.3: Predicted branching fractions for a Bulk Graviton as a function of its mass.

Part II

Experimental setup

CHAPTER 3

The Large Hadron Collider

In March 1984, the European Organization for Nuclear Research (CERN) and the European Committee for Future Accelerators (ECFA) held a workshop in Lausanne entitled “Large Hadron Collider in the LEP Tunnel”. This is history’s first written mention of the Large Hadron Collider (LHC) and the topic under discussion was exactly how to build a new type of high-energy collider, capable of bringing hadrons to collide rather than leptons. The LHC would be housed in a tunnel which, at the time, was under excavation to host the Large Electron-Positron Collider (LEP) designed to collide leptons with center-of-mass-energies up to around 200 GeV. LEP was a circular collider with a circumference of 27 km and the tunnel hosting it was located roughly 100 meters underground beneath France and Switzerland, at the outskirts of Geneva. The justification for building a machine like the LHC was that, once LEP got to its maximum center-of-mass energy, a new and more powerful collider would be needed in its place in order to probe higher energies. While collisions of electrons with positrons provide exceptionally clean and precise measurements due to their being point particles, their lightness prevent them from being accelerated to higher energies in circular colliders due to synchrotron radiation. Collisions of hadrons, however, would allow for center-of-mass energies two orders of magnitude higher than that of LEP. Therefore, after obtaining sufficient statistics when running at a center-of-mass-energy of twice the W boson mass (160 GeV) and reaching a maximum center-of-mass energy of 209 GeV, in a search for the Higgs boson, LEP was dismantled in 2000 in order to make room for the LHC.

The Large Hadron Collider first circulated protons in September 2008 and, while having the same 27-kilometer radius as the LEP collider, is capable of accelerating protons up to a center-of-mass energy of around 14 TeV, 70 times that of LEP. The accelerator consists of two oppositely circulating proton beams, isolated from each other and under ultrahigh vacuum. The protons are accelerated up to speeds close to the speed of light through radio frequency (RF) cavities, before being focused to collide at four different interaction points along the ring. These

four collision points correspond to the location of the four LHC particle detectors; ATLAS, CMS, LHCb and ALICE. While ATLAS and CMS are general-purpose detectors, built in order to study a large range of different physics processes, LHCb and ALICE are built for dedicated purposes; LHCb for b-physics processes and ALICE for heavy ion collisions. The journey of a proton from a gas to one of the LHC collision points is as follows. First, hydrogen nuclei are extracted from a small tank of compressed hydrogen gas and stripped of their electrons. The remaining protons are then injected into the LINAC2, a linear accelerator responsible for increasing the proton energy to about 50 MeV through RF cavities that push charged particles forward by switching between positive and negative electric fields. The LINAC2 additionally divides the constant stream of particles into equally spaced “bunches” by careful tuning of the frequency of the field switch. The accelerated protons are then injected into the Proton Synchrotron Booster (PSB), where their energy is increased by thirty fold, to an energy of roughly 1.4 GeV. The two final acceleration stages before the protons reach the LHC ring are the Proton Synchrotron and Super Proton Synchrotron, eventually producing protons with a total energy of 450 GeV. The protons are now ready for the final stage of their travel and are injected into the two beam pipes of the LHC in opposite directions. They are injected in trains of 144 bunches each (on order of 10^{11} protons per bunch), where each bunch is roughly 7.5 meters apart (or 25 ns). There are some larger beam gaps present in each beam in order to give special magnets sufficient time to switch on in order to inject or dump the beam. The largest beam abort gap is roughly 3 ms or 900 m long. The ring is filled with proton bunches until these are equally distributed throughout the two rings, a process taking roughly 4 minutes. This is called a “fill”. Here, the protons are accelerated to their current maximum energy of 6.5 TeV, a process taking roughly 20 minutes, through eight RF cavities. These RF cavities are also responsible for keeping the proton bunches tightly bunched. A complete sketch of the CERN accelerator complex is shown in Figure 3.1. After the beams have reached their maximum energy and are stably circulating in the LHC ring, the bunches are brought to collide. The goal of such a collision, which occurs every 25 nanoseconds, is that some of the protons in each bunch will inelastically collide, allowing the quark and gluon constituents of each proton to interact with one another and produce new and interesting particles. The number of times such an interaction will take place inside a detector per area and time is quantified through the instantaneous luminosity \mathcal{L} , which is the proportionality factor between the number of observable events per second and the cross section σ of the process you are interested in,

$$\frac{dN_{events}}{dt} = \mathcal{L}\sigma. \quad (3.1)$$

The cross section is the probability that an event (like one which would produce new and interesting

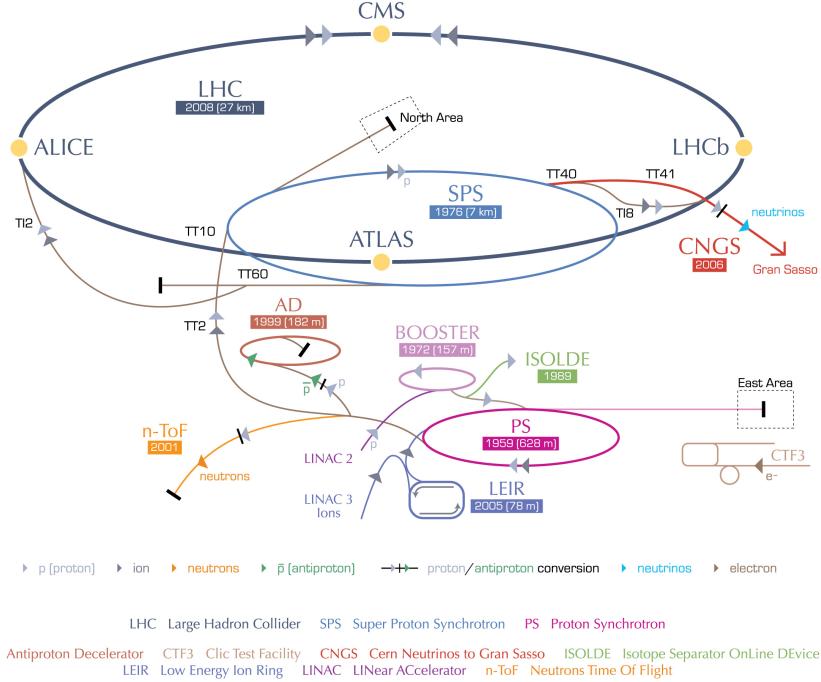


Figure 3.1: The Large Hadron Collider accelerator complex. The four collision points along the ring correspond to the location of the LHC particle detectors CMS, LHCb, ATLAS and ALICE [20].

particles) will occur and is measured in barns, where $1 \text{ barn} = 10^{-28} \text{ m}^2$. This luminosity should therefore be as high as possible. It depends only on parameters of the accelerator beams and can, in the case of the LHC, be defined through accelerator quantities as

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta_*} F, \quad (3.2)$$

where N_b is the number of particles per bunch, n_b is the number of bunches, f_{rev} is their revolution frequency, γ_r is the relativistic gamma factor, ϵ_n is the normalized transverse beam emittance (how confined the particles are in space and momentum), β_* is the beta function at the collision point (how narrow, or “squeezed”, the beam is) and F is a reduction factor to account for the case where the beams do not collide head-on but at slight crossing angles. From this, it becomes clear that the main goal of the LHC is to maximize the number of particles (N_b, n_b),

Parameter	Units	Nominal	2015	2016	2017
Energy	[TeV]	7.0	6.5	6.5	6.5
Bunch spacing	[ns]	25	25	25	25
Bunch intensity	$\times 10^{11}$ [protons/bunch]	1.15	1.15	1.15	1.2-1.45
Bunches per train		144	144	96	144
Total number of bunches		2808	2244	2220	2556
β^*	[cm]	55	80	40	27/25
Peak luminosity	$\times 10^{34}$ [cm $^{-2}$ s $^{-1}$]	1.0	0.5	1.4	2.0
Integrated luminosity			4.2	39.7	50.2

Table 3.1: Some key LHC detector parameters achieved during the first years of data taking with a center-of-mass energy of 13 TeV.

their frequency (f_{rev}) and their energy (γ_r), while at the same time ensuring the protons are packed together as tightly as possible (lower ϵ_n and β^*). Using the nominal values for the LHC, the peak instantaneous luminosity is roughly $\mathcal{L} \sim 10^{34}\text{cm}^{-2}\text{s}^{-1}$.

The peak luminosity of the LHC by the end of Run 2 in 2018 was about $2.0 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$, corresponding to 2 times the nominal design instantaneous luminosity.

To quantify the size and statistical power of a given LHC dataset, the integrated luminosity is used. This is the integral of the instantaneous luminosity over time and is defined as

$$\mathcal{L}_{int} = \int \mathcal{L} dt. \quad (3.3)$$

It is usually defined in units of inverse cross section, b^{-1} .

Despite the LHC starting up in 2008, there would be another year before data taking began due to technical difficulties with the magnets. In March 2010, the LHC saw its first collision with a center-of-mass energy of 7 TeV, and continued running at this energy, collecting around 5 inverse femtobarns of data by the end of 2011. In 2012, the energy was increased to 8 TeV and the LHC continued running until a planned long shutdown scheduled to begin in February 2013, collecting a total of $\sim 20 \text{ fb}^{-1}$, which allowed the Higgs boson to be discovered. This marked the end of Run 1 and the beginning of a two-year maintenance project intended to prepare the LHC for running at a center-of-mass energy of 13 TeV, a period referred to as Run 2.

Run 2 started in June 2015, and provides the dataset used in thesis. With the accelerator now running at 90% of its nominal energy, and with a peak luminosity between 1-2 times the design luminosity, the LHC managed to collect an impressive $\sim 160 \text{ fb}^{-1}$ with center-of-mass energy of 13 TeV up until its planned shutdown at the end of 2018. Some key LHC accelerator parameters that were in use for the datasets analyzed in this thesis are quoted in Table 3.1.

CHAPTER 4

The CMS detector

The Compact Muon Solenoid (CMS) detector is true to its name. With a diameter of 15 meters and a weight of 14000 tons, it is 60% smaller but twice as heavy as its counterpart, the ATLAS detector. Its large weight is due to its solenoid: a superconducting niobium titanium coil circulating 18500 Amps and capable of generating a magnetic field of 3.8 Tesla, the largest superconducting magnet ever built. Together with its corresponding iron return yoke, responsible for returning the escaping magnetic flux, it accounts for 90% of the total detector weight. The CMS detector is cylindrically symmetric and organized such that the inner tracking system begins at a radius of around 3 cm from the beam pipe. It consists of an inner silicon pixel detector and an outer silicon strip tracker, stretching out to a radius of roughly 1.2 meters. Following the tracker are two calorimeter layers: the electromagnetic calorimeter (ECAL) consisting of lead tungstate scintillating crystals and responsible for measuring the energy of electromagnetically interacting particles, followed by the hadronic calorimeter (HCAL) that measures the energy of hadrons. Contrary to “standard” configurations for general purpose detectors, the CMS calorimeters are located inside the superconducting solenoid. This allows the detector to be rather compact, by reducing the necessary radii of the calorimeters, and results in a strong magnetic field due to the large coil radius. The muon detectors are alternated with three layers of steel return yoke, responsible for containing and returning the magnetic field. Only muons and weakly interacting particles are expected to transverse the full detector volume without being stopped. Since the muons bend in the magnetic field of the return yoke, an additional momentum measurement can be made. A schematic overview of the CMS detector is shown in Figure 4.1. In the following, the different sub-detectors will be described in detail.

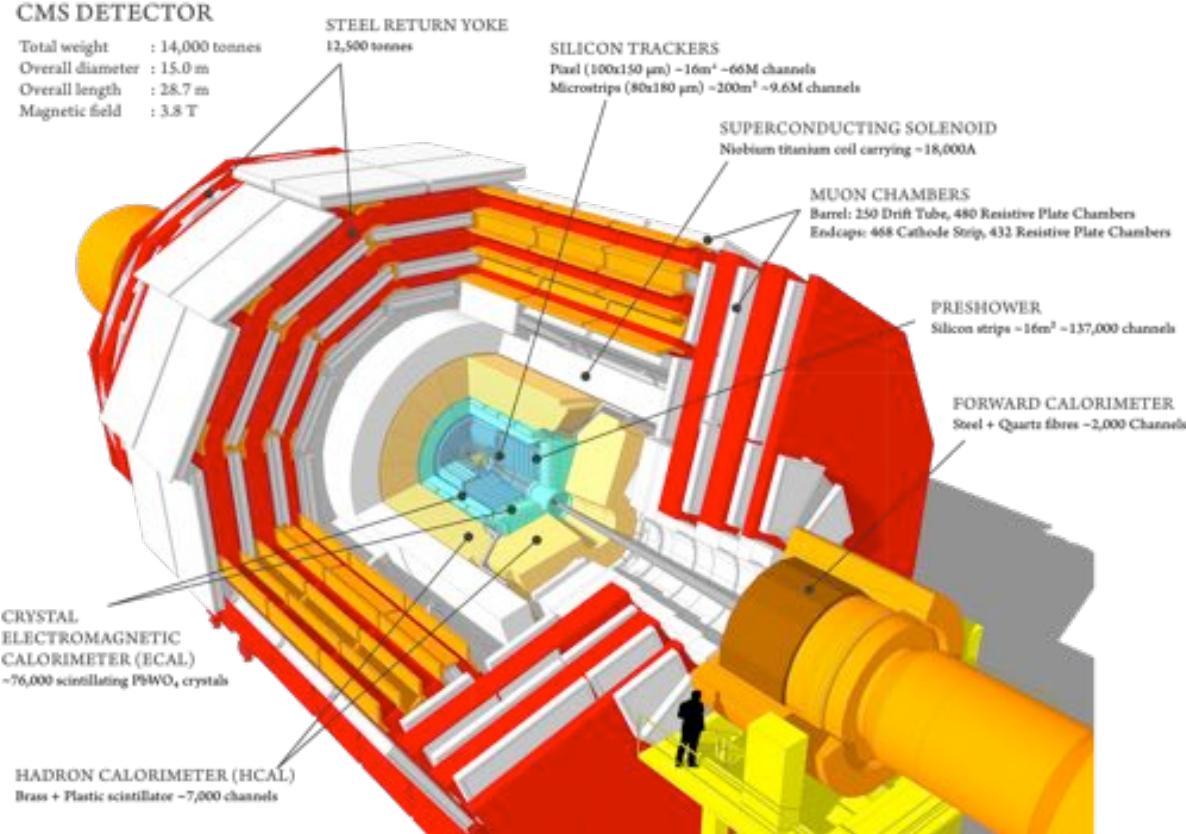


Figure 4.1: The CMS detector and its subsystems: the silicon tracker, electromagnetic and hadron calorimeters, superconducting solenoid and the muon chambers inter-layered with the steel return yoke [21].

4.1 Coordinate system

To describe locations within the CMS detector, a Euclidian space coordinate system is used. Here, the positive z axis points along the beam pipe towards the west, the positive x axis points towards the center of the LHC ring, and the positive y axis up towards the surface of the earth. Due to the cylindrical symmetry of the detector, polar coordinates are more convenient and most frequently encountered. In this scheme, the azimuthal angle ϕ is measured in the xy-plane, where $\phi = 0$ correspond to the positive x axis and $\phi = \pi/2$ correspond to the positive y axis. The polar angle θ is measured with respect to the z axis, $\theta = 0$ aligning with the positive and $\theta = \pi$ with the negative z axis. To define a particles' angle with respect to the beam line, the pseudorapidity $\eta = -\ln \tan(\theta/2)$ is preferred over θ . This is due to the fact that particle production is approximately constant as a function of pseudorapidity and, more importantly,

because differences in pseudorapidity are Lorentz invariant under boosts along the z-axis when assuming massless particles. To measure angular difference between particles in the detector, the variable $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is used, which is also Lorentz invariant under longitudinal boosts. A summary of the CMS coordinate system together with some example values are shown in Figure 4.2.

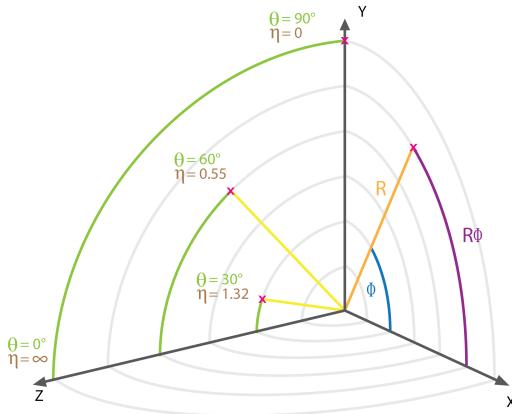


Figure 4.2: The CMS coordinate system [22], shown with some values of θ and η .

4.2 Tracking detectors

The CMS tracker is responsible for accurately reconstructing the momentum of charged particles and consists of two sub-detectors. Closest to the interaction point, and where the particle intensity is the highest, the silicon pixel detector is located. Upgraded in 2017, from the so-called Phase-0 to the Phase-1 detector, it is structured with four cylindrical barrel layers at radii 2.9, 6.8, 10.9 and 16.0 cm (the barrel pixel detector) and three disks in each of the forward regions placed at a distance from the nominal interaction point of 29.1, 39.6 and 51.6 cm (the forward pixel detector). A sketch of the current Phase-1 pixel detector compared to the Phase-0 detector is shown in Figure 4.3. The sensors located closest to the beam pipe are subject to hit intensities of $\mathcal{O}(\text{MHz/mm}^2)$ such that strict constraints on the sensor pixel size are required in order to minimize occupancy in the detector. The sensor pixels are $100 \mu\text{m} \times 150 \mu\text{m}$ with a thickness of $285 \mu\text{m}$, and when counting both barrel and forward pixel detectors, sum up to a total of 124 million channels. The sensors are mounted on detector modules each with 16 read-out chips, where the type of readout chip depends on how close the module is to the beam pipe. The inner layer uses readout chips with a rate capability of 600 MHz/cm^2 , while for the outer layers,

readout chips with a rate capability of up to $200 \text{ MHz}/\text{cm}^2$ are sufficient. Since the hit intensity

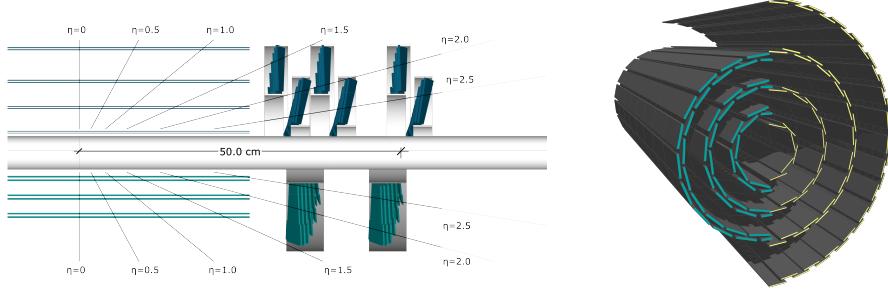


Figure 4.3: Left: The pixel detector layout before (bottom) and after (top) the Phase-1 upgrade. Right: The barrel pixel detector before (left) and after (right) the Phase-1 upgrade [23].

reduces as you go further away from the beam pipe, the pixel sensors are replaced by silicon strip sensors of larger size, making up the second of the two tracker sub-systems, the silicon strip tracker. There are ten strip layers in total, stretching out to a radius of roughly 130 cm. These are divided into four sections: The inner barrel (TIB) with four strip layers, the two inner endcaps (TID) consisting of three disks each, the outer barrel (TOB) consisting of 6 cylindrical layers, and the two endcaps (TEC) with 9 strip layers each. A schematic overview of the strip tracker layout is shown in Figure 4.4. The strips in the TIB and TID are 10 cm long, with a width of 80 μm and a thickness of 320 μm . The TOB and TEC sections consist of slightly larger strips of 25 cm x 180 μm and a thickness of 500 μm . The strip tracker has a total of 10 million detector strips and covers an area of $\sim 200 \text{ m}^2$. To prolong the silicon detector lifetime, the entire tracker (pixel and strip) is kept at a temperature of -20°C with a dedicated cooling system. The tracker has a coverage of up to $|\eta| < 2.6$ and a resolution of roughly $\sigma/p_T \approx 1.5 \times 10^{-5} p_T + 0.005$.

4.3 Electromagnetic calorimeter

Surrounding the tracking detectors is the electromagnetic crystal calorimeter (ECAL). Consisting of 75 848 laterally segmented scintillating lead tungstate (PbWO_4) crystals, it was designed to have the best possible photon energy and position resolution in order to resolve a Higgs boson decaying into two photons, one of the cleanest discovery channels of the Higgs boson. With a design energy resolution of 0.5% above 100 GeV for photons and electrons, the choice of detector material for the ECAL has been its most crucial design feature. In order to withstand the high doses of radiation and the high magnetic field present within the detector, while at the same time generating well-defined signal responses within the 25 nanoseconds between particle collisions,

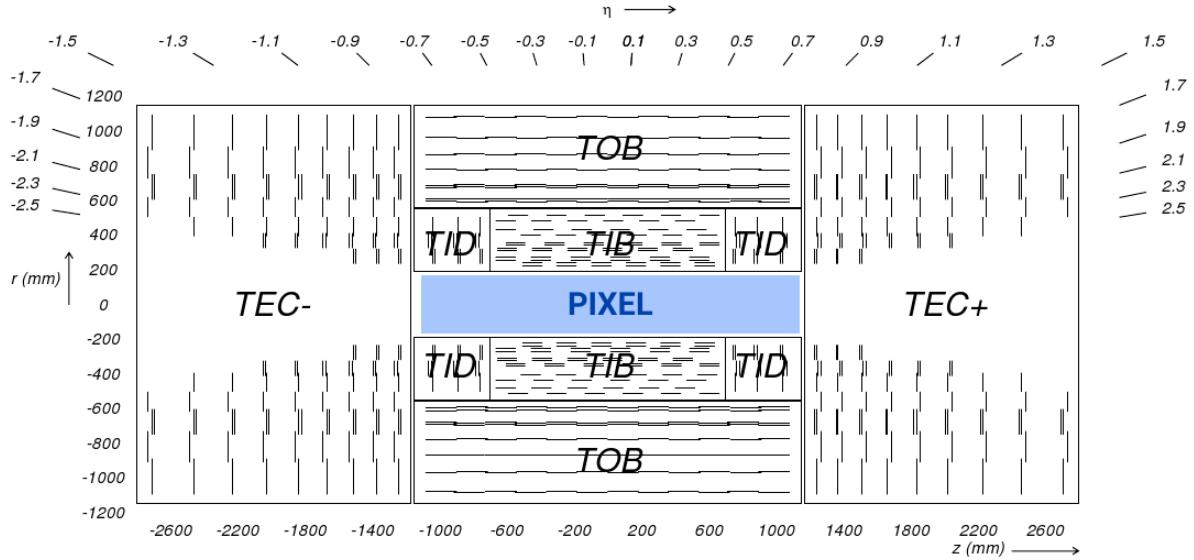


Figure 4.4: Schematic of the CMS silicon strip tracker and its four subsections: the inner barrel (TIB), inner endcaps (TID), the outer barrel (TOB), and the two endcaps (TEC) [24].

an extremely dense and transparent material capable of producing fast and clean photon bursts when hit, is required. Metal-heavy lead tungstate crystals, each taking roughly two days to artificially grow (and a total of about ten years to grow all of them), were chosen. With a density of $\delta = 8.28 \text{ g/cm}^3$ (slightly higher than for stainless steel), the crystals are compact enough to yield excellent performance without taking up too much volume, allowing the ECAL to fit within the CMS superconducting solenoid. The homogeneous medium allows for a better energy resolution as it minimizes the effects from sampling fluctuations, and it additionally contains enough oxygen in crystalline form to make it highly transparent to the entire scintillation emission spectrum. With an extremely short radiation length and small Molière radius ($X_0 = 0.85 \text{ cm}$, $R_M = 2.19 \text{ cm}$), the required homogeneity, granularity, and compactness is obtained, while at the same time emitting 80% of generated light within the 25 ns timeframe required. The largest drawbacks with a lead tungstate detector are the low light yield (100 γ per MeV), requiring dedicated avalanche photodiodes to increase the gain, and the light yield, which strongly depends on the temperature. The detector response to an incident electron changes by $3.8 \pm 0.4 \%$ per degree Celsius which requires the ECAL temperature to be kept stable at 18.0 ± 0.5 degree Celsius, which is obtained through an intricate water cooling system. The ECAL is completely hermetic and sorted into a barrel part (EB), covering pseudorapidities up to $|\eta| < 1.48$, and two endcap parts (EE) extending the total coverage to $|\eta| < 3.0$, in order to match the tracker coverage of $|\eta| < 2.5$. In order to improve the separation power between the γ and π_0 , a pre-shower detector

(ES) using lead absorbers and silicon sensors covers the forward region between $1.65 < |\eta| < 2.6$. The crystals in the barrel are organized into supermodules, each consisting of about 1700 crystals, while the endcap is divided into two half disks consisting of 3662 crystals each (so-called “Dees”). Each PbWO_4 crystal weighs around 1.5 kg and has a slightly tapered shape with a front face of $2.2 \times 2.2 \text{ cm}^2$ in the barrel and $2.86 \times 2.86 \text{ cm}^2$ in the endcaps. The crystals are 23 and 22 cm long in the barrel and endcaps, respectively. The total volume of the calorimeter including barrel and endcaps is 11 m^2 and it weighs a total of 92 tonnes. The ECAL detector layout is illustrated in Figure 4.5.

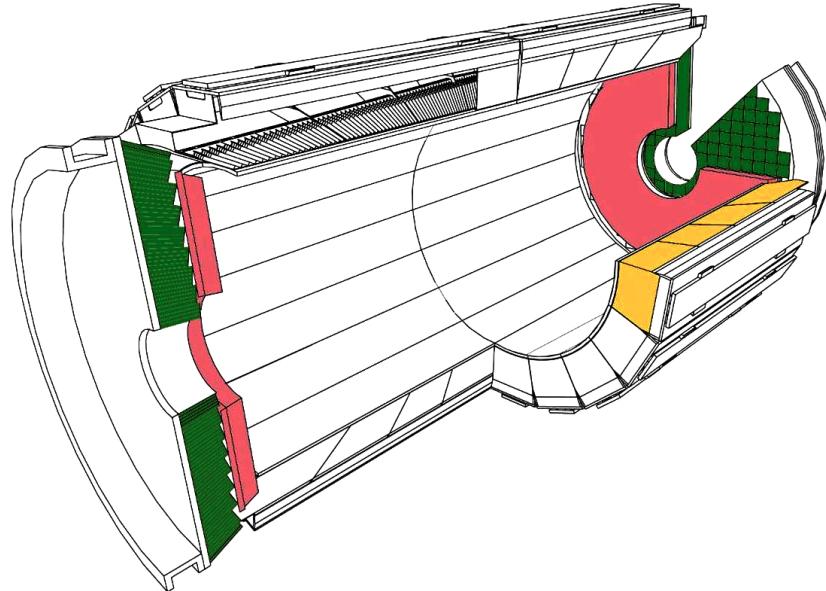


Figure 4.5: A schematic of the CMS electromagnetic calorimeter showing the barrel supermodules (yellow), the individual barrel crystals (black, top left), the endcap modules (green), and the pre-shower detectors (pink) [24].

Having no longitudinal segmentation, the ECAL relies on an accurate reconstruction of the event primary vertex, provided by the tracker, in order to reconstruct the photon angle correctly.

The obtained energy resolution of the ECAL can be parametrized in three parts: a stochastic, a noise, and a constant term [25]. It is given as

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\%,$$

in which the constant values were estimated using an electron test beam. The constant term of 0.3% is dominated by the non-uniformity in longitudinal light collection [26], and one of the

main goals of the detector design was to get this term below 1%. The energy resolution as a function of electron energy is shown in Figure 4.6.

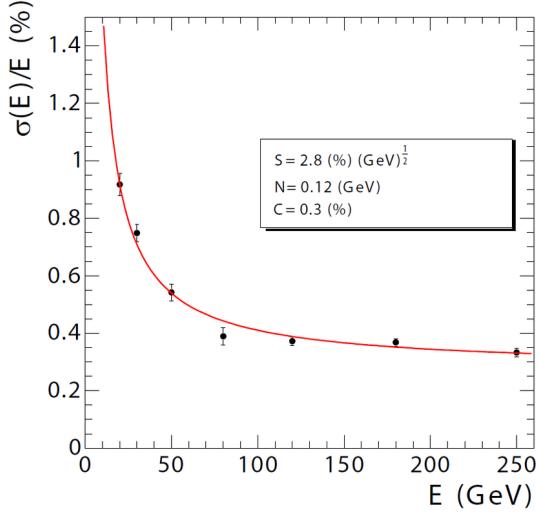


Figure 4.6: The ECAL energy resolution as a function of electron energy as measured in an electron test beam. [25]

4.4 Hadronic calorimeter

Outside the electromagnetic calorimeter is the hadronic calorimeter (HCAL). The HCAL is a sampling calorimeter, meaning it consists of alternating layers of dense brass absorber material and plastic scintillators. When a particle hits an absorber plate, it interacts with the absorber material and generates a shower of secondary particles which themselves generate new particle showers. These particles then generate light in the scintillating material which is proportional to their energy, and summing up the total amount of generated light over consecutive layers within a region, called a “tower”, is representative of the initial particles energy. It is the combined response of the ECAL and the HCAL that are responsible for measuring the energy of quarks, gluons and neutrinos through the reconstruction of particle jet energy and missing transverse energy. The hadron calorimeter is split into four regions: the inner (HB) and outer (HO) barrel, the endcap (HE) and the forward region (HF). A schematic of the CMS HCAL is shown in Figure 4.7. The inner barrel lies within the superconducting solenoid volume and covers the pseudorapidity range $|\eta| < 1.3$. It consists of 36 identical wedges, each of which weigh 26 tonnes, split into two half barrels (HB+ and HB-). A photograph of the wedges taken during installation is shown in Figure 4.8. The wedges are made up of flat brass absorber plates oriented parallel



Figure 4.7: The four regions of the CMS hadron calorimeter: the inner (HB) and outer (HO) barrel, the endcap (HE) and the forward region (HF) [24]

to the beam axis. These plates consist of a 4-cm thick front steel plate followed by eight 5-cm thick brass plates, six 5.6-cm thick brass plates and ending with a 7.5-cm thick steel back plate. The absorber plates alternate with 4-mm thick plastic scintillator tiles, which are the active medium of the detector, and which are read out using wavelength-shifting plastic fibers. The effective thickness of the barrel hadron calorimeter in terms of interaction lengths increases with the polar angle θ , starting out at about $5.8 \lambda_I$ at an angle of 90 degrees, and increasing to $10.6 \lambda_I$ at $|\eta| < 1.3$. As the energy resolution of the calorimeter depends on how much of the particles shower can be absorbed by the calorimeter, the quality of the energy measurement depends on its thickness. Due to the CMS design, the HB is confined to the volume between the ECAL (ending at a radius of 1.77 m) and the magnetic coil (starting at a radius of 2.95 m). In the central η region, the combined ECAL and HCAL interaction length is too small to sufficiently contain hadron showers. In order to ensure adequate sampling, especially of late starting showers, an additional layer of scintillator has therefore been added outside of the solenoid coil. This is the outer barrel (HO). It uses the coil itself as an absorbing material, increasing the total barrel calorimeter interaction length to $11.8 \lambda_I$. The hadron calorimeter endcaps (HE) are located in the forward region close to the beam pipe and cover the pseudorapidity range $1.3 < |\eta| < 3.0$, a region containing about 35% of the particles produced in collisions. Due to its close proximity to the beam pipe, the endcaps need to handle extremely high rates as well as have a high radiation tolerance. As the resolution in the endcap region is degraded due to pile-up and magnetic field effects, the hadron calorimeter endcaps were designed to minimize the cracks between HB and HE, rather than having the best single-particle resolution (as is the case for the barrel). The absorber plates in the endcaps are mounted in a staggered geometry rather than on top of each other as is



Figure 4.8: The installation of the barrel HCAL wedges, consisting of alternating layers of brass absorber plates and plastic scintillator, each weighing roughly 26 tonnes [27].

done in the barrel, in order to contain no dead material and provide a hermetic self-supporting construction. The HCAL is read out in individual towers with a size $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ in the barrel, and 0.17×0.17 at larger pseudorapidities. In order to obtain a completely hermetic calorimeter, an additional hadron forward calorimeter (HF) is added in the very forward region. Stretching out to a pseudorapidity of $|\eta| = 5.2$, this detector is located so close to the beam pipe that the particle rate exceeds 10^{11} per cm^2 , receiving roughly 760 GeV per proton-proton collision compared to an average of 100 GeV for the rest of the detector. It consists of two cylindrical steel structures, each with an outer radius of 130 cm and an inner radius of 12.5 cm, located 11.2 meters on either side of the interaction point. Also a sampling calorimeter, it consists of grooved 5-mm thick steel absorber plates, where the quartz fiber active medium is inserted into these grooves. The energy resolution of the CMS ECAL and HCAL detectors for pions is measured in a test beam as a function of energy and is shown in Figure 4.9. The typical electronics noise of the HCAL is measured to be 200 MeV per tower. The inclusion of the HO improves the resolution by 10% for a pion energy of 300 GeV. The final energy resolution parametrization when using ECAL+HB+HO is given by a stochastic and a constant term, as was the case for the ECAL detector, and is

$$\frac{\sigma_E}{E} = \frac{84.7\%}{\sqrt{E}} \oplus 7.4\%.$$

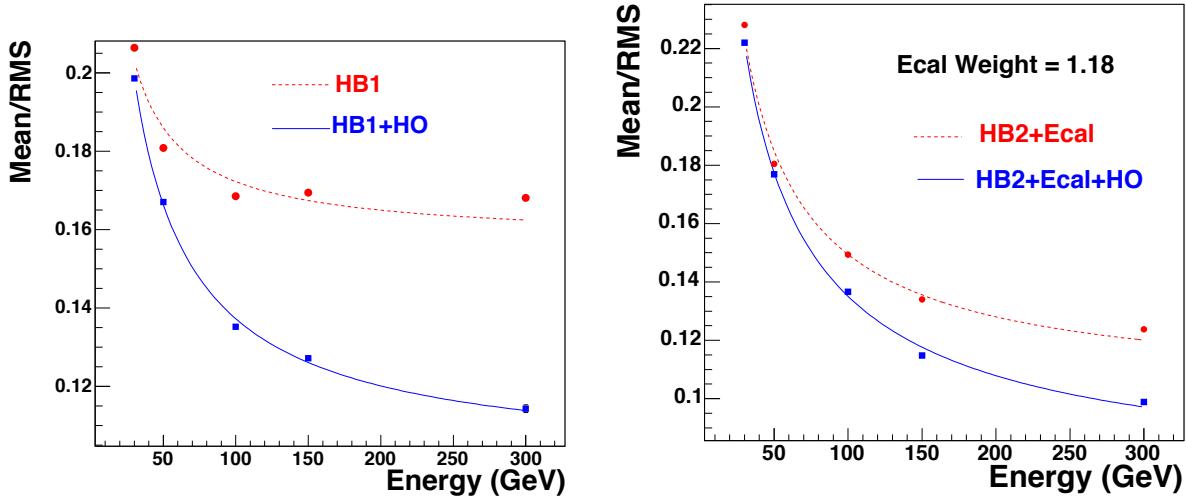


Figure 4.9: The calorimeter energy resolution as a function of pion energy using the HB only or HB+HO (left), and when adding ECAL and HCAL measurements (right) [28].

4.5 Muon chambers

The outer part of the CMS detector is dedicated to performing muon identification, momentum measurements, and triggering. The muon system is made up of three types of gaseous particle detectors: drift tube (DT) chambers, cathode strip chambers (CSCs), and resistive plate chambers (RPCs), all integrated into the magnetic return yoke structure. In the barrel region, where particle rates are low and the magnetic field uniform, DT chambers are used and cover the pseudorapidity region $|\eta| < 1.2$. In the endcap regions, however, the muon rates and background levels are considerably higher and the magnetic field itself is large and non-uniform. Therefore, CSCs with finer segmentation, higher radiation resistance, and faster signal collection are used, and cover the region $0.9 < |\eta| < 2.4$. To ensure accurate muon triggering, RPCs are used as a complimentary dedicated muon triggering system, which has been added both in the barrel and in the endcaps. These provide an excellent time resolution and cover the region $|\eta| < 1.6$. These chambers also assist in resolving ambiguities if multiple hits are present within a CSC or DT chamber. A schematic overview of the muon system is shown in Figure 4.10.

4.6 The CMS trigger system

With bunches in CMS colliding at a rate of 40 MHz, there are only 25 nanoseconds between collisions available to process event data. For typical instantaneous luminosities, one billion

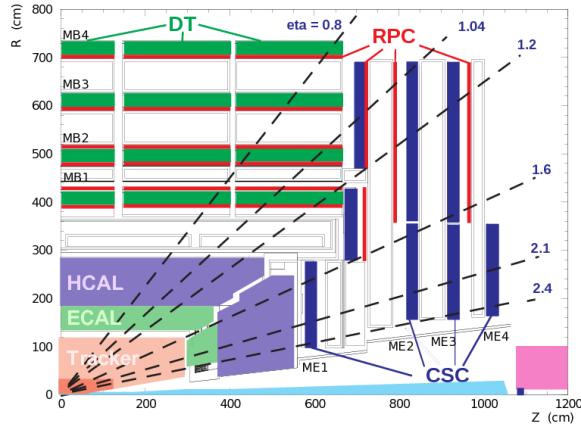


Figure 4.10: A schematic overview of the muon chambers: the DT chambers in the barrel, the CSCs in the endcaps, and the redundant RPC system stretching out to $|\eta| < 1.6$, which are used for triggering purposes [29].

collisions take place every second, and with an event size of roughly 1 MB, it is impossible for all of these events to be read out and stored to disk. The CMS triggering system is therefore designed to make ultra fast high-quality decisions about which events are interesting and which events are not. The first stage of triggering, called Level 1 (L1), is designed to reduce the event rate to a maximum of 100 kHz through custom-designed hardware. It uses coarse data from the muons system and calorimeters in order to make a decision on whether the event should be recorded or not, a decision that needs to happen within $3.2\mu s$. During this time, the full event data from each sub-detector is stored in detector front-end electronics, awaiting the L1 decision. The information used by L1 is gathered in three steps. First, trigger primitives are created. For the muon system, these consist of track segments from each of the three types of muon detectors. For the calorimeter, trigger primitives are generated by calculating the transverse energy in a trigger tower ($\Delta\eta - \Delta\phi$ of 0.087×0.087) and assigning it to the correct bunch crossing. Trigger primitives from the calorimeter are then passed on to a regional trigger, which defines electrons, muons and jet candidates. Some of this information is passed to the regional muon trigger in order to provide information about whether the particle is minimum ionizing. The global muon trigger then combines track information with calorimeter information and selects a maximum of four muon candidates and calculates their momentum, position, charge and quality. The output from the regional calorimeter trigger is also passed to a global calorimeter trigger which provides information about the jets, total transverse energy and missing energy in the event. Combining the information from the global muon trigger and the global calorimeter trigger, the L1 trigger decides whether to keep the event or not by combining several decisions by simple logic operations

(AND/OR/NOT), forming up to 128 algorithms.

If the event is accepted, the full event information is read out at a rate of 100 kHz and passed to the High Level Trigger (HLT), a farm of commercially available computers. Here, the full precision of the detector data is used in order to make decisions based on offline-quality reconstruction algorithms. The goal of the HLT is to eventually reduce the event rate to an average of 400 Hz, which will be stored on tape.

CHAPTER 5

Event reconstruction

5.1 Track and primary vertex reconstruction

The CMS tracker gets traversed by $\mathcal{O}(1000)$ charged particles at each bunch crossing, produced by several proton-proton interactions happening simultaneously. This makes track reconstructions extremely challenging, and is the reason why a high granularity of the tracker is vital. The average number of interactions per event for the whole Run 2 data taking period is shown in Figure 5.1, with an average of 34 interactions per bunch crossing. Track reconstruction

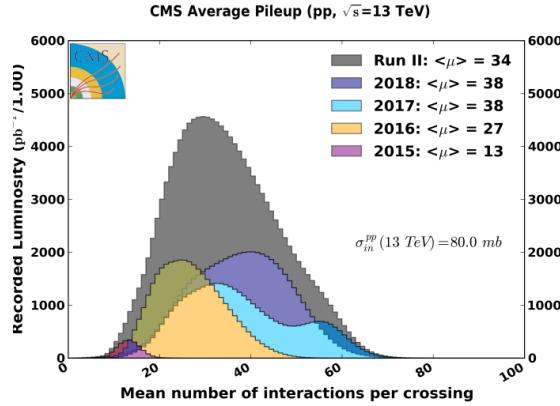


Figure 5.1: The average number of interactions per event in CMS for the whole Run 2 data taking period [30].

encompasses the process of taking hits from the pixel and strip detectors, combining them into particle trajectories, and then estimating the momentum and flight direction of the charged particle responsible for producing the hits. It is a computationally heavy process and at CMS it

is based on what is called a combinatorial Kalman filter [31]. A Kalman filter is an algorithm that uses time-dependent observations in order to estimate unknown variables, by proceeding progressively from one measurement to the next, improving the knowledge of the trajectory with each new measurement. The track reconstruction software in CMS (called the Combinatorial Track Finder (CTF)) constructs its collection of tracks by iteratively looping over the hits and reconstructing tracks, then removing those hits which are already used as inputs for a previous track. It starts from a seed in the innermost tracker layers, usually two or three hits, and then extrapolates the seed trajectories searching for additional hits to associate to that candidate. It then disregards tracks that fail certain criteria based on a χ^2 calculation, taking both hit and trajectory uncertainties into account, as well as the number of missing hits. The track reconstruction algorithm is effective over the full range of the tracker coverage, up to $|\eta| < 2.5$, and can reconstruct particles with momenta as low as 0.1 GeV and particles which are produced up to 60 cm from the beam line. In the central region, particles with a momentum of 100 GeV have a p_T -resolution of roughly 2.8 %, a transverse impact parameter resolution of 10 μm and a longitudinal impact parameter of 30 μm .

In order to define the location and uncertainty of every proton-proton interaction in an event, primary-vertex reconstruction is performed. Primary vertices lie within a radius of a few millimeters from the beam axis and are defined as the common origin of groups of tracks. The reconstruction algorithm takes as input the reconstructed tracks from the previous step which pass certain selection criteria, clusters the tracks that share a common origin, and then fit for the position of each vertex. Each track must have at least 2 hits in the pixel layers and no less than 5 hits in the pixel plus strip layers, as well as a $\chi^2 < 20$ from a fit to the particle trajectory, to be considered as input for the vertex finder. The primary vertex resolution is around 12 μm in x and 10 μm in z for vertices with at least 50 tracks.

Offline, all events are required to have at least one primary vertex reconstructed within a 24 cm window along the beam axis, with a transverse distance from the nominal interaction region of less than 2 cm. The reconstructed vertex with the largest value of p_T^2 of the summed particles is selected as the primary interaction vertex where the hard scattering process occurred.

5.2 The Particle Flow Algorithm

After track reconstruction, what remains is an incoherent collection of tracks, calorimeter clusters and hits in the muon chambers. In order to connect these, CMS uses an algorithm called Particle Flow (PF) [32] to combine the information obtained from all sub-detectors in order to infer which particles were actually produced in the event. The reconstructed physics objects in the order of which they are reconstructed are

- muons, through hits in the tracker and in the muon chambers;
- charged hadrons, through hits in the tracker and energy deposits in the calorimeters;
- neutral hadrons, through energy deposits in the calorimeters but no hits in the tracker;
- photons, through energy deposits in the ECAL but not in the HCAL, and no hits in the tracker; and
- electrons, through hits in the tracker and energy deposits in the ECAL.

How these different particles propagate through the CMS detector is illustrated in Figure 5.2.

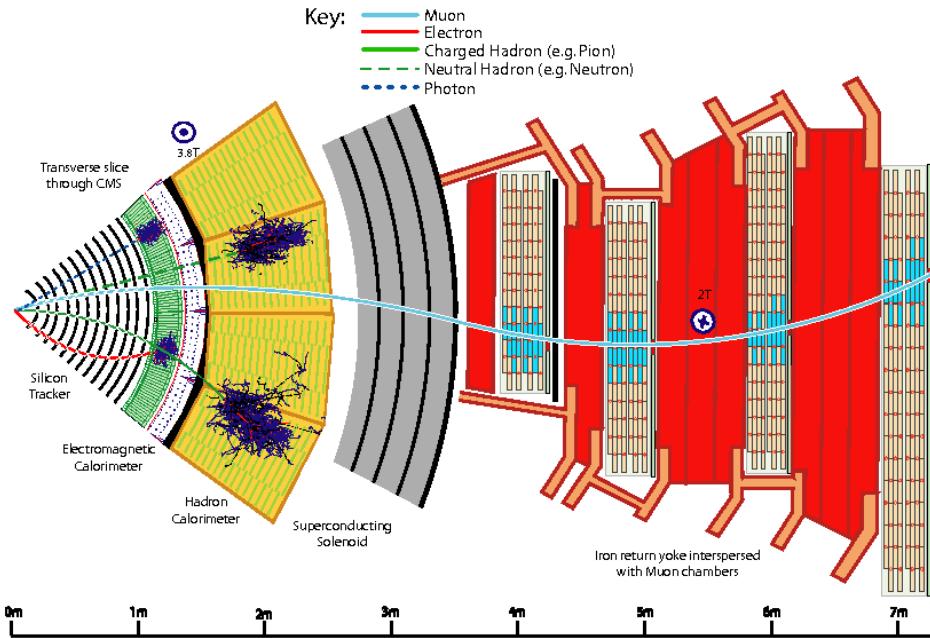


Figure 5.2: Particle interactions in the different subdetectors for a transverse slice through the CMS detector [32].

5.2.1 Reconstruction of the Particle Flow inputs

Electron tracking

Electron seeding is done in two different ways: ECAL-based or tracker-based. In the ECAL-based method, electrons are seeded from ECAL clusters with $E_T > 4 \text{ GeV}$, where the position of the cluster is used to infer which hits in the inner tracker belongs to a given electron or positron. As a large fraction of the electron/positron energy is emitted through bremsstrahlung before even reaching the ECAL, the ECAL superclusters covering a small window in η and a larger window in ϕ are defined in order to fully contain the electron as well as its bremsstrahlung photons. As these superclusters are prone to contamination, tight isolation requirements need to be applied, leading to reconstruction inefficiencies. Therefore, an additional tracker-based seeding approach has been developed. All tracks with $p_T > 2 \text{ GeV}$ are used as potential electron seeds. These tracks are then extrapolated to the ECAL and matched to the closest ECAL cluster. The ratio of the cluster energy to the track momentum is required to be ~ 1 . The electron candidates are then fit with a Gaussian-sum filter (GSF) [33] and required to pass certain criteria based on the score of a boosted-decision-tree (BDT), which combines the number of tracker hits, the χ^2 of the GSF track, the energy loss along the track, and the distance between the extrapolated track to the closest ECAL cluster.

Muon tracking

Muon tracking consists of two parts: the muon spectrometer allows muons to be identified with high efficiency over the full pseudorapidity range, while maintaining a low background due to the absorbing calorimeter layers upstream. The inner tracker on the other hand, provides an accurate measurement of the muon momentum. Three muon quality flags are defined:

- standalone muon: Muon tracks based on hits in the DT or CSC only;
- global muon: A standalone muon track matched to a track in the tracker if the track parameters of the two are compatible; and
- tracker muon: An inner track with $p_T > 0.5 \text{ GeV}$, a total momentum greater than 2.5 GeV , and at least one muon segment matching the extrapolated inner track.

Around 99% of muons produced within $|\eta| < 2.4$ are reconstructed as a global muon or a tracker muon, and very often as both. If the global and tracker muon share the same inner tracker segment, the two are combined.

Calorimeter clusters

The calorimeter clustering is performed separately for each calorimeter subdetector (ECAL barrel and endcaps, HCAL barrel and endcaps and the preshower layers). The first step is to define cluster seeds from cells with an energy exceeding some predefined threshold and larger than the energy in its neighboring cells. Topological clusters are then formed by adding cells to the seed which has at least one corner in common with a cell already in the cluster, and that has an energy which is at least twice the noise level of the detector. In Figure 5.3, an example of calorimeter clustering for a five-particle jet is shown for the HCAL (left) and ECAL (right). In the HCAL (left), two seeds have been identified (gray-filled areas) inside a topological cluster consisting of 9 cells. These are then defined as two HCAL clusters, with a position as indicated by the red circles. The green solid lines correspond to charged tracks reconstructed in the tracker, both pointing to the center of the HCAL cluster seeds. The observed deposits left by the same particles are shown on the right in Figure 5.3, where the K_L^0 , π^- and the two photons from the decay of a π^0 leave distinct clusters in the ECAL. The π^+ leaves no energy deposit in this case.

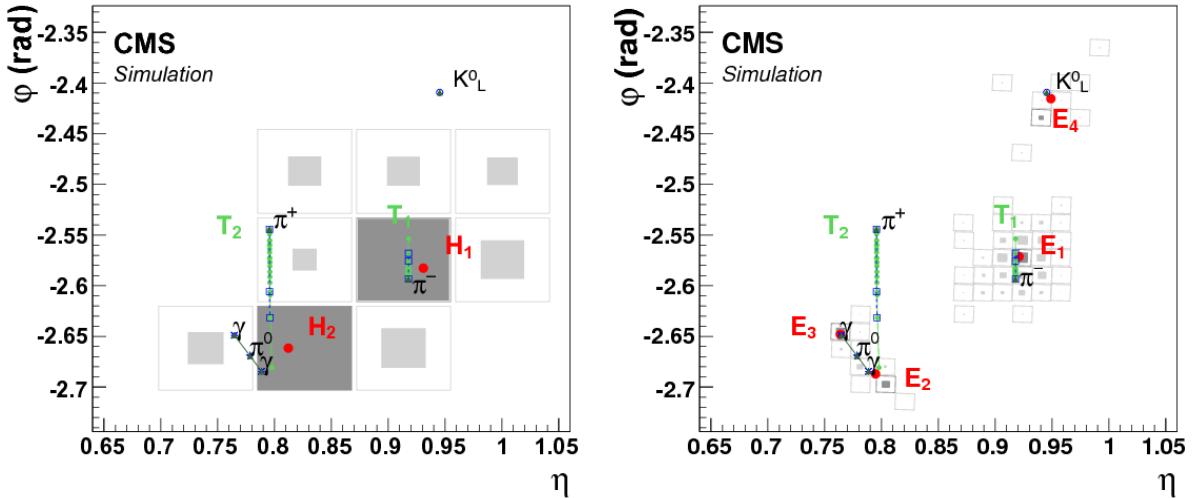


Figure 5.3: The $\eta - \phi$ views of calorimeter clusters generated by a five-particle jet in the HCAL (left) and in the ECAL (right). The squares correspond to calorimeter cells, where the inner area is proportional to the logarithm of the cell energy. Cluster seeds are depicted in dark gray. The dotted blue lines correspond to the simulated particle trajectories, while the green lines correspond to charged tracks reconstructed in the tracker [32].

5.2.2 Particle Flow identification

The link algorithm

The link algorithm is the algorithm responsible for combining the particle flow elements from different subdetectors. It can test any pair of elements in the event based on specific requirements depending on the nature of the element, but is restricted to the nearest neighbors in the $\eta - \phi$ plane. The outputs of the link algorithm are so-called *PF blocks* of linked elements, either directly linked or linked through having common elements.

- **Inner track - calorimeter cluster link:** The track is interpolated from its last hit, through the preshower layers, the ECAL, and ending in the HCAL at a nuclear interaction length depth of 1. A link is made if the track is within the cluster areas, where the area is enlarged by up to a cell in each direction to account for detector gaps. In case several ECAL/HCAL clusters are linked to the same track, only the one with the smallest distance in $\eta - \phi$ is kept.
- **Calorimeter cluster - cluster link:** A link between ECAL and HCAL clusters, as well as between ECAL and preshower clusters, is made when the cluster position of the more granular calorimeter is within the cluster envelope of the less granular calorimeter. If there are overlapping links, the one with the smallest distance is kept.
- **Inner tracker - muon chamber link:** This procedure is described in Section 5.2.1.

For each PF block, the reconstruction proceeds in the following order. First, muons are reconstructed and their corresponding PF elements removed from the PF block. This is followed by the reconstruction and subsequent removal of electrons and energetic photons. Finally, neutral and charged hadrons are reconstructed.

Muons

First, isolated global muons are selected by requiring the sum of track p_T and calorimeter energy deposits within a cone of $\Delta R = 0.3$ and not belonging to the muon track, to be smaller than 10 % of the muon p_T . If the muons are non-isolated, they are required to pass the tight muon requirement [34] and have at least three matching track segments in the muon detector or have matched calorimeter deposits compatible with being a minimum ionizing particle. Muons failing both the requirements above are kept if their standalone muon track is of high quality and have a lot of hits in the muon detectors, otherwise they are discarded. The muon momentum is defined

from the inner tracker measurement if the muon p_T is less than 200 GeV. Otherwise, it is chosen according to the track fit with the smallest χ^2 probability.

Muons used in this thesis are required to pass an isolation requirement in order to suppress the background from QCD multijet events where jet constituents are identified as muons. For this, a cone of radius $\Delta R = 0.3$ is constructed around the muon direction. The isolation parameter is defined as the scalar sum of the transverse momenta of all additional reconstructed tracks within the cone, divided by the muon p_T . Muon candidates with an isolation parameter less than 0.1 are considered isolated and are used for further analysis.

Further, the following selection criteria are applied:

- the χ^2 of the global muon track fit must be less than 10;
- at least one muon-chamber hit is included in the global-muon track fit and the global muon track fit must include at least one muon chamber hit;
- muon segments in at least two of the muon stations must be matched to the muon tracker track;
- the inner tracker track must be no more than 2 millimeters from the primary vertex in the xy plane and no more than 5 millimeters in the longitudinal direction;
- at least one hit in the pixel detector;
- at least six layers of the inner tracker must contain hits; and
- at least three matching track segments must be found in the muon detectors

Electrons

The electrons are seeded from a GSF track, as described in Section 5.2.1. To differentiate electrons from charged hadrons, the energy deposit in the HCAL within a distance of 0.15 in the $\eta - \phi$ plane of the supercluster is required to be less than 10 % of that of the supercluster. The electron candidate must further pass a requirement on the output of a dedicated electron-identification BDT, using inputs such as track-cluster distance, track χ^2 , and number of hits as input. In this step, isolated photons are also reconstructed, seeded from ECAL superclusters with $|E_T| > 10 \text{ GeV}$ and no link to a GSF track. All the tracks and calorimeter deposits used to reconstruct electrons and isolated photons are further removed from the list of PF blocks.

Only electrons passing certain quality requirements as listed in Table 5.1 are used in this thesis, with the following variable definitions:

- E_T : The supercluster energy ($x \sin(\theta_{track})$), where θ_{track} is the polar angle of the electron track as measured in the inner tracker layer and extrapolated to the interaction vertex.
- η^{sc} : η of the electron supercluster.
- **isEcalDriven**: The electron is found through ECAL requirements rather than through Particle Flow and the tracker.
- $\Delta\eta_{in}^{seed}$: difference in η between the track position as measured in the inner layer and extrapolated to the interaction vertex and then to the calorimeter, and the η of the supercluster.
- $\Delta\phi_{in}$: difference in ϕ between the track position as measured in the inner layer and extrapolated to the interaction vertex and then to the calorimeter, and the ϕ of the supercluster.
- **H/E**: Ratio of hadronic energy in the calorimeter towers, within a cone of radius 0.15 centered at the electron's calorimeter position to the electromagnetic energy of the supercluster.
- $\sigma_{in\eta\eta}$: Measure of the energy spread in η in units of crystals of electron energy in a 5×5 block centered on the seed crystal.
- **ECAL Isolation**: The transverse electromagnetic energy of all reconstructed hits (with $E > 0.08$ GeV) in a cone of radius 0.3 centered at the electron's calorimeter position, excluding those in an inner cone with a radius of 3 crystals and an η strip with a width of 3 crystals.
- **Hadronic Depth Isolation**: Defined as the transverse depth of the hadronic energy in the HCAL inside a cone of 0.3 centered on the electron calorimeter position, excluding towers in a cone of 0.15 radius.
- **Track p_T Isolation**: The sum p_T of the tracks in a ΔR cone of 0.04 to 0.3, excluding an η region of 0.015.
- d_{xy} : Transverse distance between the electron track and the primary vertex.

Hadrons

Finally, after the removal of muons and electrons, the remaining hadrons and non-isolated photons are identified. HCAL clusters with no track link are defined as neutral hadrons, while ECAL clusters with no track link are defined as photons (photons are exclusively associated to the ECAL

Variable	Barrel	Endcap
E_T	$> 35 \text{ GeV}$	$> 35 \text{ GeV}$
η range	$ \eta_{sc} < 1.4442$	$1.566 < \eta_{sc} < 2.5$
isEcalDriven	yes	yes
$\Delta\eta_{in}^{seed}$	< 0.004	< 0.006
$\Delta\phi_{in}$	< 0.06	< 0.06
H/E	$< 1/E + 0.05$	$< 5/E + 0.05$
full 5x5 $\sigma_{in in}$	n/a	< 0.03
full 5x5 E^{2x5}/E^{5x5}	$> 0.94 \text{ OR } E^{1x5}/E^{5x5} > 0.83$	n/a
EM+Had. Depth Iso.	$< 2 + 0.03 \times E_T + 0.28 \times \rho$	$E_T < 50 \text{ GeV}: < 2.5 + 0.28 \times \rho$ else: $< 2.5 + 0.03 \times (E_T - 50) + 0.28 \times \rho$
Track p_T iso.	$E_T < 100 \text{ GeV}: < 5 + 1.5 \times \rho$ else: $< 5 + 1.5 \times \rho$	$< 5 + 0.5 \times \rho$
Inner Layer Lost Hits	≤ 1	≤ 1
d_{xy}	< 0.02	< 0.05

Table 5.1: Summary of the electron requirements applied to all electrons used in this analysis.

deposits since neutral hadrons leave only 3 % of their energy in the ECAL). The remaining HCAL clusters are then linked to one or more tracks from the inner tracker. In order to determine the particle content within a cluster, the sum of track momenta and the calorimeter energy is compared. If the calorimeter energy is compatible with the sum of track momenta, a particle for each track is inferred, with its corresponding energy taken from the track momentum. If the calorimeter energy is larger than the sum of track momenta, a photon or a neutral hadron is added, together with one charged hadron for each track within the cluster area.

Missing transverse energy

Neutrinos (and other predicted, non-SM weakly interacting particles) do not interact in the detector and are instead inferred from the presence of a momentum imbalance in the detectors transverse plane. The missing transverse momentum is defined as the negative p_T vector sum of all reconstructed PF candidates in the event,

$$\vec{p}_T^{\text{miss}} = - \sum_i^N \vec{p}_{T,i}, \quad (5.1)$$

and its magnitude, $|\vec{p}_T^{\text{miss}}|$, is referred to as the missing transverse energy E_T^{miss} (which is used as a proxy for the neutrino p_T).

5.3 Pile-up removal

Particles originating from proton-proton interactions not associated with the hardest primary vertex, are denoted as pileup events. These distort observables of interest from the hard scattering event and must be mitigated through dedicated pileup removal techniques.

5.3.1 Charged Hadron Subtraction

As mentioned previously, primary vertices are reconstructed using tracks from charged hadrons. If a primary vertex does not correspond to the hard scattering vertex of the event, the charged hadrons (as reconstructed through Particle Flow) associated to this vertex (called a pileup vertex) are removed from the event collection of particles and will not participate in any further object reconstruction. This method is denoted charged hadron subtraction (CHS).

5.3.2 Pile up per particle identification (PUPPI)

CHS was the default pileup removal algorithm in CMS until very recently. In 2014, a new pileup removal algorithm with improved performance was proposed; the pileup per particle identification (PUPPI) [35] algorithm. PUPPI uses a combination of local shape information, event pileup properties, and tracking information to compute a weight describing the degree to which a given particle is likely to arise from pileup. First, a variable denoted α is computed based on the difference between soft radiation coming from pileup and the harder collinear QCD pattern. The shape of α for charged particles is then used as a proxy for all pileup particles and is used on an event-by-event basis to calculate a weight for each particle. This weight in turn describes the degree to which particles are pileup-like and are used to rescale the particle four-momenta. The shape variable for a given particle i is defined as

$$\alpha_i = \log \sum_{\substack{j \in \text{Ch,PV} \\ j \neq i}} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \Theta(R_0 - \Delta R_{ij}), \quad (5.2)$$

where Θ is the step function and j refers to the neighboring charged particles from the primary vertex within a cone of radius $R_0 = 0.4$. Charged particles are defined as coming from the primary vertex if they are associated to the leading vertex of the event or are within a distance of $d_z < 0.3$ cm from the leading vertex. In order to determine the probability that a particle comes

from pileup, a χ^2 calculation is performed. The probability is defined as

$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{RMS_{PU}^2}, \quad (5.3)$$

where $\bar{\alpha}_{PU}$ is the median value of the α_i distribution for pileup particles in the given event and RMS_{PU} is its RMS. Each particle (neutral and charged) is then assigned a weight $w_i = F_{\chi^2, NDF=1}(\chi_i^2)$, where $F_{\chi^2, NDF=1}$ is the cumulative distribution function of the χ^2 distribution with one degree of freedom. Particles with $w_i < 0.01$ are rejected. In addition, a cut on the weighted p_T of neutral particles of $w_i \cdot p_{T,i} > (A + B \cdot N_{PV})$ GeV is applied, where N_{PV} correspond to the number of reconstructed vertices in the event and A and B are tunable parameters. The performance of the PUPPI algorithm compared to CHS for jet observables is shown in Figure 5.4. The top row shows the absolute mass resolution (left), as well as the mass resolution as a function

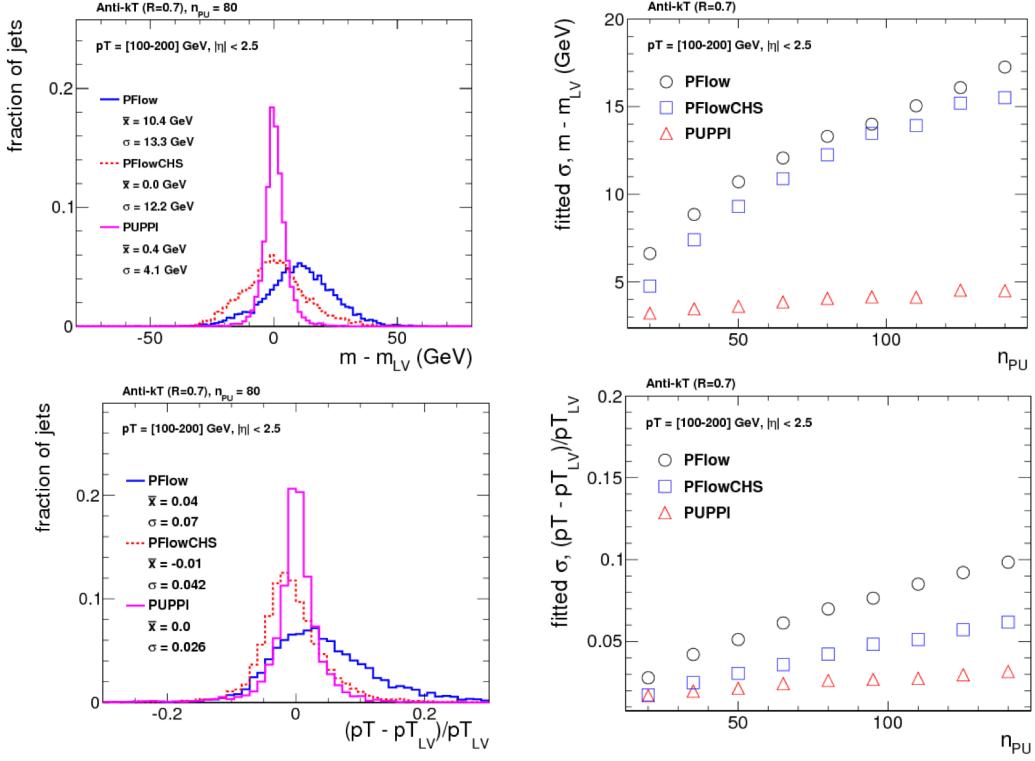


Figure 5.4: The mass (top) and p_T (bottom) resolution comparing PF-only (blue), PF+CHS (red) and PUPPI (pink) jets. The absolute resolution (left) as well as the resolution as a function of the number of reconstructed primary vertices in the event (right) is shown [35].

of N_{PV} for CHS jets (red) and PUPPI (pink) jets. The bottom row shows the corresponding quantities but for jet transverse momentum. A significantly better resolution on jet observables

can be achieved using PUPPI as compared to CHS.

5.4 Jet reconstruction

As explained in Section 1.2.1, quarks and gluons are never themselves visible in a detector. Within 10^{-23} seconds, the timescale of strong interactions, they fragment and hadronize into a collimated spray of hadrons, a so-called jet. In order to infer the properties of the original parton generating the jet, the properties of the full particle spray needs to be evaluated. Combining these particles algorithmically is non-trivial, and several algorithms jet-clustering algorithms exist. These provide a set of rules for grouping particles together into jets and are usually based on certain distance requirements between particles as well as rules for how to recombine their momenta. Thanks to Particle Flow, objects like charged hadrons, neutral hadrons and photons, together with their estimated energy and direction, are already defined and jet clustering in CMS therefore consists of associating these particles to one common origin.

5.4.1 Jet clustering

The most common jet clustering algorithms used in hadron colliders are the Cambridge/Aachen algorithm [36], the k_T algorithm [37] and the anti- k_T algorithm [38]. These are all sequential recombination algorithms, meaning they systematically go through each particle pair in the event and recombines them into one particle if the combination satisfies certain criteria. The rules, shared by all three algorithms, are as follows:

1. For each pair of particles i and j , compute the longitudinally invariant distances

$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad , \text{ with } \quad \Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2, \text{ and} \quad (5.4)$$

$$d_{iB} = p_{ti}^{2p}, \quad (5.5)$$

where d_{ij} is the so-called k_T distance between particle i and j , ΔR_{ij}^2 is the distance between them in the $\eta - \phi$ plane, ΔR^2 corresponds to a radius parameter which controls the extension of the jet and d_{iB} is the distance between the particle and the beam. The parameter p is what separates the three algorithms from one another and controls the relative power of energy versus geometrical scales. For the anti- k_T algorithm, it is defined as $p = -1$; for the k_T algorithm, $p = 1$; and in the case of the C/A algorithm, $p = 0$. The consequences of these choices are explained in detail below.

2. Find the minimum distance of d_{ij} and d_{iB} .
3. If this is d_{ij} , recombine particles i and j and return to step 1.
4. If it is d_{iB} , the particle i is defined to be a final state jet, and is removed from the list of particles. The algorithm proceeds back to step 1.
5. Repeat until no particles remain.

Infrared and collinear safety

There are two requirements that are extremely important when defining jet algorithms: They must be 1) *infrared* (IR) and 2) *collinear* (C) safe. *Infrared* safety corresponds to the requirement that if the final state particles are modified by the presence of a soft emission, and there are always soft emission in QCD events (both perturbative and non-perturbative), then the set of hard jets should remain unchanged. This is illustrated by the two left figures in Figure 5.5. Here, the

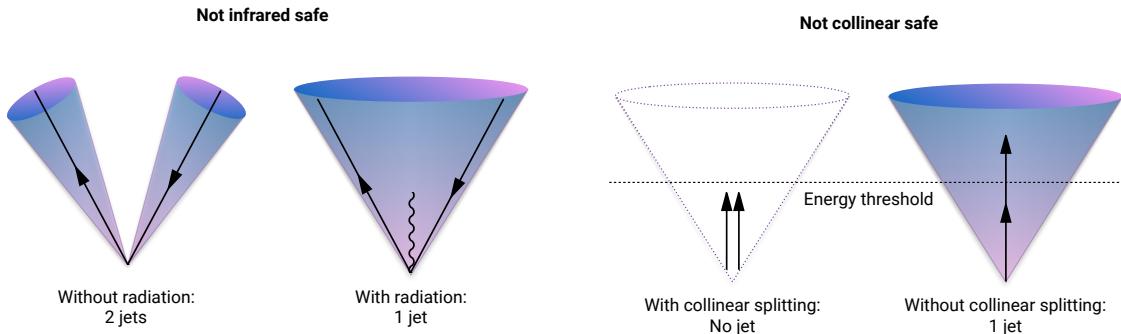


Figure 5.5: An illustration of what would happen for an infrared (left) and collinear (right) unsafe jet algorithm. If an algorithm is infrared unsafe, the presence of a soft gluon changes the jet configuration. If an algorithm is collinear unsafe, then if a hard parton undergoes a collinear splitting this will change the configuration of the jet.

algorithm is infrared unsafe: the presence of an additional soft gluon changes the jet configuration from 2 to 1 jets. If an algorithm is *collinear* unsafe, it means that the jet configuration would change if the hard parton undergoes collinear splitting (which a hard parton often does as part of the fragmentation process and which are also part of non-perturbative dynamics, like the decay of highly energetic hadrons). This is shown in the two left figures of Figure 5.5, where a hard parton undergoing collinear splitting fails to be reconstructed due to its daughters being below the energy threshold of the algorithm. All sequential recombination algorithms are trivially infrared safe.

The k_T algorithm

The k_T algorithm is the oldest of the sequential recombination algorithms and, due to its $p = 1$ definition in Equations 5.4 and 5.5, follows the QCD branching structure in both p_T and in angle (in reverse). Soft particles are clustered together first, and the final step is the clustering of the two hardest particles. A consequence of this definition is that there is nothing that keeps arbitrarily soft particles from being defined as jets, and a minimum cut on the jet p_T should be introduced. Despite several favorable qualities, the k_T algorithm is not the algorithm of choice in most hadron collider experiments due to the irregular jets it produces, a consequence of clustering soft particles first.

The Cambridge/Aachen algorithm

The Cambridge/Aachen algorithm, with $p = 0$ in Equations 5.4 and 5.5, follows the QCD branching structure only in angle as the clustering order is based solely on spatial separation. The simplest of the algorithms, it recombines all pairs close in ΔR until $\Delta R_{ij} > R$. The benefits of this is that the clustering history contains information about the presence of any geometrical substructure within a jet, a feature that will become important in Section 5.5.

The anti- k_T algorithm

The default jet clustering algorithm in CMS is the anti- k_T algorithm [38], which follows the clustering rules in Equations 5.4 and 5.5 with $p = -1$. The algorithm favors the clustering between high p_T particles, and high and low p_T particles first, disfavoring clustering between soft particles. That means the algorithm grows around a hard core, yielding jets with a well-defined cone shaped area. Since the algorithm is infrared- and collinear-safe (IRC safe) and insensitive to the underlying event (parton remnants not from the hard process) and pileup, , it is chosen as the main jet algorithm in CMS. A comparison of the resulting jet area in the $\phi - \eta$ plane after clustering with either k_T , C/A and anti- k_T , is shown in Figure 5.6. The z-axis correspond to the parton p_T . One can clearly see that when clustering with the anti- k_T algorithm, the produced jets are circular, with a radius set by R , around the hardest parton.

5.4.2 PF jets in CMS

Jet algorithms in CMS mainly use the four-vectors of PF candidates as input and a pileup removal algorithm is usually applied before clustering occurs. If using CHS (Section 5.3.1), charged hadrons not associated to the primary vertex are discarded before clustering. If PUPPI

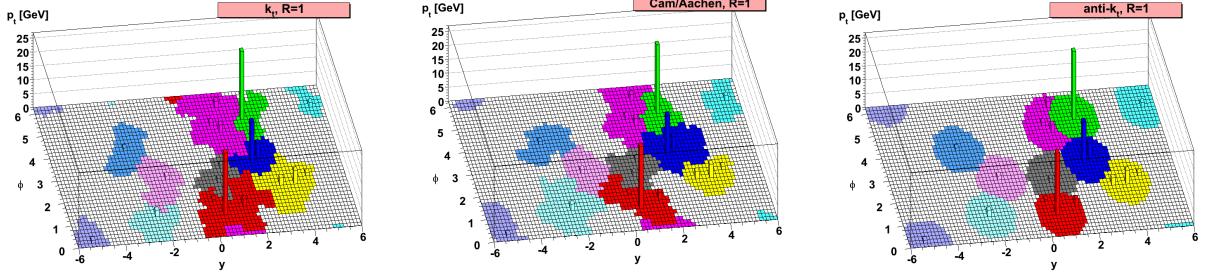


Figure 5.6: A comparison of the resulting jet cone area in the $\phi - \eta - p_T$ plane after clustering the same event with three different jet algorithms: k_T , C/A and anti- k_T . [38]

is used (Section 5.3.2), all the PF candidates are reweighted based on how likely they are to have originated from pileup. For the anti- k_T algorithm, CMS by default uses two jet cone sizes: $R=0.4$ and $R=0.8$. Jets with $R=0.4$, called PFAK4, are used for single-prong jets while the larger $R=0.8$ jets, PFAK8, are more often used when looking for jets containing multiple hard quarks/gluons in order to contain all the hadronization products.

These jets are further required to pass certain jet identification requirements in order to distinguish them from fake jets [39], called *tight ID*, which are as follows:

- the jet must contain at least two PF constituents;
- at least one of these constituents must be a charged hadron;
- the fraction of jet energy coming from neutral hadrons must be < 0.90 ;
- the fraction of jet energy coming from neutral electromagnetic energy must be < 0.90 ; and
- the fraction of jet energy coming from charged electromagnetic energy must be < 0.99 .

5.4.3 Jet energy corrections

All jets are further corrected for nonlinearities in p_T and pseudorapidity using standard CMS jet energy corrections (JEC), as described in Ref. [40]. These are intended to bring the measured jet energy closer to the true jet energy by correcting the jet energy scale (JES) and jet energy resolution (JER). The energy corrections are derived in three steps:

- L1: Energy offset corrections intended to remove pileup and electronic noise, both for data and simulation;
- L2L3: A relative (L2) and absolute (L3) correction to particle level jet response for simulation only; and

- Residual: A correction for data only meant to correct for residual differences between data and simulation.

These are illustrated in Figure 5.7.

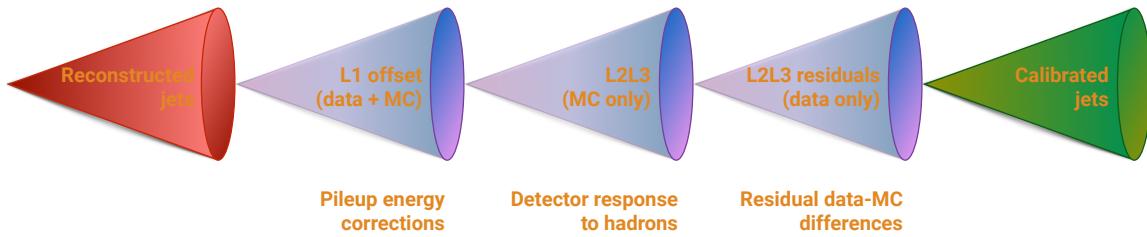


Figure 5.7: The CMS jet energy corrections are derived in three steps: A correction due to offset energy coming from pileup, applied to data and MC, a correction due to the particle-level jet response, also applied to data and MC and finally a correction to account for residual differences between data and MC.

L1 offset correction

The largest correction (for low- p_T jets) is the L1 pileup offset correction, which is meant to subtract the additional energy in a jet due to pileup. This is done on an event-by-event basis through the *jet area method* which uses the effective area of the jet multiplied by the average event energy density to calculate the size of the offset energy to be subtracted from each jet. An additional p_T - and η -dependent term is added in order to account for different pileup densities in different parts of the detector and for different jet energies. For data, an additional scale factor to account for data and simulation differences is computed. This is done by constructing a *Random Cone (RC)* centered at a given η, ϕ and dividing the energy density within that cone in data, evaluated in a dataset with no hard interactions (*Zero Bias*), by that of the true energy offset in simulation. Note that this correction is not applied to PUPPI jets since the additional energy due to pileup has already been removed by the PUPPI algorithm.

L2 relative and L3 absolute corrections

After L1 corrections are applied, corrections to account for the detector response to hadrons are derived based on the true detector response in QCD MC. The simulated particle response is defined as the ratio

$$R_{\text{particle}} = \frac{p_{T,\text{reco}}}{p_{T,\text{particle}}}. \quad (5.6)$$

These are derived in bins of particle-level p_T and reconstructed η : The L2 relative corrections are intended to make the detector response uniform and are derived as a function of η , while the L3 absolute corrections are derived as a function of jet p_T . These corrections are applied both to data and to MC.

Residual data corrections

After L1, L2, and L3 corrections are applied, two additional corrections are derived only for data in order to account for any residual discrepancies between data and MC. This is done by looking at the transverse momentum balance between a jet which is to be calibrated, and some reference object (either another jet, a Z boson, or a photon). If the jet energy scale is not equal to one, a p_T imbalance will be visible. The measurements are performed in a data sample of dijets, where the statistical uncertainty is small but the energy of the reference object poorly measured, as well as in $Z(\mu\mu) + \text{jet}$, $Z(ee) + \text{jet}$ and $\gamma + \text{jet}$ samples, where the energy of the Z and γ is very well known but the statistics are small.

The “L2 relative” residual correction is measured in dijet events by comparing the measured p_T of the reference jet, required to be central with $\eta < 1.3$, to that of the calibration jet, with an unconstrained η . This is done as a function of jet η , in bins of average jet p_T . The “L3 relative” residual correction, is instead measured in $Z/\gamma + \text{jet}$ events by comparing the measured jet p_T to the p_T of the precisely measured Z/γ , as a function of jet p_T . The response,

$$R_{\text{jet}, p_T} = \frac{p_{T,\text{jet}}}{p_{T,\text{ref}}}$$

is then evaluated in data and in simulation. The ratio of the two, $R_{\text{data}}/R_{\text{MC}}$, defines the residual corrections.

The above description of jet energy corrections in CMS is meant as a rough, instructive summary only. A full description of the measurement techniques used in CMS can be found in [40].

5.5 Jet substructure reconstruction

In analyses looking for highly energetic (“boosted”) vector bosons, the opening angle between the vector boson quark decay products becomes so small that the highly boosted boson appears as a single large jet instead of two well-separated smaller jets. The distance between the two quarks,

in the case of a hadronic decay, depends on the mass of the vector boson and its p_T and goes as

$$\Delta R = \frac{2M_V}{p_{T,V}}. \quad (5.7)$$

Above a W boson p_T of 200 GeV, the two quarks are therefore merged into a single large cone jet of size $R = 0.8$. A sketch of the two different situation is shown in Figure 5.8. If the W p_T is well below 200 GeV, its decay products are two well-defined jets (left). However, once the W boson transverse momenta is approximately 200 GeV, both the quarks are completely contained within a single jet (right), referred to as a W jet. In order to distinguish jets from hadronically decaying

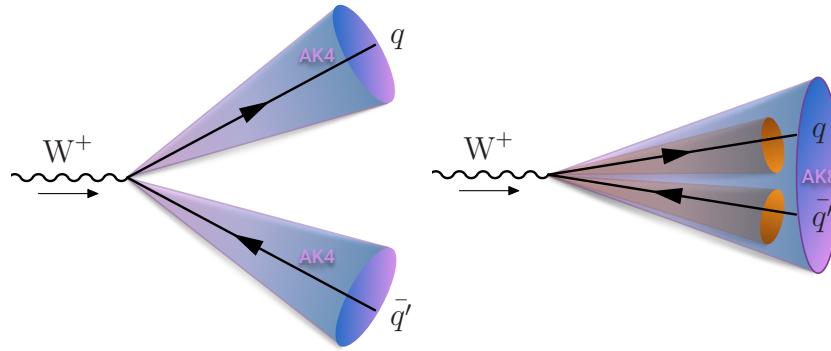


Figure 5.8: If the mass of the resonance is low enough, the quark decay products of each vector boson are well separated and clustered into distinguishable AK4 jets (left). If the transverse momentum of the vector boson is greater than 200 GeV, the vector boson decay products are merged into one single large cone AK8 jet.

vector bosons, either W or Z bosons, from those of quarks or gluons produced by QCD, the jet mass would in principle be a good discriminant since we know the W boson has a mass of around 80 GeV while the quark or gluon mass is close to zero. At very high transverse momenta, however, the width (and therefore the mass) of QCD jets may become equally large. In addition, diffuse radiation caused by the underlying event and pileup give rise to a significant number of additional particles in the event contributing to the total jet mass. Therefore, being able to accurately and efficiently separate highly boosted QCD jets from highly boosted vector bosons requires other methods. In order to remove the underlying event and pileup, algorithms like PUPPI can be used. In order to improve the mass resolution further, dedicated grooming algorithms must be applied.

5.5.1 Grooming

Grooming was introduced as a tool to improve the mass resolution of large radius jets without significantly changing the background and signal event numbers. It consists of removing the softest parts of a jet in order to resolve its “true” mass, by means of reclustering and identifying soft particles within the jet that can be removed.

Trimming

The trimming algorithm [41] is a grooming algorithm mostly used at trigger level in CMS (also where it is used in this thesis). It works in the following way: starting from a large jet clustered with either anti- k_T or C/A (in the case of CMS), it reclusters the jet using the k_T algorithm in order to create subjets of some size R_{sub} . It then proceeds to check whether each subjet has a momentum fraction above a certain threshold,

$$p_{T,i}/p_{T,jet} > p_{T,frac}.$$

If the subjet fails this requirement, it is removed. The remaining subjets are then assembled into a new “trimmed” jet. The effect of trimming on real W boson jets and QCD quark or gluon jets for different values of r_{sub} and $p_{T,frac}$ is shown in Figure 5.9. The best signal mass resolution is obtained with $r_{sub} = 0.2$ and $p_{T,frac} = 0.03$, which is also the parameter setting that provides the best signal discrimination from background by pushing the QCD jet mass closer to zero. These are the default values of the tuned parameters of the trimming algorithm in CMS ($r_{sub} = 0.2$ and $p_{T,frac} = 0.03$).

Pruning

The pruning algorithm, in addition to removing soft particles, has an additional requirement on the distance between any recombination that is at wide angle. It proceeds by reclustering the jet with the C/A algorithm, requiring at each step that

$$\frac{\min(p_{T,i}, p_{T,j})}{p_{T,P}} > z_{cut} \quad \text{and} \quad \Delta R_{i,j} < D_{cut} = \frac{2r_{cut}m_{jet}}{p_T}. \quad (5.8)$$

The first requirement is a requirement on the hardness of the combination. The variables $p_{T,i}$ and $p_{T,j}$ correspond to the transverse momenta of each protojet (single particle or group of particles already combined in a previous step) and $p_{T,P}$ is the combined p_T of the two. The protojet with the lowest transverse momenta is removed if its hardness is below z_{cut} , or if it forms an angle

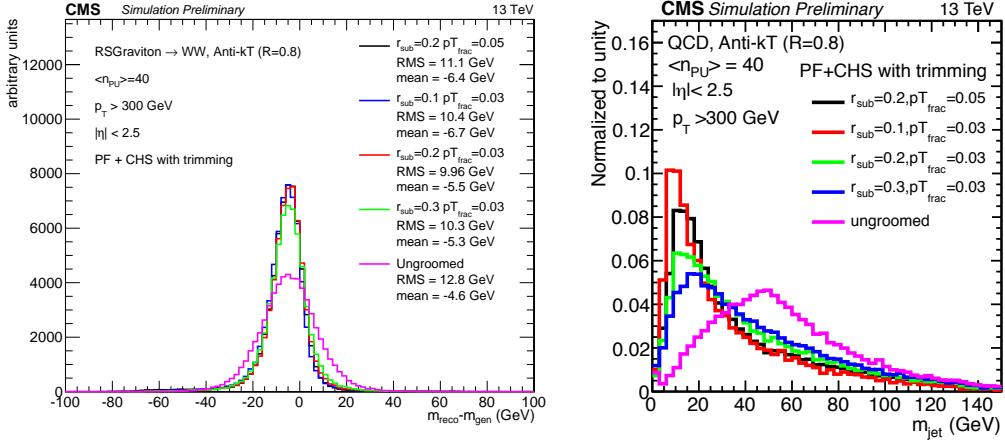


Figure 5.9: The effect of trimming on a signal jet (left) and a background jet (right) for different values of the tuned parameters r_{sub} and $pT,frac$ [42].

wider than D_{cut} relative to the axis of the recombination of the two protojets. In CMS, the tuned parameters are set to $r_{\text{cut}} = 0.5$ and where $z_{\text{cut}} = 0.1$. Figure 5.10 shows the ungroomed as well as the pruned jet mass distribution for signal (left) and background (right) jets. The highest amount of signal and background separation in CMS, is achieved with $r_{\text{cut}} = 0.5$ and $z_{\text{cut}} = 0.1$.

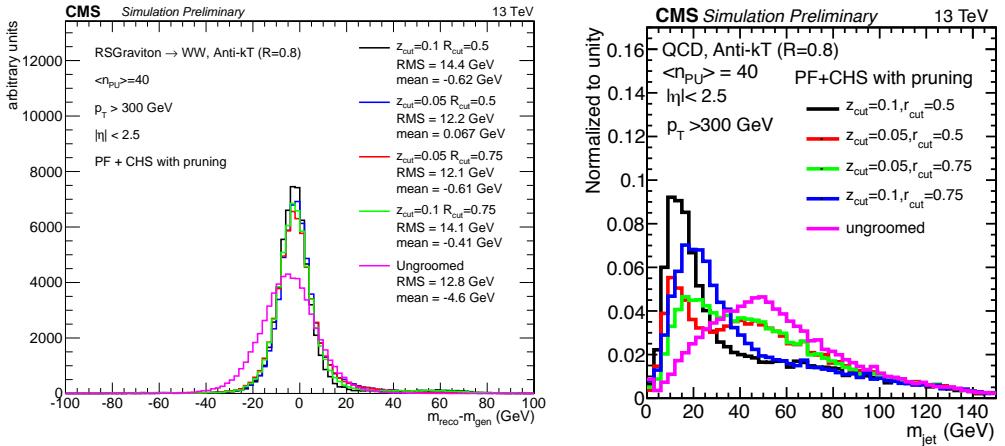


Figure 5.10: The effect of pruning on a signal jet (left) and a background jet (right) for different values of the tuned parameters z_{cut} and r_{cut} [42].

Modified Mass Drop Tagger and Soft Drop

The mass drop tagger (MDT) [43] is based on the idea that a W or Z jet is formed by two quark subjets and that, therefore, the mass of each subjet is much smaller than their combined mass (and much smaller than the mass of the boson itself). A QCD jet is, on the other hand, formed by continuous soft radiation, meaning that its heaviest subjet should be close to the mass of the jet itself. The MDT tagger therefore starts from a jet j clustered with the C/A algorithm and then declusters it again, defining subjets s_1 and s_2 , where $m_{s_1} > m_{s_2}$. If a significant mass drop occurred during declustering, $m_{s_1} < \mu m_j$, where m_j is the mass of the jet j , and the splitting is not too asymmetric, $\min(p_{T,s_1}^2, p_{T,s_2}^2) \Delta R(s_1, s_2) / m_j^2 > y_{cut}$, the jet j is selected as the tagged jet. Otherwise j is set equal to s_1 and the procedure starts over. The modified mass drop tagger (mMDT) [44] is a modification of MDT where the subjet with the largest transverse mass, $m^2 + p_T^2$, instead of bare mass, m , is followed. This fixes a flaw in the MDT tagger so that, in cases where the mass drop and asymmetry conditions are not satisfied, the more energetic rather than the heavier branch is followed. The recommended default version of mMDT additionally does not use the mass drop condition, $m_{s_1} < \mu m_j$, as it was shown to have a negligible impact on the jet mass distribution for QCD jets [44]. In addition, the mMDT algorithm by default uses a z_{cut} criteria, as for pruning, rather than a y_{cut} criteria. The modified mass drop condition is generalized through the soft drop declustering method [45], simply called Soft Drop, which allows for different types of angular requirements to enter the condition. The Soft Drop condition is the following,

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut} \frac{\Delta R_{12}^\beta}{R_0}, \quad (5.9)$$

where the asymmetry condition now is defined directly through the transverse momentum fractions of the subjets, rather than through a k_T distance to the jet mass. If the splitting is not too asymmetric, the condition is met and the full jet is deemed the softdrop jet. Otherwise only the highest- p_T subjet is kept and the declustering continues. If the jet can not be declustered any further, it can either be removed from consideration, so-called “tagging”-mode, or deemed the final soft-dropped jet, “grooming”-mode. A $\beta = 0$ corresponds to the modified mass drop tagger and removes all soft emission wider than the dominant two-prong substructure. For $\beta > 0$, soft radiation is removed, but some fraction of soft-collinear radiation is kept. Lastly, with $\beta < 0$, Soft Drop can remove soft as well as collinear radiation. The performance of Soft Drop on W jets and QCD quark/gluon jets for different values of β is shown in Figure 5.11. The modified mass drop tagger (Softdrop with $\beta=0$) with $z_{cut} = 0.1$ is the default Soft Drop settings in CMS, due to it providing the best signal/background discrimination while maintaining an excellent signal mass resolution.

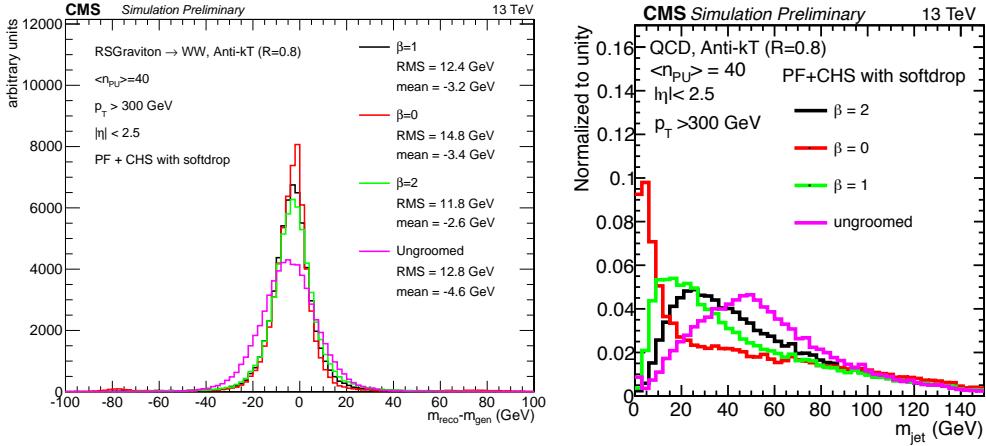


Figure 5.11: The effect of softdrop on a signal jet (left) and a background jet (right) for different values of the tuned parameters β . $\beta = 0$ corresponds to the Modified Mass Drop Tagger, which is the default Softdrop setting in CMS [42].

5.5.2 N-subjettiness

After using the algorithms above, there is still information in the jet structure itself that can distinguish W/Z jets from quark/gluon jets. A W or Z jet consists of two well-defined high- p_T subjets. A quark/gluon jet instead is made from a single parton, and consists of several large angle, asymmetric splittings, as illustrated in Figure 5.12. The N-subjettiness algorithm [46]

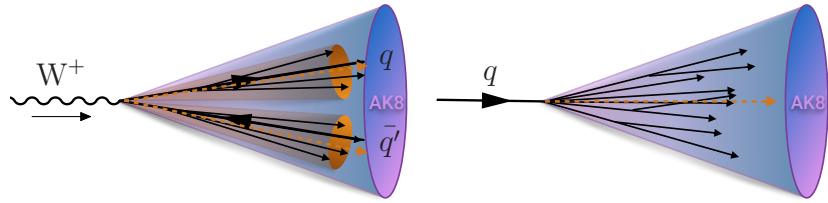


Figure 5.12: A jet stemming from the decay of a W will usually have two well-separated high- p_T subjets, while a jet with a single-prong origin consists of several large angel splittings.

takes advantage of this fact by attempting to count the number of hard sub-elements within a jet. This is quantified through the n-subjettiness variable, τ_N , defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}), \quad (5.10)$$

where k runs over all the jet constituents, $p_{T,k}$ is the constituent transverse momentum, and $\Delta R_{i,k}$ is the distance between the constituent and candidate subjet axes. These subjet axes are

obtained through a one-pass optimization procedure which minimizes N-subjettiness [47]. The normalization factor in front is given as

$$d_0 = \sum_k p_{T,k} R_0, \quad (5.11)$$

where R_0 corresponds to the cone size of the initial jet. With this definition, jets with $\tau_N = 0$ have most of their constituents aligned along the subjet axes. However, if $\tau_N \gg 0$, a large fraction of the energy is radiated away from the subjet directions and the jet is more likely to have more than N subjets. In CMS, and as recommended by the authors in [46], the ratio τ_2/τ_1 is used to discriminate W jets from QCD jets. The reason for this is that, while signal jets are expected to have a large τ_1 , quark/gluon can similarly have large τ_1 due to the diffuse radiation present. However, QCD jets with a large τ_1 tend to have an equally large τ_2 , while signal jets do not, hence the ratio of the two provides greater separation power. This can be seen from Figure 5.13. In CMS, the n-subjettiness algorithm is by default applied to ungroomed jets. The

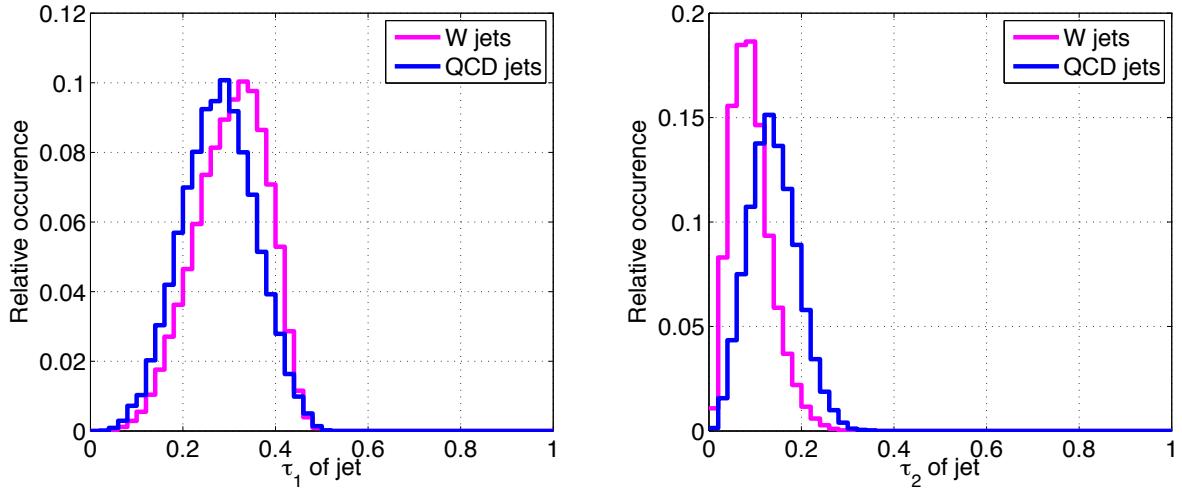


Figure 5.13: The distribution of the n-subjettiness variables τ_1 (left) and τ_2 (right) for signal jets and background jets. On their own, the τ_N variables provide little discriminating power [46].

distribution of τ_{21} for signal and background jets with different pileup subtraction algorithms applied is shown in Figure 5.14, where τ_{21} in combination with PF+PUPPI (green), yields a distribution most similar to the generated one (black).

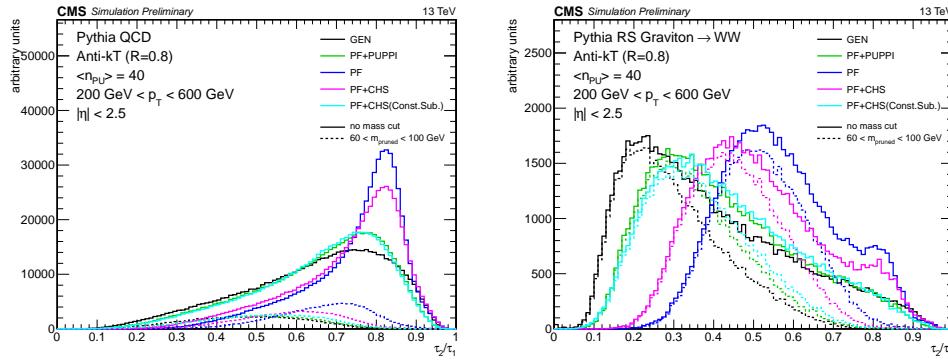


Figure 5.14: The distribution of the n-subjettiness ratio τ_{21} for signal jets (left) and background jets (right) with different combinations of pileup subtraction algorithms applied. The solid lines corresponds to the τ_{21} distribution with no mass cut applied, while the dotted lines are within a mass window of 60-100 GeV [42].

5.5.3 Decorrelated n-subjettiness

The n-subjettiness variable described above is highly correlated with the jet mass. This leads to an undesired sculpting of the groomed jet mass distribution when applying a flat cut on τ_{21} . The softdrop jet-mass distribution for gluon jets after successive cuts on τ_{21} is shown in Figure 5.15. This has the undesired effect of making the QCD background more peak-like rather

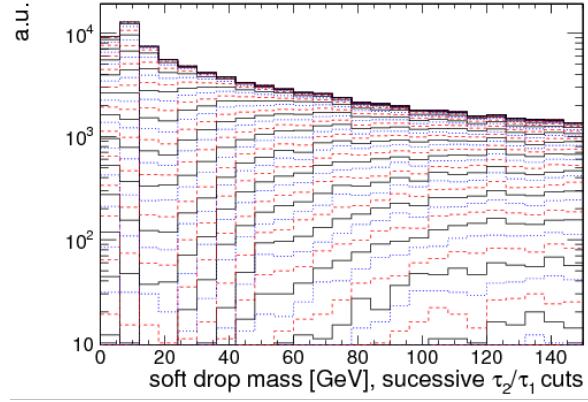


Figure 5.15: The softdrop jet-mass distribution for gluon jets after successive cuts on τ_{21} in steps of 0.02, going from 1.0 to 0.0 [48].

than smoothly falling, which consequently makes QCD more difficult to distinguish from the W-jet signal. The shape and position of this sculpted mass peak depends on the jet p_T , which makes background estimation through jet mass sidebands difficult. In order to solve both of

these issues, a tagging observable that is uncorrelated with the jet mass and that maintains a similar mass shape independent of the jet p_T was proposed, called *Designing decorrelated taggers (DDT)* [48]. This decorrelation is done by flattening the τ_{21} profile dependence on ρ' , where ρ' is defined by

$$\rho' = \log(m^2/p_T/\mu), \text{ where } \mu = 1 \text{ GeV.} \quad (5.12)$$

The τ_{21} profile distribution as a function of ρ' for W-jets and QCD-jets is shown in Figure 5.16. There is no dependence of the τ_{21} profile of QCD jets on the jet p_T when plotted versus ρ' , and

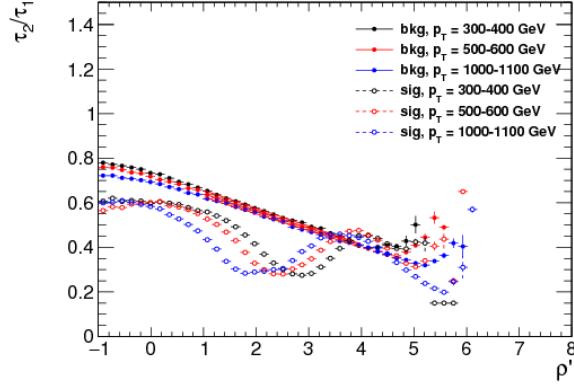


Figure 5.16: The τ_{21} profile distribution as a function of ρ' for W-jets (hollow markers) and QCD-jets (solid) in different bins of jet p_T [48].

there exists a linear relationship between τ_{21} and ρ' . In order to design a mass- and p_T -decorrelated τ_{21} variable, τ_{21}^{DDT} , a linear transformation of τ_{21} is defined as

$$\tau_{21}^{DDT} = \tau_{21} + M \times \log\left(\frac{m^2}{p_T \times 1 \text{ GeV}}\right), \quad (5.13)$$

where $M = -0.063$ is obtained from a fit of the τ_{21} profile against the variable ρ' shown in Figure 5.16. The τ_{21}^{DDT} profile distribution as a function of ρ' for W-jets and QCD-jets is shown in Figure 5.17, displaying the desired decorrelated behavior. The effect on the jet groomed mass after applying a flat cut on τ_{21}^{DDT} is shown in Figure 5.18. The sculpting of the jet mass is significantly reduced through the decorrelating procedure, resulting in a tagger with a minimal mass- and p_T -dependence.

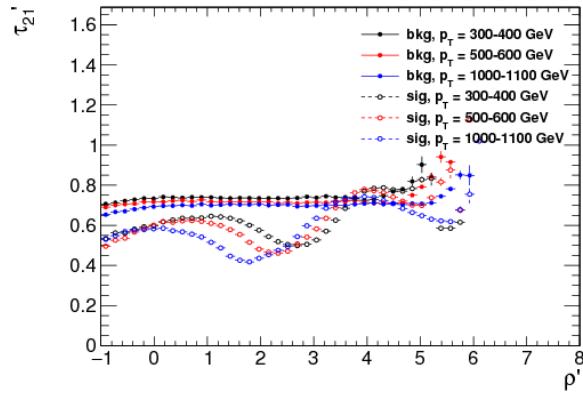


Figure 5.17: The τ_{21}^{DDT} profile distribution as a function of ρ' for W-jets (hollow markers) and QCD-jets (solid) in different bins of jet p_T [48].

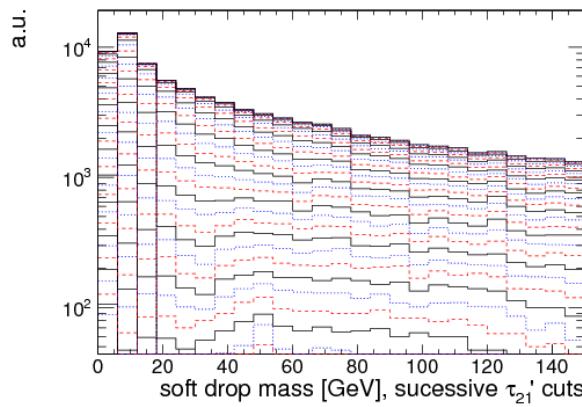


Figure 5.18: The softdrop jet-mass distribution for gluon jets after successive cuts on τ_{21}^{DDT} in steps of 0.02, going from 1.0 to 0.0 [48].

5.5.4 Vector boson tagging

In order to discriminate W and Z bosons from quark/gluon jets a combination of a groomer and shape-tagger (like n-subjettiness) is usually used. Typical values for tagging W jets is a groomed jet mass between 60 and 100 GeV and $\tau_{21} < 0.5$. The exact combination and value of cuts is analysis dependent, and has been optimized for each search presented in this thesis. The details are thoroughly explained in each section.

Polarization effects

The vector boson polarization has a significant effect on the W-tagging efficiency. The helicity angle θ^* , defined as the angle between the outgoing quark daughters of the W boson in its rest frame relative to its direction of motion [49], is very different for longitudinally polarized vector bosons, W_L , and transversely polarized vector bosons W_T [50]. Figure 5.19 shows the $\cos \theta^*$ distribution for the outgoing quarks from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (red) decays, and it can be observed that transversely polarized W bosons decay with the quarks emitted closer to the vector boson direction of motion. The consequence of this, is that there is a higher asymmetry

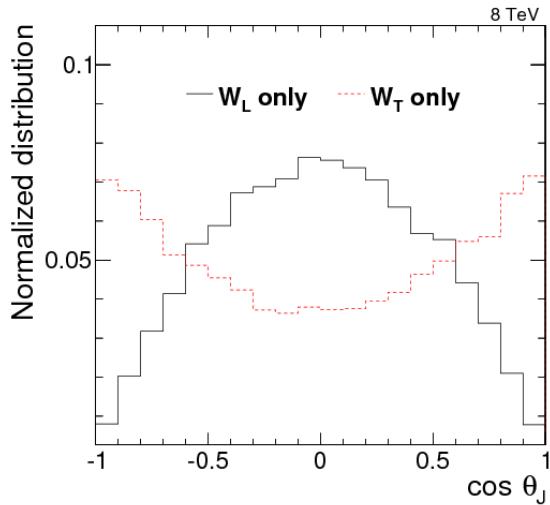


Figure 5.19: The helicity angle for generated quarks from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (red) decays [50].

in the transverse momenta of the two quarks from a W_T decay. This in turn causes grooming algorithms, designed to remove soft constituents of a jet, tend to reject particles coming from the softer quark, resulting in a lower jet mass and a drop in tagging efficiency. Figure 5.20 shows the

W-jet tagging efficiency versus q/g jet mistagging rate for a selection on the jet pruned mass of $60 \text{ GeV} < m_{\text{pruned}} < 100 \text{ GeV}$, scanning τ_{21} cuts (here for CA R=0.8 jets).

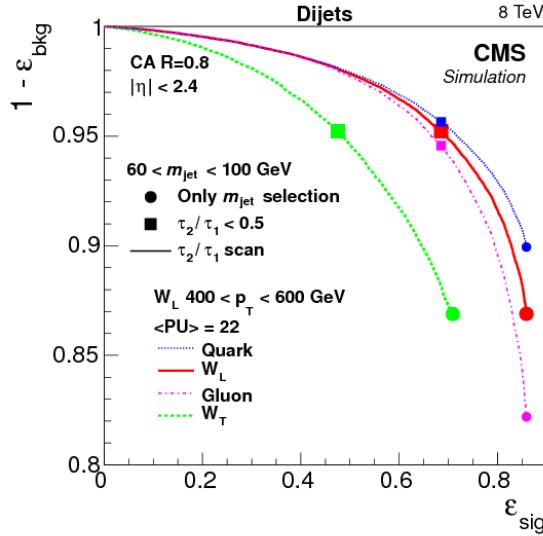


Figure 5.20: The helicity angle for generated quarks from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (red) decays [50].

The tagging efficiency for transversely polarized W bosons (green) is significantly lower than the tagging efficiency for longitudinally polarized bosons (red). This can be explained by looking at the $\cos\theta^*$ distribution on reconstructed level, using the C/A subjets, with a cut on the jet pruned mass of $60 \text{ GeV} < m_{\text{pruned}} < 100 \text{ GeV}$, as shown in Figure 5.21. When comparing to the distribution at generator level with no groomed mass window applied, Figure 5.19, one can see that the W_T jets with $\cos\theta^* \sim 1$ are completely removed. This is due to two effects: the p_T -asymmetry explained above and the fact that the ΔR distribution between the two quarks is much smaller in the case of W_L , making them more likely to be fully contained within a jet cone of R=0.8.

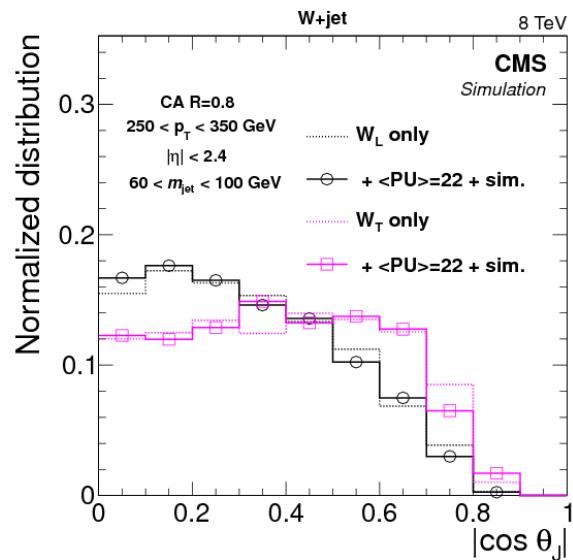


Figure 5.21: The helicity angle for subjets from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (pink) decays [50].

5.6 Monte Carlo Event Generators

Monte Carlo event generators offer a realistic estimate of high-energy collisions on an event-by-event basis, allowing us to estimate signal and background processes accurately. Simulated events are usually produced in three steps, beginning with the hard process through hadronization and decay. First, a matrix element generator simulates the hard scattering process and subsequent decays. Secondly, the showering and hadronization of unstable particles is performed and, lastly, the final-state particles are passed through a full detector simulation in order to reproduce a range of experimental effects.

General-purpose Monte Carlo (GPMC) generators, like HERWIG ++ [51] and PYTHIA 8 [52], deal with both perturbative as well as hadronization phenomena, simulating an event all the way up until detector simulation. In HERWIG ++ and PYTHIA 8, the hardest processes are only simulated at the lowest order of perturbative expansion, meaning $2 \rightarrow 2$ or $2 \rightarrow 3$ scatterings. In order to have tree-level matrix elements with an arbitrary final-state multiplicity, they can be combined with programs used to generate parton-level events at higher accuracy, which are then processed through showering and hadronization with the GPMC generators. One popular program for generating matrix elements is MADGRAPH [53]. This, however, still correspond to a tree-level (leading order) approach. To go to next-to-leading-order (NLO), meaning the inclusion of virtual corrections, two methods exist: MC@NLO [54, 55] and POWHEG [56]. These combine the full next-to-leading-order prediction for inclusive processes with the subsequent parton showers, either by a subtraction method regularizing the real contributions, or by a matrix-element correction of the parton shower branching probability. After hadronization, all final state particles are passed through a full simulation of the CMS detector. This is done with GEANT4 [57], which models the interaction and showering of particles with materials, and outputs position-dependent energy deposits.

To simulate the effect of additional proton-proton collisions within the same or adjacent bunch crossings (pileup), additional inelastic events are generated using PYTHIA and superimposed on the hard-scattering events. The simulated MC events are finally weighted to reproduce the distribution of the number of pileup interactions observed in data.

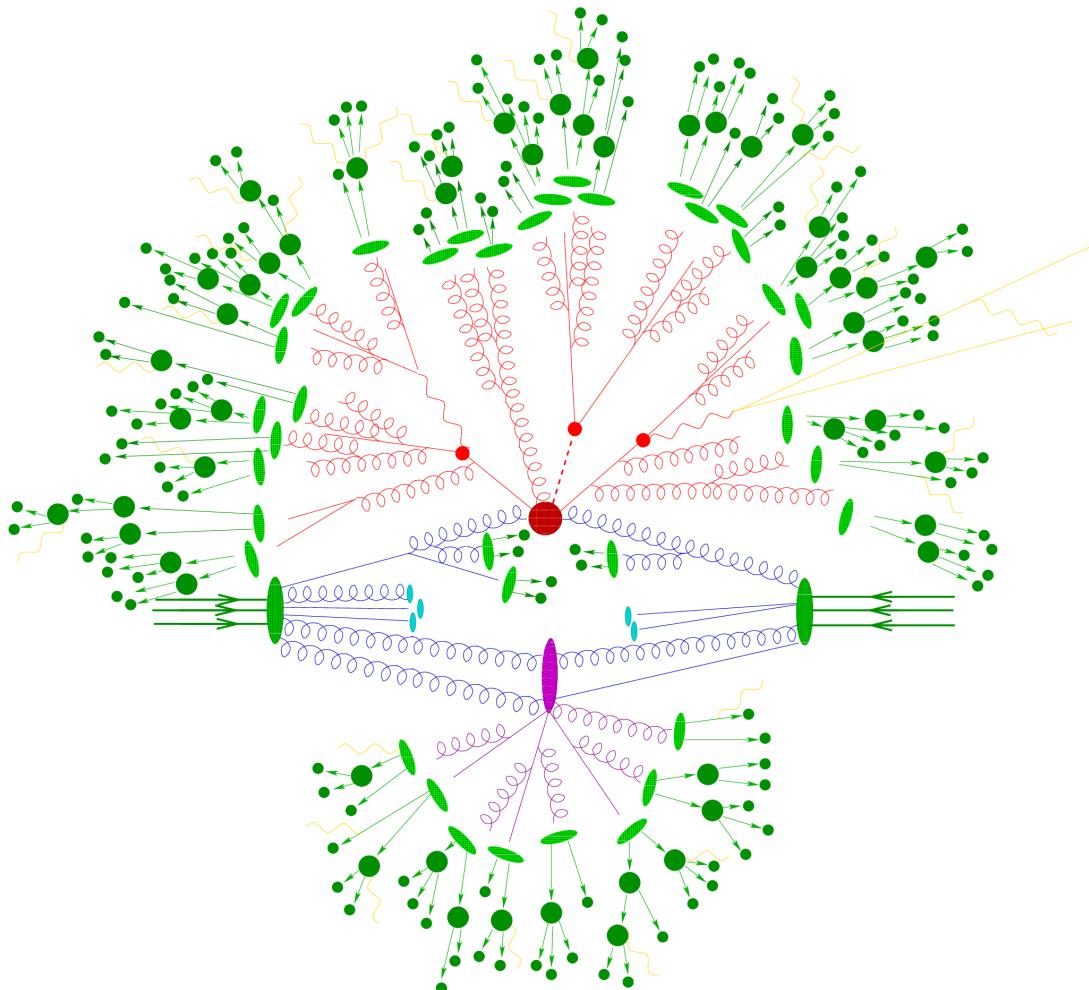


Figure 5.22: An illustration of a hadron-hadron collision as simulated by a MC generator. The hard collision center corresponds to the red circle and the branches represent parton showers. The purple circle is a second hard scattering event. The dark green circles represent the decay of hadrons and the yellow lines illustrate photons.

Part III

Diboson resonance searches in CMS

CHAPTER 6

Search I: First search for diboson resonances at 13 TeV

When the LHC started its Run II data taking period in summer 2015, it would be the first time ever for a particle collider to produce collisions with center-of-mass energies as high as 13 TeV. The Higgs boson, for which the LHC was designed to observe, had been discovered at the end of the previous data taking era, leaving us with a Standard Model that we know is either in need of extensions or only an effective theory valid in a certain energy domain. The Run II search program would therefore be oriented around two main efforts: Precision measurements of the newly discovered Higgs boson and searches for physics beyond the standard model.

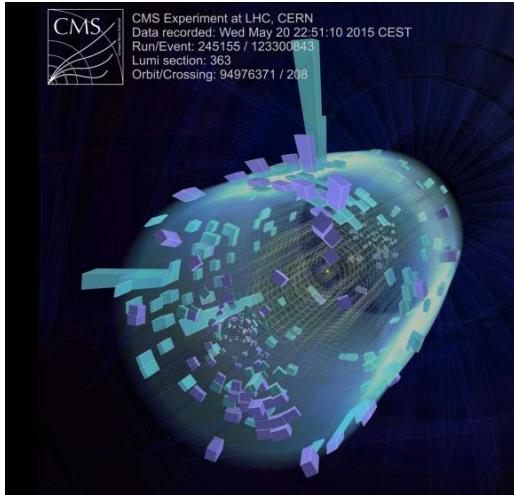
I started my PhD four months before the first 13 TeV collisions took place and had to consider the following: What was the most interesting search that could be done on a short time scale (to be presented 6 months after first collisions, which would be at the CERN end-of-year “Jamboree”), whose physics objects could be reconstructed and understood on a relatively short time scale, and would be robust enough in case there were issues with the never-before-validated 13 TeV Monte Carlo?

The attention of the high-energy physics community has in the past years been focused on certain “hot topics”: In 2018 and currently in 2019, the excitement is over leptoquarks, which could explain anomalies observed by LHCb and b-factories; in 2016 and 2017 it was diphoton resonances, with $> 3\sigma$ excesses observed at the same mass in both CMS and ATLAS. And in 2015 during the 13 TeV LHC start-up, attention was centered on diboson resonances in the all-hadronic final state. The choice was therefore clear: My first analysis would be a search for diboson resonances in the boosted dijet final state. With a background model based on a smooth fit to data in the signal region,

eliminating the need for accurate QCD MC predictions, this was a simple one-background only (QCD) analysis, feasible to finalize in one year, given dedication and sufficient effort. Despite its straightforwardness, due to observed 8 TeV excesses, it was in addition considered a high-profile analysis.

This search became one of the first “boosted” searches published with data collected with a 13 TeV center-of-mass energy, as well as the first search to take advantage of dedicated “grooming” triggers. It was published with 2.7 fb^{-1} of 2015 data.

- The author -



Published in the Journal of High Energy Physics (2017), DOI: [10.1007/JHEP03\(2017\)162](https://doi.org/10.1007/JHEP03(2017)162)

6.1 A small bump

On June 2nd, 2015, the day before CMS recorded its first ever 13 TeV event, a pre-print appeared on the arXiv titled, “Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector” [53]. It was an analysis of the full ATLAS Run 1 dataset, corresponding to 20.3 fb^{-1} , searching for heavy resonances decaying to vector bosons in the all-hadronic state. The analysis documented a 3.4σ excess for a heavy resonance decaying to WZ with a mass of around 2 TeV. The corresponding CMS analysis, published the previous year, had a 1.3σ excess at roughly the same resonance mass, but was mostly compatible with a WW final state hypothesis [58]. Figure 6.1 shows the corresponding dijet invariant mass spectrum as seen by ATLAS (left) and the upper limit on the production times the cross section for a G_{Bulk} decaying to WW (right) as documented by CMS.

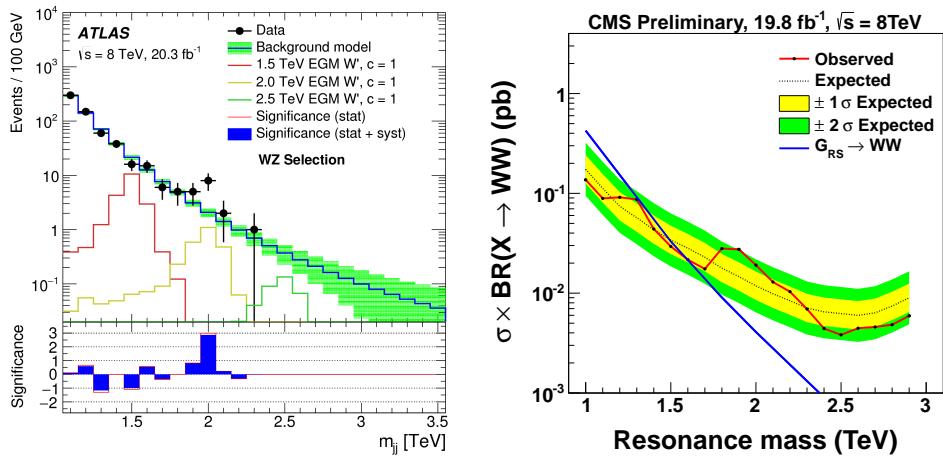


Figure 6.1: A “bump” corresponding to 3.4σ in the dijet invariant mass spectrum around 2 TeV (left) observed by ATLAS when analyzing the full 8 TeV dataset [53], together with a similar excess (1.3σ) observed in the corresponding CMS analysis [58].

The two measurements were found to be compatible, favoring a heavy resonance with a production cross section of around 5 femtobarn and a mass between 1.9 and 2.0 TeV decaying to either WW, WZ or ZZ [59]. Figure 6.2 shows the obtained p-value from the ATLAS (red) and CMS (blue) searches, as well as their combination (black). In addition to the observed excesses in the vector boson final states, another 3σ excess for a resonance with a mass of 1.8 TeV had been observed in the search for heavy resonances decaying to a W and a Higgs boson [60] at 1.8 TeV. The combination of the excesses and the timing of the ATLAS paper, naturally led to some excitement, and in the coming weeks, the arXiv was flooded with theory papers attempting an

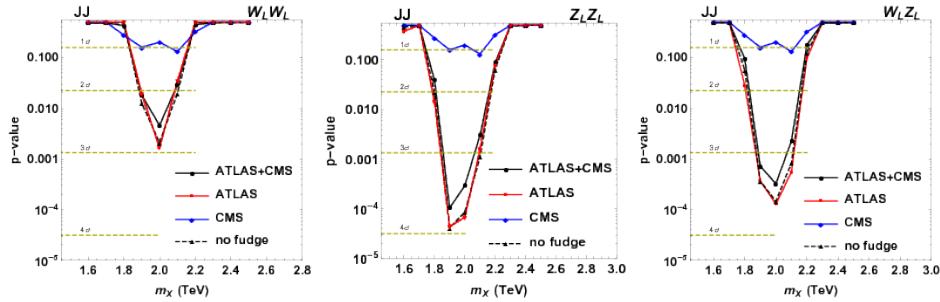


Figure 6.2: p-values as a function of resonance mass obtained with an emulation of the ATLAS (red) and CMS (blue) searches as well as the combination of the two (black). Here for a W_W (left), W_Z (middle) and Z_Z (right) hypothesis [59].

explanation of the deviations.

In addition, one of the main benefits of increasing the LHC center-of-mass energy from 8 to 13 TeV was that the partonic luminosity would increase. One could therefore expect the same number of signal events in the 20 fb^{-1} data set collected with a center-of-mass energy of 8 TeV, for a considerably smaller luminosity with a center-of-mass energy of 13 TeV. Figure 6.3 shows the system mass that can be probed with the expected 2015 integrated luminosity of 3 fb^{-1} collected with a center-of-mass energy of 13 TeV, as a function of the probe-able mass with 20 fb^{-1} of 8 TeV data for different partonic channels of qq , qg , and gg . For example, a 2 TeV mass resonance would be observable in both datasets. We therefore expected that the small excess observed in the VV all-hadronic final state would be observable in the 2015 dataset if the signal was genuine.

6.2 Analysis strategy

When a resonance X with a mass above 1 TeV decays into a vector-boson pair, the bosons have a very high energy ($\tilde{p}_T = M_X/2 = 500 \text{ GeV}$, assuming X is produced at rest) is referred to as boosted. The decay products of a hadronically decaying boosted vector boson will therefore not appear as back-to-back in the lab frame but rather be collimated, as described in Section 5.5. This results in a final state with two high-pt, large-radius jets, such that the AK algorithm with an $R=0.8$ is expected to fully contain the two quarks coming from the vector boson decay. This is illustrated in Figure 6.4. The two jets are each expected to have a mass around the W or Z boson mass, and some intrinsic substructure stemming from their two-pronged decay. The invariant mass of the dijet system, m_{jj} , should be roughly equal to the resonance mass M_X . This dijet system is the final state under scrutiny and the dijet invariant mass is the parameter of interest.

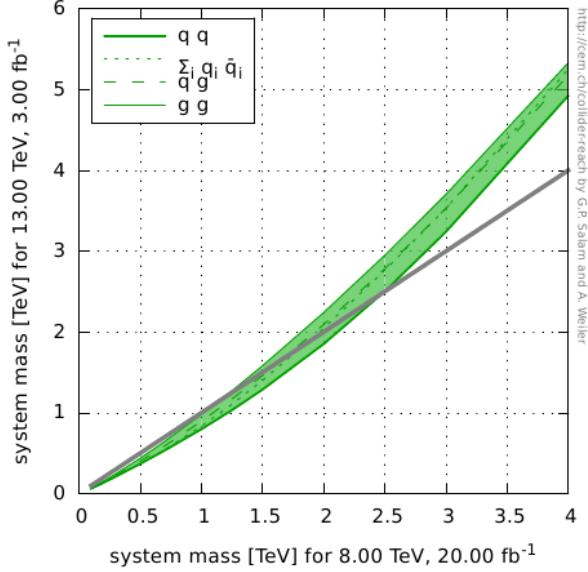


Figure 6.3: The system mass that can be probed with 3 fb^{-1} of 13 TeV data (y-axis) as a function of the probeable system mass with 20 fb^{-1} of 8 TeV data (x-axis) for different partonic channels (generated with [61]).

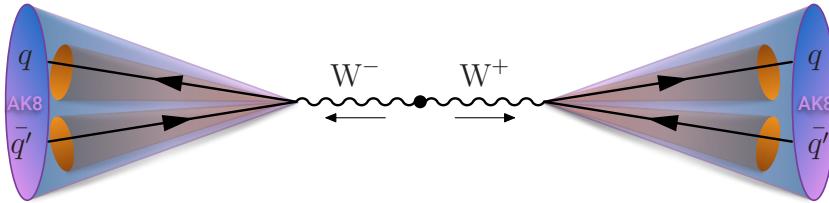


Figure 6.4: If a heavy ($> 1 \text{ TeV}$) resonance decays into vector bosons, the transverse momentum of each boson will be large and its decay products are merged into one single large cone AK8 jet.

The final states of WW, ZZ, and WZ would produce similar final states.

The main background for such an analysis is QCD multijet events. As mentioned in Section 5.5, quark/gluon jets can obtain a high mass due to diffuse radiation and QCD processes have such a large cross section that the number of QCD jets with a mass compatible with the W mass can be large. In order to discriminate between the two, we take advantage of three properties. First, the groomed mass of signal and background jets should be very different. Second, signal jets should appear two-prong like, as opposed to quark/gluon jets, and third, the dijet invariant mass for the signal process should peak around the resonance mass while the QCD spectrum is predicted to be smoothly falling. Section 6.5 explains this assumption in more detail. The strategy therefore

consists of performing a smoothness test on m_{jj} of the observed data, a so-called “bump-hunt”, by assuming that the signal will appear as a bump on top of a smooth distribution. This is illustrated in Figure 6.5. The benefit of such a method is that there is no need for a simulation

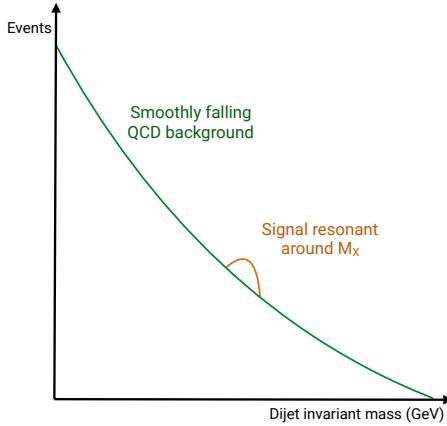


Figure 6.5: The search strategy consists of looking for signal “bumps” in the dijet invariant mass on top of a smoothly falling QCD multijet background.

of the background and the strategy is simple and robust. The disadvantage is that the analysis is intrinsically limited to regions where the dijet invariant mass spectrum is smooth and hence regions with discontinuities due to trigger turn-ons or kinematic selections must be avoided.

6.3 Data and simulated samples

The data analyzed in this search correspond to a total integrated luminosity of 2.7 fb^{-1} collected at a center-of-mass energy of 13 TeV between June and December 2015. The instantaneous luminosity of the LHC during this run was around half of the design luminosity ($0.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$), with an average number of primary vertices per event of $\langle \mu \rangle = 13$.

The bulk graviton model (see Section 2.3) and the HVT model (W' and Z' , see Section 2.2.1) are used as benchmark signal processes. In these models, the vector gauge bosons are produced with a longitudinal polarization in more than 99% of the cases, which leads to a 24% higher acceptance per boson for reasons explained in Section 5.5.4. For the HVT model, a scenario (model B) with $g_V = 3$, $c_H = -0.976243$, and $c_F = 1.02433$ is chosen, where the heavy resonance predominantly couple to bosons and the coupling to fermions is suppressed. The bulk graviton samples were generated with $\tilde{k} = 0.5$. The resonance masses considered lie in the range 1.2 to 4 TeV and are generated under the assumption of a natural width negligible with respect to

the experimental resolution (narrow-width approximation). All signal samples are generated at leading order with `MADGRAPH5_AMC@NLO v2.2.2` [62].

Simulated samples of the production of QCD multijet events are generated to leading order using PYTHIA version 8.205 [63] with the CUETP8M1 tune [64] and are used to validate the analysis procedure.

6.4 Event selection

6.4.1 Triggering

The first selection to be confronted in any analysis is the trigger selection. Due to an overwhelming QCD background in all-hadronic final states, the threshold for fully-hadronic triggers is very large in order to keep the trigger rate low (preferably around 10-30 Hz). In this analysis, we therefore decided to take advantage of triggers that place requirements on the jet’s groomed mass in addition to the “standard” jet triggers based on the scalar sum of jet transverse energy H_T . These “boosted” triggers were never before tested in data, and this analysis was the first published result taking advantage of grooming at the trigger level in CMS. The following H_T -based High Level Triggers (HLT), referred to as inclusive triggers in the following, are used:

- HLT_PFT650_WideJetMJJ900DEtaJJ1p5,
- HLT_PFT650_WideJetMJJ950DEtaJJ1p5, and
- HLT_PFT800.

Here, *PFT650* refers to a total H_T of at least 650 GeV. *WideJet* means jets reconstructed with the *wide jet algorithm* [65], an algorithm designed to improve the jet response for gluon-initiated jets by picking up final state radiation around the gluon. The two AK R=0.4 jets with the largest p_T in the event are used as seeds. Geometrically close jets are then combined into the closest jet seed if they are within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, and these two jets form a dijet system used for further selections. *MJJ900* refers to a wide-jet dijet mass of at least 900 GeV, and *DetaJJ1p5* means there is an additional cut on the $|\Delta\eta|$ between the two wide jets for reasons that will be explained below. In addition, two triggers based on jet grooming are used. These require a trimmed jet mass (see Section 3.5.1) of 30 and 50 GeV together with a selection on the jet p_T or total event H_T , yielding the triggers:

- HLT_AK8PFJet360_TrimMass30 and
- HLT_AK8PFHT700_TrimR0p1PT0p03Mass50.

The tunable parameters for the trimming algorithm are set to $r_{sub} = 0.2$ and $p_{T,frac} = 0.03$. The trigger requiring a trimmed jet mass of at least 30 GeV, in addition requires a jet p_T of at least 360 GeV, while the second trimmed mass trigger requires a total H_T above 700 GeV. The TrimMass30 trigger is seeded by single-object Level 1 triggers with jet p_T thresholds of 176 or 200 GeV, and all the remaining triggers require an online H_T of at least 150 or 175 GeV.

In order to avoid any kinks in the dijet invariant mass spectrum due to the presence of a trigger turn-on, we determine the dijet invariant mass at which the analysis triggers are fully efficient ($> 99\%$), and only consider signal events above this value. In order to estimate the trigger efficiency, we use a trigger with a lower H_T threshold of 650 GeV as a reference trigger. This trigger has a prescale of 40, meaning events are only recorded one out of 40 times. It is seeded by L1 H_T triggers with thresholds of 150 or 175 GeV. We then define the efficiency as

$$\text{Efficiency} = \frac{N_{trigger+ref}}{N_{ref}}$$

where $N_{trigger+ref}$ corresponds to the number of events passing the trigger under study as well as the reference trigger, and N_{ref} corresponds to the number of events passing the reference trigger. Figure 6.6 shows the trigger turn-on curves as a function of dijet invariant mass for jets where one of the jets is required to have a pruned mass larger than 65 GeV (in other words, compatible with a W jet). A sharp turn-on for the inclusive triggers (top left) is observed, reaching the 100% efficiency plateau for dijet masses of around 1.0–1.1 TeV. The grooming triggers, however, turn on more slowly and are not fully efficient until dijet invariant masses reach around 1.2 TeV (top right). The real power of the grooming triggers become clear when considering them in addition to the H_T -based triggers. The bottom plot in Figure 6.6 compares the trigger turn-on curves as a function of dijet invariant mass for jets passing one of the three inclusive triggers only, one of the grooming triggers only, and when combining all of them. Here, one can see that the 99% efficiency threshold is lowered by 75 GeV when including the substructure triggers, once substructure is required at analysis level. This combination of triggers allowed the analysis to consider dijet invariant masses as low as 1 TeV. As a measure of the performance of the grooming triggers, we have in addition looked at the trigger efficiencies as a function of the offline groomed mass (using the pruned and softdrop algorithms described in Sections 5.5.1 and 5.5.1), for the grooming trigger with the lowest mass threshold (30 GeV). This is shown in Figure 6.7, where an additional cut on the jet transverse momentum of one of the jets of 600 GeV is required and no other mass cut is applied. The trigger plateau is reached for offline groomed-jet masses around 50 GeV.

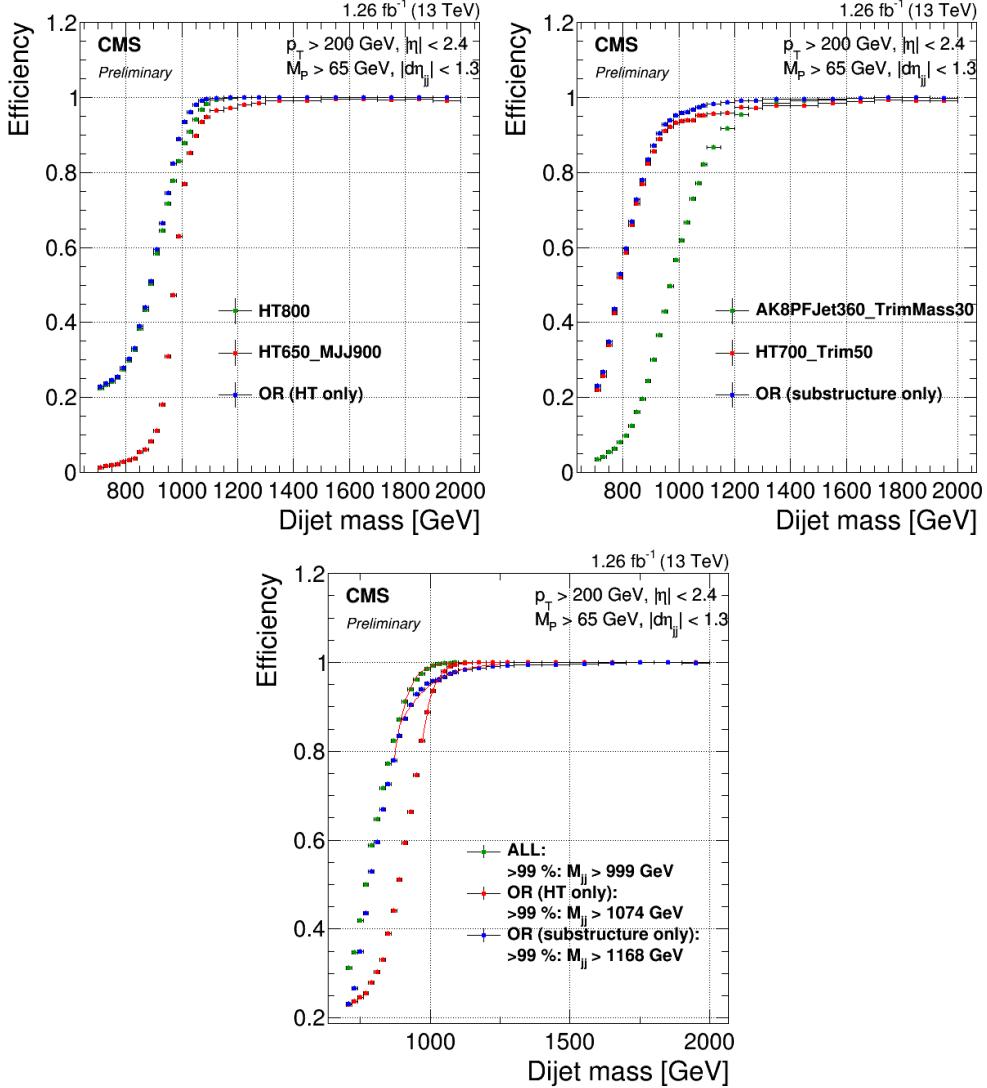


Figure 6.6: Top: Efficiency for the inclusive triggers (top left) and the grooming triggers (top right) as a function of dijet invariant mass for jet pairs where one jet has a pruned mass larger than 65 GeV. Bottom: Comparison of trigger efficiencies for jets passing one of the HT-triggers only (red), for jets passing one of the grooming-triggers only (blue) and for jets passing one of the HT-triggers or one of the grooming triggers (green). Here as a function of dijet invariant mass for all jet pairs passing loose selections and where one jet has a pruned mass larger than 65 GeV. The 99% efficiency threshold is lowered by 75 GeV when including substructure taggers.

6.4.2 Preselection

After trigger selections, and the corresponding requirement of a dijet invariant mass above 1 TeV to ensure a smooth falling background, we begin the process of maximizing the signal significance

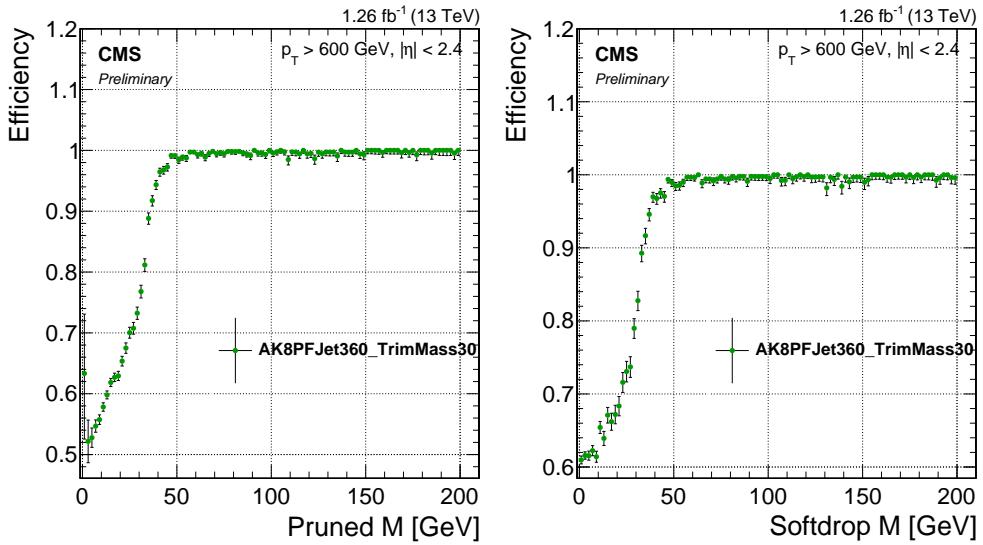


Figure 6.7: Efficiency for the lowest threshold grooming trigger as a function of pruned-jet (left) and softdrop-jet (right) mass for jets with $p_T > 600$ GeV.

while keeping the background low. This is done by optimizing the selection requirements on the jets. The jets used in this analysis are clustered with the anti- k_T jet clustering algorithm with a clustering parameter of $R = 0.8$ (see Section 5.4) to allow containment of the full vector-boson decay products. Since a minimum transverse momentum of 200 GeV is required for the decay products of a W/Z to be fully contained within an $R=0.8$ jet, events are further selected by requiring at least two jets with $p_T > 200$ GeV. These are in addition required to be central, with an $|\eta| < 2.4$. The two highest p_T jets in the event passing these criteria are selected as potential vector boson candidates. As our main background is QCD multijet events, we further take advantage of the fact that the angular distribution between these, mainly t-channel, processes are very different from the s-channel signal processes under study. The crosssection for QCD t-channel processes as a function of the opening angle with respect to the beam axis (θ^*), exhibit a pole around $\cos \theta^* = 1$, meaning QCD t-channel jets are mostly produced in the forward direction, with an opening angle with respect to the beam axis close to zero. The signal jets on the other hand, produced through an s-channel process, are concentrated in the barrel region. We therefore require the jets to have a separation of $|\Delta\eta| < 1.3$ in order to reduce the QCD multijets background. The distribution of $|\Delta\eta|$ between the two highest- p_T jets for QCD, as well as for different signal scenarios, is shown in Figure 6.8. A cut of $|\Delta\eta|_{jj} < 1.3$ removes the t-channel pole at $\cos \theta^* = 1$ and is in addition found to yield the highest signal sensitivity. In addition to these requirements on the jets themselves, a veto on jets overlapping with leptons is applied in

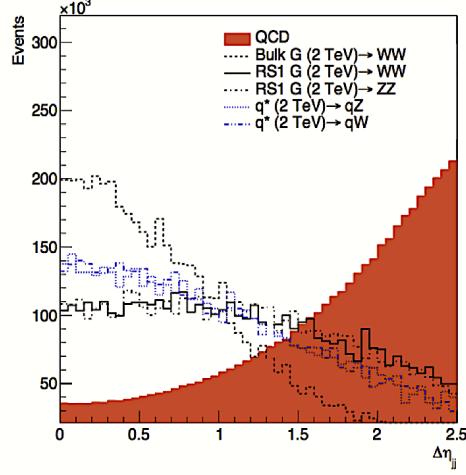


Figure 6.8: $|\Delta\eta|$ between the two highest- p_T jets for QCD jets and jets stemming from different signal scenarios.

order to ensure orthogonality between the corresponding semi-leptonic analysis. Here the overlap $\Delta R(\text{jet}, \text{lepton})$ between the jet candidate and a lepton is required to be larger than 0.8. Leptons used for this veto are required to pass the identification requirements described in Section 5.2.2 and 5.2.2, have a transverse momentum larger than 35 (30) GeV, and a pseudorapidity smaller than 2.5 (2.4) in the case of electrons (muons). The p_T , η , dijet invariant mass, and $|\Delta\eta|_{jj}$ distribution for the two leading jets in the event after the above preselections have been applied is shown in Figure 8.6.

6.4.3 Vector boson tagging

After preselections, we take advantage of the jet substructure algorithms described in Section 5.5 to further separate boosted W/Z jets from the QCD multijet background. In the 8 TeV analysis [58] published the previous year, the pruning algorithm was the chosen grooming algorithm of CMS. However, recent progress had been made in the development of alternative grooming algorithms that had favorable properties from a theoretical point of view (see Sections 5.5.1 and 7.5). We therefore studied two different grooming algorithms: pruning and softdrop (with $\beta = 0$ and $z_{cut} = 0.1$). A comparison of the jet mass for W, Z, and H jets after either the softdrop (dotted lines) or pruning (solid lines) algorithms were applied is shown in Figure 6.10. One of the first observations we made comparing the two grooming algorithms was that there appeared to be a strong dependence of the softdrop mass on the jet p_T . Figure 6.11 shows the pruned (left) and softdrop (right) mass distributions for W jets coming from the decay of a G_{bulk} with a

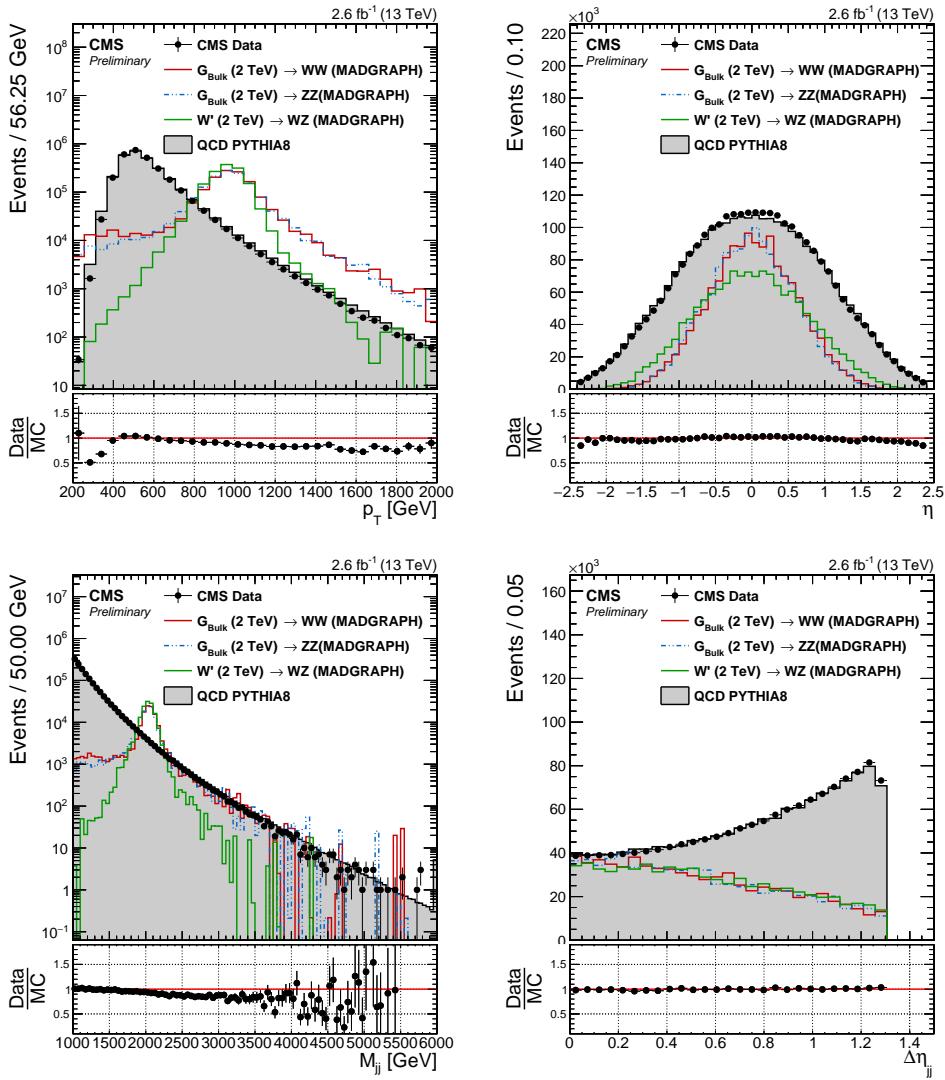


Figure 6.9: Jet p_T (top left), η (top right), dijet invariant mass (bottom left) and $|\Delta\eta|_{jj}$ (bottom right) distribution for the two leading jets in the event after loose preselections are applied. The signal is scaled by an arbitrary number.

resonance mass of $0.8 \text{ TeV} < M_X < 4 \text{ TeV}$. While the pruned jet mass mean appeared stable as the jet transverse momenta of the jet increased ($p_T \sim M_X/2$), the mean of the softdrop jet mass shifted towards lower values as jet p_T increased. In order to investigate whether this was a reconstruction effect or an algorithmic effect, we additionally looked at the pruned and softdrop mass for generator-level jets (jets clustered with generator-level particles not passed through the detector simulation). Figure 6.12 shows the reconstructed (solid line) and generator-level (dotted

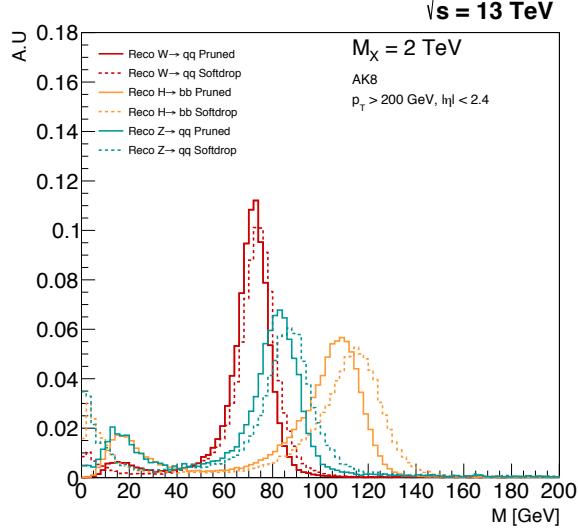


Figure 6.10: The softdrop (dotted lines) and the pruned (solid lines) jet mass for W, Z and H jets.

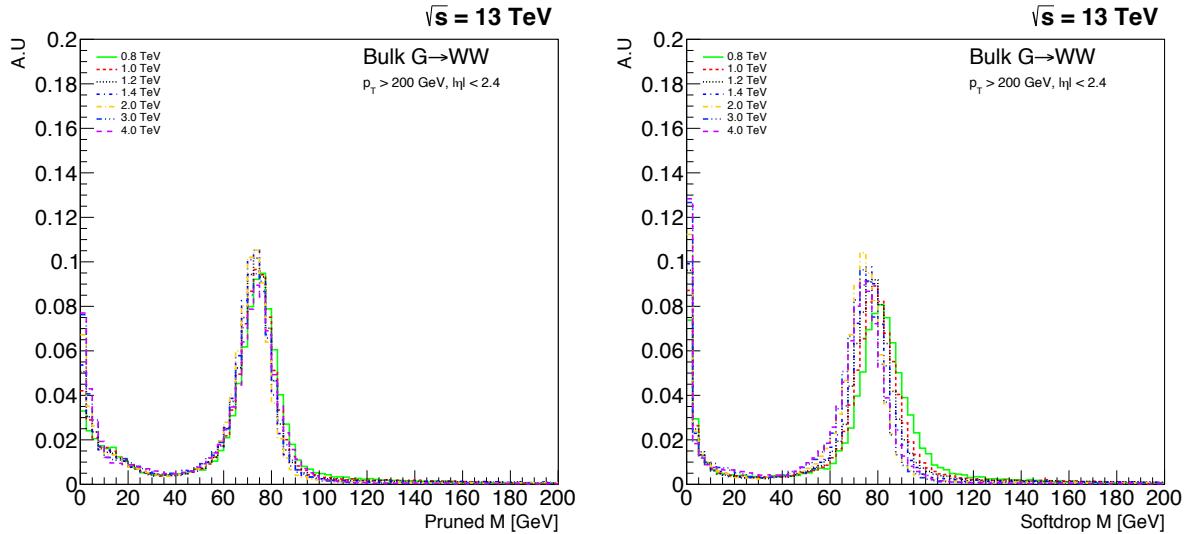


Figure 6.11: The jet mass distribution for W jets coming from a G_{bulk} of masses in the range $0.8 \text{ TeV} < M_X < 4 \text{ TeV}$ decaying to WW , here with pruning applied (left) and softdrop (right). A strong shift in the jet mass mean as a function of p_T ($\sim M_X/2$), is observed for jets groomed with the softdrop algorithm. Charge hadron subtraction is applied to all jets before clustering.

line) jet mass distributions after pruning (left) or softdrop (right) have been applied. Again,

the distributions are compared for jets with very different p_T profiles, here for W jets coming from a $G_{\text{bulk}} \rightarrow WW$ of mass 0.8 TeV (red) yielding a p_T of about 400 GeV, and a mass of 2.0 TeV, yielding a p_T of about 1 TeV. Interestingly, we observe a p_T -dependent mass shift already for generator level softdrop jets (comparing the dotted lines in the right plot), an effect further enhanced at reconstruction level. This effect is not present for pruned jets, for either generator level or reconstruction level. The observed p_T -dependence of the softdrop mass was problematic

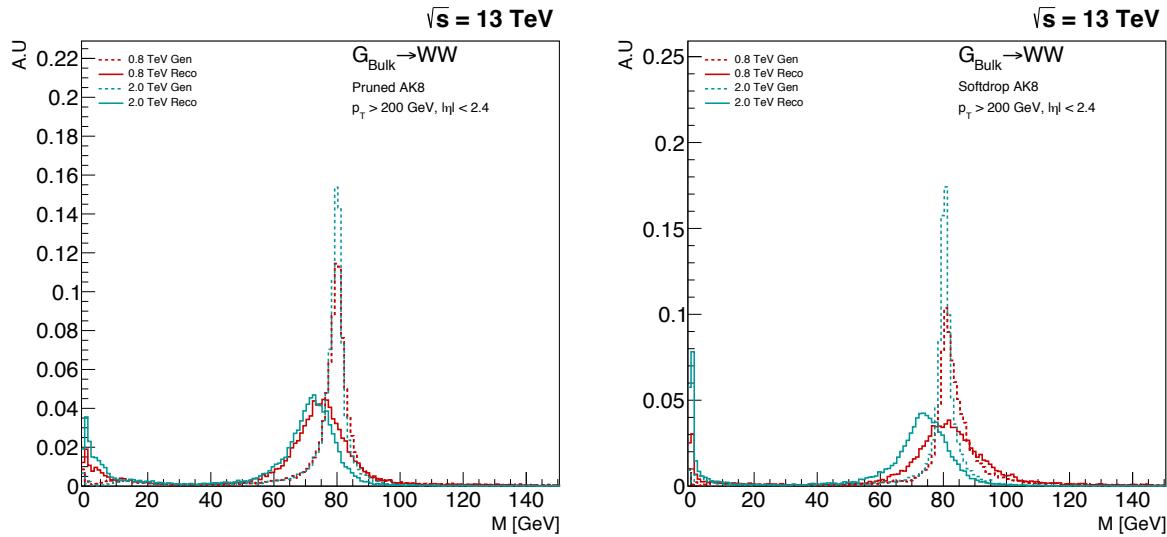


Figure 6.12: The reconstructed (solid line) and generator level (dotted line) jet mass distribution for W jets coming from a $G_{\text{bulk}} \rightarrow WW$ of mass $M_X = 0.8 \text{ TeV}$ (red), roughly $p_T \sim 400 \text{ GeV}$, and $M_X = 2.0 \text{ TeV}$ (blue), $p_T \sim 1 \text{ TeV}$. Here for the pruned (left) and softdrop (right) jet mass.

due to the fact that it would require a p_T -dependent mass window. This would again require several different measurements to produce an efficiency scale factor between simulation and data, for each mass window, or a significantly higher uncertainty on the signal yield. Due to these findings, the grooming algorithm of choice for this analysis is pruning, with the signal selection window defined as $65 \text{ GeV} < m_p < 105 \text{ GeV}$. The above findings will be important for subsequent analyses and are revisited in Section 7.

The variable used to determine the substructure of the V jets is the n-subjettiness ratio τ_{21} , as described in Section 5.5.2. The τ_{21} variable is correlated to the pruned jet mass, however, it still provides additional signal discrimination when applied after the pruned jet mass selection. Figure 6.13 shows the τ_{21} distribution for the QCD background and W jets from a signal decay before (left) and after (right) a pruned mass cut of $65 \text{ GeV} < m_p < 105 \text{ GeV}$ has been applied. We perform a cut optimization on τ_{21} after all analysis selections, including the pruned mass

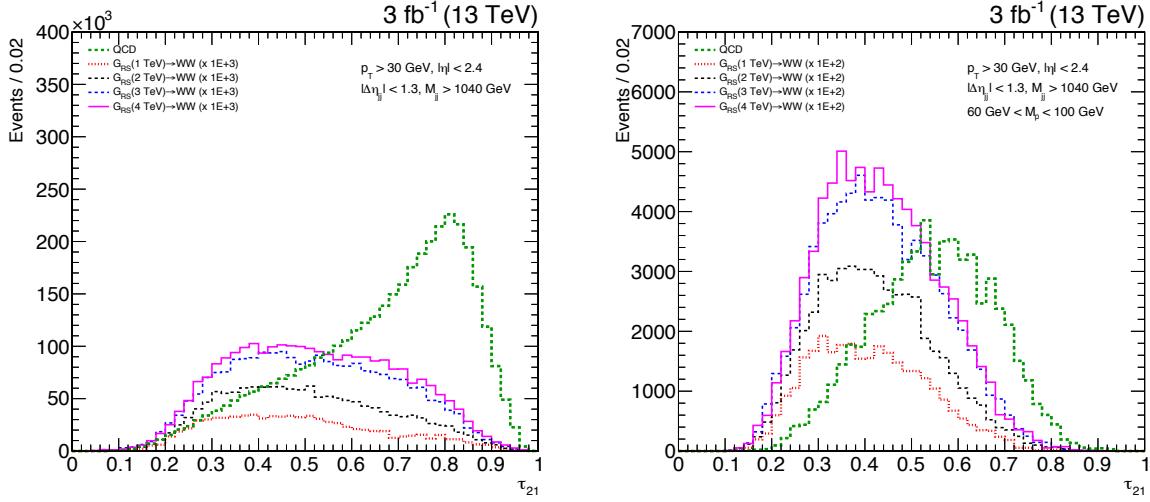


Figure 6.13: The τ_{21} distribution for QCD background and signal jets before (left) and after (right) a pruned mass window is applied. The discriminating power of τ_{21} is strongly reduced after grooming.

window of $65 \text{ GeV} < m_p < 105 \text{ GeV}$, have been applied. This is done by scanning over thresholds for the τ_{21} variable, and for each threshold, computing the Punzi significance [66] defined as

$$S = \frac{\epsilon_S}{1 + \sqrt{B}},$$

where ϵ_S is the signal efficiency and B is the total number of background events. The selection with the highest significance is defined as the optimal value. The signals under consideration are W jets coming from the decay of a G_{bulk} with $1 \text{ TeV} < m_X < 4 \text{ TeV}$, against a background of light-flavored QCD jets. Only jets with a dijet invariant mass in a 20% window around the resonance mass are considered. The Punzi significance as a function of the upper cut value on τ_{21} is shown on the left in Figure 6.14. The optimal cut gets looser as the dijet invariant mass increases, something which can be understood when looking at the QCD dijet invariant mass spectrum in Figure 8.6. The number of QCD jets falls off exponentially with m_{jj} , meaning that the background at 4 TeV is considerably lower than at 1 TeV. This allows for a looser cut on τ_{21} as m_{jj} increases. In order to choose a single cut which works reasonably well for all values of resonance mass, we look at the ratio of a given τ_{21} cut over the significance of the best cut at that mass value. This is shown in the right plot of Figure 6.14. Choosing signal events with a $\tau_{21} < 0.45$ yields the most stable performance out of the investigated τ_{21} selection requirements and also maintains low background rates at low m_{jj} . This selection is therefore used for our

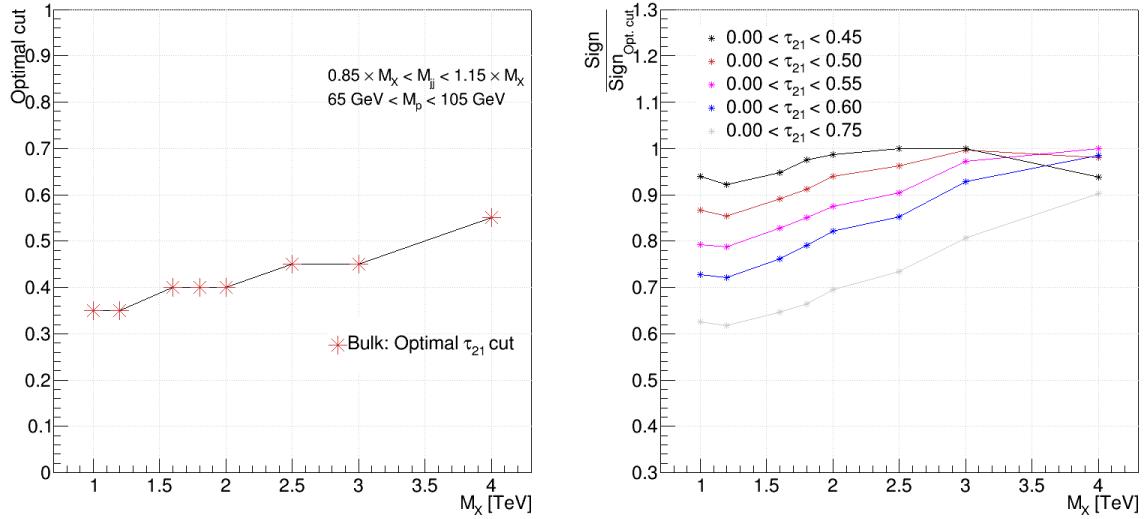


Figure 6.14: Left: Optimal upper value of τ_{21} for selecting the signal as a function of G_{bulk} mass. Right: The ratio of a given τ_{21} cut over the significance of the best cut at that mass value.

main analysis category. In order to account for the fact that background rates are lower for higher values of m_{jj} , we add an additional analysis category, $0.45 < \tau_{21} < 0.75$, which contains $> 95\%$ of the signal and enhances the analysis sensitivity in the case when the background is low. These categories are hereafter referred to as the *high-purity* (HP) category, for jets with $0 < \tau_{21} \leq 0.45$, and the *low-purity* (LP) category, for jets with $0.45 < \tau_{21} \leq 0.75$. QCD jets from light-flavor quarks and gluons (u,d,s,g) can be incorrectly identified as W-jets, and these events are referred to as “mistags” of the W-tagger. The W-tagging efficiency and QCD light-flavored jet mistagging rate for a W-tagger consisting of $0 < \tau_{21} \leq 0.45$ and $65 \text{ GeV} < m_p < 105 \text{ GeV}$ is shown in Figure 6.15, both as a function of jet p_T and as a function of number of primary vertices in the event. The signal efficiency when applying only the pruned jet mass selection is around 80% with a mistag rate of $\sim 15\%$. After applying the τ_{21} selection, the signal efficiency drops to around 55% and the mistagging rate to $\sim 2\%$. Another interesting feature is the dependence of τ_{21} on jet p_T and pileup, compared to the resilience of the groomed mass as a function of the same variables. This will be another feature we explore in the second analysis (Section 7). Figure 6.16 shows the pruned jet mass (left) and the τ_{21} distribution (right) for signal and background Monte Carlo, as well as the distributions measured in data.

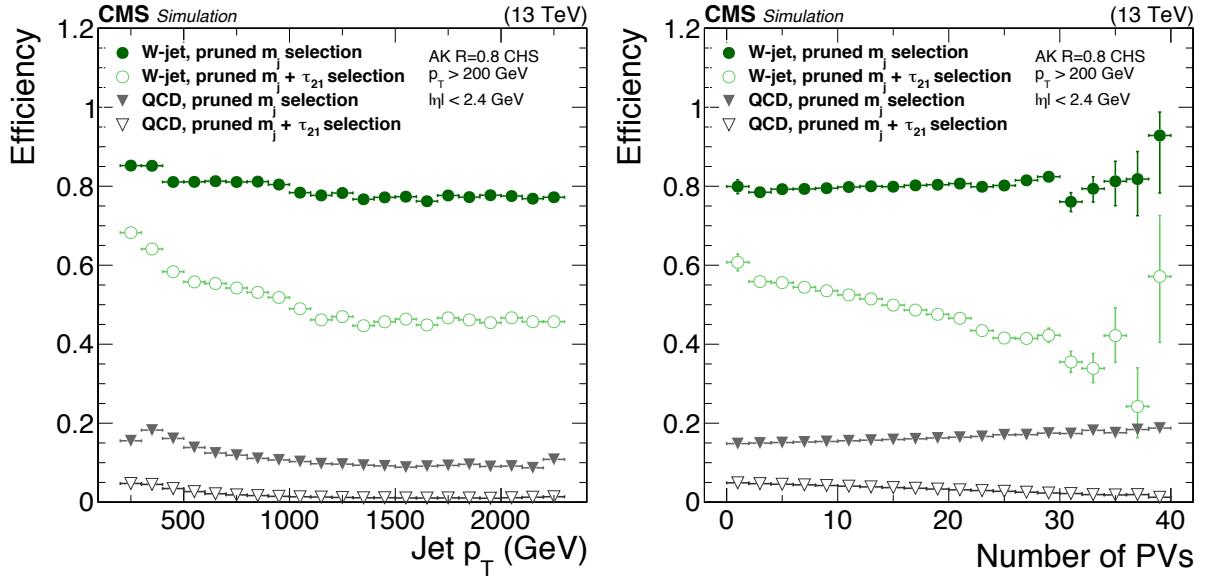


Figure 6.15: The W-tagging efficiency (green) and light jet mistag rate (grey) for a selection based on either the pruned jet mass or the pruned jet mass and τ_{21} as a function of p_T (left) and the number of primary vertices (right).

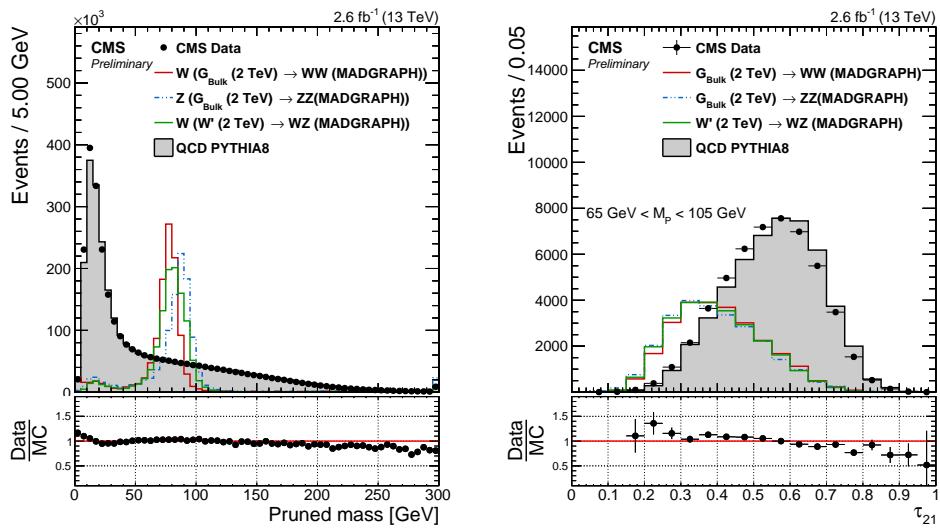


Figure 6.16: Pruned jet mass (left) and τ_{21} (right) distributions for data and simulated samples. Simulated samples are scaled to match the distribution in data. The τ_{21} distribution is shown for jets after a cut on the pruned jet mass of $65 \text{ GeV} < m_p < 105 \text{ GeV}$ has been applied.

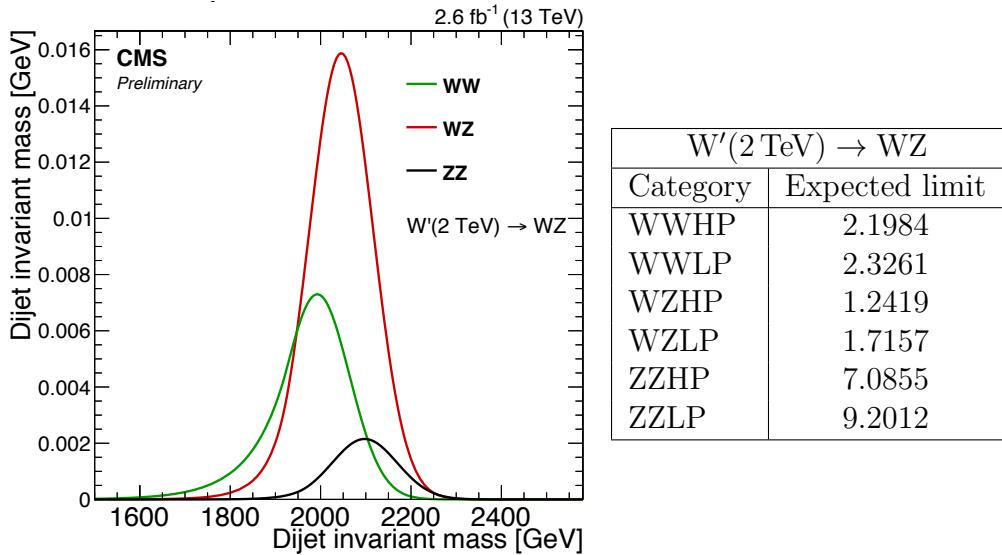


Figure 6.17: The expected signal yield per mass category for a W' (2 TeV) decaying to a W and a Z (left) together with the expected limit per mass category for the same signal (right).

6.4.4 Analysis categorization

As the analysis requires two W/Z -tags, we always require one HP-tagged jet and then divide into LP and HP categories depending on whether the other jet is of high or low purity. In addition, in order to further enhance the analysis sensitivity, we further split the pruned jet mass window into a W and a Z boson window where the W window is defined as $65 \text{ GeV} < m_p < 85 \text{ GeV}$ and the Z boson window as $85 \text{ GeV} < m_p < 105 \text{ GeV}$. This has the added benefit of allowing us to discriminate between a G_{bulk} decaying to WW or ZZ , and a W' decaying into WZ by separating these events into categories. The signal yield will be higher in the WZ category for a W' decaying to a W and Z boson than for a G_{bulk} decaying to WW or ZZ . Figure 6.17 shows the relative expected signal yield (left) and expected limits (left) in the different mass categories for a W' with a mass of 2 TeV. All categories are combined in the end, leading to the same or better sensitivity at each resonance mass value than when using the whole pruned mass window. Figure 6.18 shows the expected 95% CL upper limits on the production cross section of a W' decaying to WZ (left) and a G_{bulk} decaying to WW (right) as a function of the resonance mass in the HP category. The blue line corresponds to the expected limits obtained when not splitting into mass categories and the red line corresponds to the limit using the combination of two categories. The dotted and solid black lines are the limits for the W and Z categories, respectively. The combination of two mass categories leads to slightly better (10%) or similar sensitivity as when using one large

mass window. The real benefit of splitting into mass categories becomes obvious when defining

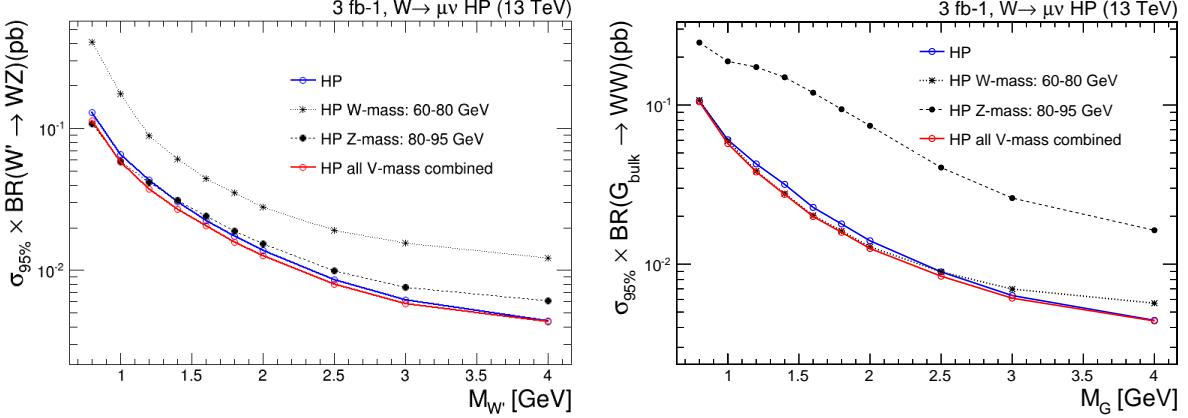


Figure 6.18: Expected 95% CL upper limits on the production cross section of a W' (left) and G_{bulk} (right) signal as a function of the resonance mass for the different mass categories for events passing the high-purity τ_{21} selection.

a test statistics based on the likelihood ratios of each signal hypothesis, $q = -2 \ln(L_{G_{\text{bulk}}}/L_{W'})$, shown in Figure 6.19. For a signal with a signal strength corresponding to a $3-4\sigma$ excess, the test statistics for each signal hypothesis are well separated ($\sim 3.5\sigma$), allowing us to make a statement of whether a possible signal is more likely due to a G_{bulk} or a W' particle. With the high-purity

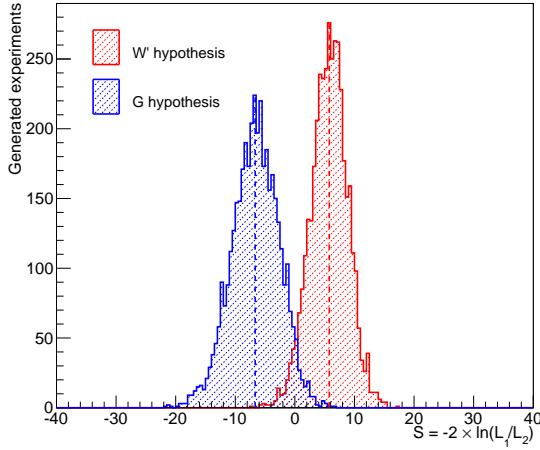


Figure 6.19: Distribution of the test statistic $q = -2 \ln(L_{G_{\text{bulk}}}/L_{W'})$ for a G_{bulk} (blue) and W' signal hypothesis.

and low-purity categories as defined above for each mass window combination, this leaves us with

six different signal categories: HP with 3 mass categories corresponding to WW, WZ, and ZZ, and the same 3 mass categories for LP. In parallel to the mass-category based analysis, we perform an analysis without categorization in mass (similar to the 8 TeV analysis) as a cross-check. We found the sensitivity with mass categories to be higher and hence will not present these studies here. The final tagging efficiency for different signal hypotheses (top) together with the QCD mistag rate (bottom) in the different signal categories is shown in Figure 6.20. The solid lines represent the tagging efficiency in the full mass window ($65 \text{ GeV} < m_p < 105 \text{ GeV}$) before splitting into mass categories. A lower signal efficiency in the ZZ mass category is observed in all cases. This can be explained from the pruned jet mass distribution on the left in Figure 6.16, where a lower pruned jet mass selection of 85 GeV leaves a large fraction of the Z peak in the W mass window. The main benchmark models under consideration preferably decay to W bosons, since in the Bulk Graviton model the branching ratio $\text{BR}(G_{\text{Bulk}} \rightarrow \text{WW}) = 2 * \text{BR}(G_{\text{Bulk}} \rightarrow \text{ZZ})$ and in the HVT model $\text{W}'/\text{Z}' \rightarrow \text{WZ}/\text{WW}$ (but not ZZ). Therefore, the tagging procedure is optimized for a high efficiency to tag W bosons. In the limit-setting procedure all the categories are combined and the overall signal efficiency is maintained. For the combined mass-categories (solid line) the signal efficiency is between 16 and 23% in the double-tag categories, and between 20 and 34 % in the single-V tag categories. The mistagging rate is below 1% in the high-purity category. The full analysis selections and final signal categories are listed in Table 6.1, where m_{jet_1} and m_{jet_2} refers to the jet with the highest and second highest jet p_T , respectively.

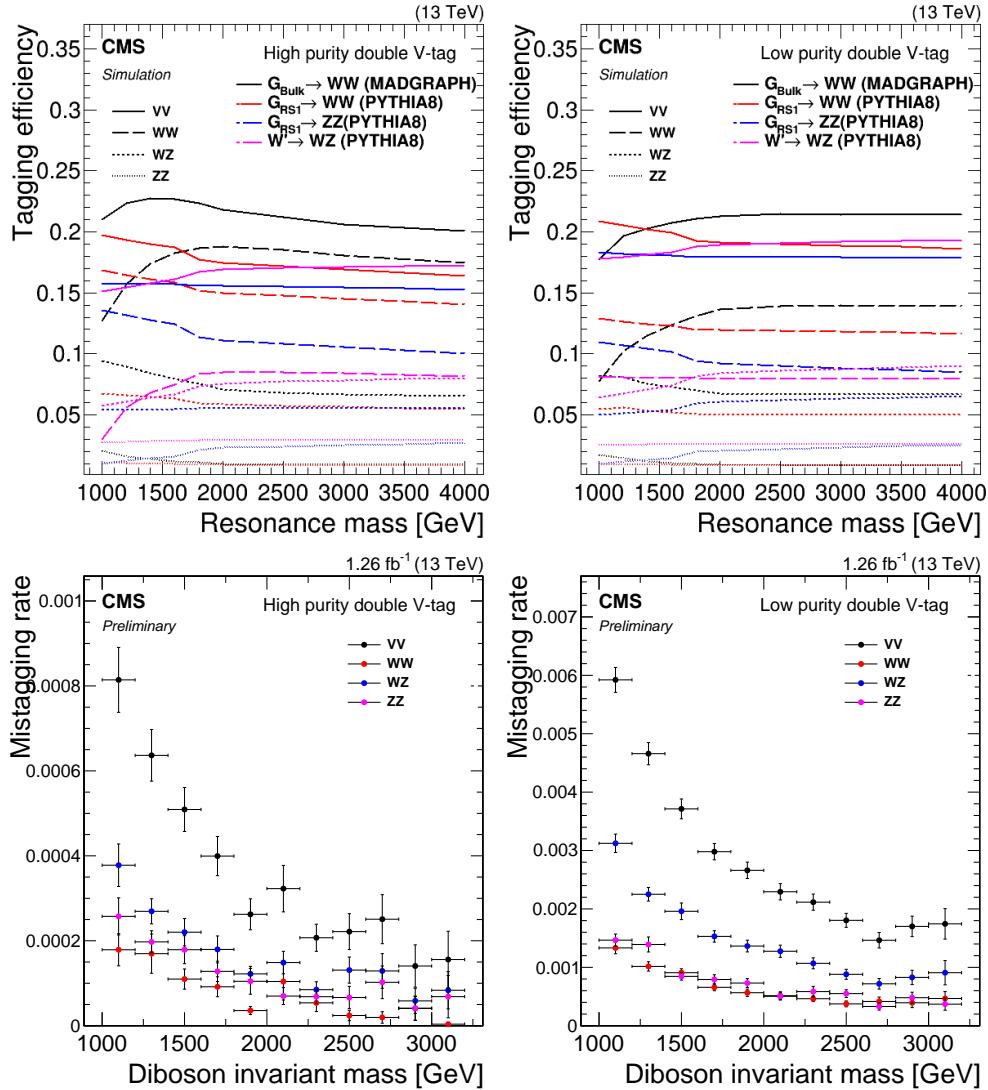


Figure 6.20: Tagging efficiency (top) and mistagging rate (bottom) in the different pruned mass categories in the high-purity category (left) and in the low-purity category (right).

Selection	Value
Boson selections	
$V \rightarrow q\bar{q}$ (2 AK8 jets)	$p_T > 200 \text{ GeV}$ $ \eta < 2.4$
Pruned jet mass	$65 < m_{\text{jet}_1}, m_{\text{jet}_2} < 105 \text{ GeV}$
Topology	$ \Delta\eta_{jj} < 1.3$
Dijet invariant mass	$m_{jj} > 1 \text{ TeV}$
2- to 1-subjettiness ratio	$\tau_{21} < 0.75$
m_{jet} categories	
WW	$65 < m_{\text{jet}_1} < 85 \text{ GeV}, 65 < m_{\text{jet}_2} < 85 \text{ GeV}$
WZ	$65 < m_{\text{jet}_{1/2}} < 85 \text{ GeV}, 85 < m_{\text{jet}_{2/1}} < 105 \text{ GeV}$
ZZ	$85 < m_{\text{jet}_1} < 105 \text{ GeV}, 85 < m_{\text{jet}_2} < 105 \text{ GeV}$
τ_{21} categories	
High-purity	$\tau_{21,\text{jet}1} \leq 0.45, \tau_{21,\text{jet}2} \leq 0.45$
Low-purity	$\tau_{21,\text{jet}1/\text{jet}2} \leq 0.45, 0.45 < \tau_{21,\text{jet}2/\text{jet}1} \leq 0.75$

Table 6.1: The full analysis selections, and the signal categories based on the pruned jet mass and τ_{21} values.

6.5 Background modeling

We assume that the QCD multijets background can be described by a smooth, monotonically decreasing function, which is parametrizable. This is similar to what is done in previous CMS analyses looking for bumps in the dijet invariant mass spectrum [67, 68]. The search is then performed by fitting the sum of the functions for background and signal to the dijet invariant mass spectrum in data. Neither data control regions nor simulated samples are used directly by this method. The background functions are of the following form:

$$\frac{dN}{dm_{jj}} = \frac{P_0}{(m_{jj}/\sqrt{s})^{P_2}} \quad \text{and} \quad \frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2}}, \quad (6.1)$$

where m is the dijet invariant mass, \sqrt{s} is the center-of-mass energy, P_0 is a normalization parameter for the probability density function, and P_1 and P_2 describe the shape. The number of fit parameters is decided through a Fisher's F-test [69]. In this test, we start from the 2-parameter function and compare the goodness of fit (quantified through the sum of residuals) when fitting the data signal region with a 2, 3, 4 and 5 parameter function. The Fishers test statistic is

$$\text{Fishers}_{12} = \frac{\frac{RSS_1 - RSS_2}{p_2 - p_1}}{\frac{RSS_2}{n - p_2^2}}, \quad (6.2)$$

where RSS_i is the residual sum of squares of model i , p_2 is the number of parameters of model 2 and p_1 is the number of parameters of model 1 with $p_2 > p_1$, and n correspond to the number of data points. The model with more parameters will always fit the data at least as well as the model with less parameters, but the F statistic defines whether model 2 yields a significantly better fit to the data, quantified through a confidence level (CL) of the simpler function being correct. In our case, if the confidence level is less than 10%, we add additional parameters to model the background distribution. The 4- and 5-parameter functions are

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2 + P_3 \times \log(m_{jj}/\sqrt{s})}} \quad \text{and} \quad (6.3)$$

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2 + P_3 \times \log(m_{jj}/\sqrt{s}) + P_4 \times \log(m_{jj}/\sqrt{s})^2}}, \quad (6.4)$$

where P_3 and P_4 are additional free parameters. As an additional crosscheck, an alternative fit function is also tested:

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s} + P_3(m_{jj}/\sqrt{s})^2)^{P_1}}{(m_{jj}/\sqrt{s})^{P_2}}. \quad (6.5)$$

The fit range is chosen such that it starts where the trigger efficiency has reached its plateau to avoid bias from trigger inefficiency, and extends to the bin after the highest m_{VV} mass point. The binning chosen for the fit follows the detector resolution as in Refs. [67, 68], and is referred to as the “dijet binning”. In units of GeV, the binning is as follows:

Dijet binning = 1000, 1058, 1126, 1181, 1246, 1313, 1383, 1455, 1530, 1607, 1687, 1770, 1856, 1945, 2037, 2132, 2231, 2332, 2438, 2546, 2659, 2775, 2895, 3019, 3147, 3279, 3416, 3558, 3704, 3854, 4010, 4171, 4337, 4509, 4686, 4869, 5058, 5253, 5500.

Before unblinding the signal region, we check that the QCD dijet invariant mass spectrum is expected to be smooth from the distribution in QCD MC as well as exercise the F-test in QCD MC and in a data sideband. The fits to data in the signal region using the different fit functions are shown in Figure 6.21, and the corresponding F-test results are given in Tables 6.2 through 6.4. The findings can be summarized as follows: for the WW enriched category, a 2-parameter fit is sufficient to describe the data in both the high- and low-purity categories. In the WZ category, a 2-parameter fit is sufficient in the high-purity category, while three parameters are needed for the low-purity category. For the ZZ category, a 3-parameter fit is needed for both purity categories. The 2- and 3-parameter fit functions as defined in Equation C.2 will therefore be used to model the background component in the simultaneous signal and background fit.

WW-enriched, HP				WW-enriched, LP			
Function	Residuals	χ^2	ndof	Function	Residuals	χ^2	ndof
2 par	0.034	9.279	11	2 par	0.270	13.462	17
3 par	0.034	9.160	10	3 par	0.300	13.819	16
4 par	0.040	8.030	9	4 par	0.324	13.680	15
Fishers ₂₃	-0.053	CL	1.0	Fishers ₂₃	-1.723	CL	1.0
Fishers ₃₄	-1.456	CL	1.0	Fishers ₃₄	-1.191	CL	1.0

Table 6.2: Residuals, χ^2 , degrees of freedom, and F-test statistic, with its corresponding confidence level, for the WW-enriched HP and LP categories. A 2-parameter fit is needed to describe the data in both categories.

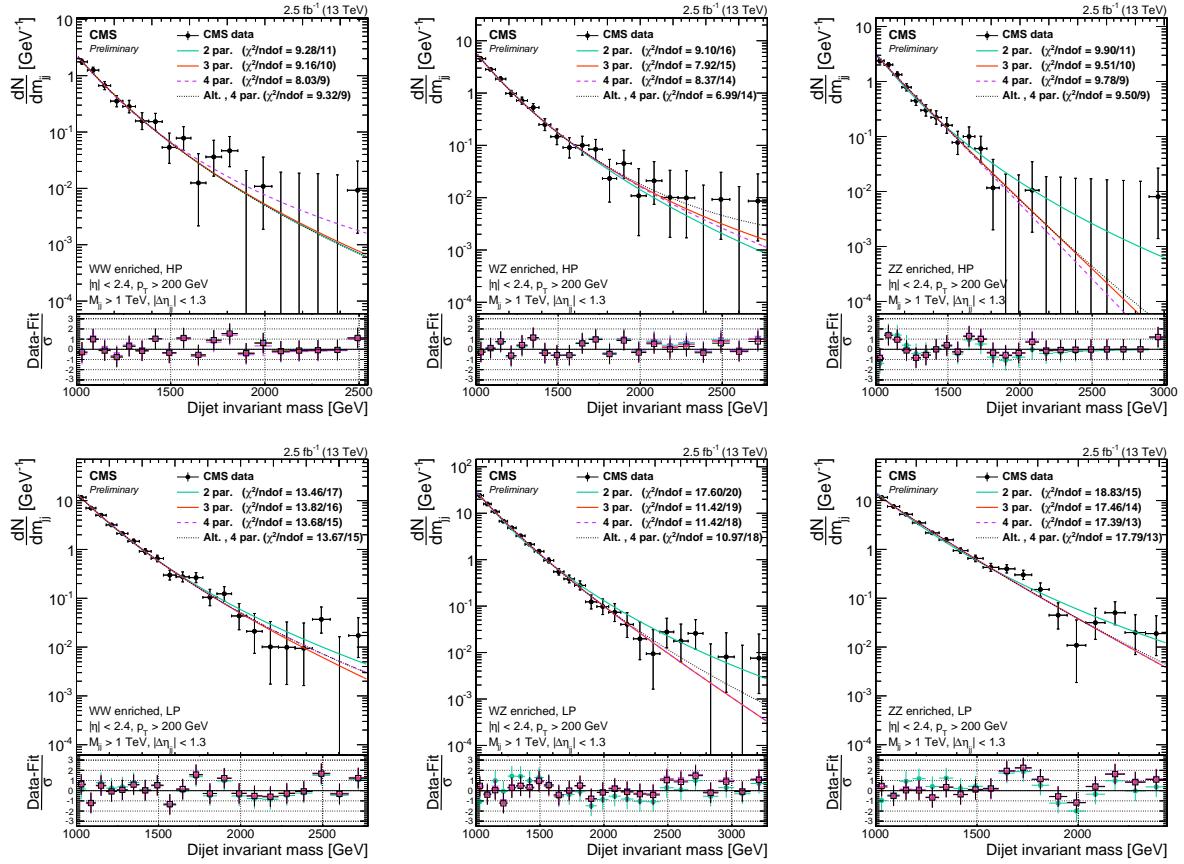


Figure 6.21: Fitted dijet mass spectrum in the different mass and purity categories in data. A 2-parameter fit is sufficient to describe the data in the WW- (HP and LP) and WZ-enriched (LP) categories. For the ZZ- (HP and LP) and WZ-enriched (HP) categories, a 3-parameter fit is needed.

WZ-enriched, HP				WZ-enriched, LP			
Function	Residuals	χ^2	ndof	Function	Residuals	χ^2	ndof
2 par	0.039	9.105	16	2 par	1.016	17.602	20
3 par	0.047	7.915	15	3 par	0.270	11.424	19
4 par	0.048	8.370	14	4 par	0.269	11.421	18
Fishers ₂₃	-2.598	CL	1.0	Fishers ₂₃	55.258	CL	0.0
Fishers ₃₄	-0.491	CL	1.0	Fishers ₃₄	0.078	CL	0.783

Table 6.3: Residuals, χ^2 , degrees of freedom, and F-test statistic, with its corresponding confidence level, for the WZ-enriched HP (left) and LP (right) categories. A 2-parameter fit is sufficient to describe the data in the high-purity category, while three parameters are needed for the low-purity category.

ZZ-enriched, HP				ZZ-enriched, LP			
Function	Residuals	χ^2	ndof	Function	Residuals	χ^2	ndof
2 par	0.220	9.901	11	2 par	0.448	18.832	15
3 par	0.140	9.511	10	3 par	0.121	17.463	14
4 par	0.124	9.781	9	4 par	0.118	17.394	13
Fishers ₂₃	6.302	CL	0.029	Fishers ₂₃	40.438	CL	0.0
Fishers ₃₄	1.246	CL	0.290	Fishers ₃₄	0.356	CL	0.56

Table 6.4: Residuals, χ^2 , degrees of freedom, and F-test statistic, with its corresponding confidence level,m for the ZZ-enriched LP and HP categories. A 3-parameter fit is sufficient to describe the data in both categories.

6.6 Signal modeling

The signal shape is extracted from signal MC with resonance masses in the range from 1 to 4 TeV. A linear interpolation provides shapes for the mass points in between in steps of 100 GeV. From these shapes, signal shape models are constructed as composite models with a Gaussian core due to detector resolution and an exponential tail to account for parton distribution function effects. Parametric shape uncertainties due to jet energy scale and resolution uncertainties are inserted by variations of the Gaussian peak position and width. The dijet invariant mass shape for different benchmark model signals is shown in Figure 6.22. The signal and background components are then simultaneously fitted to the data points.

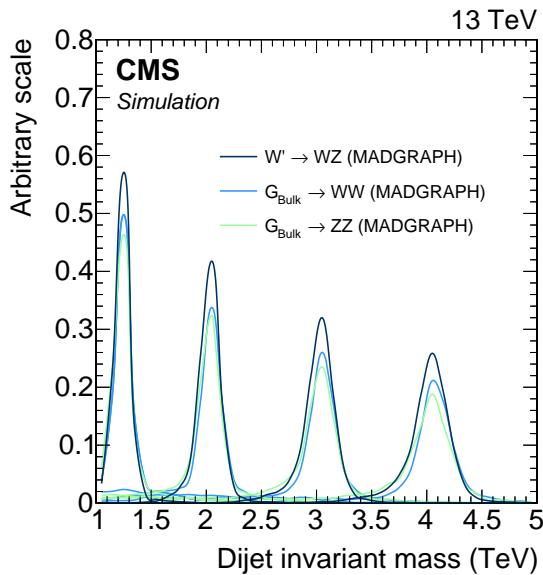


Figure 6.22: Dijet invariant mass from signal MC used to extract the signal shape, shown here for resonances with masses of 1.2, 2, 3, and 4 TeV.

6.7 V-tagging scale factors

As seen in Figure 6.16, a discrepancy is observed in the τ_{21} distribution between data and MC. This could lead to a bias in the signal efficiency estimation so we must measure the real signal efficiency in data in an orthogonal data sample. The W-tagging efficiency is measured using real boosted W-jets in a data sample enriched in $t\bar{t}$ decays with a hadronically decaying W boson. This region is mainly quark-enriched, as opposed to the gluon-enriched QCD region previously studied, and substructure variables are better described here. The sample is obtained by requiring a final state compatible with two b-jets and two W bosons, where one of the bosons decays leptonically and the other one hadronically. There are several good reasons to use this channel: Top-quark pair production events are plentifully produced at the LHC; we can ensure a high purity of the sample by requiring a high-energy lepton, b-tag and missing energy requirements; and lastly, we can ensure that the W jets are boosted by requiring the leptonic leg, together with the hadronic W boson candidate, to have high transverse momentum. The final state is illustrated in Figure 6.23, with the object of interest being the AK R=0.8 jet containing the two quark daughters of the hadronically decaying W boson.

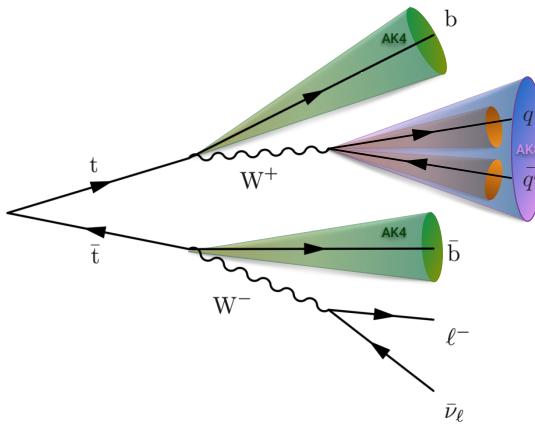


Figure 6.23: A top quark pair decaying into two b quarks and two W bosons, one of which decays leptonically and one of which decays hadronically.

6.7.1 Event selection

The W boson can decay to a neutrino and either an electron or muon, and both final states (“channels”) are used in the analysis. We select events by triggering and event selection on the leptonic W decay. First, we require a high-energy lepton at trigger level, with an online

p_T above 45 GeV for the muon or 135 GeV for the electron. This requires an offline muon (electron) p_T threshold of 53 (120) GeV. The leptons are further required to pass the lepton requirements defined in Section 5.2.2, and events containing additional leptons (passing the same ID requirements, but looser cuts as defined in Table 6.5) are vetoed. Offline, we further require a high missing energy of 40 (80) GeV in the muon(electron) channel. To ensure a high signal purity of boosted, hadronically decaying W bosons, the four-vector of the leptonically decaying W boson is reconstructed such that we can put tight momentum requirements on the leptonic W boson (ensuring that both tops, and therefore vector bosons, have a high momentum). The leptonic W boson is reconstructed in two steps: First, the unknown z component of the neutrino momentum must be solved for through a second order equation assuming the real W boson mass

$$M_W^2 = m_\ell^2 + 2(E_\ell E_\nu - p_{x_\ell} p_{x_\nu} - p_{y_\ell} p_{y_\nu} - p_{z_\ell} p_{z_\nu}) = (80.4)^2.$$

This results in a completely defined neutrino four-vector, which is then added to the lepton four-vector. The sum of the two defines the leptonic W boson four-momentum, and its momentum is required to be greater than 200 GeV. Further, we require at least one AK R=0.4 jet to be b-tagged with the Combined Secondary Vertex (CSV) algorithm [70, 71]. This algorithm exploits the relatively long lifetime and large mass of b hadrons that leads to the presence of a displaced vertex in order to distinguish between jets originating from b quarks and those originating from light-flavor quarks. More information on the CSV algorithm can be found in [70, 71]. The reason for requiring only one b-tagged jet is to ensure a high selection efficiency. Finally, we require at least one AK R=0.8 jet in the event with a momentum greater than 200 GeV, which will be the hadronic W boson candidate. Its pruned jet mass is required to be between 40 GeV and 150 GeV. After reconstructing and selecting all our objects, a set of angular selections are applied to ensure a diboson-like topology. These are the following:

- the ΔR between the lepton and the hadronic W boson candidate must be $< \pi/2$,
- the $\Delta\phi$ between the hadronic W boson candidate and the E_T^{miss} must be > 2 , and
- the $\Delta\phi$ between the hadronic W boson candidate and the lepton must be > 2 .

With these requirements, we have a nearly pure sample of $t\bar{t}$ events, with a small contamination from single top-quark production, W+jets, and VV events. A summary of the final selection criteria is presented in Table 6.5. The pruned jet mass and τ_{21} variables in data and in MC are shown in Figure 6.24.

Selection	Value	Comments
Tight Lepton selection		
Electron p_T	$p_T > 120 \text{ GeV}$	
Muon p_T	$p_T > 53 \text{ GeV}$	
Electron η	$ \eta _{\text{SC}} < 2.5$ except [1.4442, 1.566]	Veto ECAL barrel-endcap transition.
Muon η	$ \eta < 2.1$	
Loose Lepton selection		
Electron p_T	$p_T > 35 \text{ GeV}$	
Muon p_T	$p_T > 20 \text{ GeV}$	
Electron η	$ \eta _{\text{SC}} < 2.5$ except [1.4442, 1.566]	Veto ECAL barrel-endcap transition.
Muon η	$ \eta < 2.4$	
AK8 jet selections		
Jet p_T	$p_T > 200 \text{ GeV}$	
Jet η	$ \eta < 2.4$	For hadronic W reconstruction
AK4 jet selections		
Jet p_T	$p_T > 30 \text{ GeV}$	
Jet η	$ \eta < 2.4$	Used for b-tag jet selection
E_T^{miss} selections		
E_T^{miss} (electron channel)	$E_T^{\text{miss}} > 80 \text{ GeV}$	
E_T^{miss} (muon channel)	$E_T^{\text{miss}} > 40 \text{ GeV}$	
Boson selections		
Pruned jet mass	$40 < m_p < 150 \text{ GeV}$	
Leptonic W p_T	$p_T > 200 \text{ GeV}$	
Hadronic W p_T	$p_T > 200 \text{ GeV}$	
Veto		
Number of loose electrons	0	
Number of loose muons	0	
Number of b-tagged jets	> 0	CSV medium working point
Angular selections		
$\Delta R(l, W_{\text{AK8}})$	$> \pi/2$	
$\Delta\phi(W_{\text{AK8}}, E_T^{\text{miss}})$	> 2	
$\Delta\phi(W_{\text{AK8}}, W_{\text{lep}})$	> 2	

Table 6.5: Summary of the final semi-leptonic $t\bar{t}$ selections.

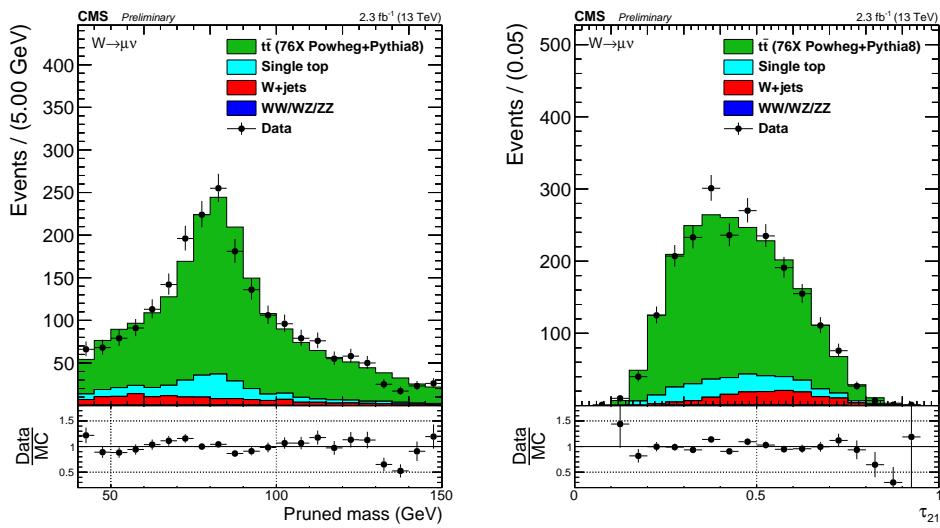


Figure 6.24: Distribution of pruned jet mass (left) and n-subjettiness (right) in the $t\bar{t}$ control sample.

6.7.2 Fitting procedure

For this measurement, what we are interested in is to extract and compare the W-tagging efficiency of the combined selection of jet mass and τ_{21} in data and MC. We are additionally interested in the difference in jet-mass scale (mean of the W-jet mass peak) and jet-mass resolution (width of the W-jet mass peak), as this also affects the shape of the signal jet mass and therefore the tagging efficiency. In order to study these variables, we look at the pruned jet mass spectrum between 40 and 150 GeV in two regions:

- Pass region: $0 < \tau_{21} \leq 0.45 \sim$ high purity , and
- Fail region: $0.45 < \tau_{21} \leq 0.75 \sim$ low purity.

Our goal is to understand what the real fraction of merged W jets is in the pass category and in the fail category, assuming that the sum of the two correspond to a 100% selection efficiency (the amount of W boson jets falling outside of this region is negligible). The strategy is as follows: We first derive probability density functions (PDFs) which describe the distribution of fully merged W boson jets and non-W boson jets in $t\bar{t}$, both in the pass and in the fail region. The PDFs describing real W jets and non-W jets are added with a fraction which is left floating: the fit decides what the fraction of real W to non-W jets is in the pass and in the fail region. A simultaneous fit of pass and fail regions is then performed (using the two composite PDFs), where the fraction of real W jets in both pass and fail regions is constrained such that, if the signal efficiency in the pass region is ϵ_S , the signal efficiency in the fail region is $(1 - \epsilon_S)$. This is done by letting the normalization of the PDF describing real W jets in the pass category be defined as the total real W boson yield in the pass and fail regions combined, multiplied by some fraction, ϵ_S . The normalization of the PDF describing real W boson jets in the fail category is then the total real W boson yield in the pass and fail regions combined, multiplied by $(1 - \epsilon_S)$.

To understand which part of the $t\bar{t}$ jet mass distribution contains real, merged W boson jets and which are only pure combinatorial background, non-W jets, we start from $t\bar{t}$ MC. By matching the AK8 jet with quarks coming from the hadronic W boson at generator level, in a cone of $\Delta R < 0.8$, we can access the real merged W and non-merged W softdrop mass shapes. The real W and non-W PDFs for jets that pass and fail the N-subjettiness selection $\tau_{21} < 0.45$, are found to be well described by the following functions:

$$f_{\text{bkg}}(m_j) = F_{\text{ExpErf}} = e^{c_0 m_j} \cdot \frac{1 + \text{Erf}((m_j - a)/b)}{2} \quad \sim \text{for non-W jets in both pass and fail}$$

$$f^{\text{sig}}(m_j) = F_{\text{Gaus}}(m_j) + F_{\text{ExpErf}}(m_j) \quad \sim \text{for real W jets in both pass and fail}$$

Figure 6.25 shows the fitted pruned jet mass spectrum for real W jets (upper) and non-W jets (lower) in $t\bar{t}$ for jets that passed (left column) and failed (right column) the N-subjettiness selection $\tau_{21} < 0.45$.

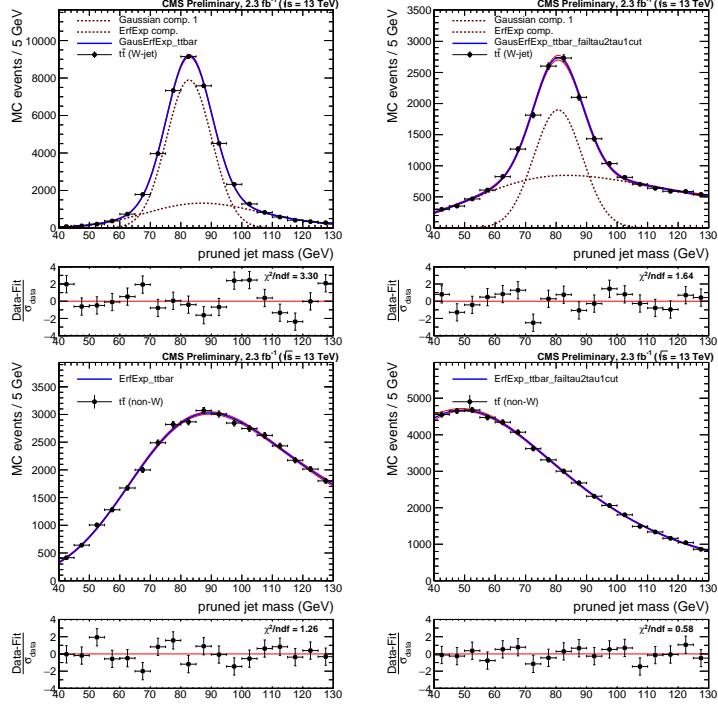


Figure 6.25: Fit to the real W (upper) and non-W (lower) boson pruned jet mass distribution for jets that pass (left) and fail (right) the cut on $\tau_{21} < 0.45$.

These shapes constitute the fit functions used for the simultaneous fit. As can be seen from the fit to real W jets in the pass region, the distribution is not purely Gaussian and has a tail at higher groomed masses. This tail depends on the matching requirements used to define real merged W jets and is unphysical. We therefore assume that the distribution of real W jets can be described by a Gaussian only, allowing the exponential error function used to describe non W-jets to cover the contribution from the tails, thereby taking the number of real W jets as the integral of the Gaussian shape only. This eliminates two additional fit functions, corresponding to six free parameters from the fit. In older estimations of the W-tagging scale factor based on the same procedure [72]), the functions used to describe the tail of the real W-jet distributions were also taken into account as contributing to the real W-jet tagging efficiency. These two calculations test two extremes: the new method assumes a Gaussian peak, absorbing the tails into the background function making the fit more robust, while the old method assumes a Gaussian peak with tails estimated from matched MC. The latter uses a more precise definition of real W jets, but yields a

less robust fit. Both methods were investigated and we found that the absorption of tails into the background function resulted in a decrease in the relative uncertainty on the final scale factor of 50% and an overall improvement on the fit quality, reducing the fit χ^2 by 15%. The fit parameters of the functions used to describe non-W jets in both the pass and in the fail region are further constrained using the values obtained from matched $t\bar{t}$ MC. The W-tagging scale factors (SF_{HP}) for the high purity selection ($\tau_{21} < 0.45$) are then extracted by estimating the selection efficiency (ϵ_{HP}) for both data and simulated samples by fitting, simultaneously, the pass and fail categories:

$$L_{\text{pass}} = \prod_i^{N_{\text{evt}}^{\text{pass}}} \left[N_W \cdot \epsilon_{HP} \cdot f_{\text{pass}}^{\text{sig}}(m_j) + N_2 \cdot f_{\text{pass}}^{\text{bkg}}(m_j) + \sum_{j=\text{ST,VV,WJet}} N_{\text{pass}}^j \cdot f_{\text{pass}}^j \right],$$

$$L_{\text{fail}} = \prod_i^{N_{\text{evt}}^{\text{fail}}} \left[N_W \cdot (1 - \epsilon_{HP}) \cdot f_{\text{fail}}^{\text{sig}}(m_j) + N_3 \cdot f_{\text{fail}}^{\text{bkg}}(m_j) + \sum_{j=\text{ST,VV,WJet}} N_{\text{fail}}^j \cdot f_{\text{fail}}^j \right],$$

where N_W is the number of real W boson jets, N_2 and N_3 are the number of combinatorial background events passing and failing the τ_{21} selection, respectively, and N_j and f_j , with $j = \text{ST, VV, WJet}$, are the normalizations and shapes of the minor backgrounds (single top, VV, W+jets), which are fixed from simulation. The fit functions used are:

$$f_{\text{pass}}^{\text{Top}} = F_{\text{ErfExpGaus}}(x) = \frac{1 + \text{Erf}((x - a)/b)}{2} \cdot e^{-(x - x_0)^2/2\sigma^2}$$

$$f_{\text{fail}}^{\text{Top}} = F_{\text{ExpGaus}}(x) = e^{ax} \cdot e^{-(x - b)^2/2s^2},$$

$$f_{\text{pass}}^{\text{VV}} = F_{\text{ExpGaus}}(x) = e^{ax} \cdot e^{-(x - b)^2/2s^2},$$

$$f_{\text{fail}}^{\text{VV}} = F_{\text{ExpGaus}}(x) = e^{ax} \cdot e^{-(x - b)^2/2s^2},$$

$$f_{\text{pass}}^{\text{wjet}} = F_{\text{ErfExp}}(x) = e^{c_0 x} \cdot \frac{1 + \text{Erf}((x - a)/b)}{2},$$

$$f_{\text{fail}}^{\text{wjet}} = F_{\text{ErfExp}}(x) = e^{c_0 x} \cdot \frac{1 + \text{Erf}((x - a)/b)}{2},$$

with the corresponding distributions shown in Figure 6.26. The floating parameters of the fit (besides the PDF shape parameters themselves) are the rates N_W , N_2 and N_3 , and the mean and sigma of the W-mass distribution defined in $f_{\text{pass}}^{\text{sig}}(m_j)$ and $f_{\text{fail}}^{\text{sig}}(m_j)$. The ratio between data and simulation efficiencies is then taken as the W-tagging scale factor:

$$SF_{HP} = \frac{\epsilon_{HP}(\text{data})}{\epsilon_{HP}(\text{sim})}. \quad (6.6)$$

Considering that, both for data and simulation, $\epsilon_{HP} + \epsilon_{LP} + \epsilon_{fail} = 1$, the scale factor for low

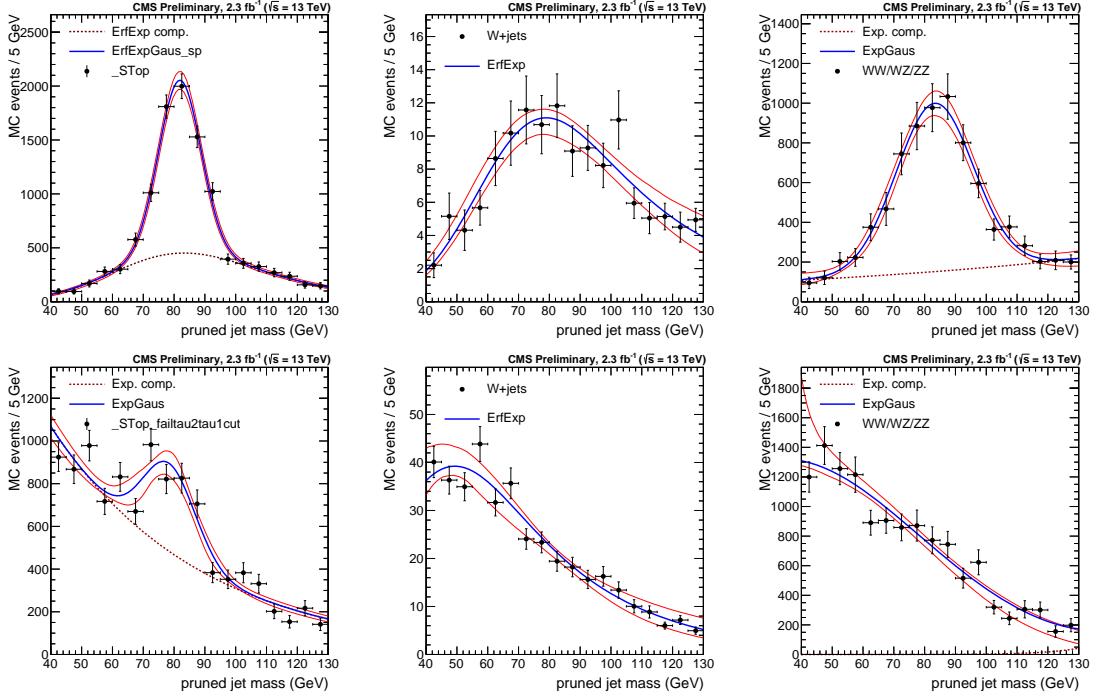


Figure 6.26: Fits to the pruned jet mass spectrum for the non-dominant backgrounds (Single top, W+jets and VV respectively) in the pass (upper) and fail (lower) regions.

purity category can be defined as:

$$SF_{LP} = \frac{1 - \epsilon_{HP}(\text{data}) - \epsilon_{fail}(\text{data})}{1 - \epsilon_{HP}(\text{sim}) - \epsilon_{fail}(\text{sim})},$$

where ϵ_{fail} is the ratio between the number of events with $\tau_2/\tau_1 > 0.75$ and the total number of events. As mentioned previously, the number of real W jets with $\tau_2/\tau_1 > 0.75$ is negligible and the definition of the low purity scale factor simplifies to

$$SF_{LP} = \frac{1 - \epsilon_{HP}(\text{data})}{1 - \epsilon_{HP}(\text{sim})}. \quad (6.7)$$

6.7.3 Systematic uncertainties

As systematic uncertainties, we consider effects due to differences between $t\bar{t}$ simulation as well as effects due to the choice of fit method. The former is evaluated by comparing the extracted scale factor when using $t\bar{t}$ MC samples produced with three different combinations of matrix element (ME) and shower generators: POWHEG (NLO) interfaced with PYTHIA8, MADGRAPH (LO) QCD

interfaced with HERWIG++, and POWHEG interfaced with HERWIG++. The uncertainty due to different ME generators (POWHEG versus MADGRAPH) corresponds to 3(13)% and is listed in Table 6.6 as the first quoted systematic uncertainty. The uncertainty due to parton showering (PYTHIA8 versus HERWIG++) is 8.6%, but is not relevant for analyses where no HERWIG++ based simulation is used, as is the case for the search presented in this chapter. For the systematic uncertainty accounting for effects due to choice of fit method, we compare the estimated extracted efficiency in $t\bar{t}$ MC using the two different fit models described above: the new model, where the signal is modeled by a Gaussian peak and the tails of the distribution are absorbed in the background fit model, and the old model, including the tails when calculating the fraction of real W jets. Figure 6.27 shows the fits obtained in the pass and fail regions using the two different models. With the new model only the Gaussian component of the fit contributes to the W-tagging efficiency while, with the old model, a Chebyshev component is additionally contributing to the total W-tagging efficiency. The estimated efficiencies obtained using both methods, after

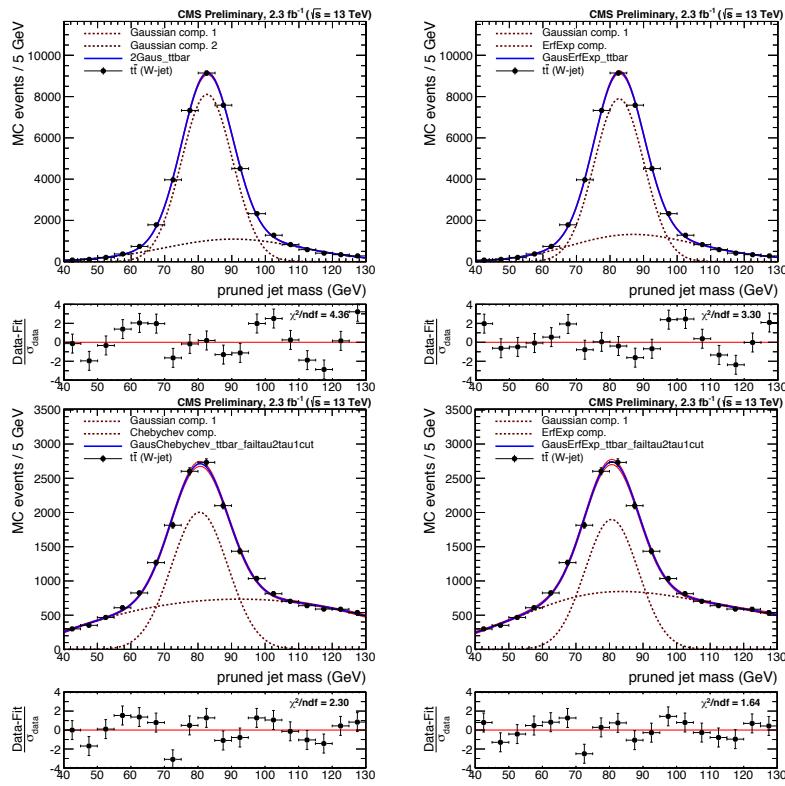


Figure 6.27: Fits obtained in the pass (top) and fail (bottom) regions using two different models: An alternative model with tails (top and bottom, left) where the tail component is contributing to the total W-tagging efficiency. When using the default model (top and bottom, right), only the Gaussian component of the fit contributes to the W-tagging efficiency.

being corrected for the fraction of W jets in the tails, agree within 0.3(0.8)% and are listed as a systematic uncertainty in Table 6.6.

One additional uncertainty is added. As the W-tagging scale factor is evaluated in a $t\bar{t}$ sample, the transverse momentum range is rather limited. When the W-jet p_T reaches ~ 400 GeV, the AK8 jet becomes a fully merged top jet with a mass of 170 GeV and a scale factor measurement becomes impossible. However, the jets used in the analyses presented in this thesis have very high transverse momenta, up to 2-3 TeV, and we therefore need an estimate of how the uncertainty on the W-tagging scale factor changes as a function of p_T . This is estimated by comparing the difference in tagging efficiency between $G_{\text{bulk}} \rightarrow WW$ signal MC showered by PYTHIA8 as compared to HERWIG++ as a function of jet p_T , relative to the difference in tagging efficiency between the two at a $p_T \sim 200$ GeV. This measurement was performed by a separate analysis team, and found to be $5.90\% \times \ln(p_T/200 \text{ GeV})$. Systematic uncertainties from other sources (lepton identification, b tagging, etc.) are less than 0.5% and therefore negligible.

6.7.4 Fit results

The simultaneous fit as described above is then performed both for data and for simulation, where we take the ratio of data and MC efficiencies as efficiency scale factors. The corresponding fits are shown in Figure 7.17, with the corresponding extracted efficiencies and scale factors summarized in Table 6.6. We additionally extract the jet mass scale and jet mass resolution, used to scale

	Working point	Eff. data	Eff. simulation	Scale factor
HP	$\tau_{21} < 0.45$	0.775 ± 0.041	0.822 ± 0.033	$0.94 \pm 0.05 \text{ (stat)} \pm 0.03 \text{ (sys)} \pm 0.003 \text{ (sys)}$
LP	$0.45 < \tau_{21} < 0.75$	0.225 ± 0.041	0.178 ± 0.033	$1.27 \pm 0.25 \text{ (stat)} \pm 0.13 \text{ (sys)} \pm 0.008 \text{ (sys)}$

Table 6.6: Efficiencies in data and in MC together with the corresponding W-tagging scale factors for the high purity and low purity categories.

and smear the jet mass signal shape in the limit setting procedure. These values are taken from the mean $\langle m \rangle$ and width σ of the Gaussian component of the simultaneous fit in the pass region and are summarized in Table 6.7. Both the jet mass scale as well as the jet mass resolution is larger in simulation than in data with a relative difference of 2 and 10%, respectively. However, the jet mass resolution scale factor has a large uncertainty attached to it and is statistically insignificant (in agreement with unity within uncertainty).

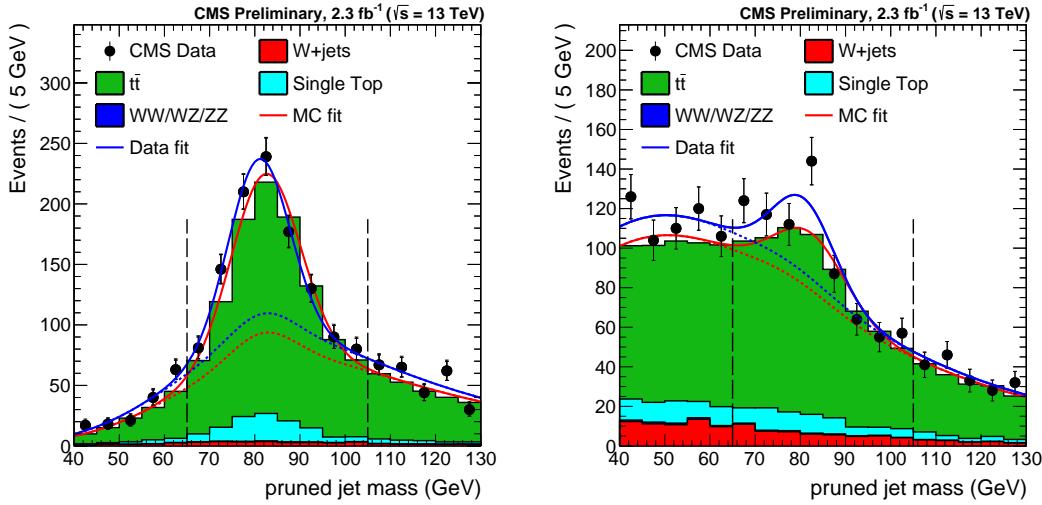


Figure 6.28: Pruned jet mass distribution that pass (left) and fail (right) the $\tau_{21} < 0.45$ selection. Results of both the fit to data (blue) and simulation (red) are shown. The background components of the fit are shown as short-dashed lines.

Parameter	Data	Simulation	Data/Simulation
Pruning $\langle m \rangle$	80.9 ± 0.6 GeV	82.5 ± 0.1 GeV	0.980 ± 0.007
Pruning σ	6.7 ± 0.7 GeV	7.5 ± 0.3 GeV	0.89 ± 0.10

Table 6.7: Jet mass scale and resolution in data and in simulation together with the relevant data-simulation scale factors.

6.7.5 Impact on search variables

The obtained W-tagging scale factors are used as a scale of the signal yield. As we require two W-tagged jets, either HP or LP, the actual scale factors for the high-purity signal yield is $SF_{HP} \times SF_{HP}$ and for the low-purity category $SF_{HP} \times SF_{LP}$. The signal yields are then

$$N_S^{HP} = N_{\text{HP tot. yield}} \times SF_{HP} \times SF_{HP}$$

$$N_S^{LP} = N_{\text{LP tot. yield}} \times SF_{HP} \times SF_{LP}$$

The uncertainties on the scale factors are considered as anti-correlated between the HP and the LP categories. The jet mass scale and resolution are used to scale and smear the signal Monte Carlo. An uncertainty on the signal yield based on the uncertainty on jet mass scale and

resolution is also considered by scaling and smearing the jet mass up and down within the quoted uncertainties and then recomputing the signal efficiency. The results are listed in Table 6.8.

6.8 Systematic uncertainties

The uncertainty on the background parametrization is statistical only and is taken as the covariance matrix of the dijet fit function. As demonstrated in the F-test, we study different background parameterizations and we have found these to be within the fit uncertainty of the nominal fit. The remaining uncertainties concern the signal shape and yield and are listed in Table 6.8. Jet reconstruction uncertainties affect both the signal yield and the signal shape. These are evaluated by rescaling the jet four-momenta according to uncertainties on the jet energy scale and resolution and recomputing the signal efficiency. The difference in efficiency with and without smearing/scaling is taken as a systematic uncertainty, as described above. The jet mass/energy scale and resolution also affect the signal shape, and are added as uncertainties in the peak position and width of the Gaussian component of the signal PDFs.

Source	Relevant quantity	HP uncertainty (%)	LP uncertainty (%)
Jet energy scale	Resonance shape	2	2
Jet energy resolution	Resonance shape	10	10
Jet energy and m_{jet} scale	Signal yield	0.1–4	
Jet energy and m_{jet} resolution	Signal yield	0.1–1.4	
Pileup	Signal yield	2	
Integrated luminosity	Signal yield	2	
PDFs (W')	Signal yield	4–19	
PDFs (Z')	Signal yield	4–13	
PDFs (G_{bulk})	Signal yield	9–77	
Scales (W')	Signal yield	1–14	
Scales (Z')	Signal yield	1–13	
Scales (G_{bulk})	Signal yield	8–22	
Jet energy and m_{jet} scale	Migration	1–50	
V tagging τ_{21}	Migration	14	21
V tagging p_T -dependence	Migration	7–14	5–11

Table 6.8: Summary of the systematic uncertainties and the quantities they affect. Migration uncertainties result in events switching between the purity/mass categories and change the efficiency in each category, but do not affect the total signal efficiency.

6.9 Results

The background fits for each analysis category in the data signal region are shown in Figure 6.29. Here a background-only fit is performed while, as described above, a simultaneous fit is used for the limit-setting procedure. The filled area corresponds to the 1σ error band of the background fit, obtained using linear error propagation. No excess is observed, and we proceed by setting upper limits on the cross section times branching ratio of the process $X \rightarrow VV$, using the asymptotic CL_S method [73]. The binned likelihood is defined as

$$L = \prod_i \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}, \quad (6.8)$$

with

$$\mu_i = \sigma \cdot N_i(S) + N_i(B). \quad (6.9)$$

Here σ is the signal strength scaling the expected number of signal events in the i -th dijet invariant-mass bin $N_i(S)$, $N_i(B)$ is the expected number of background events in dijet invariant-mass bin i and n_i is the observed number of events in the i th dijet invariant-mass bin. The background per bin $N_i(B)$ is estimated from the background component of the best signal and background fit to the data points, with the signal cross section set to zero. The number of signal events in the i -th dijet invariant mass bin, $N_i(S)$, is estimated from the signal templates, considering only dijet invariant massed in a 20% window around the resonance mass, containing most of the signal contribution while making sure to keep a good description of the core.

6.10 Limits

6.10.1 All-hadronic analysis only

As mentioned in Section 6.3, we set limits on three different signal scenarios: $G_{\text{bulk}} \rightarrow WW$, $G_{\text{bulk}} \rightarrow ZZ$, and $W' \rightarrow WZ$, with a $\tilde{k} = 0.5$ for the G_{bulk} . Figure 6.30 shows the asymptotic limits at 95 % confidence level on the signal cross section as a function of its resonance mass obtained with 2.7 fb^{-1} of 13 TeV CMS data after combining all mass and purity categories (top). The corresponding p-values are shown in the bottom panel. The statistics are too low to exclude the excess around 2 TeV observed in the corresponding Run 1 analysis and, in addition, an under-fluctuation in data is present in this region. The largest excess of 2.8σ is observed for a $G_{\text{bulk}} \rightarrow ZZ$ hypothesis at a resonance mass of 2.8-3 TeV. This is driven by the high-purity ZZ

mass category, the category with the lowest statistics, and is caused by one event. A 3-parameter fit is the default background fit function for this category, however, a 2-parameter fit could also be used to describe these data. In Figure 6.31 we compare the limits and p-values obtained using a 2-parameter and a 3-parameter fit to describe the background in this category. The significance at 3 TeV is reduced from 2.8 to 1.5 σ with a 2-parameter fit, reflecting the fact that the fit is poorly constrained in the high mass tail due to low statistics. The fit to data using both a 2 and 3-parameter fit in the ZZ high-purity category is shown in Figure 6.32, where we see that the 2-parameter fit lies within the fit uncertainties of the nominal fit. The lack of constraint on the fit in the dijet invariant mass tail when statistics are very low is a drawback of a method relying fully on a parametric fit and reduces the analysis sensitivity in the high- m_{jj} region. In Search II (Section 7) we will keep taking advantage of the dijet fit, however, the integrated luminosity is ~ 15 times higher, resulting in more datapoints in the m_{jj} tail which further constrains the fit. In Search III (Section 8), we will explore alternate methods which allow more control over the background shape across the full mass spectrum.

6.10.2 Semi-leptonic and all-hadronic combination

To maximize the search sensitivity, we combine the results obtained above with those of the corresponding semi-leptonic analysis. We assume the uncertainties on luminosity, V-tagging efficiency, and jet mass scale and resolution to be fully correlated. The obtained exclusion limits in Figure 6.33 show the resulting expected and observed exclusion limits after combining the two analyses. As before, we consider a scenario where only either a W' or Z' resonance is expected, called the singlet hypothesis (upper two plots). In addition, we set limits on the triplet hypothesis, assuming the W' and Z' bosons to be degenerate in mass (bottom left plot). The all-hadronic analysis sets stronger upper limits than the semi-leptonic analysis above 1.7 TeV for a Z' and > 1.3 TeV for a W' . The improvement in sensitivity is due to the higher branching fraction, such that $\mathcal{B}(WW \rightarrow q\bar{q}q\bar{q}) = 44\%$, $\mathcal{B}(WW \rightarrow \ell\nu q\bar{q}) = 31\%$, $\mathcal{B}(WZ \rightarrow q\bar{q}q\bar{q}) = 46\%$, and $\mathcal{B}(WZ \rightarrow \ell\nu q\bar{q}) = 16\%$. The analysis sensitivity for G_{bulk} is too weak to set limits, but cross sections between 3–1200 fb are excluded. For the HVT models A and B, a W' is excluded below 2.0 and 2.2 TeV, respectively, and Z' resonances are excluded below 1.7 (1.6) TeV assuming a HVT model A(B). Assuming a HVT Model A(B) triplet hypothesis, resonances below 2.3 (2.4) TeV are excluded. The combined results would therefore just exclude a W' with a mass around 2 TeV, the favored candidate to explain the diboson excess observed in 8 TeV data. However, Bulk Graviton signals were still far from excluded and, with the expected increase in luminosity of a factor of ten in 2016, we were excited to keep on searching.

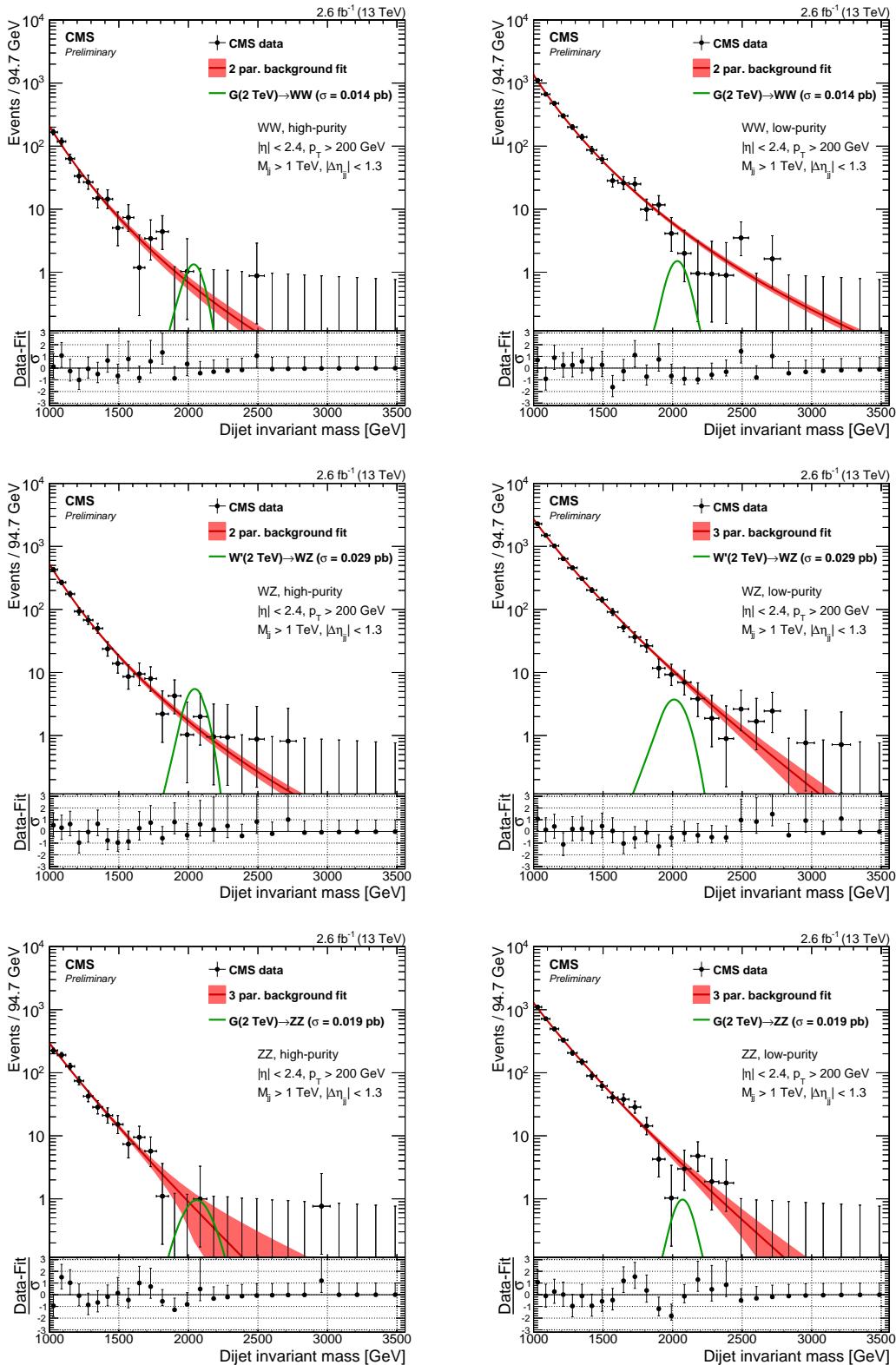


Figure 6.29: Fit to data in the signal region using the background fit only for the different mass and purity categories. The filled red area correspond to the 1σ statistical error of the fit.

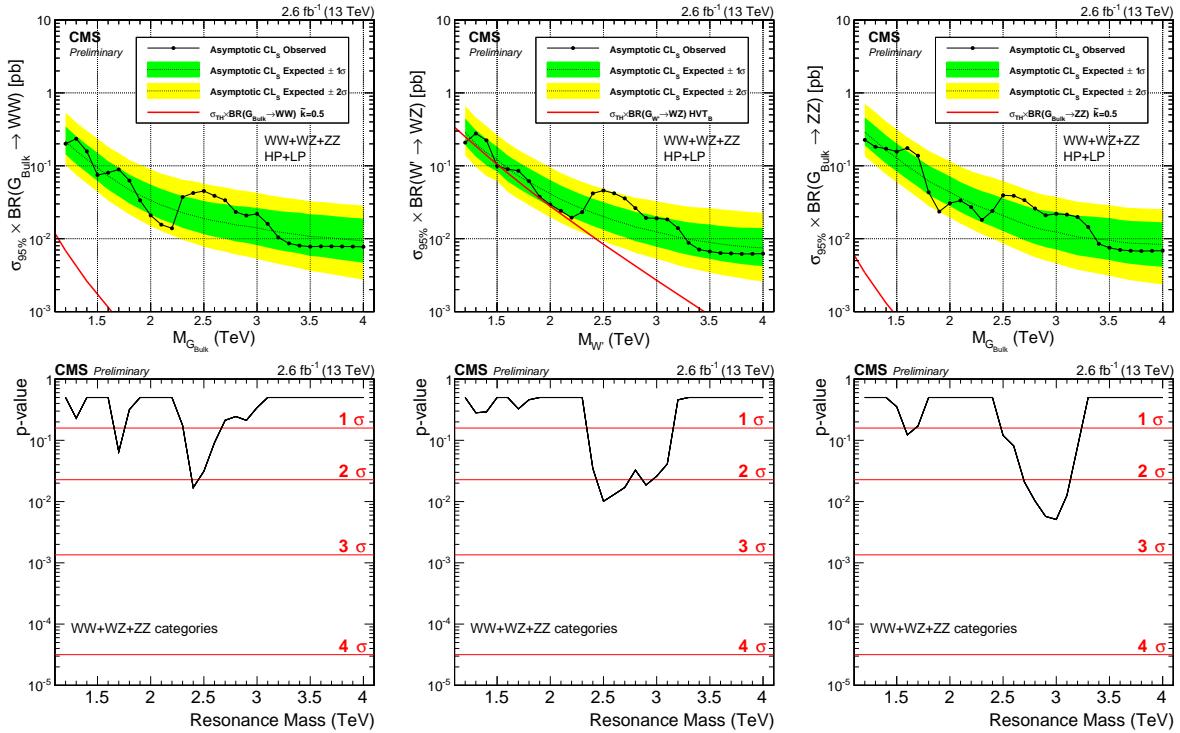


Figure 6.30: Expected and observed limits with corresponding p-values obtained using 2.7 fb^{-1} of CMS data after combining all mass and purity categories. Here for a Bulk $G \rightarrow WW$ (left), $W' \rightarrow WZ$ (middle), and $G \rightarrow ZZ$ (right) signal.

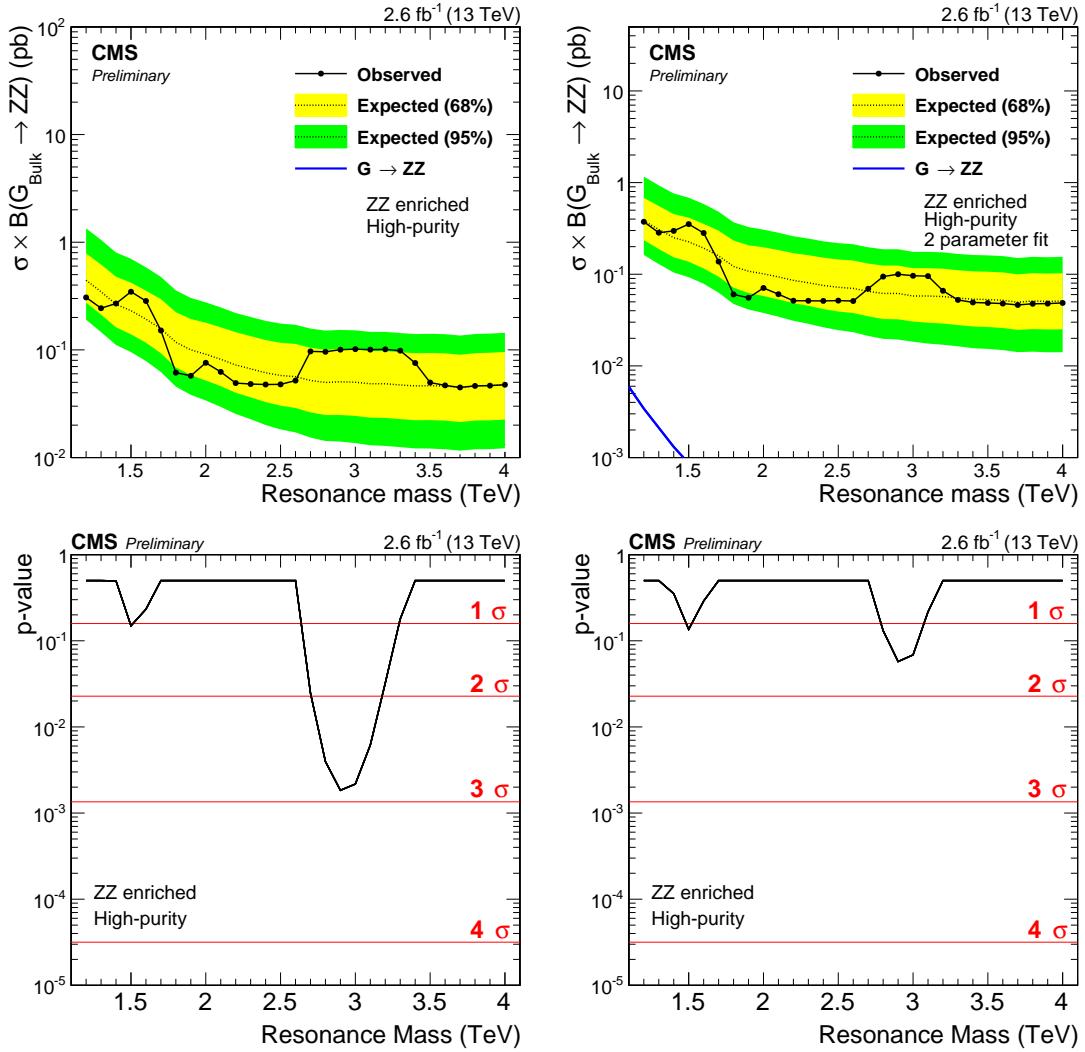


Figure 6.31: Expected and observed limits (top) and the corresponding p-values (bottom) obtained in the ZZ high-purity category using a 3- (left) and 2-parameter (right) fit to describe the background. The significance at 3 TeV is reduced from 2.8 to 1.5 σ .

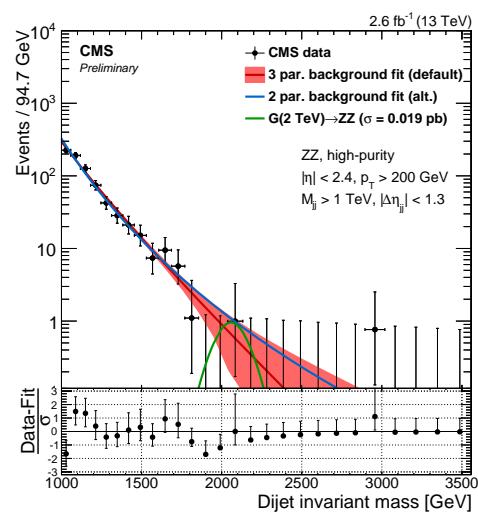


Figure 6.32: Background fit to data in the ZZHP category using the nominal 3- (red) and an alternate 2-parameter (blue) fit to describe the background.

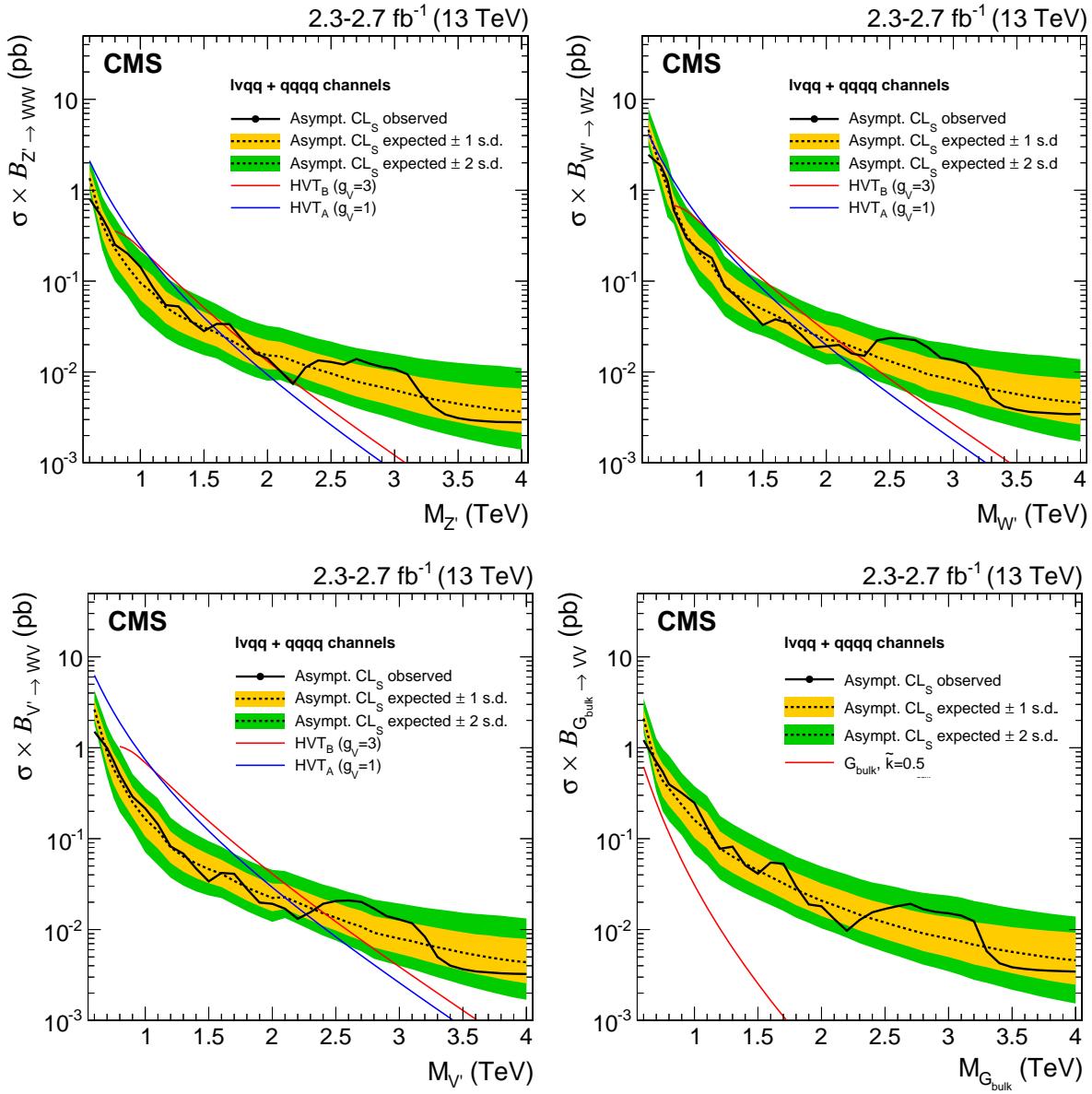


Figure 6.33: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of a narrow-width resonance decaying to a pair of vector bosons for different signal hypotheses. In the upper plots, limits are set in the context of a spin-1 neutral Z' (left) and charged W' (right) resonance, and compared with the prediction of the HVT Models A and B. In the lower left plot, limits are set in the same model under the triplet hypothesis (W' and Z' of degenerate mass). In the lower right plot, limits are set in the context of a bulk graviton with $\tilde{k} = 0.5$ and compared with the prediction.

CHAPTER 7

Search II: A new pileup resistant and perturbative safe tagger

With the first diboson resonance search with a center-of-mass energy of 13 TeV published, we concluded that more data would be needed in order to fully exclude the excess observed in Run 1. Data from 2016 was right around the corner and, with the LHC planning to reduce β^ from 80 to 40 cm, the machine was expected to deliver an instantaneous luminosity three times that of the peak luminosity in 2015. This higher instantaneous luminosity, however, meant double the pileup.*

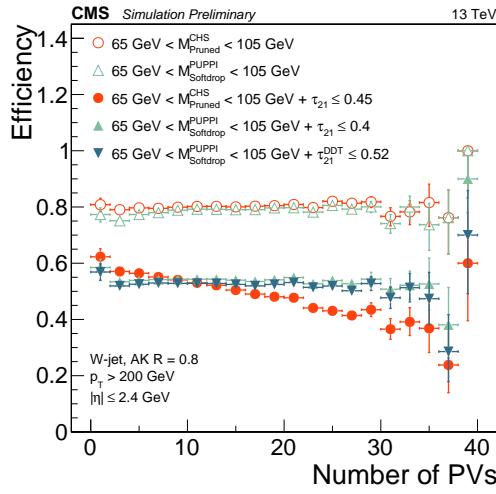
We knew that a novel pileup-subtraction algorithm had been developed, which provided far better pileup rejection than the current default algorithm that used charged hadron subtraction. We also knew that there had been made progress on the theory side in the development of a groomer which was completely insensitive to correlated soft emissions (non-global logarithms), allowing jet grooming to be accomplished in a theoretically calculable way, softdrop with $\beta = 0$ (mMDT).

With more time at hand than in 2015, we therefore decided to pursue a novel W-tagger for this second search. This meant studying and optimizing new approaches for pileup rejection and grooming, developing dedicated jet-mass corrections, as well as validating this new tagger in MC and data. This tagger, together with the mass corrections, became the default and recommended W-tagging algorithm of CMS.

Search II was the first published analysis to use a novel combination of PUPPI and softdrop grooming algorithms, now default for W-tagging in CMS. Through this search, the tagger was optimized, commissioned and validated, making it available for several analysis to come. In addition, the search was extended to consider three additional signal hypothesis. Two of these

were in a final state never before explored with a center-of-mass energy of 13 TeV, the excited quark scenario where q^* decays to qW or qZ , and the vector boson is boosted and identified through V -tagging. It was published with 35.9 (12.9) fb^{-1} of 2016 data.

- The author -



Published in PRD, DOI: 10.1103/PhysRevD.97.072006; CMS-PAS-B2G-16-021; CMS-PAS-JME-16-003

7.1 Towards robust boosted jet tagging

When we first studied W tagging with 13 TeV data, in the context of the analysis conducted with the 2015 dataset, Section 6.4.3, two interesting correlations were observed:

- 1) a strong dependence of the jet mass on the jet p_T , when using the AK8 jet clustering algorithm with CHS pileup removal and softdrop with $\beta = 0$ (mMDT) grooming, and
- 2) a strong dependence of the τ_{21} selection efficiency on pileup, when using the AK8 jet clustering algorithm with CHS pileup removal.

The reason we studied the softdrop algorithm with $\beta = 0$, hereafter just referred to as “softdrop”, as an alternative to the pruning algorithm used in the 2015 analysis was that, besides the possibility that it might perform better, we knew it had certain favorable qualities compared to other groomers; the softdrop algorithm, in addition to removing sensitivity to the soft divergences of QCD as all groomers do (they are IRC safe, see Section 5.4.1), eliminates all correlated soft emissions in the jet which are wider than the dominant two-prong substructure, leading to no non-global logarithmic terms (NGLs) in the jet mass [44]. NGLs arise from configurations where, for instance, a soft gluon is radiated into the jet cone, as illustrated in Figure 7.1. The benefit of

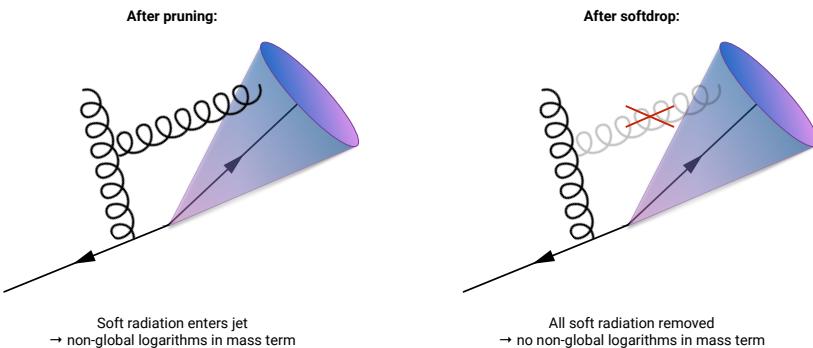


Figure 7.1: The pruning algorithm does not remove all soft emission and therefore has non-global logarithmic terms in the jet mass. Softdrop with $\beta = 0$ removes all soft emissions wider than the dominant two-prong substructure and is therefore free of non-global logarithms.

being NGL-free, is that one can calculate the softdrop jet mass to a significantly higher precision than what is possible for other grooming algorithms or for the plain jet mass (NGLs are the main reason a full resummation of the plain jet mass beyond NLL accuracy does not exist). There

were therefore theoretically well-motivated reasons for wanting the baseline CMS vector boson tagger to be softdrop-based.

However, as was discussed in Section 6.4.3, the softdrop jet-mass of signal jets displayed a large dependence on jet p_T , both for reconstructed and generator-level jets. One hypothesis was that this was due to pileup and that perhaps softdrop was more sensitive to contribution from additional interaction vertices than pruning. A study in Ref. [74] had also shown mMDT to be more sensitive to the underlying event, but using a different asymmetry criteria than softdrop (see Appendix A.1).

In parallel to the ongoing theoretical work on groomers, a novel pileup removal algorithm had been proposed: Pileup per particle identification (PUPPI) [35]. Described in detail in Section 5.3.2, PUPPI considers not only charged pileup, but also reweights neutral particles in the jet with its probability of arising from pileup. PUPPI had proven itself far superior to the current CHS algorithm in terms of jet observables for large radius jets, and therefore seemed like a good candidate to address both issues listed above: the softdrop p_T dependence and the strong pileup dependence of τ_{21} . The focus of Search II would therefore be on the commissioning of a novel W-tagger. There are interesting additions to the analysis strategy as well: the inclusion of a $Z' \rightarrow WW$ signal hypothesis and the addition of a completely new analysis, the single V-tag analysis.

7.2 Analysis strategy

The analysis strategy for this search is conceptually the same as for Search I. In addition, we take advantage of the n-subjettiness categorization and do an additional analysis in parallel: a search for excited quark resonances q^* [75, 76] decaying to qW or qZ . We call this the single V-tag analysis, and the analysis selection only differs in that one jet is not required to pass the V-tag selection (groomed mass and n-subjettiness). The VV analysis is hereby referred to as the double V-tag analysis. The difference between the two analyses is illustrated in Figure 7.2. In addition, limits are set on a $Z' \rightarrow WW$ signal hypothesis in the double V-tag analysis, another first with data collected at 13 TeV center-of-mass energy. This analysis was published in two steps: an early CMS physics analysis summary document (PAS) based on 12.9 fb^{-1} of 2016 data [72], describing the new PUPPI softdrop based V tagger as well as the single V-tag analysis, and a second analysis considering the full 2016 dataset [77]. The commissioning of the new V tagger has also been documented in a jet performance PAS [78]. In this chapter, the main emphasis will be on the work presented in CMS-PAS-B2G-16-021 [72] since this was where the new V tagger and single V-tag analysis was first presented. The second part of the results chapter, Section 7.10.2,

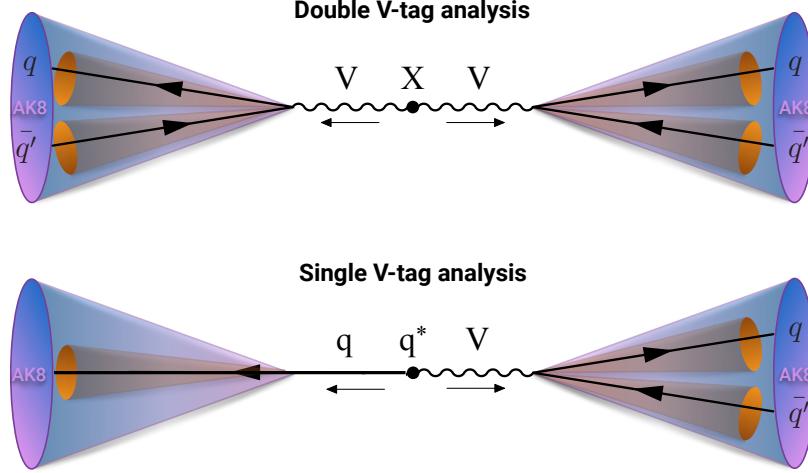


Figure 7.2: The double (top) and single (bottom) W/Z-tag analysis.

includes the results obtained using the full 2016 dataset of 35.9 fb^{-1} .

7.3 Data and simulated samples

As mentioned above, the analysis of the 2016 dataset was done in two steps corresponding to two different datasets: one analysis based on 12.9 fb^{-1} of early 2016 data, describing the new W-tagger and single V-tag category, and a second paper considering the full 2016 dataset, corresponding to 35.9 fb^{-1} . Both were collected at a center-of-mass energy of 13 TeV.

The G_{bulk} and HVT signal samples are modeled in the same way as in 2015. The q^* samples are simulated assuming unpolarized bosons and a compositeness scale Λ set equal to the resonance mass. These are generated at leading order using PYTHIA version 8.212 [63].

The standard model background processes of QCD, W+jets, and Z+jets are all simulated to leading order. The W and Z+jets backgrounds are simulated with MADGRAPH5_AMC@NLO [62, 79]. Three different combinations of matrix element and shower generators are used for QCD as these predictions are known to differ: PYTHIA-only, the leading order mode of MADGRAPH5_AMC@NLO matched with PYTHIA, and HERWIG++ 2.7.1 [51] with the CUETHS1 tune [64].

7.4 Event selection

7.4.1 Triggering

The triggers used in this analysis are the same ones as in 2015 (see Section 6.4.1), however, due to the new single V-tag analysis, the trigger turn-ons have been reevaluated separately with the requirement that either one or two jets have an offline softdrop jet mass above 65 GeV. Figure 7.3 shows the trigger turn-on curves as a function of dijet invariant mass for jets passing one of the H_T -based triggers only (pink markers), one of the grooming triggers only (green markers), and when combining all of them (purple markers). The turn-on curves are shown for all jet pairs passing loose selections as described in Section 6.4.2 (top), and for jet pairs where one (bottom left) or both (bottom right) of the jets additionally are required to have a softdrop mass larger than 65 GeV. Including the grooming triggers lowers the 99% trigger efficiency threshold by around 50 (80) GeV in the single-tag (double-tag) category once substructure is required on the analysis level. Using a combination of either of the triggers, we are safely on the trigger plateau for dijet invariant masses above 955 (986) GeV in the double (single) tag category, so that the dijet invariant-mass threshold can be set at 955 GeV for the double-tag analysis and 990 GeV for the single-tag analysis. For control plots, where no groomed-mass window is applied, a trigger threshold of 1020 GeV is used. Figure 7.4 shows the trigger efficiency as a function of the offline PUPPI softdrop-jet mass (left) and pruned-jet mass (right) for the trigger requiring an online trimmed-jet mass of at least 30 GeV. Here, the jet transverse momentum of one of the jets is required to be at least 600 GeV and no other mass cut is applied. The trigger turn-on is sharp and the plateau is reached for offline groomed-jet masses around 50 GeV.

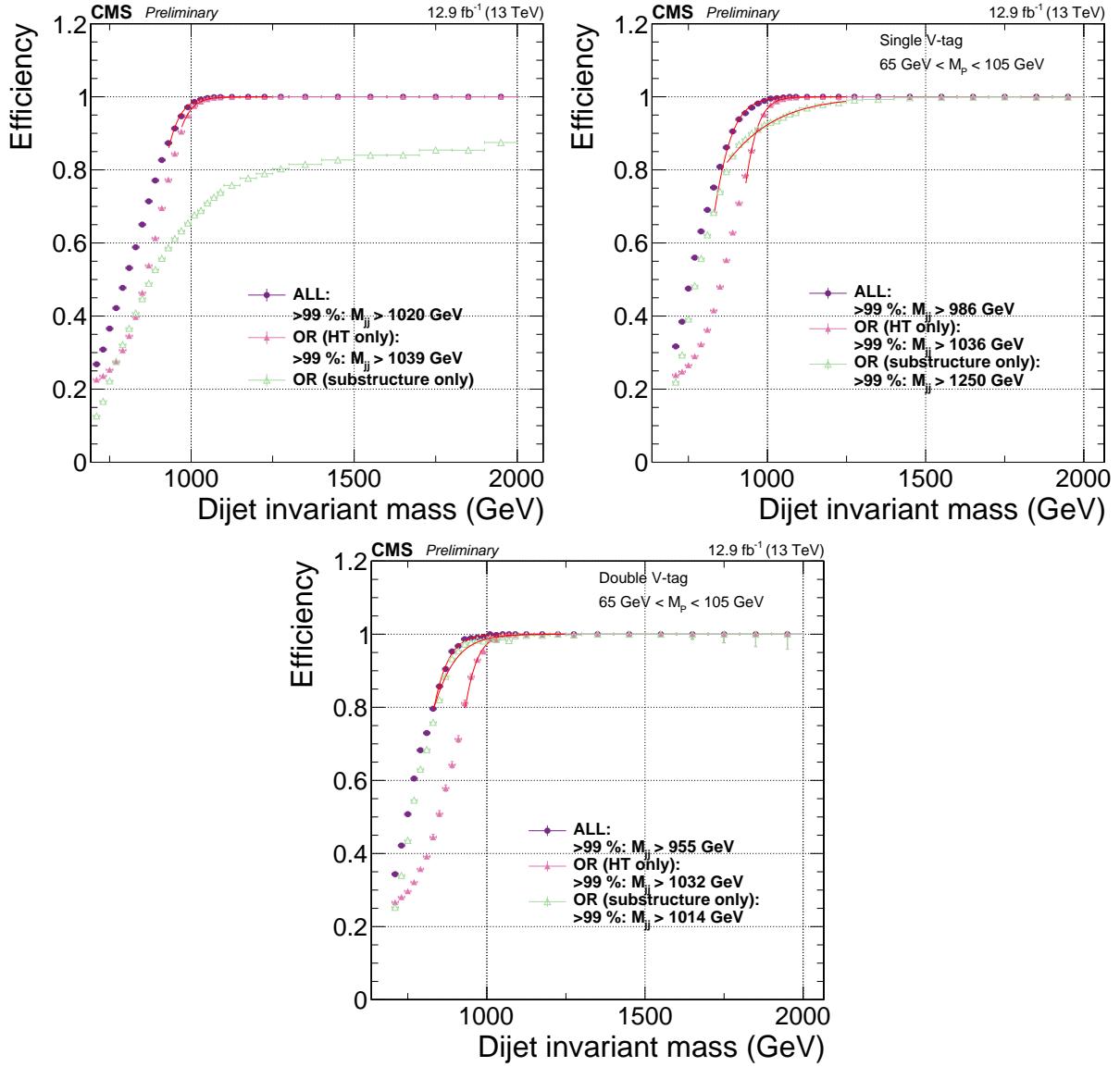


Figure 7.3: Comparison of trigger efficiencies for jets passing one of the HT-triggers only (pink), for jets passing one of the grooming-triggers only (green), and for jets passing one of the HT-triggers or one of the grooming triggers (purple). It is shown here as a function of the dijet invariant mass for all jet pairs passing loose selections (top), where one jet additionally is required to have a softdrop mass larger than 65 GeV (bottom left), and where both jets are required to have a softdrop mass larger than 65 GeV (bottom right).

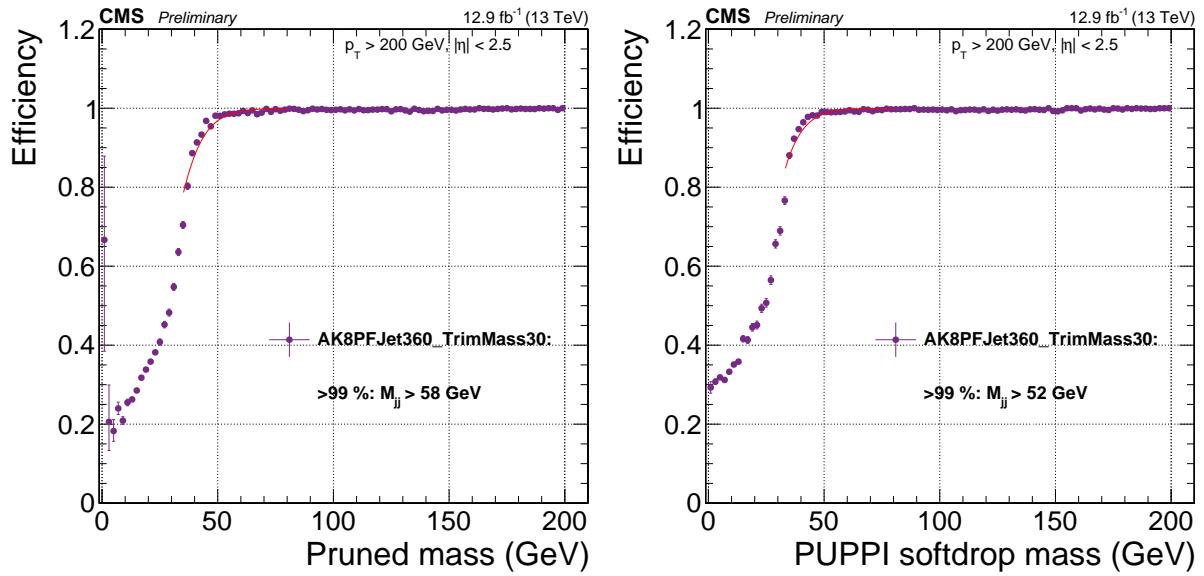


Figure 7.4: Efficiency for the trigger requiring an online trimmed jet mass of at least 30 GeV as a function of the pruned-jet (left) and softdrop-jet (right) mass for jets with $p_T > 600$ GeV.

7.4.2 Preselection

The same preselections as in Search I, described in Section 6.4.2, are used; we require two AK8 jets with CHS pileup subtraction applied, and that are required to pass the tight jet-ID requirement, $p_T > 200 \text{ GeV}$ and $|\eta| < 2.5$. The same selection requirement that suppresses t-channel QCD production is required, $|\Delta\eta| < 1.3$, together with thresholds on the dijet invariant mass of 955 GeV for the double-tag and 990 GeV for the single-tag analyses. The jet p_T (top left), η (top right), $\Delta\eta_{jj}$ (bottom left) and dijet invariant mass (bottom right) for the two leading jets in the event after loose preselections are applied is shown in Figure 7.5. A large difference in slope in

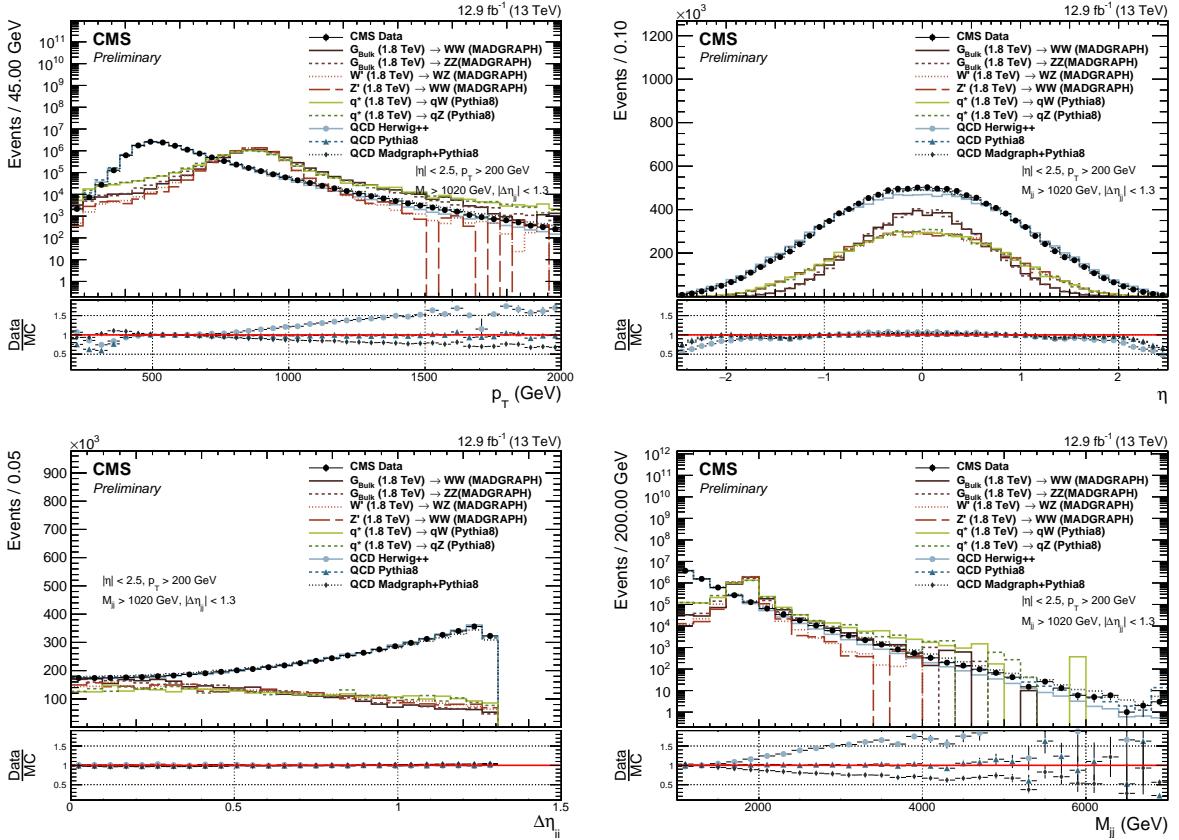


Figure 7.5: Jet p_T (top left), η (top right), $\Delta\eta_{jj}$ (bottom left) and dijet invariant mass (bottom right) for the two leading jets in the event after loose preselections are applied. The signal is scaled by an arbitrary number.

the jet p_T and dijet invariant mass spectrum depending on the QCD matrix element or shower generator is observed. Pure PYTHIA QCD MC describes the data best, while HERWIG++ tends to underestimate and MADGRAPH5_AMC@NLO+PYTHIA tends to overestimate the number of high

p_T/m_{jj} jets. Comparing these distributions to the ones in Figure 8.6, we see that PYTHIA8 now yields a far better description of the jet p_T and dijet invariant mass spectrum. This is most likely due to the change in tune from CUETP8M1 to CUETHS1 [64]. Pure PYTHIA8 QCD MC is used for all background checks in this analysis.

7.5 Developing a new W-tagger

As mentioned in the introduction to this chapter, studies had shown that the PUPPI pileup subtraction algorithm yielded superior resolution on large-cone jet observables like the jet mass. We therefore wanted to check whether the softdrop jet mass, and its observed sensitivity to the underlying event and pileup, would be improved if a better pileup subtraction algorithm was applied pre-clustering.

Two interesting observations were made. Softdrop used together with PUPPI pileup subtraction displayed a much smaller p_T -dependent shift than the combination of CHS and softdrop. Figure 7.6 shows the PUPPI softdrop jet mass for W-jets from a 1 TeV resonance, corresponding to a jet $p_T \sim 500$ GeV, and a 4 TeV resonance, corresponding to a jet $p_T \sim 2$ TeV, exhibiting the desired reduced p_T dependence on jet mass scale. However, when applying L2 and L3 jet

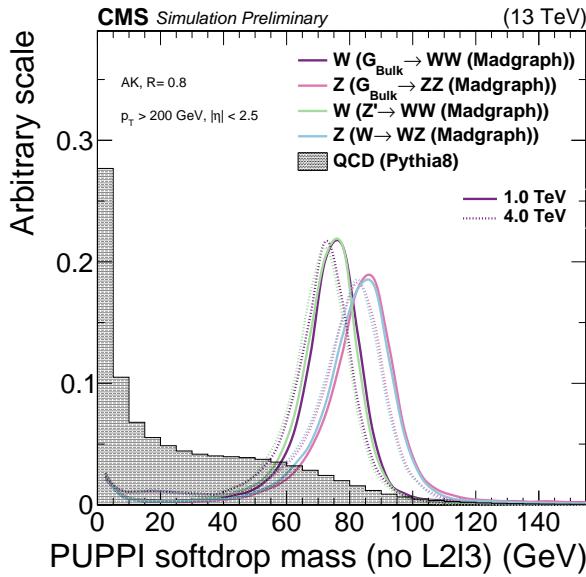


Figure 7.6: The PUPPI softdrop jet mass distribution for W boson jets coming from a 1 and 4 TeV resonance assuming different signal hypothesis. The QCD multijet background is shown in gray. No jet energy corrections have been applied to the jets.

energy corrections (see Section 5.4.3) to the jet groomed mass, a strong p_T dependent shift is

re-introduced. This effect is not present for the pruned jet mass. Figure 7.7 shows the softdrop (top left) and pruned (top right) jet mass distributions with L2L3 corrections applied. Here, the PUPPI softdrop jet mass shift as a function of p_T is significantly increased with respect to what was observed for the uncorrected mass, while the CHS pruned jet mass remains stable. This points to the PUPPI jet energy corrections not being optimal for PUPPI soft-dropped jets. Indeed, there is no reason to expect the CMS JEC corrections to work for groomed jets at all as they were derived from un-groomed PUPPI or CHS jets. The fact that they work well for CHS pruned jets is not a given. The jet energy corrections derived for CHS and PUPPI jets as a function of jet p_T are shown in the bottom plot in Figure 7.7. A significant slope in JEC as a function of p_T is measured for PUPPI, while not present for CHS.

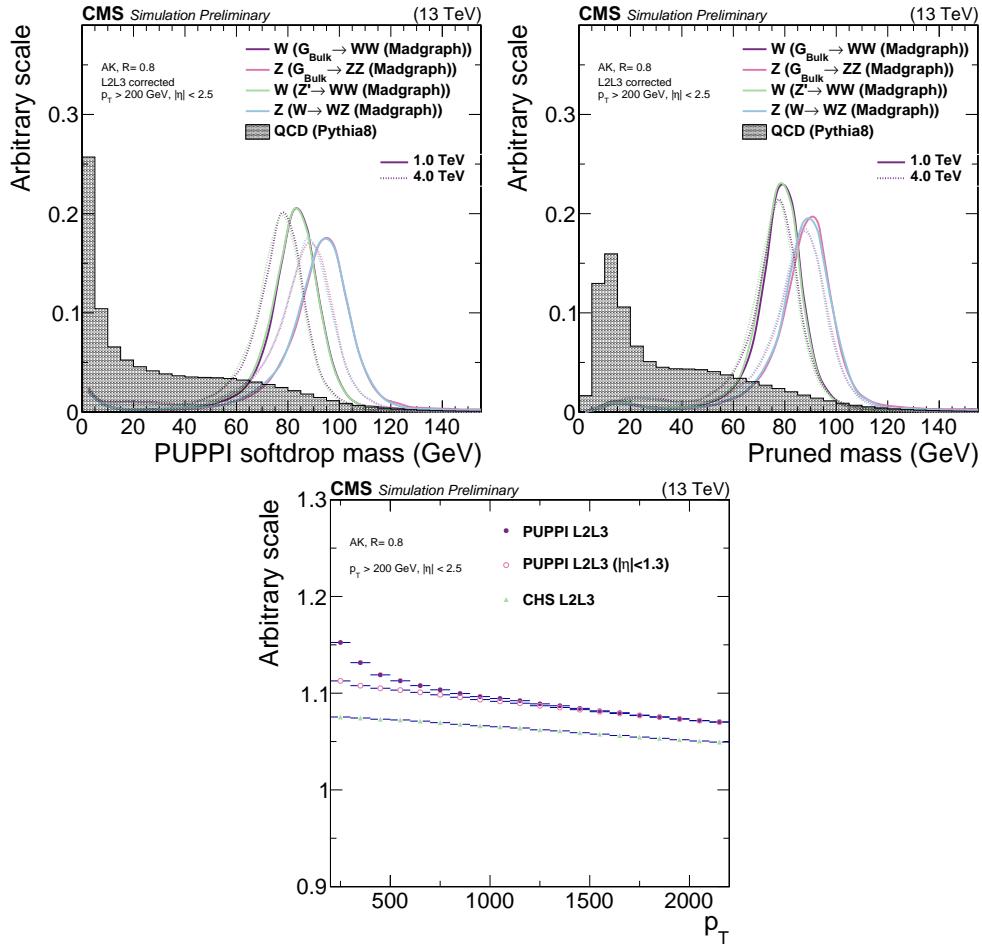


Figure 7.7: Top: PUPPI softdrop jet mass distribution (top left) and pruned jet mass distribution (top right) with L2 and L3 corrections applied. Bottom: The projection of CHS and PUPPI jet energy corrections versus jet p_T .

7.5.1 Dedicated PUPPI softdrop jet mass corrections

In order to remove the PUPPI softdrop jet mass dependence on jet p_T , all jet energy corrections to the softdrop jet mass are removed. However, this still leaves a residual p_T -dependent shift and, in addition, the uncorrected mass does not peak near the correct W-mass of 80.4 GeV. Figure 7.8 shows the mean of a Gaussian fit to the uncorrected PUPPI softdrop mass as a function of jet p_T in two different η bins (smaller or greater than $|\eta| = 1.3$) for W-jets coming from a Bulk Graviton signal sample. A mass shift both as a function of η and p_T is observed, together with an average mean shifted significantly lower than the W boson mass. In order to use PUPPI softdrop for

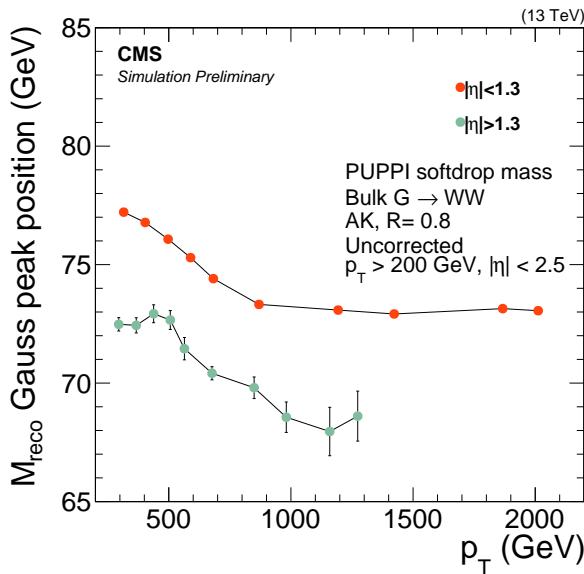


Figure 7.8: The mean of a Gaussian fit to the W-jet PUPPI softdrop mass peak as a function of jet p_T in two different η bins (smaller or greater than $|\eta| = 1.3$). No jet energy corrections have been applied to the softdrop mass.

W tagging, we therefore derive dedicated jet-mass corrections to compensate for two factors: a generator-level p_T -dependence, as first observed in Section 6.4.3, and a reconstruction-level p_T - and η -dependence due to detector effects (recall, the PUPPI jet energy corrections have been removed). Figure 7.9 shows the mean of the generated softdrop jet mass (left) and the normalized difference in reconstructed and generated softdrop jet mass (right) as a function of jet p_T . The shift in generated softdrop mass at lower p_T is of the order of 2-3% while the difference between reconstructed and generated softdrop mass is a 5-10% effect. The mass shift introduced at generator-level is corrected by a fit to M_{PDG}/M_{GEN} as a function of jet p_T , where $M_{PDG} = 80.4$ GeV and M_{GEN} is the fitted mean of the generator-level mass as shown in the left

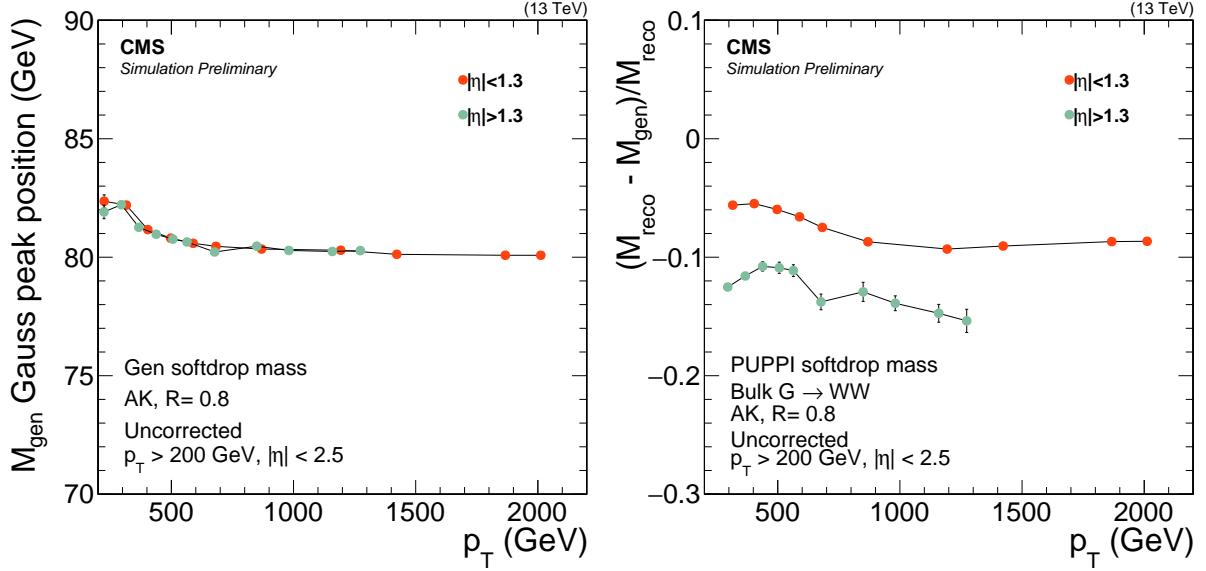


Figure 7.9: The mean of the fitted generator level W-jet softdrop mass distribution as a function of jet p_T (left) and the normalized difference in reconstructed and generated softdrop jet mass (right).

plot in Figure 7.9. To correct for the residual shift between generated and reconstructed softdrop mass, a fit to $(M_{\text{RECO}} - M_{\text{GEN}})/M_{\text{RECO}}$, where M_{RECO} is the reconstructed mass shown in the right plot in Figure 7.9 and M_{GEN} is as defined above, as a function of jet p_T in two η bins (smaller or greater than $|\eta| = 1.3$) is performed. Polynomial fit functions of the following forms are used:

$$\begin{aligned} w(p_T) &= A + B(x^2)^{-C} && \sim \text{"gen correction"} \text{ and} \\ w(p_T) &= A + Bx + Cx^2 + Dx^3 + Ex^4 + Fx^5 && \sim \text{"reco correction"}. \end{aligned}$$

The distribution and corresponding parametrization of the two corrections is shown in Figure 7.10 for the “gen correction” (left) and “reco correction” (right). The two corrections are then applied to the uncorrected PUPPI softdrop mass both in data and in MC as

$$M_{SD} = M_{\text{SD,uncorr}} \times w_{\text{GEN}} \times w_{\text{RECO}} \quad (7.1)$$

where w_{GEN} and w_{RECO} correspond to the generator and reconstructed level corrections, respectively, and $M_{\text{SD,uncorr}}$ is the uncorrected PUPPI softdrop mass.

A closure test is performed in order to check that the corrected PUPPI softdrop W-jet mass peaks at 80.4 GeV and is stable with respect to the jet p_T and η . The fitted mean of the corrected

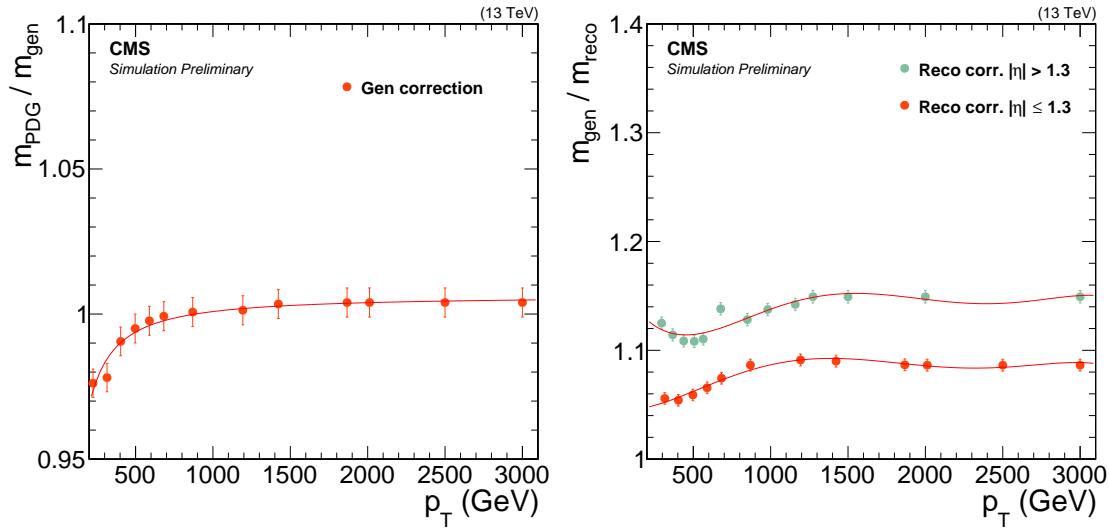


Figure 7.10: Left: fit to $M_{\text{PDG}}/M_{\text{GEN}}$ as a function of jet p_T , where $M_{\text{PDG}} = 80.4$ GeV and M_{GEN} is the fitted mean of the generator level mass. Right: fit to $(M_{\text{RECO}} - M_{\text{GEN}})/M_{\text{RECO}}$, where M_{RECO} is the reconstructed softdrop mass, as a function of jet p_T in two η bins.

PUPPI softdrop mass peak as a function of jet p_T in two different η bins is shown in Figure 7.11. Good closure is observed, with the corrected mass peaking around 80 GeV independent of the jet p_T and η . The PUPPI softdrop jet mass peak for W, Z, and Higgs boson jets from different signal samples after jet mass corrections have been applied is shown in Figure 7.12, for resonances with a mass of 1 and 4 TeV. The corrections yield a mass stable with p_T , peaking around the boson mass, for all three boson-jets under consideration.

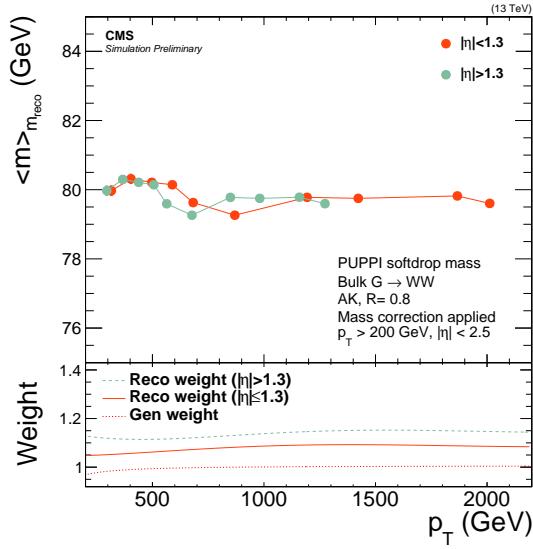


Figure 7.11: The mean of a Gaussian fit to the corrected PUPPI softdrop mass peak for real W-jets as a function of jet p_T in two different η bins.

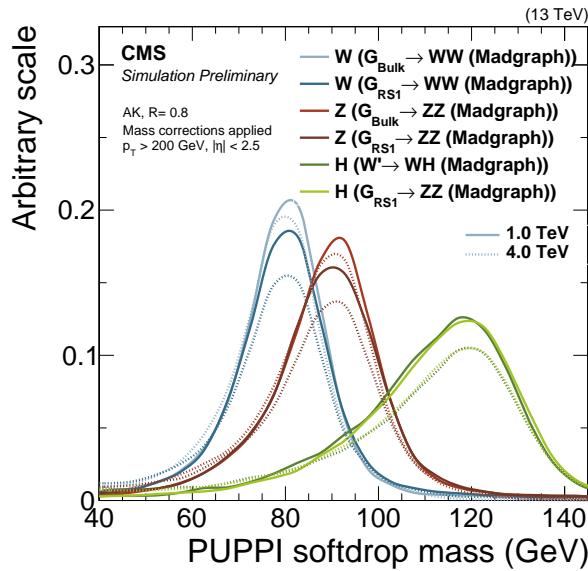


Figure 7.12: The corrected PUPPI softdrop jet mass for W, Z, and H jets from different signal samples with masses of 1 and 4 TeV.

7.5.2 W-tagging performance

The new PUPPI softdrop based V tagger uses a softdrop jet mass window of $65 \text{ GeV} < m_{SD} < 105 \text{ GeV}$ in combination with a selection of PUPPI $\tau_{21} < 0.4$. The CMS default is to compute the τ_{21} variable starting from un-groomed jets, but it can also be computed from already groomed jets. I briefly studied calculating τ_{21} starting from soft-dropped jets, summarized in Appendix A.2, where I found a small gain in performance when calculating τ_{21} starting from jets groomed with softdrop $\beta = 0$. This is something that needs further study, and is left for future work. We compare the performance of the PUPPI softdrop based V tagger used together with PUPPI τ_{21} or a “DDT-transformed” τ_{21} , as described in Section 5.5.3, to that of the CHS pruning-based tagger used in Search I.

The background rejection efficiency for QCD light-flavor jets as a function of W-jet signal efficiency is shown in Figure 7.13. The efficiency is measured after requiring a softdrop jet mass selection of $65 \text{ GeV} < m_{SD} < 105 \text{ GeV}$, while scanning the cut on τ_{21} . The general performance

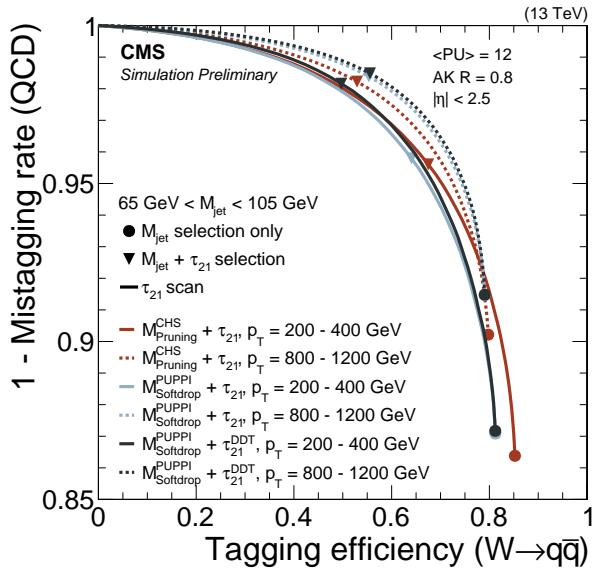


Figure 7.13: The background rejection efficiency for QCD light-flavor jets as a function of W-jet signal efficiency. A cut on CHS pruned or PUPPI softdrop jet mass of $65 < m_{p/sd} < 105 \text{ GeV}$ is applied while scanning the cut on τ_{21} . The cuts corresponding to $\tau_2/\tau_1 < 0.45$ for CHS pruned jets, PUPPI $\tau_{21} < 0.4$ or $\tau_{21}^{DDT} < 0.52$ for PUPPI softdrop jets are indicated with triangles, while the solid circles represent the efficiency and mistag rate for a mass cut only.

of each tagger is very similar, with the PUPPI softdrop-based taggers displaying a slightly higher signal efficiency for a given mistag rate at high p_T and CHS pruning-based taggers performing slightly better at low p_T . In order to better understand the difference between each tagger, we

look at the tagging performance as a function of jet p_T and pileup, as shown in Figure 7.14 and 7.15. In Figure 7.14 we observe that the signal efficiency for a PUPPI softdrop or a CHS pruned jet mass selection is flat and around 80% as a function of jet p_T . The QCD mistagging rate, on the other hand, falls off as a function of p_T , with a 1-3% lower mistagging rate when using PUPPI softdrop jet mass than when using CHS pruned jet mass. Once applying an n-subjettiness

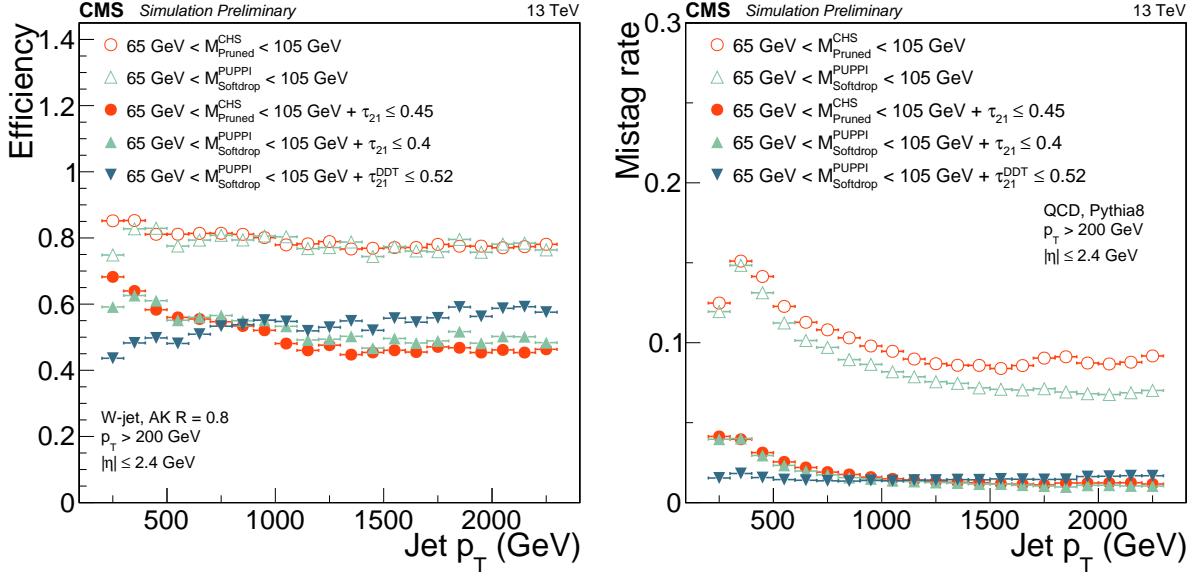


Figure 7.14: W-jet efficiency (left) and QCD light-flavor jet mistag rate (right) for a PUPPI softdrop or CHS pruned jet mass selection only (hollow circles) and the combined $m_{p/\text{sd}}$ + (PUPPI) τ_{21} (DDT) selection (solid circles) as a function of jet p_T .

cut, the signal efficiency as well as the mistag rate for the τ_{21} -based taggers becomes smaller as a function of jet p_T , with an average signal efficiency of around 50% for a $\sim 2\%$ mistag rate. An interesting behavior is observed for the τ_{21}^{DDT} tagger: while the mistagging rate is flat as a function of p_T , by construction, the signal efficiency gets higher as the p_T increases, outperforming the other taggers above 1 TeV.

Figure 7.15 shows the W-jet tagging efficiency (left) and QCD light-flavor jet mistagging rate (right) as a function of the number of interaction vertices in the event. The tagging efficiency of the CHS pruned jet tagger (red solid circles) falls off steeply versus the number of primary vertices in the event, while the PUPPI softdrop taggers (light and dark blue solid circles) are more or less insensitive to pileup. The per-jet efficiency of the PUPPI softdrop and PUPPI τ_{21} tagger is around 50-55% for a 1-2% mistag rate. Based on the observed general performance, tagging stability versus pileup, and due to theoretical considerations, PUPPI softdrop jet mass with dedicated mass corrections applied, together with PUPPI τ_{21} , is chosen as the W tagger for

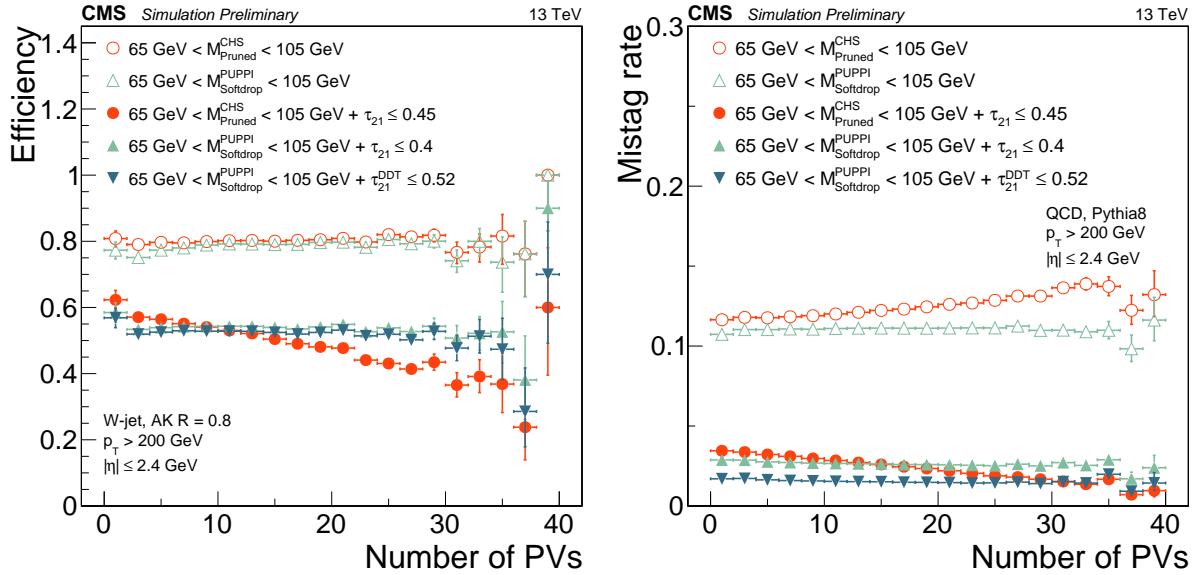


Figure 7.15: W-jet efficiency (left) and QCD light-flavor jet mistag rate (right) for a PUPPI softdrop or CHS pruned jet mass selection only (hollow circles) and the combined $m_{\text{p/sd}} + (\text{PUPPI}) \tau_{21}$ (DDT) selection (solid circles) as a function of the number of interaction vertices in the event.

this analysis.

7.5.3 Efficiency scale factors, jet mass scale and resolution

In order to measure the W-tagging efficiency, and jet-mass scale and resolution for the new PUPPI softdrop jet mass tagger, we use the same procedure as outlined in Section 6.7. The first commissioning of the new tagger was done using 2.3 fb^{-1} of data collected in 2015, and was published in a jet algorithms performance note [78]. The measurements were then redone using 12.9 and 35.9 fb^{-1} of data collected in 2016 for the two analyses presented in this chapter (the latter measurement was performed by a separate analysis team). The results shown in the following will be those obtained when the V tagger was first presented in an analysis context, corresponding to the PAS based on 12.9 fb^{-1} of data collected in the beginning of 2016. Since the fit method is outlined in detail in Section 6.7, fits to matched $t\bar{t}$ MC and minor backgrounds for the PUPPI softdrop based tagger are skipped here and can be found in Appendix C.1.

The PUPPI softdrop jet mass and the PUPPI τ_{21} variable in data and MC are shown in Figure 7.16. These can be compared to the corresponding plots for the CHS pruned jet mass and τ_{21} distributions in Figure 7.16. The softdrop and pruned jet mass distributions, as well as

the PUPPI and CHS τ_{21} variables, are very similar and the variables are described equally well in simulation. Following what was done in Section 6.7, we extract and compare the W-tagging

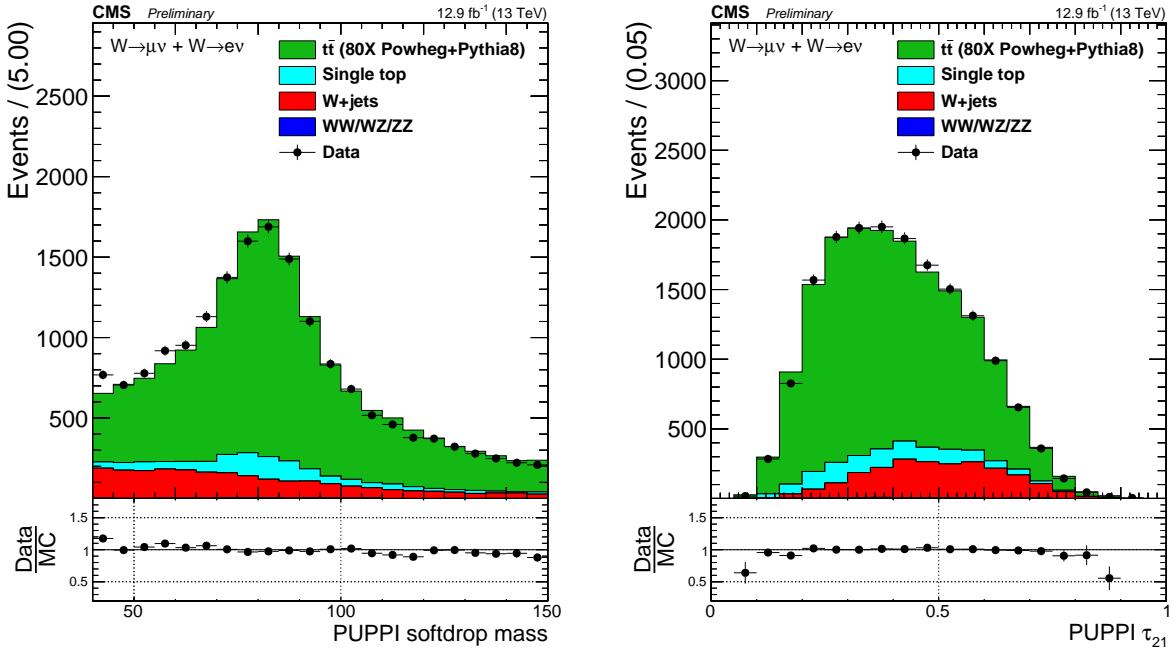


Figure 7.16: Distribution of the PUPPI softdrop mass (left) and PUPPI n-subjettiness (right) distribution in the $t\bar{t}$ control sample.

efficiency, jet mass scale and jet mass resolution of the combined jet mass and τ_{21} selection in data and in MC. This is done through a simultaneous fit of the the softdrop jet mass spectrum between 40 and 150 GeV in two regions:

- pass region: $0 < \tau_{21} \leq 0.40 \sim$ high purity, and
- fail region: $0.40 < \tau_{21} \leq 0.75 \sim$ low purity.

The fits are shown in Figure 7.17, with the corresponding extracted efficiencies from the Gaussian component of the total fit and scale factors summarized in Table 7.1. The quoted systematic uncertainties are evaluated in the same way as was described in Section 6.7.3, and correspond to the uncertainty due to the choice of ME generator and due to the choice of fit method. Both scale factors are compatible with unity within uncertainties. We additionally extract the jet-mass scale and jet-mass resolution from the mean and width of the Gaussian component of the total fit in the pass region. These are summarized in Table 7.2. We find that the W-jet mass scale for PUPPI softdrop jet mass is identical in simulation and in data, whereas for pruning

	Working point	Eff. data	Eff. simulation	Scale factor
HP	$\tau_{21} < 0.4$	0.839 ± 0.020	0.817 ± 0.012	$1.03 \pm 0.03 \text{ (stat)} \pm 0.04 \text{ (sys)} \pm 0.06 \text{ (sys)}$
LP	$0.4 < \tau_{21} < 0.75$	0.154 ± 0.020	0.176 ± 0.012	$0.88 \pm 0.12 \text{ (stat)} \pm 0.17 \text{ (sys)} \pm 0.12 \text{ (sys)}$

Table 7.1: W-tagging scale factors for both categories the high purity and low purity categories for two taggers: Pruned jet mass + τ_{21} and PUPPI softdrop jet mass + PUPPI τ_{21} .

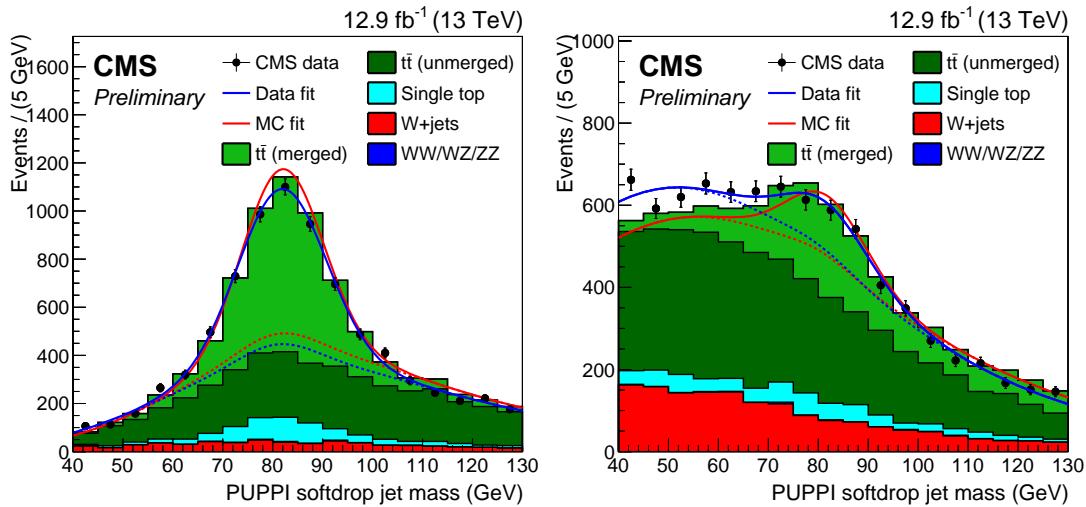


Figure 7.17: PUPPI softdrop jet mass distribution that pass (left) and fail (right) the PUPPI $\tau_{21} < 0.40$ selection. Results of both the fit to data (blue) and simulation (red) are shown and the background components of the fit are shown as short-dashed lines.

(Table 6.7) the difference was around 2%, The jet mass resolution, on the other hand, is larger in data for PUPPI softdrop jet mass, by roughly 8%, whereas for pruning the resolution is larger in simulation (11%). However, both are statistically insignificant and compatible with unity within uncertainties. The W-tagging efficiency scale factors, jet-mass and resolution scales affect the

Parameter	Data	Simulation	Data/Simulation
PUPPI softdrop $\langle m \rangle$	81.9 ± 0.3 GeV	82.0 ± 0.2 GeV	$0.999 \pm 0.004 \text{ (stat)} \pm 0.0006 \text{ (sys)}$
PUPPI softdrop σ	8.9 ± 0.4 GeV	8.3 ± 0.3 GeV	$1.08 \pm 0.07 \text{ (stat)} \pm 0.08 \text{ (sys)}$

Table 7.2: Summary of the fitted W-jet mass peak parameters.

signal yield and are included in the analysis in the same way as was described in Section 6.7.5, by scaling the total signal yield and incorporating an uncertainty on the signal efficiency due to a shift and broadening of the W-jet mass peak.

7.5.4 W-tagging mistag rate measurement

In order to understand how the modeling of substructure variables differ between MC generators, the W-tagging light-flavor jet mistagging rate is measured in a QCD dijet-enriched region in data, and is compared to the prediction from QCD MC using three different combinations of generators: HERWIG++, PYTHIA8, and MADGRAPH+PYTHIA8. Figure 7.18 shows the mistagging rate as a function of jet p_T for three different taggers: CHS pruning and τ_{21} , PUPPI softdrop and τ_{21} , and PUPPI softdrop and τ_{21}^{DDT} . We find a substantial difference in the modeling of substructure variables between the different generators, most likely coming from their very different description of gluon radiation (dominant in QCD multijet events). The best description is obtained with HERWIG++, while all three generators model the p_T -dependence of the tagger well.

We additionally study how a selection on τ_{21} and τ_{21}^{DDT} affect the total quark and gluon content of the QCD background. Figure 7.19 shows the stacked relative quark and gluon content in a PYTHIA8 QCD dijet sample for selection requirements based on PUPPI τ_{21} and τ_{21}^{DDT} . We see that the quark content increases as a function of jet p_T when applying a selection on τ_{21}^{DDT} , while it decreases when applying a selection on τ_{21} . This can be attributed to the fact that the distribution of jet mass divided by the jet p_T , m/p_T , for quark and gluon jets are very different from one another, and these differences increase as the jet p_T increases. Figure 7.20 shows the mass divided by p_T for jets originating from a quark (blue) or a gluon (red), for jets with a p_T of 200 GeV (left) or 1600 GeV (right). We see that the jet mass divided by the jet p_T peaks at significantly higher values for gluon jets than for quark jets. From the definition of the τ_{21}^{DDT} tagger in Equation 5.13, it can therefore be seen that a selection on τ_{21}^{DDT} will act more aggressively on jets with a high m/p_T , effectively removing more gluon jets.

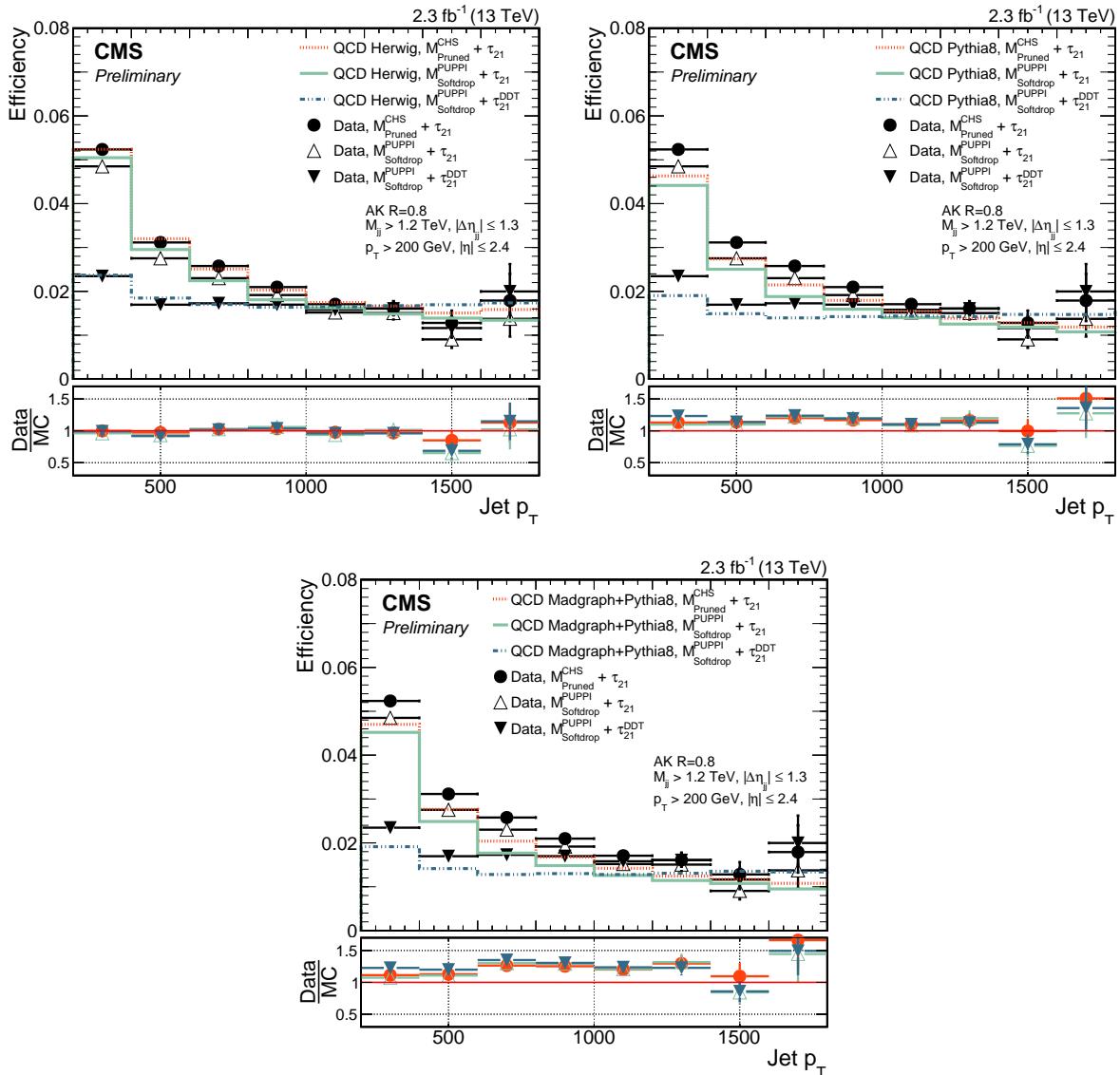


Figure 7.18: The fraction of jets that pass the $m_{p/\text{sd}}$ and τ_{21} selections in a dijet enriched sample for data and for simulation as a function of jet p_T . Here the comparison is between HERWIG++ (top left), PYTHIA8 (top right), and PYTHIA8 with MADGRAPH as the matrix-element generator (bottom).

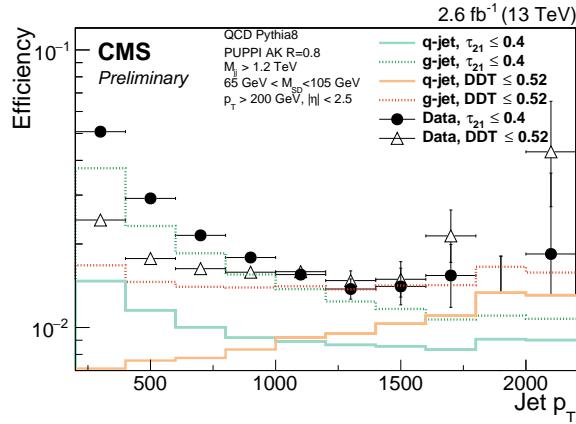


Figure 7.19: The fraction of jets that pass the PUPPI softdrop jet mass selection and a selection on τ_{21} (turquoise) or τ_{21}^{DDT} (orange) in a dijet-enriched sample. The jets from QCD MC are split into two contributions: jets originating from gluons (dotted line) and jets originating from quarks (solid line).

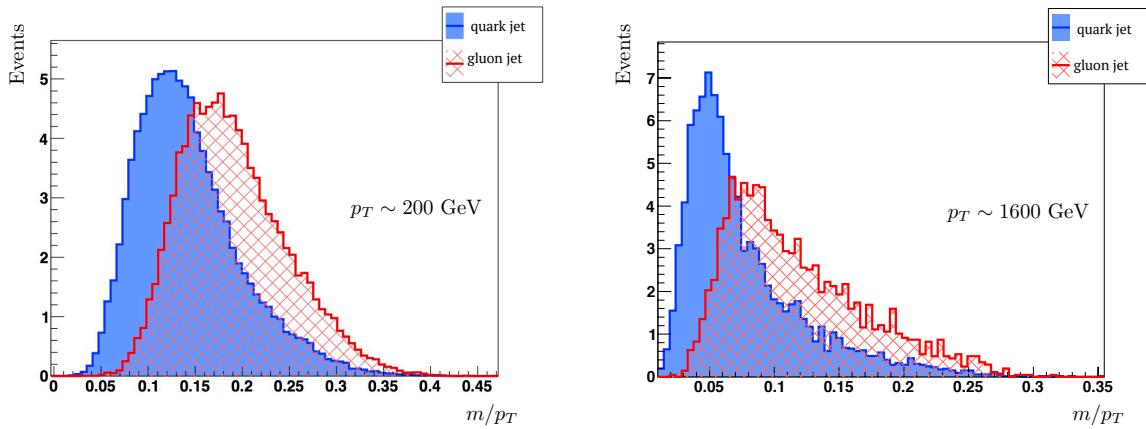


Figure 7.20: The jet mass divided by the jet p_T for quark (blue) and gluon (red) jets for a jet p_T of 200 (left) and 1600 GeV (right). Generated using [80].

7.6 Mass and purity categorization

The PUPPI softdrop jet mass and PUPPI τ_{21} distributions after loose analysis preselections, as outlined in Section 7.3, have been applied are shown in Figure 7.21. We see some disagreement between data and MC, especially in the high-purity region ($\text{PUPPI } \tau_{21}^{DDT} < 0.4$), confirming what was observed in Section 7.5.4. As this analysis is sensitive to both heavy resonances

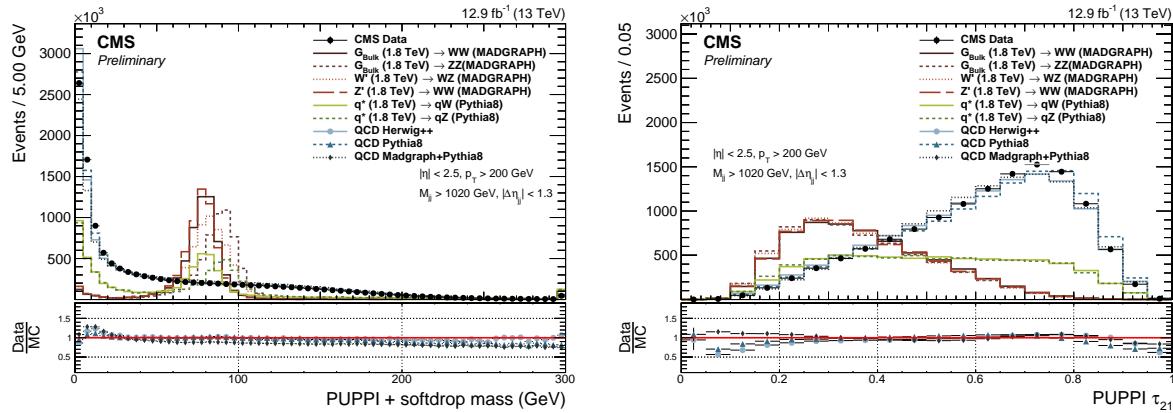


Figure 7.21: PUPPI softdrop jet mass (left) and PUPPI n-subjettiness τ_{21} (right) distribution for data and simulated samples. Simulated samples are scaled to match the distribution in data.

decaying into two vector bosons and excited quark resonances q^* decaying to qW and qZ, we look for events with both a single V-tag and events with two V-tags. Vector boson candidates are selected with a PUPPI softdrop jet mass of $65 \text{ GeV} < m_{sd} < 105 \text{ GeV}$. Further, and similar to what was done in Search I, we select “high-purity” (HP) W and Z jets by requiring $\text{PUPPI } 0 < \tau_{21} \leq 0.40$ and “low-purity” (LP) W and Z jets with $0.40 < \tau_{21} \leq 0.75$. The events with one W/Z-tag are classified in HP and LP categories according to the two categories described above. Events with two W/Z-tagged jets are always required to have one HP tagged jet, and are further divided into LP and HP categories depending on whether the other jet is of high or low purity. We additionally split events into two mass categories in order to enhance the analysis sensitivity, with the W-window defined as $65 \text{ GeV} < m_{sd} < 85 \text{ GeV}$ and the Z boson window as $85 \text{ GeV} < m_{sd} < 105 \text{ GeV}$. This results in ten different signal categories: for the double W/Z-tag analysis, there are 3 mass categories corresponding to WW, WZ, and ZZ for HP, and the same 3 mass categories for LP. For the single W/Z-tag analysis, there are 2 mass categories corresponding to qW and qZ in HP, and the same for LP.

7.7 Background modeling

The background is modeled using a parametric fit of the dijet invariant mass spectrum in the data signal region in the same way as described in Section 6.5. We determine the number of necessary fit parameters in order to describe the background through a Fishers F-test, comparing the same fit functions as in Section 6.5. This test is first exercised in QCD MC and then in a data sideband before the final determination is performed in the data signal region. As the F-test method was presented in detail in the context of Search I, only a brief summary of the findings as well as a presentation of the fits in the single-tag categories will be presented here. Fits and F-test results for all categories can be found in Appendix C.3. A two- or three-parameter fit is sufficient to describe the background for all the double-tag categories: a two-parameter fit is sufficient for the “high-purity” WZ and ZZ categories, as well as the “low-purity” WW category, while the remaining analysis categories require a three-parameter background fit. From the fits of the single-tag categories, shown in Figure 7.22, a three-parameter fit is sufficient for all categories except for the “high-purity” qW category. In the qW category, the improvement in fit quality when increasing the number of parameters is so large that adding an additional fit parameter is justified, and we continue by using a 5-parameter fit for this category. A summary of what fit functions are used for each analysis category is listed in Table 7.3.

Mass category	N pars.	
	HP	LP
WW	3	2
WZ	2	3
ZZ	2	3
qW	5	3
qZ	3	3

Table 7.3: The number of fit parameters used in the background fit to the dijet invariant mass distribution for each analysis category as determined through an F-test.

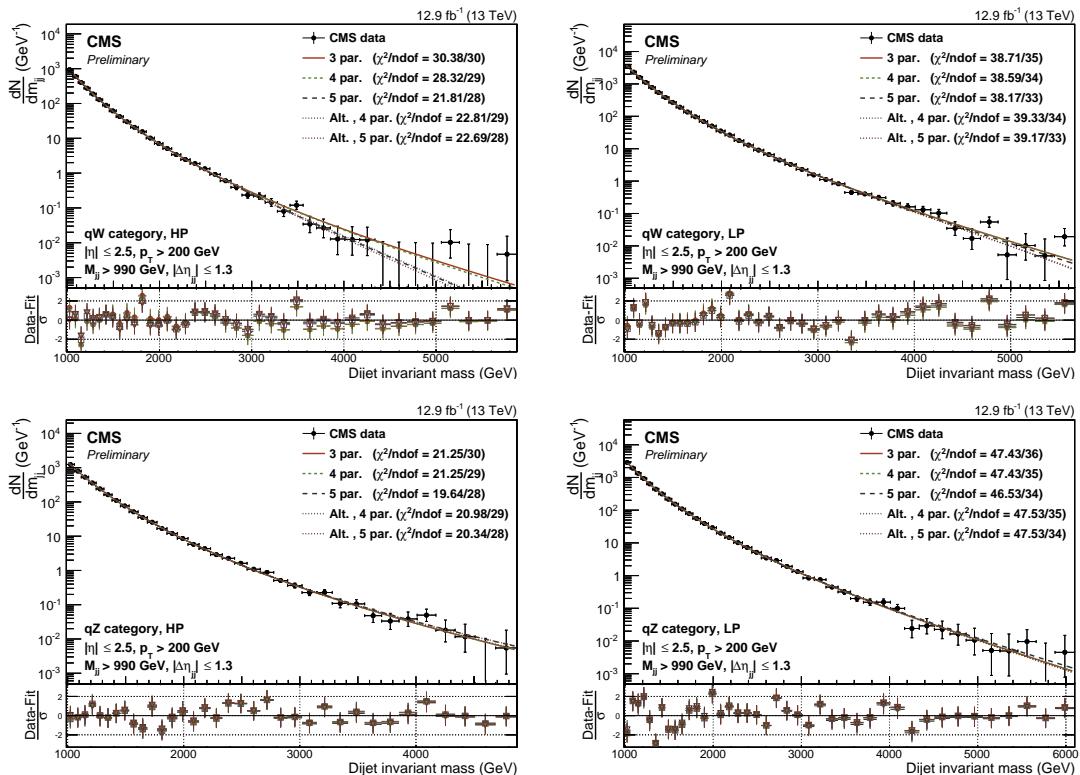


Figure 7.22: Background-only fit of the dijet invariant mass distribution in the data signal region used to establish the number of parameters of the background shape for the single-tag analysis. Here for the high- (left) and low-purity (right) single V-tag categories qW (top) and qZ (bottom).

7.8 Signal modeling

The dijet invariant mass distribution for signal is modeled from MC in the same way as in Section 6.6, where we assume the shape can be described by a Gaussian core and an exponential tail. The interpolated signal shapes for $q^* \rightarrow qW$ and $q^* \rightarrow qZ$ in their most sensitive analysis categories (qW and qZ , respectively) are shown in Figure 7.23. The signal shapes for the double-tag categories can be compared to those in Figure 6.22.

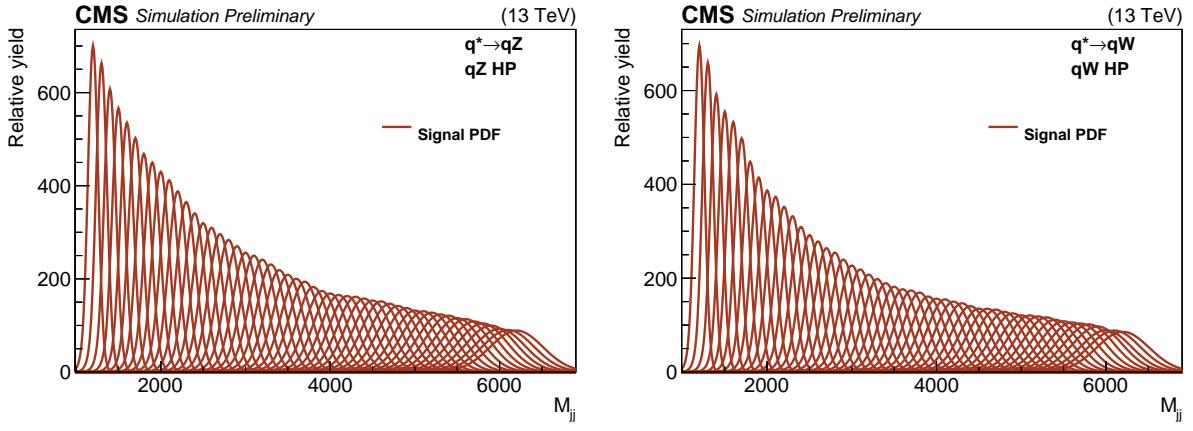


Figure 7.23: Interpolated signal shapes for a $q^* \rightarrow qZ$ (left) and $q^* \rightarrow qW$ (right) signal.

7.9 Systematic uncertainties

The largest sources of systematic uncertainty for this analysis are, as for Search I, related to the signal modeling and are due to the uncertainty in the tagging efficiency of the V-tagger, the jet energy and mass scale, the jet energy and mass resolution and the integrated luminosity. The V-tagging uncertainty is estimated as described in Section 7.5.3, and yield uncertainties on the scale factors for the HP and LP tagging categories. The p_T - and η -dependent jet energy scale and resolution uncertainties on the resonance shape are approximated by a constant 2% and 10%. The jet energy scale and resolution uncertainty are taken into account as shape uncertainties by shifting and widening the signal resonance model, while all other signal uncertainties only affect the yield. The most relevant systematic uncertainties are listed in Table 7.4.

Source	Relevant quantity	HP+HP unc. (%)	HP+LP unc. (%)
Jet energy scale	Resonance shape	2	2
Jet energy resolution	Resonance shape	10	10
Jet energy scale	Signal yield	<0.1–4.4	
Jet energy resolution	Signal yield	<0.1–1.1	
Jet mass scale	Signal yield	0.02–1.5	
Jet mass resolution	Signal yield	1.3–6.8	
Pileup	Signal yield	2	
Integrated luminosity	Signal yield	6.2	
PDFs (W')	Signal yield	4–19	
PDFs (Z')	Signal yield	4–13	
PDFs (G_{bulk})	Signal yield	9–77	
Scales (W')	Signal yield	1–14	
Scales (Z')	Signal yield	1–13	
Scales (G_{bulk})	Signal yield	8–22	
Jet mass scale	Migration	<0.1–16.8	
Jet mass resolution	Migration	<0.1–17.8	
W-tagging τ_{21}	Migration	15.6	21.9
W-tagging p_T -dependence	Migration	7–14	5–11

Table 7.4: Summary of the systematic uncertainties on the signal and their impact on the event yield in the signal region and on the reconstructed dijet invariant mass shape (mean and width).

7.10 Results

As mentioned in the introduction to this chapter, the analysis of the 2016 dataset was done in two steps: one based on 12.9 fb^{-1} of data collected early in 2016, demonstrating the new PUPPI softdrop V-tagger and the new single-tag analysis categories, and one utilizing the full 35.9 fb^{-1} of data collected in 2016. The results from both will be presented in the following sections.

7.10.1 Early analysis

The dijet invariant mass distributions in the analysis signal regions using 12.9 fb^{-1} of data collected in 2016, are shown in Figure 7.24 for the double-tag and in Figure 7.25 for the single-tag analysis. No significant excess is observed and we proceed to set upper limits on the signal cross section using the method described in Section 6.9. Exclusion limits are set in the context of the bulk graviton model, the HVT model B, and excited-quark resonance models, assuming the resonances have a natural width negligible with respect to the experimental resolution (as in Search I). Figure 7.26 shows the expected and observed upper exclusion limits at 95% confidence level (CL) on the signal cross section times the branching fraction as a function of the resonance mass for the different signal hypotheses in the double-tag analysis. The limits are compared with the theoretically predicted cross section times branching fraction to WW and ZZ for a bulk graviton with $\tilde{k} = 0.5$, and with the cross section times branching fraction to WZ and WW for spin-1 particles predicted by the HVT model B. For the HVT model B, we exclude W' (Z') resonances with masses below 2.7 (2.6) TeV. The signal cross section uncertainties are displayed as a red checked band and result in an additional uncertainty on the resonance mass limits of 0.05 (0.04) TeV. The cross section limits for $Z' \rightarrow WW$ and $G_{\text{bulk}} \rightarrow WW$ are not identical due to the different acceptance for those two signal scenarios. Figure 7.31 shows the corresponding exclusion limits for excited quarks decaying to qW and qZ, excluding a q^* decaying into qW and qZ with masses below 5.0 and 3.9 TeV, respectively. The signal cross section uncertainties are displayed as a red checked band and result in an additional uncertainty on the resonance mass limits of 0.1 TeV.

7.10.2 Full 2016 dataset

Figure 7.28 shows the dijet invariant mass distributions in the analysis signal regions using the full dataset of 35.9 fb^{-1} collected in 2016 for the double-tag and in Figure 7.29 for the single-tag analysis. No significant excess is observed and we proceed with setting upper limits on the signal cross section times branching ratio as in Section 6.9.

For a G_{bulk} we exclude production cross sections in a range from 36.0 fb, at a resonance mass of 1.3 TeV, to 0.6 fb at a resonance mass of 3.6 TeV. The W' (Z') resonances are excluded with masses below 3.2 (2.7) TeV for the HVT model B, in addition to W' resonances with a mass between 3.3 and 3.6 TeV. For excited quark resonances, we can exclude the production of q^* decaying to qW or qZ for masses below 5.0 and 4.7 TeV, respectively. Figure 7.30 and 7.31 show the resulting expected and observed upper limits at 95% confidence level on the signal cross section times branching ratio as a function of the resonance mass for VV and qV resonances, respectively.

The sensitivity improvement in terms of mass reach with respect to Search I for a W' resonance of the HVT model B, improved from excluding resonances below 2 TeV to excluding resonances below 3.2 TeV.

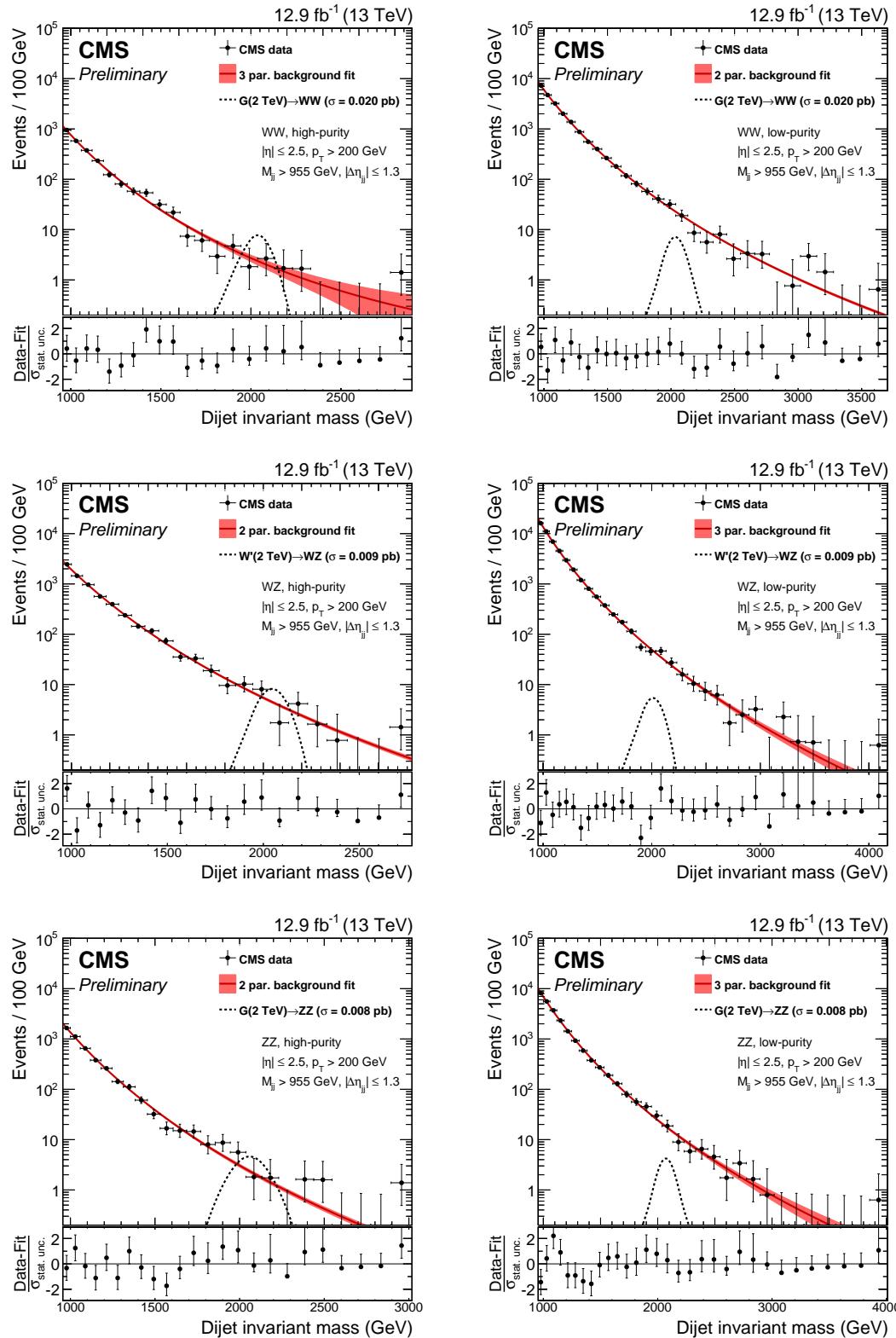


Figure 7.24: Observed dijet invariant mass spectrum for the double-tag analysis in the high-purity (left) and low-purity (right) categories using 12.9 fb^{-1} of data collected in 2016.

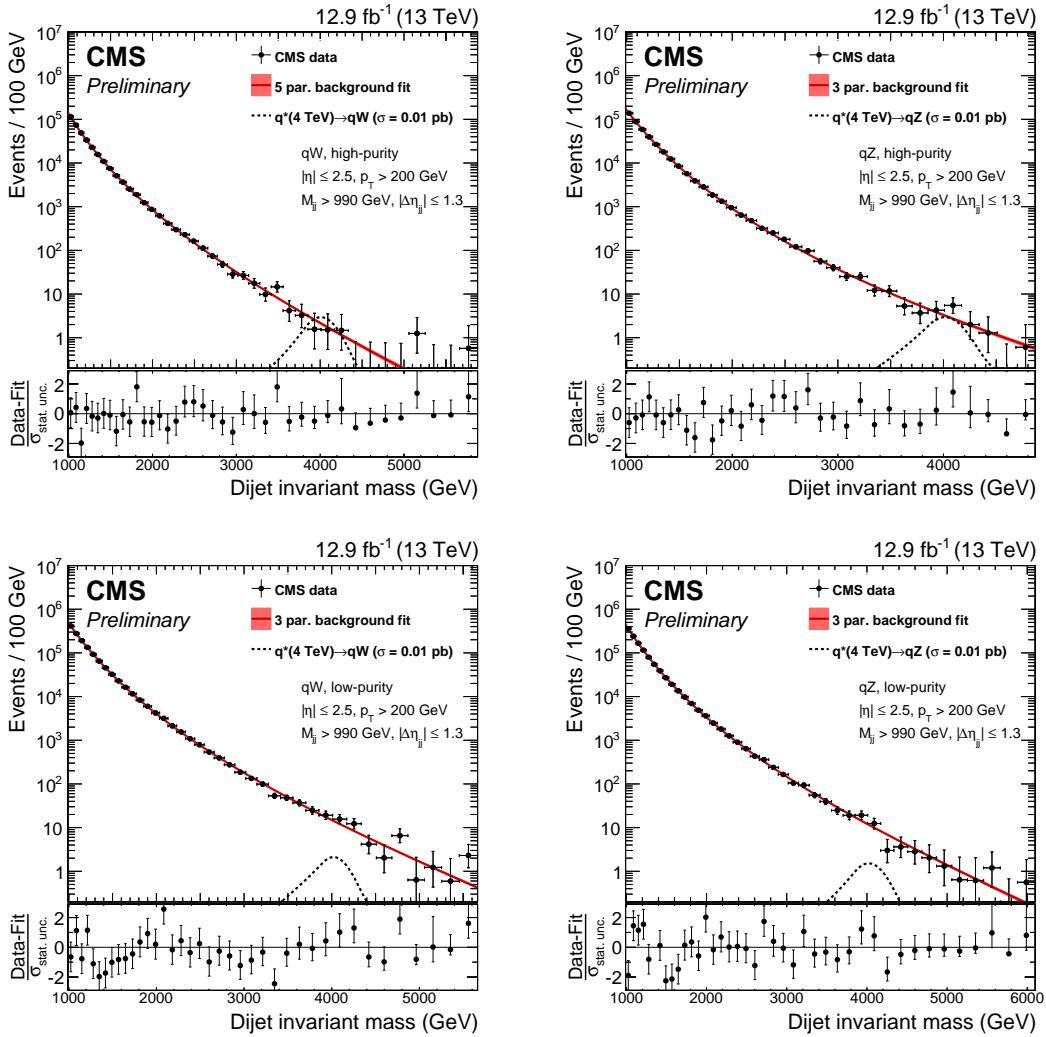


Figure 7.25: Observed dijet invariant mass spectrum for the single-tag analysis in the high-purity (top) and low-purity (bottom) categories using 12.9 fb^{-1} of data collected in 2016.

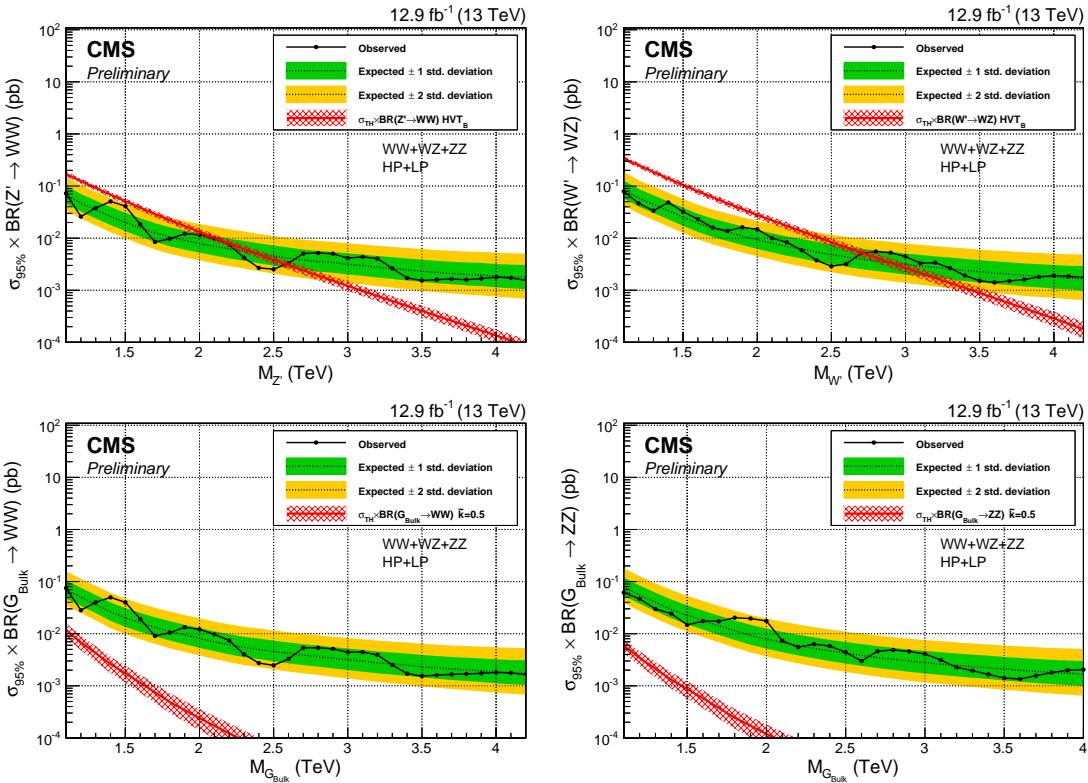


Figure 7.26: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of a narrow-width resonance decaying to a pair of vector bosons for different signal hypotheses. Limits are set in the context of a spin-1 neutral Z' (left) and charged W' (right) resonance, and compared with the prediction of the HVT model B. On the bottom, limits are set in the context of a bulk graviton decaying into WW (left) and ZZ (right) with $\tilde{k}=0.5$ and compared with the model prediction. Signal cross section uncertainties are displayed as a red checked band.

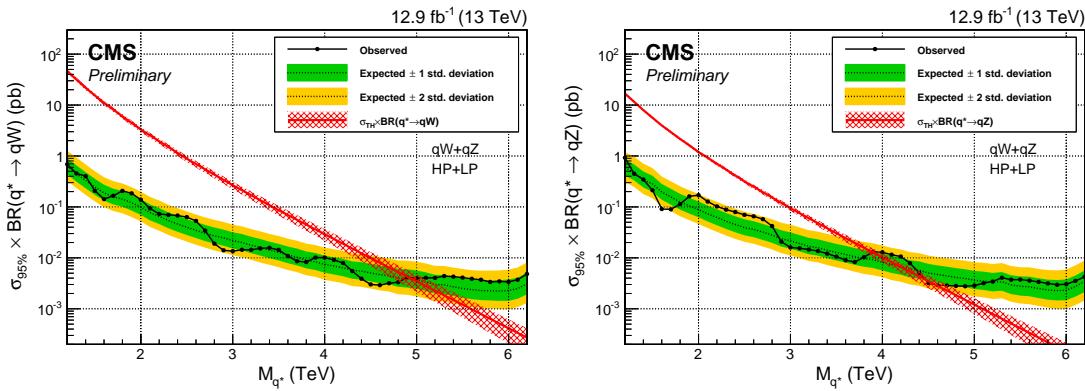


Figure 7.27: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of an excited quark resonance decaying into qW (left) or qZ (right). Signal cross section uncertainties are displayed as a red checked band.

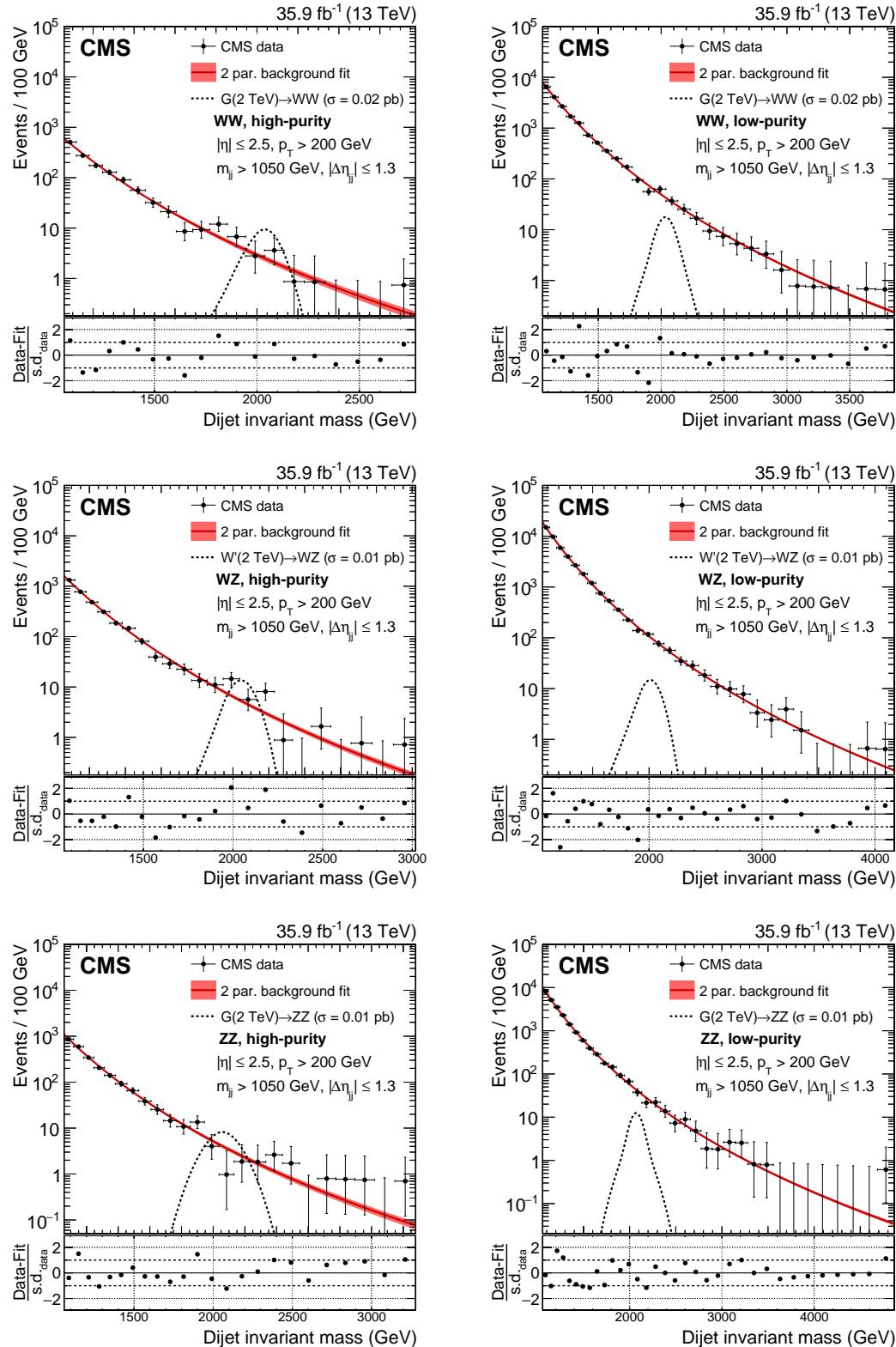


Figure 7.28: Observed dijet invariant mass spectrum for the double-tag analysis in the high-purity (left) and low-purity (right) categories using 35.9 fb^{-1} of data collected in 2016.

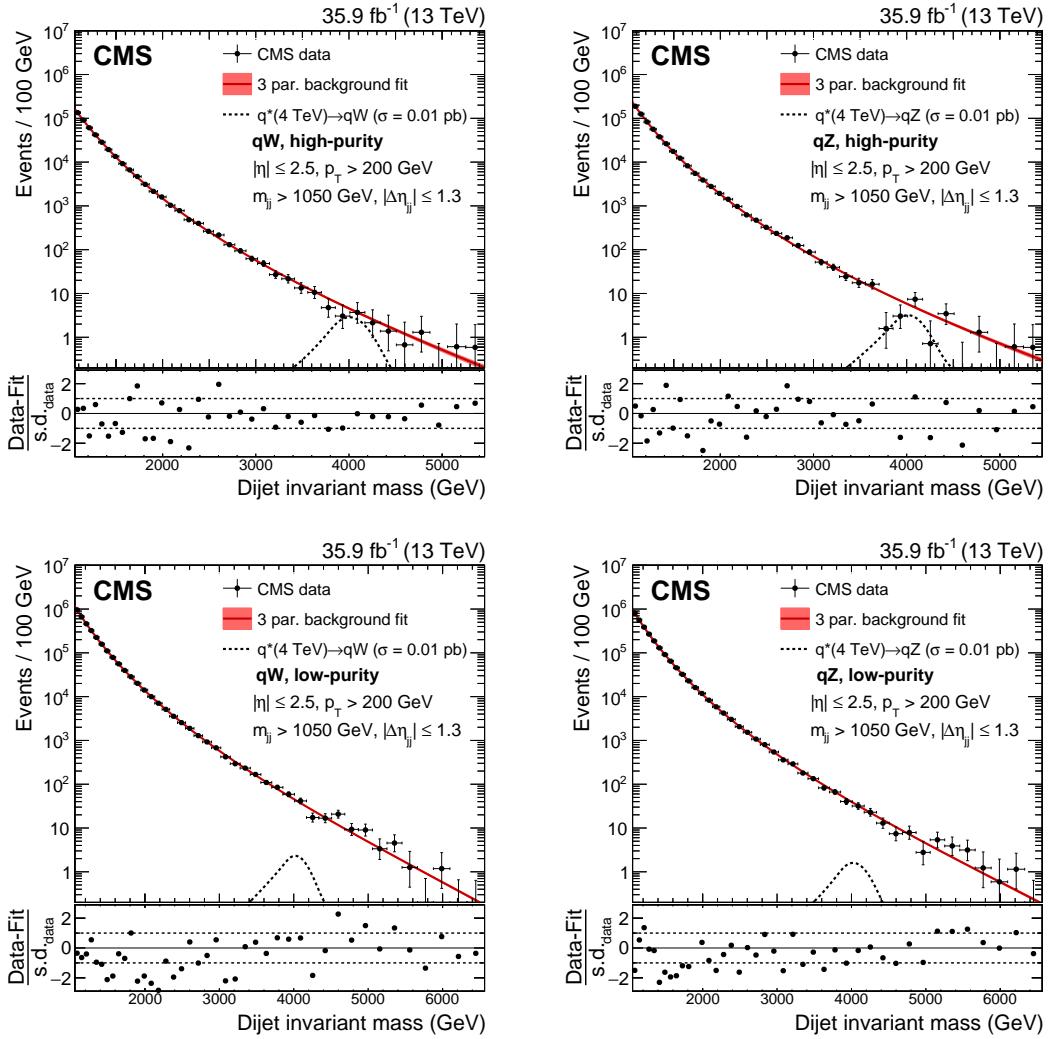


Figure 7.29: Observed dijet invariant mass spectrum for the single-tag analysis in the high purity (top) and low purity (bottom) categories using 35.9 fb^{-1} of data collected in 2016.

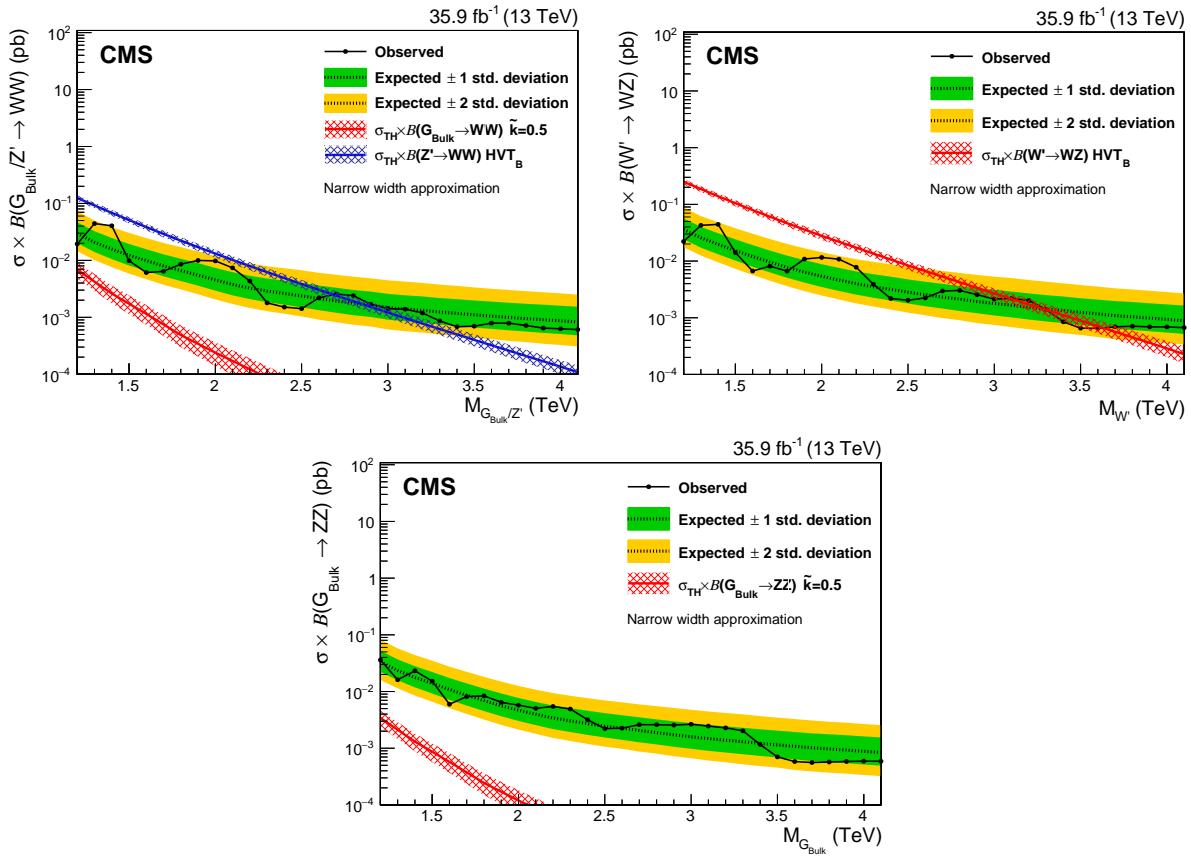


Figure 7.30: Observed (solid line) and expected (dashed line) 95% CL upper limits on the production cross section of a narrow resonance decaying into two vector bosons for different signal hypotheses: a Z' or G_{bulk} resonance decaying into WW (top left), a Z' decaying into WZ (top right) and a bulk graviton decaying into ZZ' (bottom).

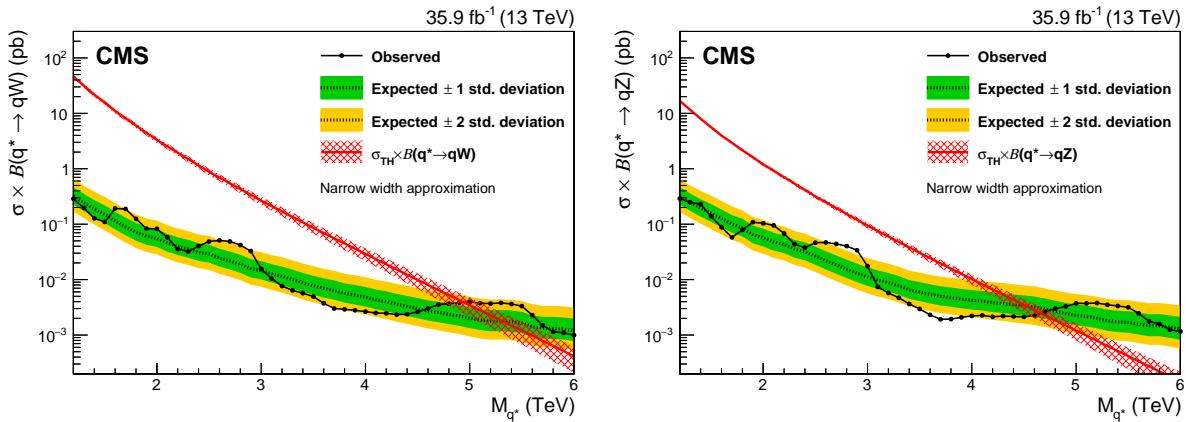


Figure 7.31: Observed (solid line) and expected (dashed line) 95% CL upper limits on the production of an excited quark resonance decaying into qW (left) or qZ (right).

CHAPTER 8

Search III: A novel framework for multi-dimensional searches

After two successful analysis using the LHC data collected at a center-of-mass energy of 13 TeV, there was no confirmation of the excess observed at a center-of-mass energy of 8 TeV, and the prospect of observing new physics in diboson analyses was bleak. However, this was not a source of concern to some theorists, who considered whether the small bumps that had been observed were due to us observing the tail of another type of boson with a mass slightly different from that of a W or a Z boson, and that perhaps these bosons were not 2-prong, but 4-prong objects.

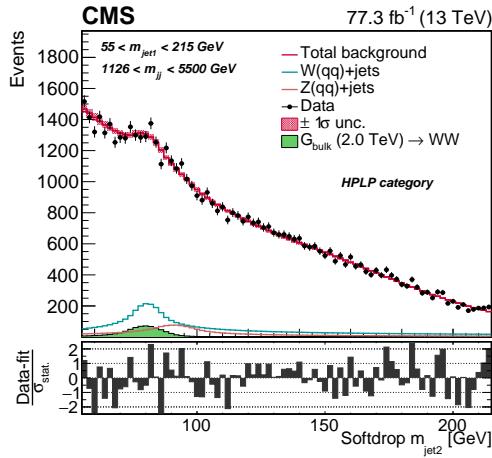
With no significant excess observed with the 2016 dataset of 36 fb^{-1} , we were expecting to collect a total dataset of 150 fb^{-1} in Run 2 (2015-2018) that would allow us to probe alternative BSM models, and we wanted to do so as efficiently as possible. More specifically, we wanted a way of looking for any heavy resonance decaying to any two jets with substructure, with masses anywhere in the jet mass spectrum. We decided to do so by taking advantage of the fact that we are looking for bumps in a three-dimensional space: the mass of the two jets as well as their dijet invariant mass. The one-dimensional dijet fit can therefore be replaced by a three-dimensional fit, looking for bumps in all three dimensions simultaneously. This would allow us to easily search for resonances decaying to any object peaking in jet mass; $W(qq)$, $Z(qq)$, $H(qq)$, as well as non-SM bosons; in one common framework.

An additional benefit of the method, was that the modeling of the QCD multijet background would start from simulation rather than relying on a parametric fit to data. This had the benefit of allowing more control over the background shape across the full dijet invariant mass spectrum, making the fit less prone to fluctuations in the low statistics tail of the distribution, something we

found to be problematic with the 1D dijet fit method.

Search III introduces a novel three-dimensional fit method that can be used to search for heavy resonances decaying to any two jets with substructure peaking in the jet groomed mass spectrum. It is validated in the context of a diboson resonance search in the all-hadronic final state, but is easily extendable to other signals. The fit method has allowed for the first measurement of the jet mass scale and resolution simultaneously from a $W(qq)+\text{jets}$ and $Z(qq)+\text{jets}$ mass peak, and could also allow for the extraction of the SM $W/Z+\text{jets}$ cross section. It was published with data collected in 2016 and 2017, corresponding to a total integrated luminosity of $\sim 80 \text{ fb}^{-1}$.

- The author -



In progress. To be submitted to The European Physical Journal C

8.1 Small bumps and tri-bosons

In addition to the observation of a 3.4σ excess in the search for diboson resonances in the all-hadronic final state by ATLAS [53], not confirmed by subsequent searches, several little bumps near 2 TeV were observed in data collected at center-of-mass energies of 8 and 13 TeV, as illustrated by the dijet invariant mass distributions observed by ATLAS and CMS in Figure 8.1. These were not statistically significant, and are expected due to statistical fluctuations, however, only a coherent analysis would be able to determine if a new physics signal could be partly responsible for any of these excesses in the dijet invariant mass of V jets. Due to their small size

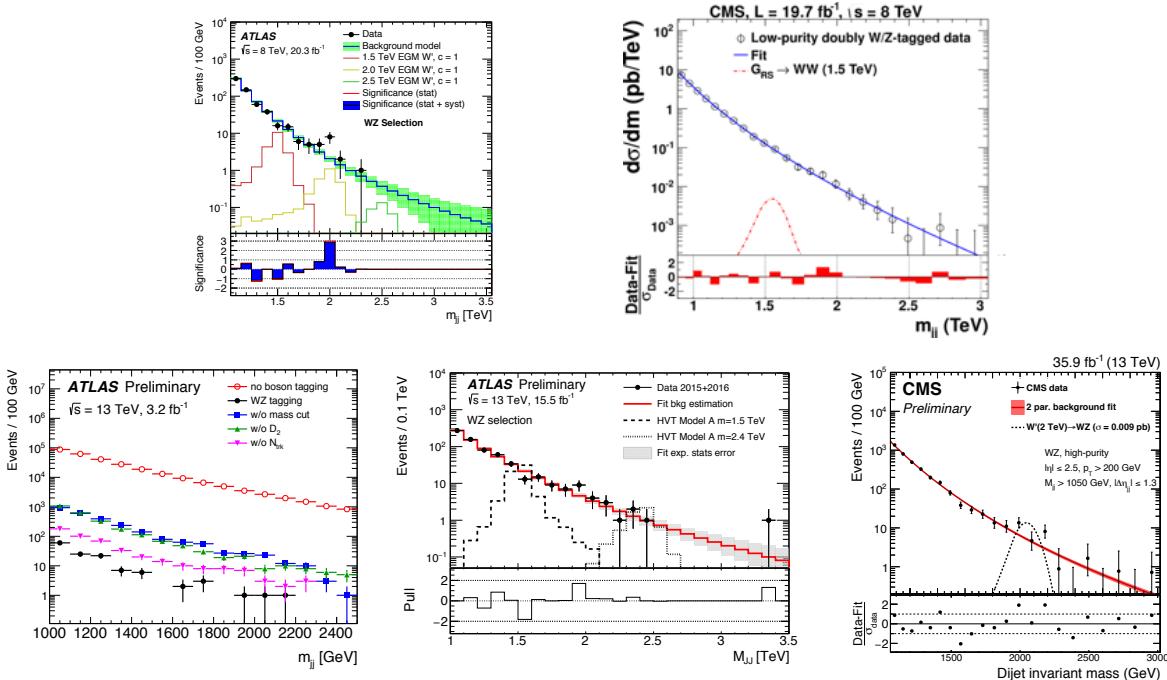


Figure 8.1: Several small bumps have been observed in diboson resonance searches in the all-hadronic final state, both in ATLAS and in CMS [81].

and the way the excesses appeared to slightly shift around, these were obviously not diboson resonances. However, a proposal was made that they could be caused by the observation of a non-SM boson, such that jets with a mass several standard deviations away from the new particle's mean, its so-called *tail*, were being observed, as illustrated in Figure 8.2. Further, these could be 4-pronged objects rather than 2-prong, which would cause the excess to vary in size depending on the 4-prong efficiency of the analysis specific W-tagger in use. An explanation for the observed excesses was proposed in [82]. This paper pointed out that, if particles like W' and

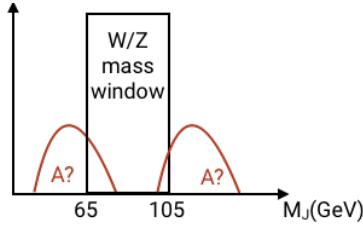


Figure 8.2: The small excesses in diboson searches could be caused by the observation of the tail of a non-SM object, peaking at a groomed jet mass slightly higher or lower than that of a W or Z boson.

Z' exist, an extended scalar sector is needed in order to give mass to the vector bosons. These heavy scalars would decay to lighter bosons, if kinematically allowed, leading to a cascade of decays that produced signals of multiple bosons in a single jet. Some example signatures are illustrated in Figure 8.3. Signatures like these would peak in the groomed-jet mass spectrum and,

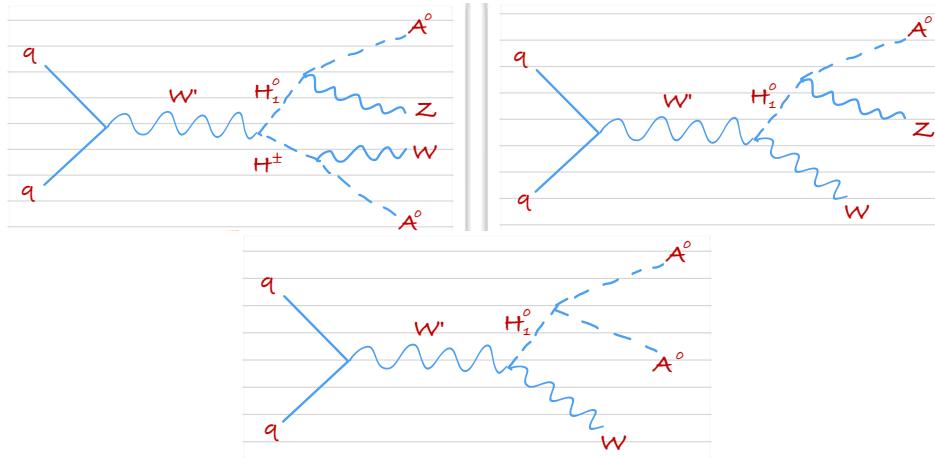


Figure 8.3: A W' decaying to a neutral H^0 and a charged H^\pm scalar particle leading to a quadriboson final state (left), and a W' decaying to a neutral scalar particle H^0 and a W leading to a triboson final state (middle and right) [82].

depending on what the final bosons decay into, have very different substructure profiles.

In order to efficiently search for such types of signals, or any signal peaking in the jet mass spectrum, we decided to build a generic framework that would allow searching for peaks anywhere in the jet-mass and dijet invariant mass spectrum. Rather than selecting jets with a groomed mass between 65 and 105 GeV and then searching for resonances peaking in the dijet invariant mass, we would attempt to look for resonances peaking anywhere in the three-dimensional space formed by the groomed mass of each jet and their dijet invariant mass, scanning the full groomed mass spectrum in a single analysis. We would first demonstrate the new method in context of

the diboson all-hadronic search since this was an analysis we were familiar with and which would allow for a straight forward comparison of the obtained results.

8.2 Analysis strategy

The background estimation used in Search I and Search II relies on a one-dimensional fit of the dijet invariant mass of the signal region after a tight selection on the jet mass has been applied. Since the signal will peak in all three dimensions, mainly the dijet invariant mass M_{VV} , and the mass of jet 1 and jet 2 (M_{jet1} and M_{jet2}), we will extract the signal from the three-dimensional plane of M_{VV} - M_{jet1} - M_{jet2} . The benefits of doing so are that we can search for

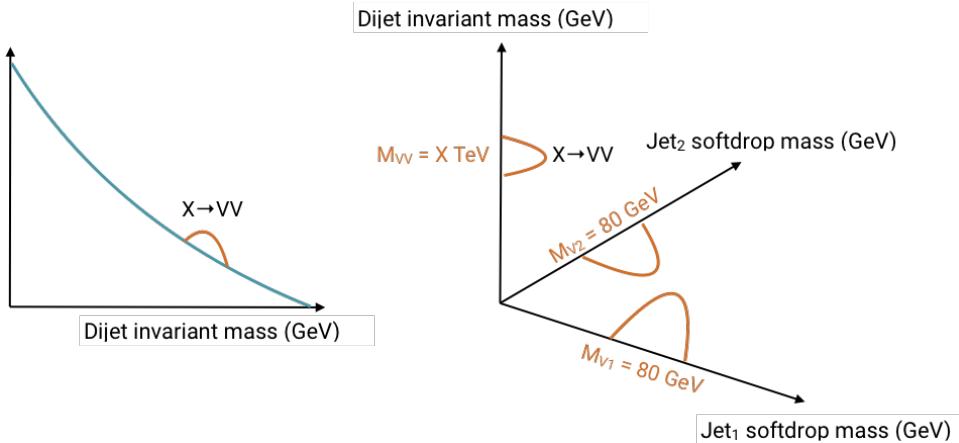


Figure 8.4: The one dimensional diboson analysis versus the three dimensional fit.

resonances decaying to VV ($V=W/Z$), VH ($H=$ Higgs), HH, VX, VH, XX, or XY, where X and Y are new hypothetical bosons, in the same analysis. Additionally, a jet-mass selection is no longer needed as we fit the full jet mass line-shape to extract the signal. This effectively increases the signal statistics since a large fraction of the W and Z signal falls outside the mass window. Fitting the groomed-jet mass and resonance mass together also allows for the addition of nuisance parameters that simultaneously affect both in order to fully account for the correlation between the variables. Finally, we would model the background starting from simulation, rather than from a dijet fit to data. This allows the background shape to assume non-smooth distributions, and could allow the search to probe lower dijet masses, something we will discuss further in Section 8.5. Replacing the parametric fit by a simulation-based model would also reduce the fit sensitivity to background fluctuations in the extreme tails of the dijet invariant mass spectrum.

This chapter presents a novel three-dimensional fit method in the search for diboson resonances

in the all-hadronic final state, and is based on 77.3 fb^{-1} of data collected in 2016 and 2017. The data collected in 2016 was already analyzed in the context of Search II, but is reanalyzed here in order to compare the performance of the three-dimensional fit to that of the one dimensional search.

8.3 Data and simulated samples

The data analyzed in this search consists of 35.9 fb^{-1} of data collected in 2016 and 41.4 fb^{-1} of data collected in 2017, yielding a total of 77.3 fb^{-1} .

The simulated samples are the same as those described in Section 7.3.

8.4 Event selection

Events are selected following the same criteria as in Search I and Search II (see Section 6.4.2) and can be summarized as follows: we require two AK8 jets with PUPPI pileup subtraction applied, and with $|\eta| < 2.5$ and $p_T > 200 \text{ GeV}$. The two jets are further required to be separated by $|\Delta\eta|_{jj} < 1.3$. The two jets with the highest transverse momentum in the event are selected as potential vector boson candidates. In order to avoid any bias in the jet mass shapes when performing the three-dimensional fit, the two jets are randomly sorted after the above selections have been applied. In addition, the dijet invariant mass is required to be greater than 1126 GeV in order to be on the trigger plateau. As already mentioned in the introduction, the background modeling used in this analysis is capable of modeling trigger turn-ons. We explored the possibility of starting the analysis below the trigger plateau, however, while the background modeling was found to be reliable, the extraction of a signal peaking on top of a turn-on was difficult, and we therefore had to abandon the modeling of the trigger turn-on for this first demonstration of the method. More details will be given in Section 8.5. The dijet invariant mass and $|\Delta\eta|_{jj}$ distribution for the two leading jets in the event after the above preselections have been applied are shown in Figure 8.5 and the jet p_T and η distributions for signal and for background is shown in Figure 8.6.

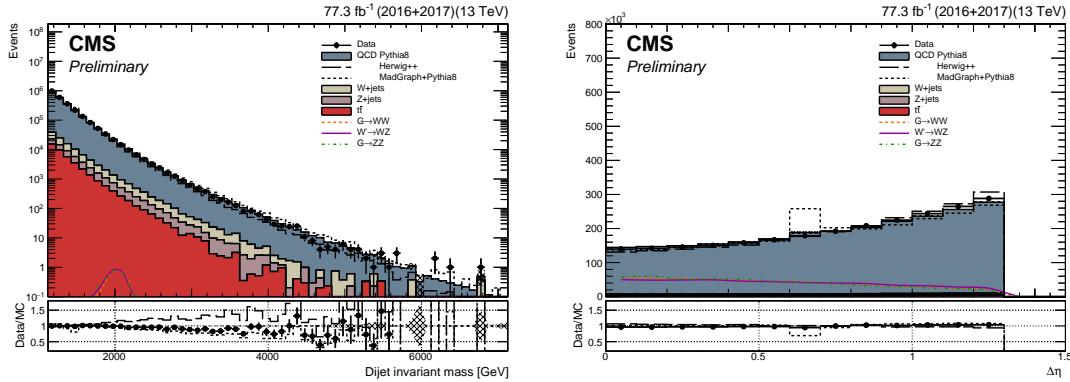


Figure 8.5: The dijet invariant mass (left) and $|\Delta\eta|_{jj}$ (right) for the two leading jets after preselections are applied. The signal is scaled by an arbitrary number to be visible.

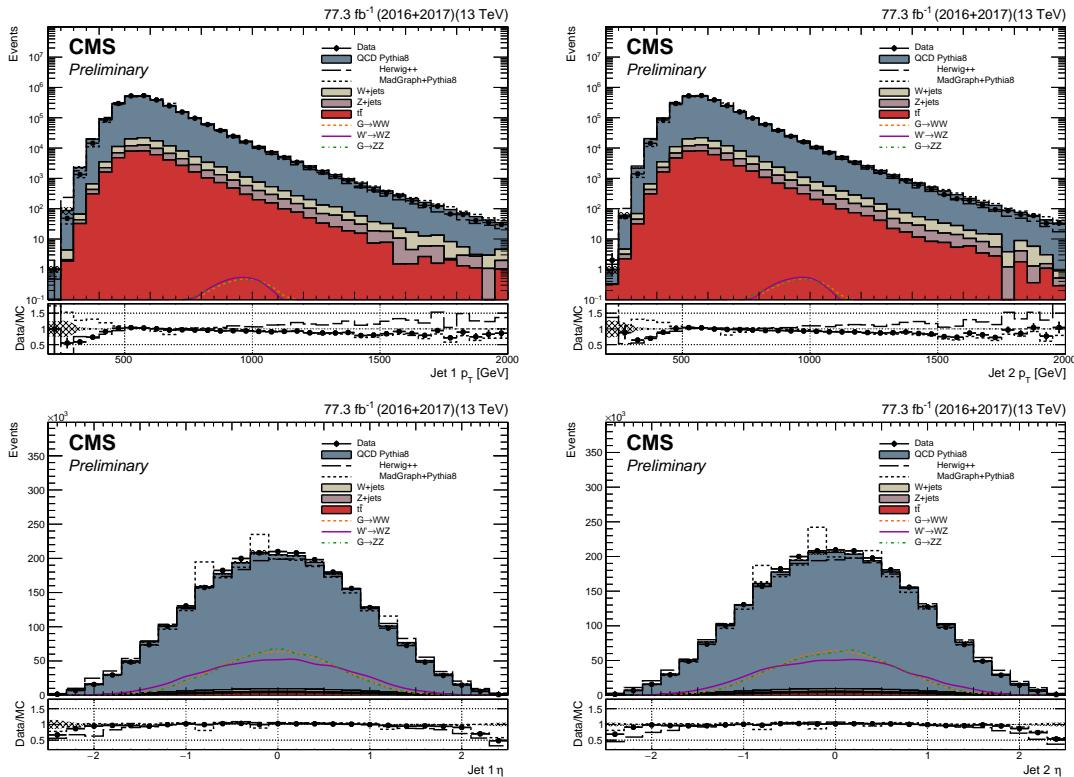


Figure 8.6: Jet p_T (top) and η (bottom) of the first (left) and second (right) selected jet in the event. The signal is scaled by an arbitrary number.

8.5 Triggering

The triggers used for the data collected in 2016 are the same as in Section 6.4.1, while the thresholds in 2017 have increased in order to push the trigger rate to a level acceptable for the increased luminosity. The triggers used for 2017 data are

- HLT_PFHT1050
- HLT_AK8PFJet500
- HLT_AK8PFJet360/380/400/420_TrimMass30
- HLT_AK8PFHT750/800/850/900_TrimMass50.

As in Section 6.4.1, the trigger names indicate the primary selection criteria, with the terms “PF”, “HT”, and “AK8” being previously explained. The term “JetXXX” refers to the lower value of the p_T selection criteria of the jet, “HTXXX” refers to the HT requirement, and “TrimMassYY” refers to the trimmed-jet mass selection applied to one jet. For the results presented here, the analysis threshold is set by the value of M_{VV} such that the combination of all triggers are greater than 99% efficient. The trigger turn-on is evaluated in an orthogonal data set in which the presence of a single muon with a p_T above 27 or 50 GeV is required. The trigger efficiency as a function of dijet invariant mass using a combination of all triggers (left), and as a function of the jet soft drop mass for the grooming-based triggers only (right) are shown in Figure ???. The grooming-based triggers in the full 2017 data set (solid yellow markers) do not fully reach the trigger plateau as a function of the jet groomed mass, since the triggers were unavailable during one data taking period where 4.8 fb^{-1} of data were collected. The trigger efficiencies excluding this run is shown with hollow yellow markers. The combination of all triggers are $> 99\%$ efficient above a dijet invariant mass of 1126 (990) GeV for the 2017 (2016) dataset. The point of full efficiency is higher in the 2017 dataset due to the increased trigger thresholds and this therefore sets the analysis threshold.

8.5.1 Trigger turn-on modeling

The threshold in the dijet mass at which resonances can be searched for depends on the value which the trigger can efficiently select events. The threshold used in the previous analyses was required to be high enough so that the dijet mass spectrum could be described with a smoothly falling function. Since the background modeling for this analysis does not depend on a smoothly falling dijet invariant-mass spectrum, as will be described in detail in Section 8.7, and in order to compensate for a loss in acceptance due to the increased trigger thresholds, we sought to model

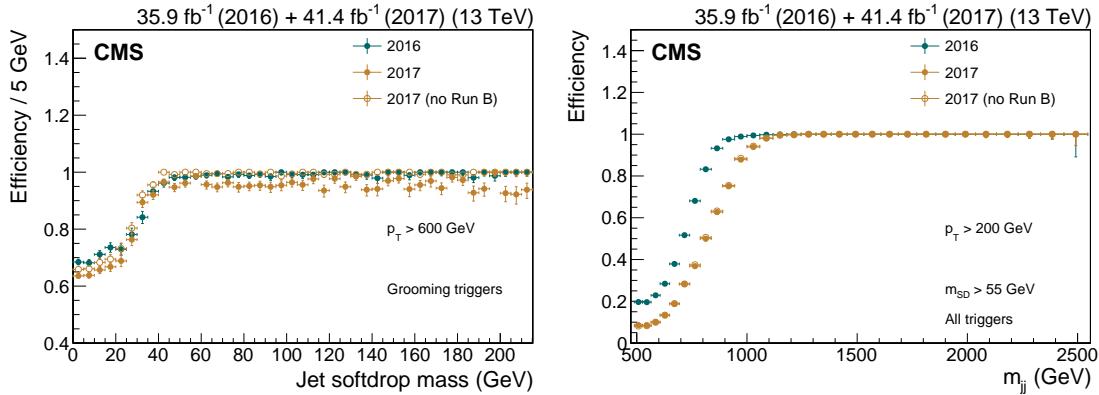


Figure 8.7: Left: trigger efficiency as a function of the dijet invariant mass using a combination of all analysis triggers. Right: trigger efficiency as a function of the jet softdrop mass for triggers requiring an online trimmed mass of at least 30 or 50 GeV.

the trigger turn-on directly from data. This would allow us to search for diboson resonances with masses below the value at which the trigger is fully efficient. To model the trigger turn-on we derive a three dimensional histogram of the trigger efficiency versus dijet invariant mass (M_{VV}) and the groomed-jet mass of jet 1 and jet 2 (M_{jet1} and M_{jet2}), where each bin corresponds to the trigger efficiency for a small range of values of M_{VV} , M_{jet1} and M_{jet2} . This histogram is obtained as follows. From the one-dimensional trigger turn-on histograms shown in Figure 8.7, values of M_{VV} , M_{jet1} and M_{jet2} at which the signal efficiency is above 99% are determined. For every bin above this threshold, the trigger efficiency is fixed to one, corresponding to a 100% efficiency. For all bins below this threshold, a trigger weight is derived. Since the trigger efficiency falls below 50 percent around a dijet invariant mass of 800 GeV, as seen from the right-hand plot in Figure 8.7, searching for resonances with masses below this point is not feasible due to low acceptance. In addition, the full signal shape needs to be contained within the dijet invariant mass spectrum, requiring the search to consider dijet invariant masses 10% lower than the actual value of the lowest resonance mass being probed. These two factors dictate that the minimum resonances mass we can search for is 1 TeV, and that the background modeling must begin around 900 GeV in order to fully contain the signal. Starting from a dijet invariant mass of $M_{VV} = 893$ GeV, the ‘‘dijet bin’’ closest to 900 GeV (where the ‘‘dijet binning’’ is as described in Section 6.5), and a jet mass of $M_{jet} = 40$ GeV, a coarsely binned three-dimensional histogram is filled with the fraction of events that pass one of the signal triggers,

$$w_{ijk}^{Bin} = \frac{\text{PASS}(m_{jj}^i, m_{j1}^j, m_{j2}^k)}{\text{ALL}(m_{jj}^i, m_{j1}^j, m_{j2}^k)}.$$

From this three-dimensional coarse histogram, a finer binned histogram is achieved through interpolation. In bins of $M_{\text{jet}1}$ and $M_{\text{jet}2}$, each slice in M_{VV} is fitted with a sigmoid function,

$$s(x) = \frac{1}{1 + e^{-p_1(x-p_2)}},$$

and the trigger weight is extracted in dijet invariant mass bins of 10 GeV. This fine binning in dijet invariant mass is sufficient to yield a smooth distribution after trigger reweighting, as will be demonstrated below, and no additional interpolation is done. Figure 8.8 shows the total projections on the M_{VV} , $M_{\text{jet}1}$ and $M_{\text{jet}2}$ axes for the full trigger weight histogram. The M_{jet} and

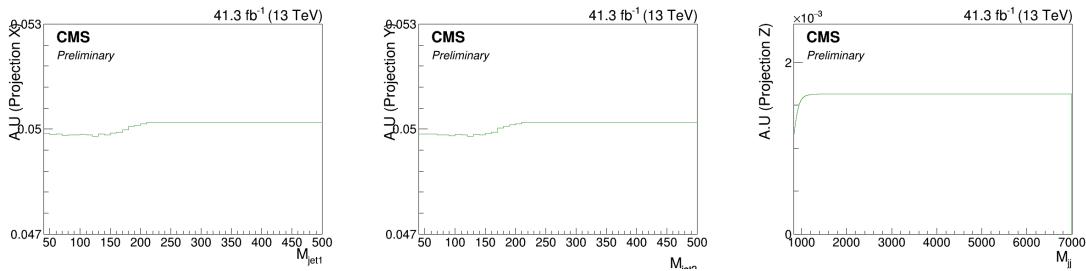


Figure 8.8: One-dimensional projections of the trigger weight histogram for $M_{\text{jet}1}$, $M_{\text{jet}2}$ and M_{VV} , respectively.

M_{VV} spectra for the signal sample with the lowest resonance mass, and for the QCD background before and after trigger weights have been applied, are shown in Figure 8.9 and are compared to data. A greater than 95% the signal efficiency is retained after the trigger weights have been applied, and the reweighted QCD simulation agrees well with data. The modeling of the trigger

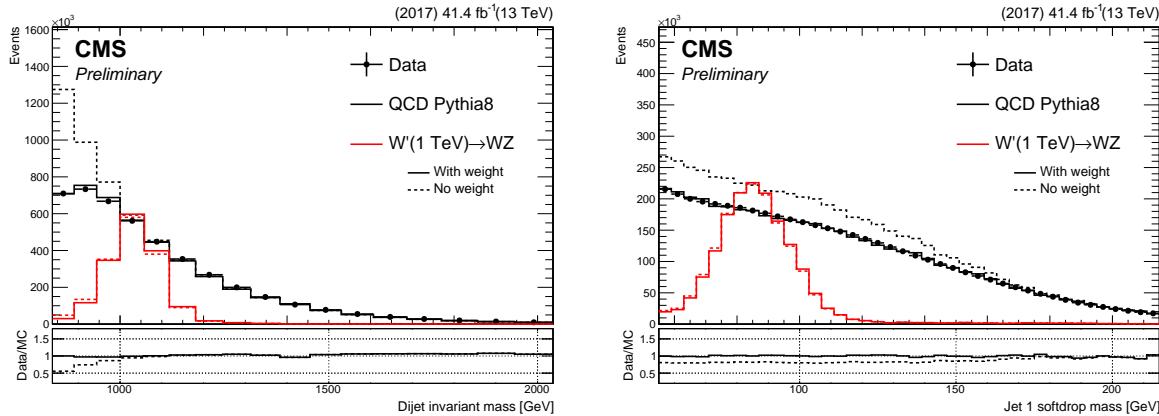


Figure 8.9: The M_{VV} (left) and M_{jet} (right) spectra for signal and background before and after trigger weights have been applied.

turn-on was then implemented in the background fit method in the 3D analysis and we found that the method could model the turn-on well. However, when studying the bias on the extracted signal rate for a possible signal in this turn-on region, we found this to be large due to the difficulties involved with fitting a signal peak on top of a background that is peaking. Since the intention of the analysis presented here was to introduce the 3D fit method and allow it to become available as soon as possible, we leave the modeling of the trigger turn-on as an opportunity for future analyses, going beyond this thesis.

8.6 A mass- and p_T -decorrelated tagger

In order to identify W-jet and Z-jet candidates we utilize the softdrop algorithm and calculate the n-subjettiness ratio τ_{21} for AK8 jets that are clustered with PUPPI constituents. As before, the softdrop jet mass is used to improve the mass resolution of the jet, while n-subjettiness serves as a discriminant by yielding a probability of how compatible the jet is with having N axes. For this search we require the softdrop jet mass to be between $55 < m_{sd} < 215$ GeV, removing as much as possible of the low-mass QCD background and going high enough to encompass all searches for resonances decaying to W, Z, H or tops (the SM particles known to produce jets with substructure in the final state). Since the analysis is done in bins of jet mass and dijet invariant mass, the background must be modeled for a very wide range of mass and transverse momenta. It is therefore desirable that the QCD background spectrum is minimally sculpted as a function of p_T and mass, such that the background shapes in all regions are similar to one another and remains smoothly falling in all three analysis dimensions and in all bins of M_{VV} , M_{jet1} and M_{jet2} . In order to ensure minimal sculpting, we therefore decorrelate the τ_{21} variable from the jet softdrop mass and the jet p_T following what was done in Ref. [48], and described in Section 5.5.3. This decorrelation is performed by flattening the τ_{21} profile dependence on $\rho' = \log(m^2/p_T/\mu)$, where $\mu = 1$ GeV. Figure 8.10 shows the profile distribution of τ_{21} as a function of ρ' for QCD jets after applying the preselections listed above together with a softdrop mass selection of $55 < m_{sd} < 215$ GeV. A linear transformation is then defined as

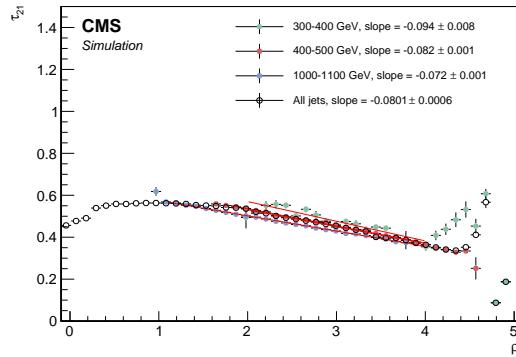


Figure 8.10: Profile distributions of τ_{21} as a function of $\rho' = \log(m^2/p_T/\mu)$ after applying a softdrop mass selection of $55 < m_{sd} < 215$ GeV.

$$\tau_{21}^{DDT} = \tau_{21} - M \times \rho', \quad (8.1)$$

where the slope M is fitted from the linear part of the τ_{21} profile versus ρ' for the inclusive p_T spectrum, shown in Figure 8.10. The resulting slope is $M = -0.080$. The profile of the retuned τ_{21}^{DDT} versus ρ' is shown in Figure 8.11, exhibiting the desired flattened spectra for QCD jets versus ρ' . Which selections on τ_{21}^{DDT} to be used are chosen in the following way. First, we

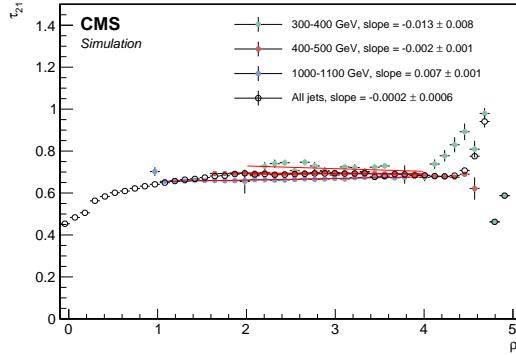


Figure 8.11: Profile distributions of τ_{21}^{DDT} as a function of $\rho' = \log(m^2/p_T/\mu)$, where $\mu = 1$ GeV.

check which τ_{21}^{DDT} selection corresponds to the highest Punzi significance as a function of the resonance mass for different signal samples, as shown in Figure 8.12. All other analysis selections, including a selection on the groomed-jet mass of $55 < m_{sd} < 215$ GeV, have been applied. The

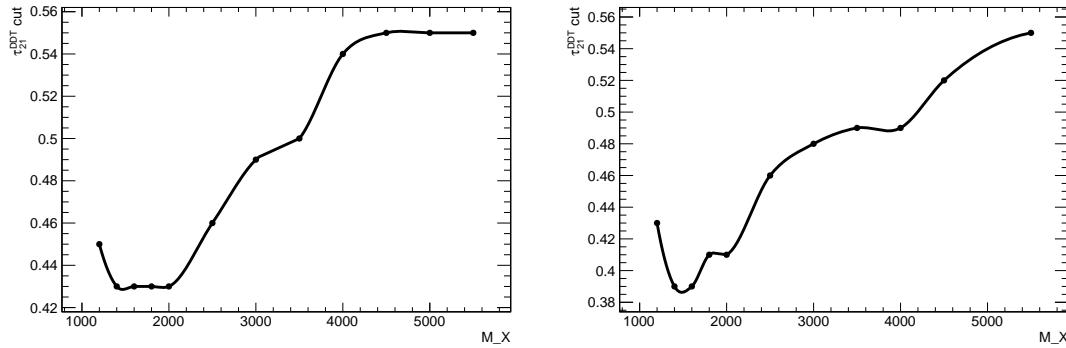


Figure 8.12: The τ_{21}^{DDT} cut corresponding to the highest Punzi significance for a given signal resonance mass, here for a Bulk $G \rightarrow WW$ (left) and a Bulk $G \rightarrow ZZ$ (right) signal.

cut maximizing the Punzi significance at lower resonance mass, where the background is highest, is chosen as the “high-purity” (HP) working point. This corresponds to $\tau_{21}^{DDT} \leq 0.43$. We then proceed by finding which selections on τ_{21}^{DDT} contains at least 95% of the signal, and out of those

select the one that optimizes the Punzi significance. This is found to be $0.43 < \tau_{21}^{DDT} \leq 0.79$, and is classified as the low purity (LP) category. The purpose of this category is to enhance the overall sensitivity of the analysis, especially where the background is low.

Figure 8.13 shows the performance of τ_{21} and τ_{21}^{DDT} (2016 and 2017 tune) in the background-signal efficiency plane (top). We observe a significant gain in signal efficiency at a fixed mistag rate with the decorrelated taggers, when not using a tight jet groomed mass selection (recall: a window of $55 \text{ GeV} < M_{SD} < 215 \text{ GeV}$ is used in this analysis). That is because we are taking advantage of far more information when computing the DDT: subjetiness, and the ratio of jet groomed mass and momentum. All these have a distribution which is different between quark/gluon jets and W-jets, leading to a larger separation between signal and background, as can be seen when comparing the distributions of the different taggers in the left plot of Figure 8.14. The distribution of $\log(m_{SD}^2/p_T)$ is shown on the right in Figure 8.14, and it is clear that the variable adds discriminating power. In addition to cutting on τ_{21}^{DDT} , a selection on $\rho = \log(m^2/p_T^2) < -1.8$

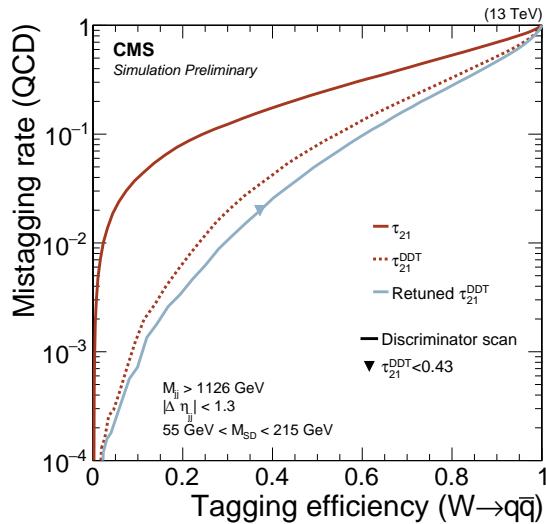


Figure 8.13: Performance of τ_{21} and τ_{21}^{DDT} (2016 and 2017 tune) in the background-signal efficiency plane.

is applied (here, the QCD scaling variable ρ is used rather than ρ' . The reason why ρ' is used for decorrelating τ_{21}^{DDT} , is that this variable was observed to better eliminate the dependence of τ_{21} on p_T). This is in order to avoid configurations where the jet mass is high, but the jet transverse momentum is low. In these cases, the AK8 cone size is too small to contain the full jet, affecting both the jet mass resolution and the τ_{21}^{DDT} tagging efficiency. These edge effects are difficult to model in simulation and we therefore avoid this region in the following. Figure 8.15 shows the profile distribution of τ_{21}^{DDT} versus ρ for QCD jets, where a clear deviation from the observed

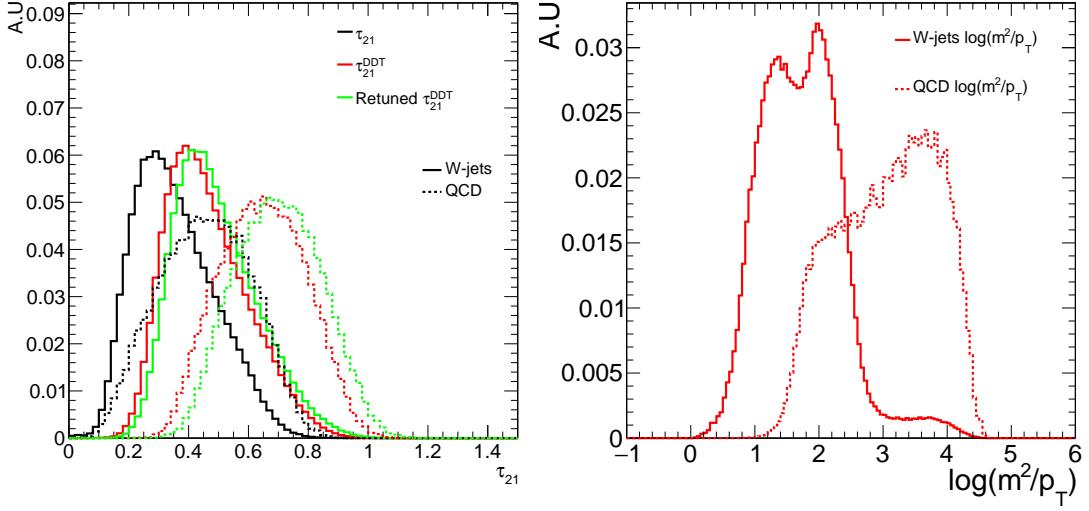


Figure 8.14: Distribution of τ_{21} and τ_{21}^{DDT} with the old and the new tune (left) and comparison of $\log(m^2/p_T)$ for signal and background jets (right).

linear spectra is observed above a ρ of -1.8. Note that this has a negligible effect on signal jets,

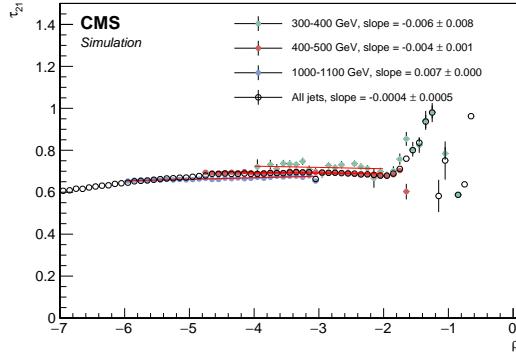


Figure 8.15: Profile distributions of τ_{21}^{DDT} as a function of $\rho = \log(m^2/p_T^2)$, where $\mu = 1$ GeV. The DDT breaks down above $\rho = -1.8$.

which mainly peak around 80 GeV and have a high jet transverse momenta. Figure 8.16 shows the distributions of the PUPPI softdrop jet mass, the τ_{21}^{DDT} , and the τ_{21} variable for simulation and for data. The softdrop jet mass in the signal distribution peaks nicely around the W mass, while the multijets background spectrum is peaked at lower softdrop masses.

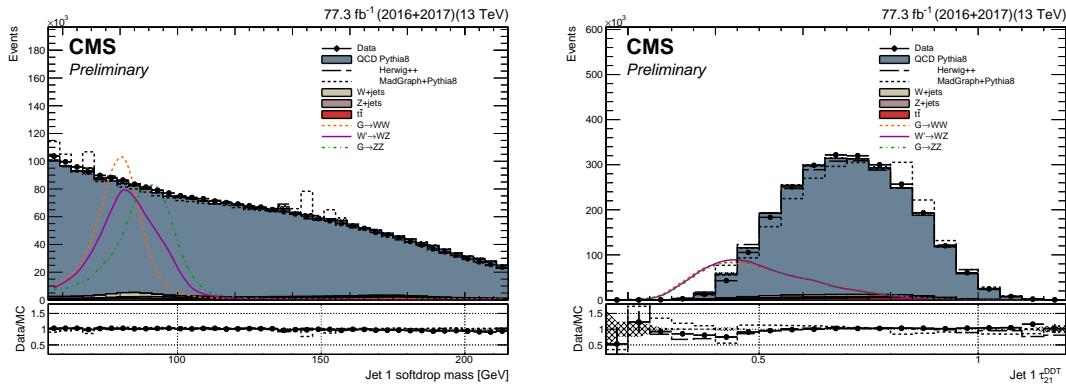


Figure 8.16: The softdrop jet mass (left) and τ_{21}^{DDT} (right) distribution in data and in simulation. Signal is scaled by an arbitrary number.

8.6.1 Data to simulation scale factors

Following what was done in Section 6.7 and 7.5.3, we derive W-tagging scale factors in order to correct the efficiency in MC of the selection on τ_{21}^{DDT} , by estimating the ratio of the selection efficiencies in data and in simulation. The softdrop jet mass range is fitted from 55 to 215 GeV, and the two categories of purity are:

- pass region: $0 < \tau_{21}^{DDT} \leq 0.43 \sim$ high-purity,
- fail region: $0.43 < \tau_{21} \leq 0.79 \sim$ low purity.

The obtained scale factors, jet-mass scale and resolution together with their statistical uncertainty are listed in Tables 8.1 and 8.2 for 2016 and 2017 data, respectively. The corresponding simultaneous fits are shown in Figure 8.17. The W-tagging efficiency in this region is lower than what is expected for signal jets, around 7% compared to 35%. This is expected as the measurement in $t\bar{t}$ is dominated by W-jets with a p_T of around 200 GeV, just where the W decay products merge into a single jet, whereas the signal jets mostly have a transverse momentum above 600 GeV. The signal efficiency for τ_{21}^{DDT} increases at a function of jet p_T , as can be seen from Figure 8.25, whereas the background efficiency is flat by design. In addition to the statistical uncertainty, two systematic uncertainties are evaluated: one due to generator differences and one due to NNLO corrections. The former is evaluated by comparing the resulting scale factors when using $t\bar{t}$ simulation produced with different generators. The latter is evaluated by comparing the extracted efficiency with and without top- p_T reweighting. This weight is derived from data in order to better describe the observed p_T distribution in $t\bar{t}$ and is calculated for each top jet as $w = \exp^{0.0615 - 0.0005 * p_{T,top}}$. The W tagging scale factors, jet mass scale, and jet mass resolution, with their total uncertainty after adding systematics, are listed in Table 8.3. As before, the scale factor is used to scale the signal yield and the jet mass scale and resolution are used to smear the MC. The difference in jet mass scale between data and simulation is around 2%, and the jet mass resolution scale factor is roughly 8%.

$\tau_{21}^{DDT} < 0.43$	m [GeV]	σ [GeV]	W-tag efficiency
Data	81.999 ± 0.454 GeV	7.148 ± 0.544 GeV	0.080 ± 0.008
Simulation	80.890 ± 0.160 GeV	6.579 ± 0.149 GeV	0.085 ± 0.003
Data/simulation	1.014 ± 0.006	1.086 ± 0.086	0.937 ± 0.094
$0.43 < \tau_{21}^{DDT} < 0.79$			
Data			0.920 ± 0.008
Simulation			0.915 ± 0.003
Data/simulation			1.006 ± 0.009

Table 8.1: Jet mass scale, jet mass resolution, and τ_{21}^{DDT} scale factors, as evaluated in the full 2016 single-muon dataset.

$\tau_{21}^{DDT} < 0.43$	m [GeV]	σ [GeV]	W-tag efficiency
Data	80.784 ± 0.391 GeV	7.694 ± 0.445 GeV	0.065 ± 0.006
Simulation	82.208 ± 0.293 GeV	7.127 ± 0.284 GeV	0.068 ± 0.005
Data/simulation	0.983 ± 0.006	1.080 ± 0.076	0.955 ± 0.113
$0.43 < \tau_{21}^{DDT} < 0.79$			
Data			0.935 ± 0.006
Simulation			0.932 ± 0.005
Data/simulation			1.003 ± 0.008

Table 8.2: Jet mass scale, jet mass resolution, and τ_{21}^{DDT} scalefactors as evaluated in the full 2017 single-muon dataset.

	$SF \pm \sqrt{\text{Stat.} \pm \text{Sys}_{\text{Generator}} \pm \text{Sys}_{\text{NNLO}}}$	$SF \pm \text{Total Unc.}$
$HPSF_{DDT}^{2017}$	$0.955 \pm \sqrt{0.113^2 \text{ (stat.)} + 0.003^2 \text{ (sys.)} + 0.043^2 \text{ (sys.)}}$	0.955 ± 0.121
$HPSF_{DDT}^{2016}$	$0.937 \pm \sqrt{0.094^2 \text{ (stat.)} + 0.003^2 \text{ (sys.)} + 0.043^2 \text{ (sys.)}}$	0.937 ± 0.103
$LPSF_{DDT}^{2017}$	$1.003 \pm \sqrt{0.008^2 \text{ (stat.)} + 0.003^2 \text{ (sys.)} + 0.02^2 \text{ (sys.)}}$	1.003 ± 0.008
$LPSF_{DDT}^{2016}$	$1.006 \pm \sqrt{0.009^2 \text{ (stat.)} + 0.003^2 \text{ (sys.)} + 0.02^2 \text{ (sys.)}}$	1.006 ± 0.009
JMS^{2017}	$0.983 \pm \sqrt{0.006^2 \text{ (stat.)} + 0.002^2 \text{ (sys.)} + 0.001^2 \text{ (sys.)}}$	0.983 ± 0.007
JMS^{2016}	$1.014 \pm \sqrt{0.006^2 \text{ (stat.)} + 0.002^2 \text{ (sys.)} + 0.001^2 \text{ (sys.)}}$	1.014 ± 0.007
JMR^{2017}	$1.080 \pm \sqrt{0.076^2 \text{ (stat.)} + 0.027^2 \text{ (sys.)} + 0.001^2 \text{ (sys.)}}$	1.080 ± 0.081
JMR^{2016}	$1.086 \pm \sqrt{0.086^2 \text{ (stat.)} + 0.027^2 \text{ (sys.)} + 0.001^2 \text{ (sys.)}}$	1.086 ± 0.090

Table 8.3: Final jet mass scale, jet mass resolution, and τ_{21}^{DDT} scalefactors with statistical and systematic uncertainties.

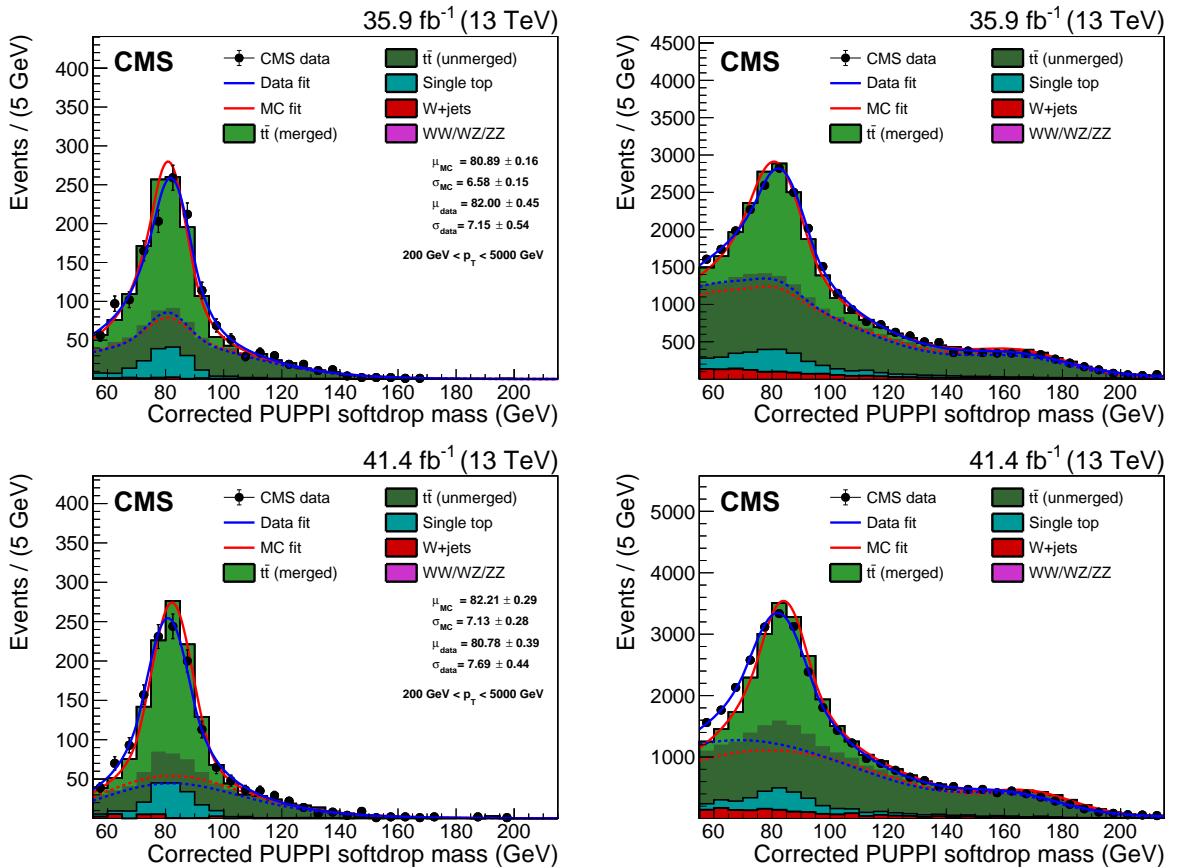


Figure 8.17: PUPPI softdrop jet mass distribution for events that pass (left) and fail (right) the $\tau_{21}^{DDT} < 0.43$ selection in the $t\bar{t}$ control sample. The result of the fit to data and simulation is shown by the solid blue and solid red line, respectively. The background components of the fit are shown as dashed-dotted lines. The fit to 2016 data is shown in the upper panels and the fit to 2017 data in the lower panels.

8.6.2 p_T -dependence

The τ_{21}^{DDT} selection criteria used for this search is restrictive to favor better signal significance, but results in small event samples of the data and simulation needed to produce the scale factors. Due to this, measuring the W-tagging scale factor in bins of p_T is unfeasible for this tagger. In order to estimate the dependence of the W-tagging efficiency scale factor, and the PUPPI softdrop jet-mass scale (JMS) and resolution (JMR) on the transverse momentum of the jets, we use the looser softdrop + τ_{21} based tagger and divide the validation sample into 4 different jet- p_T bins. The jet-mass scale and resolution are expected to be more or less the same for a τ_{21} or τ_{21}^{DDT} selection, and the JMS and JMR scale factors extracted using a selection on τ_{21} can therefore be directly applied to this analysis. The efficiency measurement is not directly transferable, however, they give an idea of how the uncertainty on the measurements change with momentum. The systematic uncertainties are evaluated the same way as above (one due to top- p_T reweighting and one comparing different $t\bar{t}$ samples). Figure 8.18 shows the extracted W-tagging efficiency for data (black markers) and for simulation (red markers) using a tagger based on PUPPI softdrop with a requirement of $\tau_{21} < 0.4$, as a function of jet p_T . The inclusive efficiency measurement is marked with triangles. The lower panel shows the efficiency ratio of data over simulation, corresponding to the W-tagging scale factor. All scale factors are compatible with unity, but the uncertainty on the measurement grows as statistics decrease. The corresponding extracted scale factors are listed in Table 8.4. With the observation of this clear trend of an uncertainty increase

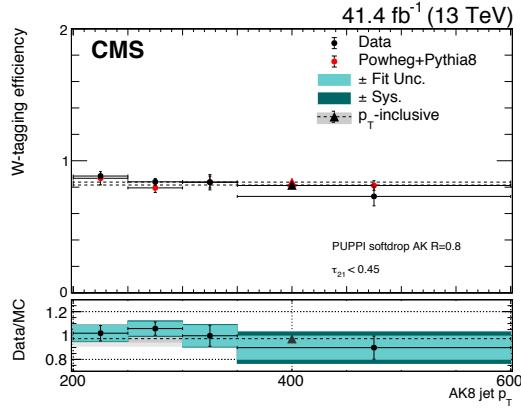


Figure 8.18: The W-tagging efficiency in data (black circles) and in simulation (red circles) as a function of jet p_T . The p_T -inclusive measurement is marked with triangles. The lower panel shows the efficiency in data divided by the efficiency in simulation, corresponding to the W-tagging uncertainty. The blue bands mark the fit and systematic uncertainties.

as a function of p_T , a p_T -dependent W-tagging scale factor uncertainty is needed. As τ_{21}^{DDT} is a

Bin	$SF \pm \sqrt{\text{Stat.} + \text{Sys}_{\text{Generator}} + \text{Sys}_{\text{NNLO}}}$	$SF \pm \text{Total Unc.}$
200 - 250 GeV	$1.019 \pm \sqrt{0.064^2 + 0.005^2 + 0.022^2}$	1.02 ± 0.07
250 - 300 GeV	$1.058 \pm \sqrt{0.055^2 + 0.033^2 + 0.002^2}$	1.06 ± 0.06
300 - 350 GeV	$0.998 \pm \sqrt{0.087^2 + 0.035^2 + 0.007^2}$	1.00 ± 0.09
350 - 600 GeV	$0.898 \pm \sqrt{0.097^2 + 0.089^2 + 0.007^2}$	0.90 ± 0.13
≥ 200 GeV	$0.974 \pm \sqrt{0.029^2 + 0.055^2 + 0.015^2}$	0.97 ± 0.06

Table 8.4: The data to simulation scale factors in bins of jet p_T for a tagger based on PUPPI softdrop with a requirement of $\tau_{21} < 0.4$. All scalefactors are compatible with unity.

linear transformation of τ_{21} , using the measurement above and extrapolating the uncertainty to high p_T would be an opportunity but requires further investigation into the differences between the two as a function of jet transverse momentum. This is left for future studies. For this analysis we therefore rather estimate the uncertainty using an established method, which is the following: using signal Monte Carlo generated with two different shower generators, PYTHIA8 and HERWIG++, we compute the difference in tagging efficiency between the two samples at low- p_T , where we have a real measurement in data, and compare that to the difference in tagging efficiency between the two at high- p_T . In other words, we take a double ratio

$$\sigma_{p_T, \text{Bin}=i} = \frac{\left(\frac{\epsilon_{\text{HERWIG}}}{\epsilon_{\text{PYTHIA}}}\right)_{p_T, \text{Bin}=i}}{\left(\frac{\epsilon_{\text{HERWIG}}}{\epsilon_{\text{PYTHIA}}}\right)_{p_T=300 \text{ GeV}}} - 1, \quad (8.2)$$

where $\sigma_{p_T, \text{Bin}=i}$ is the uncertainty on the scale factor on p_T bin i . This uncertainty grows logarithmically with p_T and the uncertainty σ_{p_T} and resulting parametrization is shown in Figure 8.19. In addition to measuring the dependence of the tagging efficiency on jet p_T , we also extract the change in softdrop jet-mass scale and resolution as a function of p_T in data. We find that the jet-mass scale in data divided by MC ranges between 0.5 and 2.5% for the p_T bins considered, and the jet-mass resolution varies between 4 and 10%, the latter measurement not being statistically significant as the uncertainties are large, around 10%. For this analysis, we therefore use a fixed uncertainty of 2 and 10% for the softdrop jet-mass scale and resolution, respectively, to cover any broadening or shift at high p_T not described by the simulation.

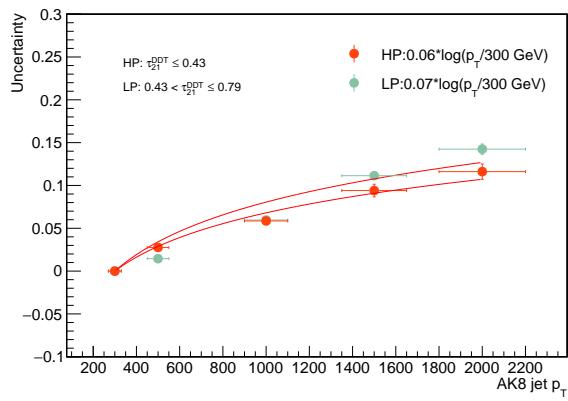


Figure 8.19: The parametrized uncertainty on the W-tagging efficiency scale factor as a function of the jet p_T .

8.7 The multidimensional fit

The three-dimensional fit method is based on the assumption that the signal peaks in three dimensions: the dijet invariant mass (M_{VV}), the groomed jet mass of the first jet ($M_{\text{jet}1}$), and the groomed jet mass of the second jet ($M_{\text{jet}2}$), in order to extract a possible signal from the three-dimensional $M_{\text{jet}1}$ - $M_{\text{jet}2}$ - M_{VV} (x,y,z) plane.

In order to obtain a complete three-dimensional model, four different types of probability density

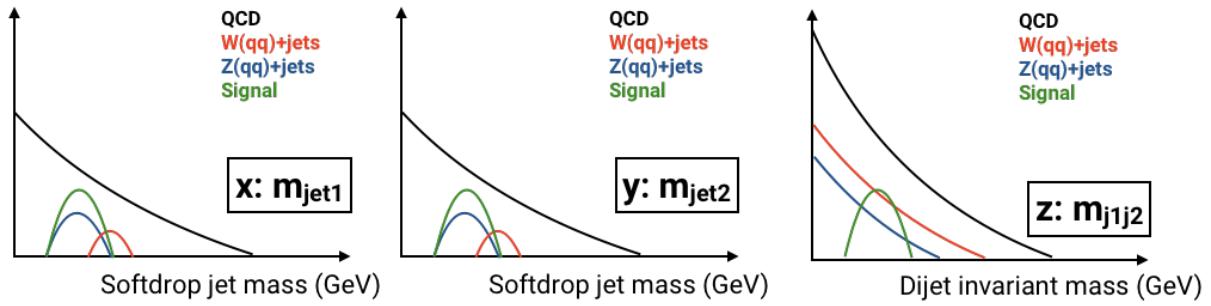


Figure 8.20: An illustration of the shape of the signal and the relative background contributions in the three dimensions $M_{\text{jet}1}(x)$, $M_{\text{jet}2}(y)$ and $M_{VV}(z)$.

functions (PDF's) need to be derived:

- **signal:** a PDF resonant in $M_{\text{jet}1}$, $M_{\text{jet}2}$ and M_{VV} ,
- **non-resonant background:** PDF for the dominant QCD multijet background. Non-resonant in $M_{\text{jet}1}$, $M_{\text{jet}2}$ and M_{VV} ,
- **resonant background:** PDF describing of $W+jets$ and $Z+jets$. Resonant in $M_{\text{jet}1}$ and $M_{\text{jet}2}$, but smoothly falling in M_{VV} , and
- **alternative shapes:** five additional shape uncertainties for the non-resonant background.

These are illustrated in Figure 8.20 and will be described in detail in the following.

8.7.1 Modeling of the signal

The signal shape in three dimensions is defined as the product of the resonance mass shape and the jet mass shapes:

$$P_{sig}(M_{VV}, M_{jet1}, M_{jet2} | \theta(M_X)) = P_{VV}(M_{VV} | \theta_1(M_X)) \times P_{j1}(M_{jet1} | \theta_2(M_X)) \times P_{j2}(M_{jet2} | \theta_2(M_X)). \quad (8.3)$$

The shapes for M_{VV} , M_{jet1} and M_{jet2} all depend on the hypothesized mass of the new particle (M_X) and a set of parameters $\theta = (\theta_1, \theta_2)$ that in principle depend on M_X . The modeling of M_{jet} is done separately for the high-purity and the low-purity category, while the modeling of M_{VV} is done with no selection on τ_{21}^{DDT} applied (M_{VV} was found to be decorrelated from τ_{21}^{DDT} for signal jets).

The signal is parametrized by fitting the resonance mass and jet-mass line shapes for each resonance mass value, extracting the fitted parameters and then interpolating these as a function of the resonance-mass hypothesis. For the resonance mass M_{VV} , the sum of a crystal-ball function and a Gaussian shape is used for each mass point, following the shapes used in Search II. Figure 8.21 shows the derived parameters and interpolation as a function of resonance mass, and the final M_{VV} shapes as extracted from the parametrization are shown in Figure 8.22.

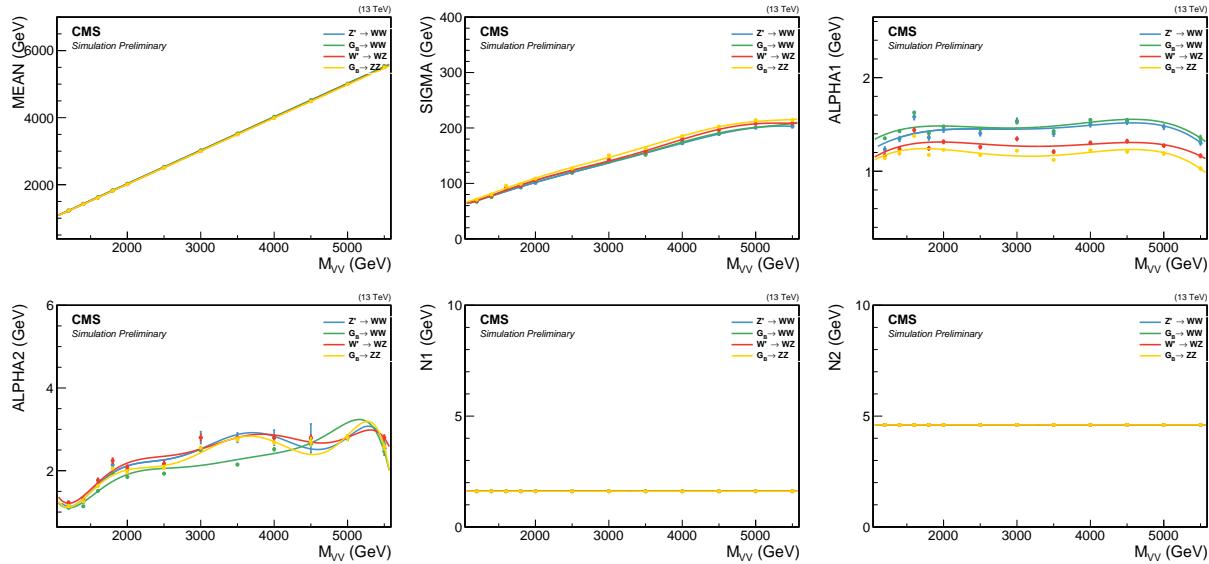


Figure 8.21: The interpolated fit parameters of a crystal-ball fit to the dijet invariant mass as a function of M_X . The small variations for ALPHA2 have been shown to have no effect on the overall modeling.

The same procedure is used to model the jet mass; the M_{jet} spectrum for each resonance mass

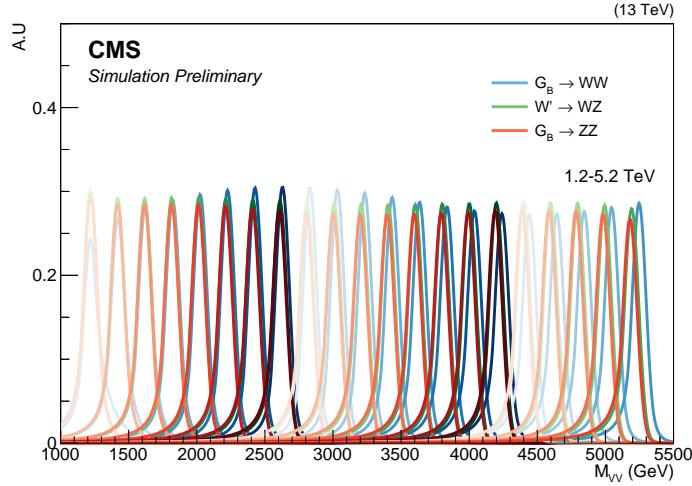


Figure 8.22: Final M_{VV} signal shapes extracted from the parametrization. Here for a G_{bulk} decaying to WW (blue), ZZ (red), and for a W' decaying to WZ (green).

hypothesis is fitted using a double Crystal-ball function, and the fitted parameters are extracted and interpolated as a function of the resonance mass. This is done separately for $M_{\text{jet}1}$ and $M_{\text{jet}2}$. The fitted parameters and interpolations are shown in Figure 8.23 for $M_{\text{jet}1}$, and the corresponding distributions for $M_{\text{jet}2}$ can be found in Appendix D.1. The final $M_{\text{jet}1}$ shapes, as extracted from

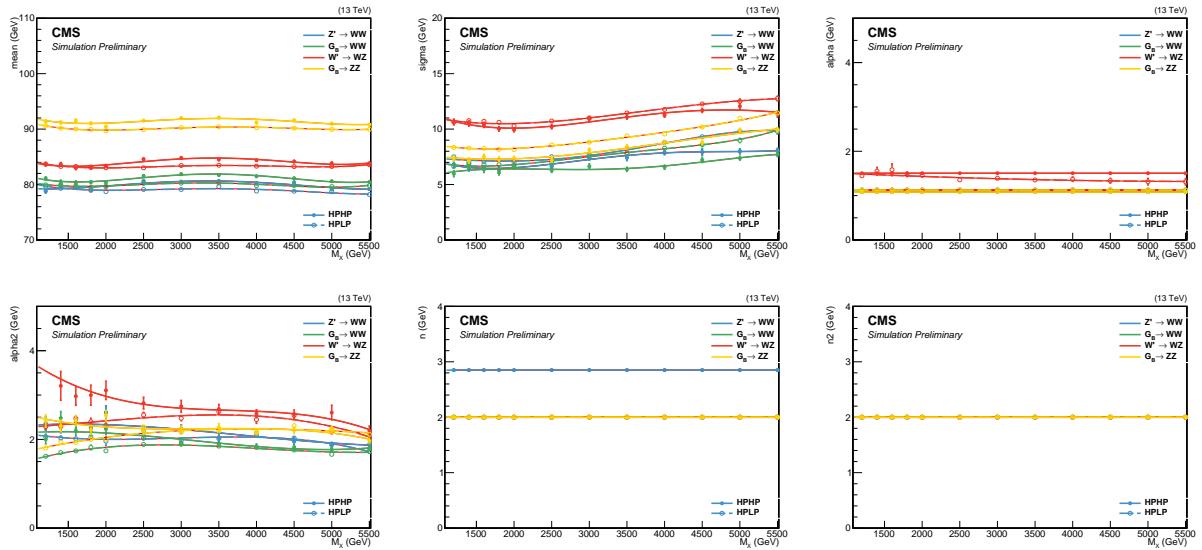


Figure 8.23: The interpolated double Crystal-ball parameters for the softdrop jet mass as a function of M_X . To improve the stability of the fit some parameters are set constant.

the parametrization are shown in Figures 8.24. Finally, the signal yield is parametrized as a

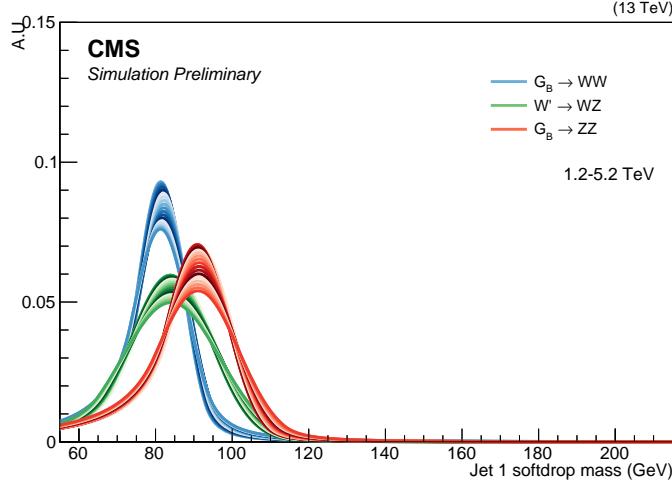


Figure 8.24: Final M_{jet} signal shapes extracted from the parametrization for a G_{bulk} decaying to ZZ (red), a G_{bulk} decaying to WW (blue), and for a W' decaying to WZ (green).

function of the resonance mass. For each mass point M_X and each purity category, the signal yield per picobarn of cross section is calculated as the integral of the resulting Monte Carlo histogram after all analysis selections have been applied. The yields are then interpolated as a function of M_X . The signal efficiency as a function of resonance mass is shown in Figure 8.25.

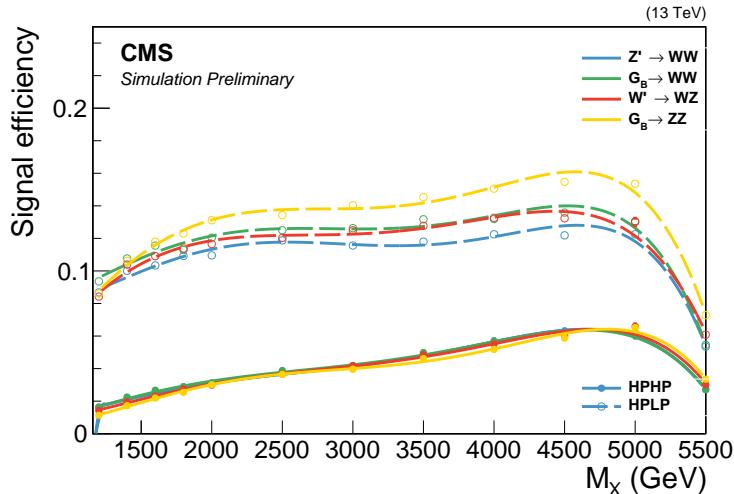


Figure 8.25: Signal efficiency as a function of resonance mass.

8.7.2 Modeling of the non-resonant background

In order to model the QCD multijets background in the three-dimensional M_{VV} - M_{jet1} - M_{jet2} plane, we use the following conditional product:

$$P(M_{VV}, M_{jet1}, M_{jet2}) = P_{VV}(M_{VV}|\theta_1) \times P_{cond,1}(M_{jet1}|M_{VV}, \theta_2) \times P_{cond,2}(M_{jet2}|M_{VV}, \theta_2). \quad (8.4)$$

This probability density requires a computation of the conditional two-dimensional shapes of M_{jet1} or M_{jet2} , given M_{VV} , as well as a one dimensional shape of the M_{VV} distribution. The reason for requiring a conditional jet-mass shape, is due to the high correlation between the jet mass and the dijet invariant mass for quark and gluon jets. The modeling is done separately for the high-purity and the low-purity category.

The following fit ranges are used for the three axes: the jet masses M_{jet1} and M_{jet2} are fitted from 55 to 215 GeV using 2 GeV bins, and the dijet invariant mass M_{VV} is fitted from 1126 to 5500 GeV. The lower M_{VV} bound is chosen to avoid complications in the fitting procedure due to trigger turn-on effects, while the upper M_{VV} bound is chosen by considering the highest dijet invariant mass event found in data. For M_{VV} , the “dijet binning”, corresponding to the dijet mass resolution and described in Section 6.5, is used. In units of GeV, the binning is:

Dijet binning = 1126, 1181, 1246, 1313, 1383, 1455, 1530, 1607, 1687, 1770, 1856, 1945, 2037, 2132, 2231, 2332, 2438, 2546, 2659, 2775, 2895, 3019, 3147, 3279, 3416, 3558, 3704, 3854, 4010, 4171, 4337, 4509, 4686, 4869, 5058, 5253, 5500.

The background model is built starting from simulation using a “forward-folding” approach. To build the one-dimensional template for the dijet invariant mass, $P_{VV}(M_{VV}|\theta_1)$, a 1D Gaussian kernel is built starting from generator level quantities where, for each MC event, a Gaussian contributes to the total one-dimensional probability density function. To build the two-dimensional conditional templates, $P_{cond,1}(M_{jet1}|M_{VV}, \theta_2)$ and $P_{cond,2}(M_{jet2}|M_{VV}, \theta_2)$, a two-dimensional Gaussian kernel is created for each MC event, where each 2D kernel contributes to the total conditional PDF.

In order to define the Gaussian kernel, the dijet invariant mass scale and resolution, and the softdrop jet-mass scale and resolution, must be derived. This is done by comparing the generated jet-mass $M(\text{gen})$ to the reconstructed jet-mass $M(\text{reco})$. The M_{jet} and M_{VV} scale and resolution are extracted from a Gaussian fit to either $M_{jet}(\text{reco})/M_{jet}(\text{gen})$ or $M_{VV}(\text{reco})/M_{VV}(\text{gen})$, in bins of generator-level jet p_T . Figure 8.26 shows the fit to $M_{jet}(\text{reco})/M_{jet}(\text{gen})$ (left) and $M_{jj}(\text{reco})/M_{jj}(\text{gen})$ (right) for an arbitrary p_T bin. The mean of the Gaussian yields the jet-mass scale and the width yields the jet-mass resolution, for a given bin in generator-level jet p_T .

The M_{jet} and M_{VV} jet mass scale and resolution as a function of generator jet p_T is shown in

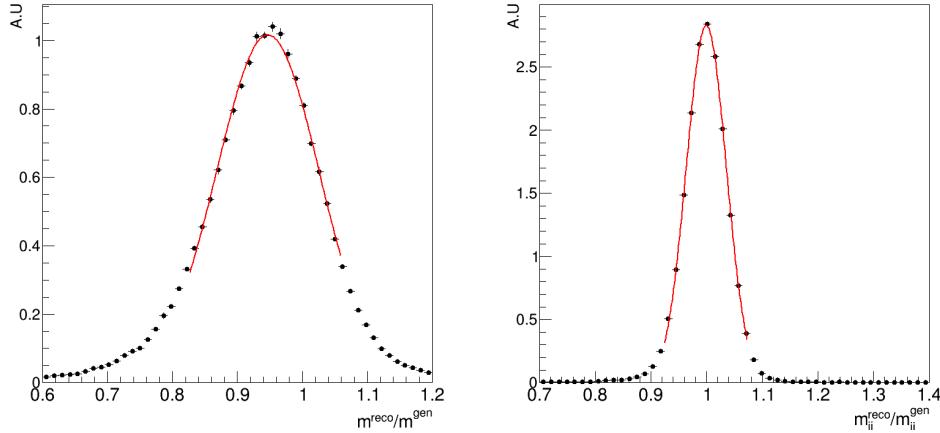


Figure 8.26: Left: fit to $M_{\text{jet}}(\text{reco})/M_{\text{jet}}(\text{gen})$. Right: fit to $M_{\text{VV}}(\text{reco})/M_{\text{VV}}(\text{gen})$. The mass resolution is taken as the width of the fitted Gaussian, while the Gaussian mean yields the mass scale.

Figure 8.27, and the total projection of these resolution functions are shown in Figure 8.28. The jet-mass and the dijet invariant mass scale and resolution are then used to populate the conditional 2D histogram as follows. Each generated event is smeared with a 2D Gaussian kernel of the form

$$k(M_{\text{jet}}, M_{\text{VV}}) = \frac{w_i}{\sqrt{2\pi} r_{M_{\text{VV}},i} \cdot r_{M_{\text{jet}},i}} \exp \left(-\frac{1}{2} \left(\frac{M_{\text{VV}} - s_{M_{\text{VV}},i}}{r_{M_{\text{VV}},i}} \right)^2 - \frac{1}{2} \left(\frac{M_{\text{jet}} - s_{M_{\text{jet}},i}}{r_{M_{\text{jet}},i}} \right)^2 \right), \quad (8.5)$$

where s_i, r_i are the scale and the resolution derived in the previous step and w_i is the event weight product due to effects such as PU-reweighting and cross-sections. The resulting kernel values are then filled into a 2D histogram. This procedure is performed separately for $M_{\text{jet}1}$ and $M_{\text{jet}2}$. To build the one-dimensional template for the dijet invariant mass, the same procedure as above is used with the exception that the smearing is done with a one-dimensional Gaussian kernel only depending on M_{VV} . The three templates are then added together to form a three-dimensional PDF, and this PDF is fitted to QCD MC in order to remove any residual bias. The result is a smooth shape, covering the entire search range, that can be used in place of the prediction from simulation.

As the high-purity category is limited by statistics, we instead build this template starting from the low-purity templates. This is done by fitting the low-purity 3D template, obtained using the method described above, to QCD MC in the high-purity category. Figure 8.29 shows the final

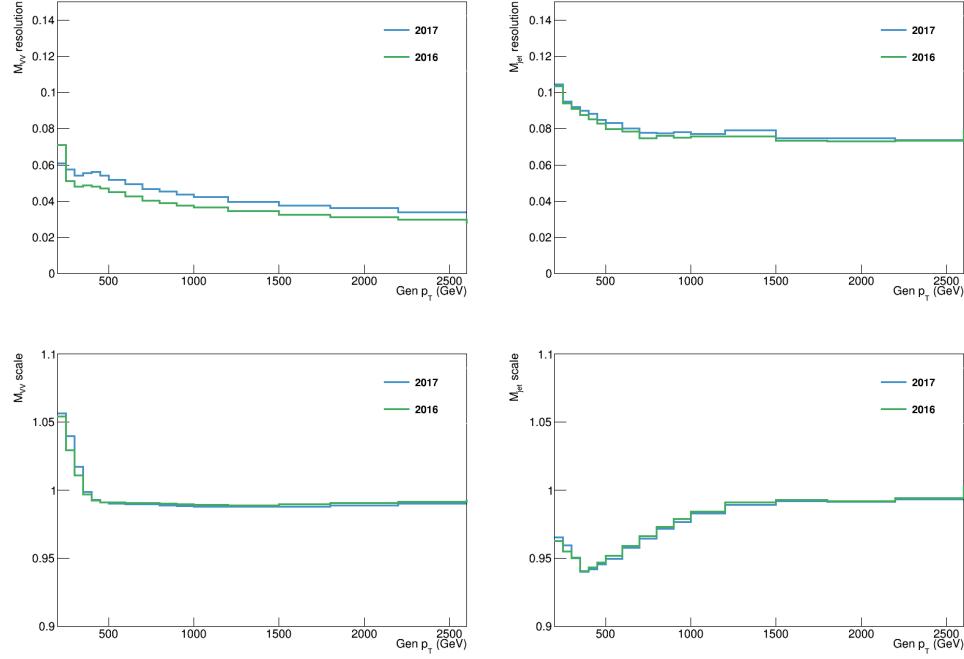


Figure 8.27: The M_{VV} resolution (top left) and scale (bottom left), and the M_{jet} resolution (top right) and scale (bottom right) as a function of generator-level jet p_T .

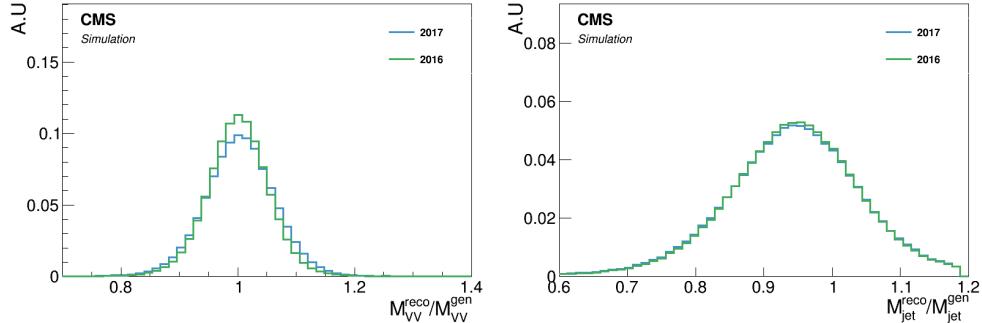


Figure 8.28: Projections of the resolution functions for all generator-level jet p_T bins for M_{VV} (left) and M_{jet} (right).

templates (solid lines) together with the QCD MC (data points) derived from the 2017 MC in the low-purity (top) and high-purity (bottom) category. Good agreement between simulation and templates in all three dimensions is observed, within statistical uncertainties. The corresponding distributions for 2016 MC can be found in Appendix D.2.

In order to validate the kernel-transfer method, we check that we can fit a higher-statistics

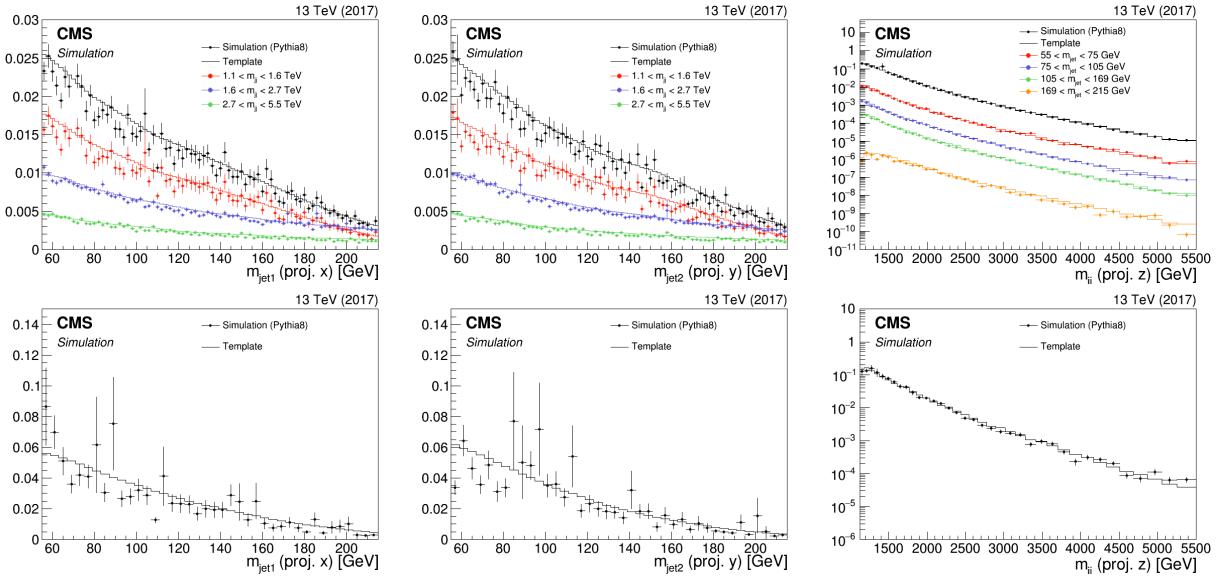


Figure 8.29: Comparison between QCD MC simulation (markers) and kernels derived from generator-level quantities (lines) for the HP (top) and LP (bottom) categories using 2017 MC. The kernels are shown for $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{VV} (right).

high-purity region by loosening the τ_{21}^{DDT} cut to 0.49. This results in 12 times more background events, and should uncover whether any degeneracy is present in the fits themselves and whether the low-purity kernel indeed is capable of modeling the high-purity region. The resulting kernel-versus-MC spectra are shown in Figure 8.30. Good closure is observed in all three dimensions, demonstrating that the LP kernels adapt well to the HP MC data points even when statistics are sufficient, and we therefore consider the method to be sound.

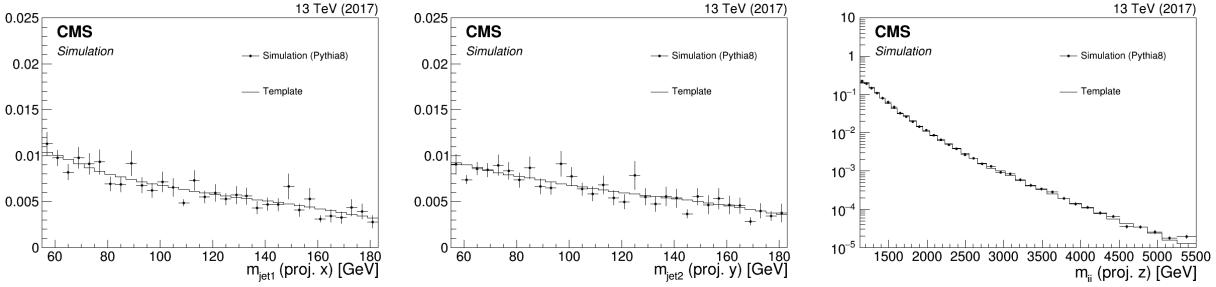


Figure 8.30: Comparison between QCD MC simulation (markers) and kernels derived from generator-level quantities (lines) in the HP category, using a looser cut on τ_{21}^{DDT} .

8.7.3 Modeling of the resonant background

The resonant background is defined as the background with a merged W or Z boson jet and is dominated by W+jets and Z+jets, with a minor contribution from top pair production. As the jets are randomly sorted, each jet mass distribution contains two contributions: a resonant part consisting of real vector-boson jets, peaking around the W or Z boson mass; and a non-resonant part composed of jets originating from a quark or a gluon, similar to the QCD-multijets background. These two contributions are modeled separately since the non-resonant part of the jet mass spectrum is correlated with the dijet invariant mass (as was the case for QCD), while the resonant part is not (as was the case for the signal PDF). Figure 8.31 shows the distribution of $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{VV} (right) for the resonant backgrounds under consideration.

. We want to encode the information that these backgrounds in reality only have one real vector

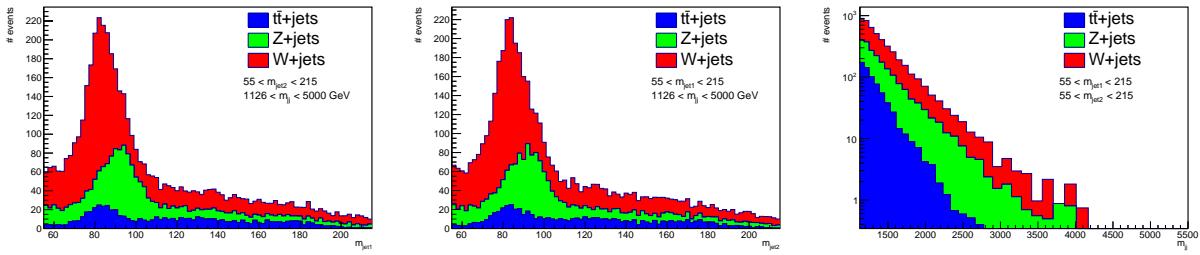


Figure 8.31: Projections of the sub-dominant backgrounds on the jet mass axes $M_{\text{jet}1}$ (left) and $M_{\text{jet}2}$ (middle), as well as on the dijet invariant mass M_{VV} (right). Here in the low-purity category.

boson and therefore only peak in one jet-mass dimension, by requiring the PDFs to consist of a resonant part on one axis, and a non-resonant part on the other axis, related by a fraction f . Its value can vary by 10% around $f = 0.5$, the expected median when using a random jet sorting. A three-dimensional PDF for the resonant backgrounds is built as a product of three one-dimensional PDFs as follows:

$$\begin{aligned} P_{V\text{jets}}(M_{\text{jet}1}, M_{\text{jet}2}, M_{\text{VV}}) = f \times & (P_{\text{dijet}}(M_{\text{VV}}) \times P_{\text{res}}(M_{\text{jet}1}) \times P_{\text{non-res}}(M_{\text{jet}2})) \\ & + (1 - f) \times (P_{\text{dijet}}(M_{\text{VV}}) \times P_{\text{res}}(M_{\text{jet}2}) \times P_{\text{non-res}}(M_{\text{jet}1})) . \end{aligned} \quad (8.6)$$

The contribution of W+jets and top pair production are considered as one combined background shape, due to both being resonant at the W boson mass peak, while the Z+jets contribution is fitted separately. The top contribution is less than 1% and negligible for this analysis. The first step is to derive the resonant mass shape P_{res} by fitting a double Crystal Ball to the jet mass

spectrum, performed separately for $M_{\text{jet}1}$ and $M_{\text{jet}2}$. The resonant events are separated from the non-resonant ones when building the templates by requiring that there is a generated boson in a cone of $\Delta R = 0.8$ around the reconstructed merged jet. A double-sided crystal-ball function, the same shape used for the modeling of the signal M_{jet} shapes, is fitted to the resonant spectrum for W+jets and Z+jets separately. This treatment allows to fully correlate the uncertainties on the mean and width of the M_{jet} distribution with the signal, since these uncertainties affect all jets generated from real vector bosons in the same way. Figure 8.32 shows the final fit of a double-sided crystal-ball function to the resonant part of the V+jets spectrum. The small enhancement around 170 GeV is caused by fully merged top jets, but is so small ($< 2\%$ in the low-purity and $< 0.5\%$ in the high-purity category) that we do not take it into account in the final PDF. The non-resonant component of the jet mass shape is fitted with a Gaussian function and

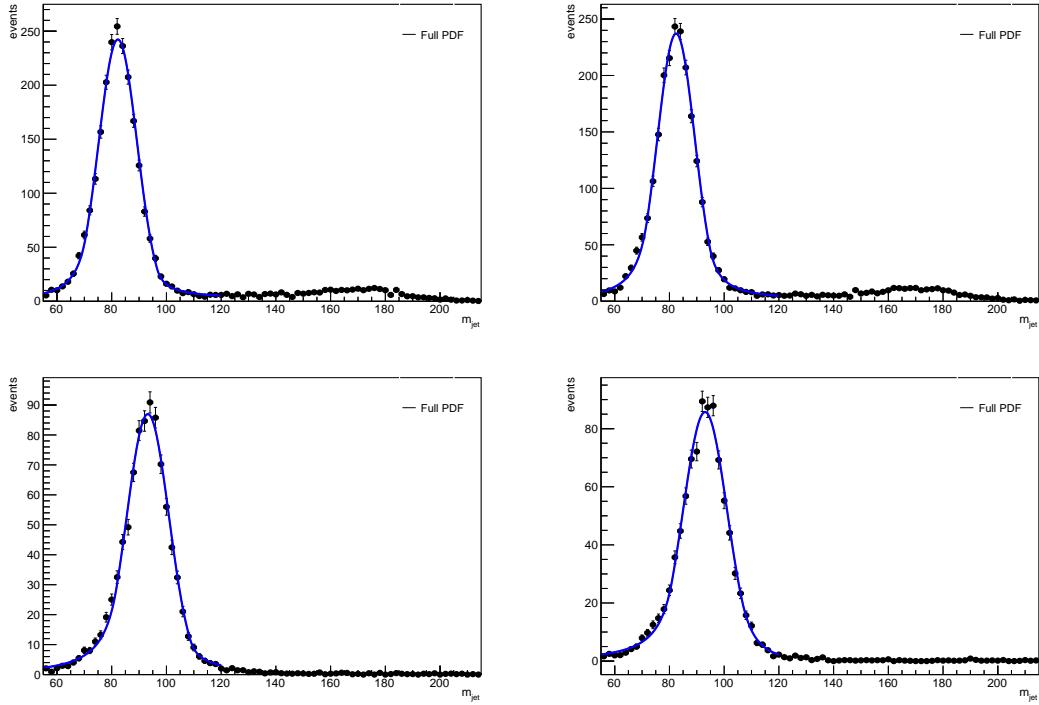


Figure 8.32: Final fits for a double-sided crystal-ball function (blue line) to the resonant part of the V+jets spectrum (black markers) for W+jets (top) and Z+jets (bottom). The left figure shows $M_{\text{jet}1}$ the right $M_{\text{jet}2}$, for low purity selections.

is shown in Figure 8.33. Exhaustive tests have shown that correlations between the jet-mass and the dijet invariant mass are small enough to be neglected in the modeling of these backgrounds, i.e. the shape of the jet mass spectrum does not significantly depend on the momentum of the

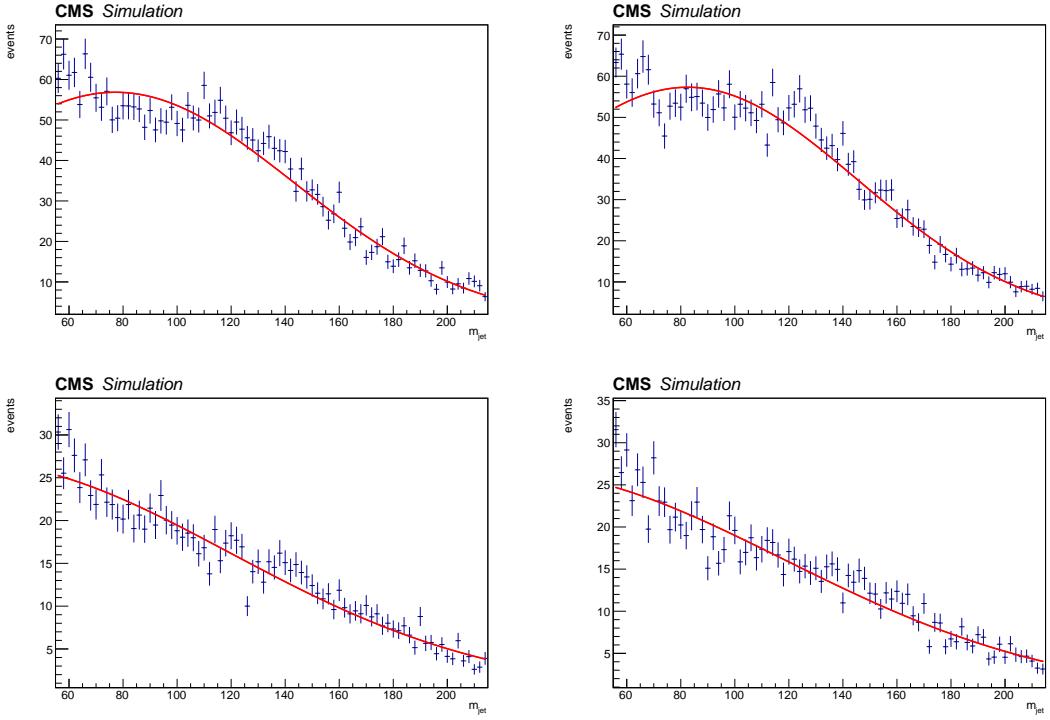


Figure 8.33: Final fits for a Gaussian function (red line) to the non-resonant part of the $V + \text{jets}$ spectrum (markers) for $W + \text{jets}$ (top) and $Z + \text{jets}$ (bottom). The left figure shows $M_{\text{jet}1}$ the right $M_{\text{jet}2}$, for low purity selections.

jets. For the non-resonant dijet invariant mass shape, P_{nonres} , the same kernel approach as for the non-resonant QCD background is used with one minor difference: due to the low statistics in the high- M_{VV} tail, an additional smoothing of the jet mass tail using a leveled exponential of the form

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/s)^{P_1}}{(m_{jj}/s)^{P_2}} \quad (8.7)$$

is performed. Here, s is the center-of-mass energy. The function is fitted to the spectrum starting from a dijet mass of 1.1 (2.1) TeV for the high-purity (low-purity) category. Two p_T -dependent corrections are added to the W and $Z + \text{jets}$ crosssection which affect the dijet invariant mass slope. One corrects the leading-order p_T -distribution of the vector bosons to the one predicted at next-to-leading order in QCD (*NLO k-factor*), and the other accounts for electroweak effects at high p_T (*EWK correction*). These higher-order corrections were calculated in Refs. [83–85] and are derived as a function of the generated boson p_T . Two uncertainties on the M_{VV} shape are added in order to accommodate possible MC mis-modeling: one proportional to M_{VV} and one proportional to $1/M_{VV}$. The resulting M_{VV} kernels (solid lines) for the $W + \text{jets}$ background are

shown in Figure 8.34 and are compared to MC simulation (markers). The blue line corresponds to the nominal shape, while the red and green lines correspond to the uncertainties proportional to M_{VV} and $1/M_{VV}$, respectively. The corresponding distributions for the Z+jets background are shown in Appendix D.3.

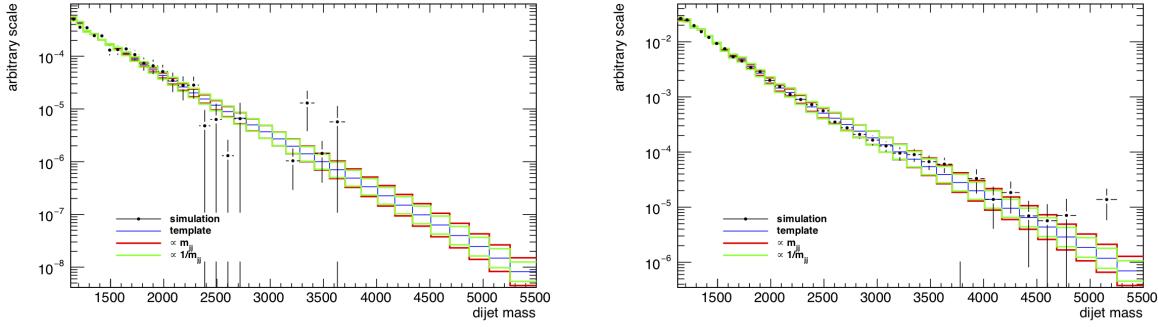


Figure 8.34: One-dimensional M_{VV} kernels (solid line) compared to MC (markers) for the W+jets background in the HP (left) and LP (right) categories. The blue line corresponds to the nominal shape, while the red and green lines correspond to uncertainties proportional to M_{VV} and $1/M_{VV}$, respectively.

As a closure test, the three-dimensional kernel as defined in Equation 8.6 is fitted to the V+jets and $t\bar{t}$ simulation. Figure 8.35 shows the fitted kernel (red) together with the MC data points (black markers) in the low-(top) and high-purity (bottom) categories. Mostly good agreement is observed along all dimensions, with some deviations in the very high M_{VV} tail in the high-purity category. This is, however, a region populated with very few events and is completely overwhelmed by the QCD multijets background that has the same shape.

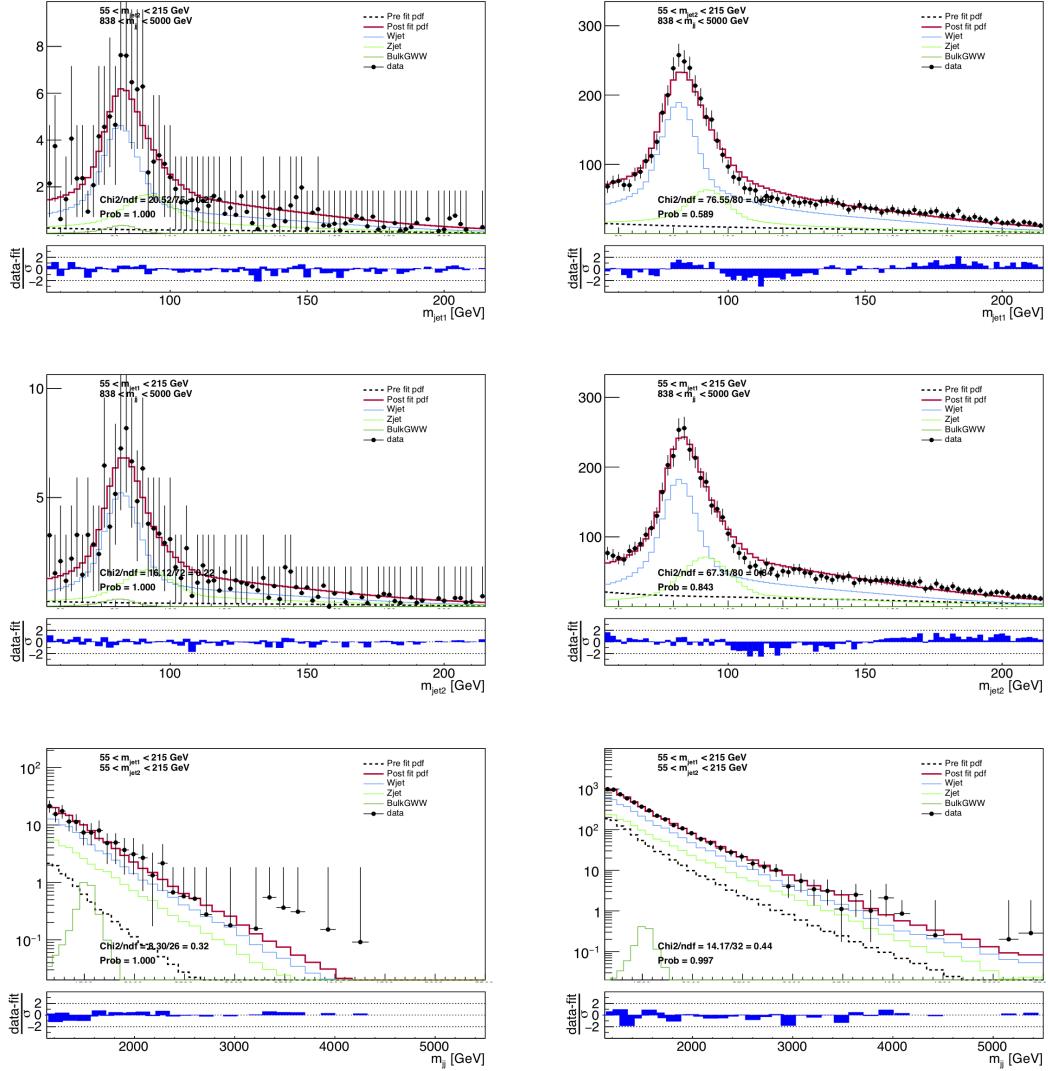


Figure 8.35: Fits using the complete resonant background model (red line) to the V+jets MC simulation (black markers) for the high- (left) and low-purity (right) category. Here for $M_{\text{jet}1}$ (top), $M_{\text{jet}2}$ (middle) and M_{VV} (bottom).

8.8 Systematic uncertainties

The background estimate for each signal mass hypothesis is obtained by fitting the signal and background three-dimensional probability density functions obtained above, to the observed data in each analysis category. We follow the modified frequentist prescription (CLs) criterion, evaluated using asymptotic expressions described in Ref. [73]. Systematic uncertainties are treated as nuisance parameters and profiled in the statistical interpretation using log-normal priors, while Gaussian priors are used for shape uncertainties. The uncertainties entering the fit are listed in Table 8.5. Here, V+jets ratio describes the ratio between W+jets and Z+jets events

Source	Relevant quantity	HPHP unc. (%)	HPLP unc. (%)
PDFs	Signal yield	3	
W-tagging efficiency	Signal+ V+jets yield	25 (21)	13 (11)
W-tagging p_T -dependence	Signal+ V+jets yield	8-23	9-25
Integrated luminosity	Signal+ V+jets yield	2.3 (2.6)	
QCD normalization	Background yield	50	
V+jets normalization	Background yield	10	
V+jets ratio	Migration	10	
PDFs	Signal M_{VV}/M_{jet} shape	< 1	
Jet energy scale	Signal M_{VV} shape	2	
Jet energy resolution	Signal M_{VV} shape	5	
Jet mass scale	Signal + V+jets M_{jet} shape	1	
Jet mass resolution	Signal + V+jets M_{jet} shape	8	
QCD HERWIG++	QCD shape	33	
QCD MADGRAPH+PYTHIA8	QCD shape	33	
p_T -variations	QCD shape	33	
Scale-variations	QCD shape	33	
High- M_{jet} turn-on	QCD shape	33	
p_T -variations	V+jets M_{VV} shape	33	

Table 8.5: Summary of the systematic uncertainties and the quantities they affect. Numbers in parenthesis correspond to uncertainties for the 2016 analysis when these differ from those in 2017.

and is allowed to float by 10%. The W-tagging efficiency scale factor is fully anti-correlated between the HP and LP categories (3–10%), and fully correlated between signal and +jets. The p_T -dependence uncertainty of the scale factor arises from the extrapolation of the W-tagging efficiency scale factors, which are measured in $t\bar{t}$ events where the jet has a p_T around 200 GeV, towards higher transverse momenta. This uncertainty is estimated in signal MC and is based on the difference in tagging efficiency between PYTHIA and HERWIG++ as a function of p_T relative to the difference at 200 GeV. This is considered as correlated between the τ_{21}^{DDT} categories and is defined in Section 8.6.2. The shape uncertainties on M_{jet} are considered fully correlated between signal and V+Jets, allowing for the data to constrain these parameters. These affect the mean

and the width of the signal and V+jets PDFs.

Uncertainties on the background shape are added as alternate PDFs to the fit through vertical template morphing. This creates nuisance parameters for each shape that simultaneously affect all three dimensions. We define five shape nuisance parameters in total. The first effect corresponds to a variation of the underlying transverse momentum spectrum and the corresponding template is obtained by simultaneously varying jet masses and M_{VV} by a quantity proportional to M_{VV} and M_{jet} . The second effect is a variation of the scale and is taken into account through an alternate shape obtained by simultaneously varying the jet masses and M_{VV} by a quantity proportional to $1/M_{VV}$ and $1/M_{jet}$. Three additional alternate shapes that simultaneously affect resonance mass and jet groomed mass are also added in order to take into account differences in MC generation and modeling of parton shower. Since the choice of QCD MC used to generate the nominal

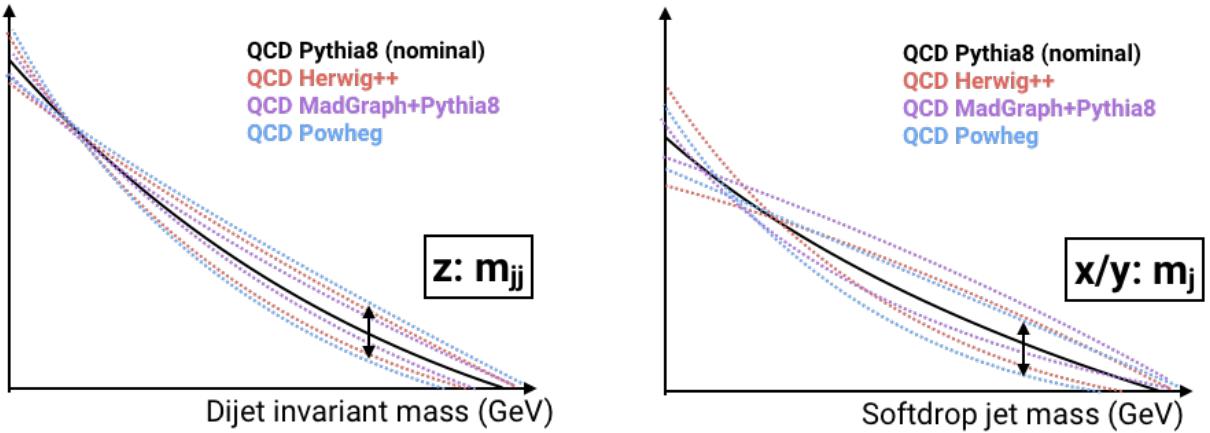


Figure 8.36: QCD Monte Carlo predictions differ significantly between generators. Since we do not know a-priori whether Nature behaves more like PYTHIA8, HERWIG++, MADGRAPH+PYTHIA8 or POWHEG, we allow the 3D template to assume the shape predicted by any of them by adding three additional shapes to the fit, corresponding to the 3D templates obtained using different QCD MC generators.

template is random and we do not know a-priori whether Nature behaves more like PYTHIA8, HERWIG++, or MADGRAPH+PYTHIA8, we insert additional templates derived using all available QCD MC, as illustrated in Figure 8.36. This allows us to include all background knowledge we have available into the fit. Finally, in order to account for an expected M_{VV} turn-on in the extreme large- M_{jet} ($> 175 \text{ GeV}$) and low- M_{VV} ($< 1200 \text{ GeV}$) region, an additional shape uncertainty parametrizing any discrepancy between the three-dimensional template and QCD MC is added to the fit. Note that the latter shape uncertainty only affects the region above $M_{jet} > 175 \text{ GeV}$ and below $M_{VV} < 1200 \text{ GeV}$, far from where signal is expected. The above nuisance parameters are assigned very large pre-fit uncertainties, allowed to float within 33–50%,

effectively allowing the simulation enough flexibility to adapt to data. The alternate shapes described above are shown in Figure 8.37.

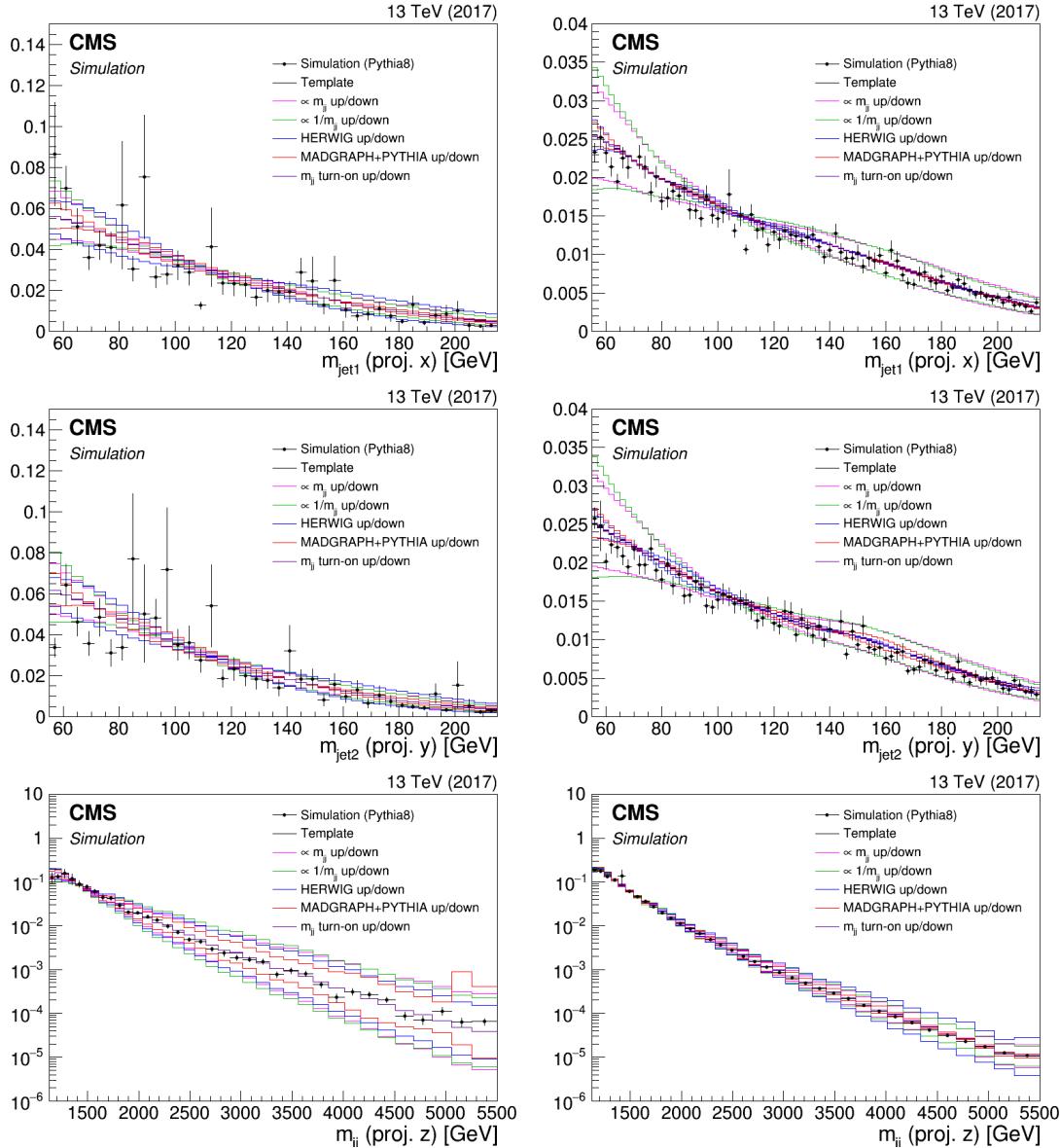


Figure 8.37: The nominal MC data (markers) and smooth nominal kernel obtained from PYTHIA8 (black line), together with the five alternate shapes added to the fit as nuisance shape parameters. Here for the high- (left) and low-purity (right) categories.

8.9 Background model validation

In order to test the robustness of the fit, several checks are performed in simulation and in a data control region before unblinding the data signal region.

8.9.1 Variations of QCD multijet predictions

First, we check that the main three-dimensional background template (generated starting from QCD PYTHIA8 MC) can fit predictions from alternate QCD multijets samples in order to demonstrate that systematic uncertainties in the modeling of parton showers are covered by the relevant nuisance parameters. Fits to three different QCD samples are compared: PYTHIA8 (nominal), HERWIG++ (alternate shape 1) and MADGRAPH+PYTHIA8 (alternate shape 2). To ensure a smooth data distribution not affected by low statistics present in the MC spectra, we generate a toy dataset from the three-dimensional templates rather than using MC events directly. The post-fit distributions after fitting each toy is shown in Figures 8.38-8.40 for the high- (left) and low-purity (right) category. The fit quality of these are quantified using a goodness-of-fit (based on a saturated model, which is valid when data are non-Gaussian), shown in Figure 8.41. The test statistics are Gaussian distributed and the toy dataset is in good agreement with the background only hypothesis, demonstrating the fits ability to account for differences in QCD multijet predictions.

8.9.2 Fit to data control region

To validate the method in data, the 3D fit procedure is tested in a data control region where both jets are required to have $0.43 < \tau_{21}^{DDT} \leq 0.79$, the so-called LPLP category. The templates are built following the procedure described in Section 8.7.2. In this category, the contribution of the resonant background is negligible with respect to the dominant QCD multijet background and is removed from the fit. We use this control region (where no signal is expected) to quantify which alternate shapes are needed to describe the data. This is done by adding in alternate shapes one by one and checking the improvement in the goodness-of-fit. This is quantified through a p-value which describes the agreement between the observed likelihood value and the distribution of the toy likelihoods, where 0.5 indicates perfect agreement. The p-values obtained with the nominal model (using all alternate shapes) are compared with the values using alternative models (removing or adding alternate shapes) in Table 8.6. From the table we conclude that the best agreement is found when including all shape uncertainties and hence we decide to keep all of them included in the nominal background model for the final fit to data in the signal region.

The post-fit distributions in the data control region using all alternative shapes are shown in

Table 8.6: P-values obtained by comparing the distributions of the test statistic for toys and the observed value for the nominal background model and alternative models.

model	HERWIG shape	MADGRAPH shape	$\propto (1/M_{VV}, 1/M_{jet})$ shapes	$\propto (M_{VV}, M_{jet})$ shapes	p-value
Nominal	yes	yes	yes	yes	0.30
Model 1	yes	yes	yes	no	0.28
Model 2	yes	yes	no	yes	0.27
Model 3	yes	yes	no	no	0.26
Model 4	no	yes	yes	yes	0.16
Model 5	yes	no	yes	yes	0.13
Model 6	no	no	yes	yes	0.09
Model 7	no	no	no	yes	0.03

Figures 8.42–8.44 for different jet-mass and dijet invariant mass bins. The post-fit value and uncertainty of each nuisance parameter involved in the fit is also studied, where deviations from the pre-fit value is quantified through the “pull”, defined as $p_\theta = (\theta - \theta_{in})/\sigma_\theta$, where, θ and σ_θ are the post-fit value of the nuisance parameter and its uncertainty, and θ_{in} the pre-fit value. The pulls for all nuisance parameters are shown in Figure 8.45, where the uncertainty is defined as the ratio between the post- and pre-fit uncertainties. The post-fit values show a reasonable deviation from the chosen pre-fit values of 0. The uncertainties are strongly constrained by data, as expected given the large pre-fit uncertainty assigned to let the shapes adjust to the real data. In addition to the goodness-of-fit check in the LPLP data sideband region, we perform the same test on the data signal region. This is shown in Figure 8.46. Good agreement between the observed fit result in data and the background only hypothesis from MC simulation is observed.

8.9.3 Bias test in pseudodata

Finally, we study whether there is any bias on the extracted signal rate due to the background model. A $G_{\text{bulk}} \rightarrow WW$ signal is injected on top of a toy dataset generated under either the PYTHIA8 or the HERWIG++ template, and is fit using the full background model. The test is done for four different signal mass values and with a signal strength chosen such that it corresponds to a significance of 4-4.5 standard deviations for each signal mass which is tested. An example distribution after injecting a 3 TeV signal is shown in Figure 8.47 for the high-purity category. For each tested signal mass point, a signal plus background fit is performed. The signal normalization is free to float in the fit, which determines the signal strength μ and its uncertainty σ_μ . For

each toy, the pull of the signal strength $p_\mu = (\mu - \mu_{in})/\sigma_\mu$ is calculated, where μ_{in} is the pre-fit value of the signal strength. The procedure is repeated ~ 1000 times for each category, and the cumulative distribution of the pulls is fitted with a Gaussian function to determine the mean and its uncertainty, which represent the bias. This is shown in Figure 8.48. Ideally, the distribution should be Gaussian distributed with a mean around 0 and a width of 1. The bias for a toy dataset generated under the PYTHIA8 (red markers) or the HERWIG++ (black markers) template, as shown in Figure 8.49, is consistently below 50% and therefore no additional correction/bias term is introduced. In addition, we calculate the pull for each nuisance parameter and fit the cumulative distribution with a Gaussian, in order to determine the mean and its uncertainty, representing the shift of the parameters θ with respect to the pre-fit values. The results are shown in Figure 8.50. The mean of the cumulative distributions of the pulls is found to be zero for all nuisance parameters, as expected. The width of the distribution can range from very small values to about 1, depending on the assigned pre-fit uncertainty and the power of the data to constrain it.

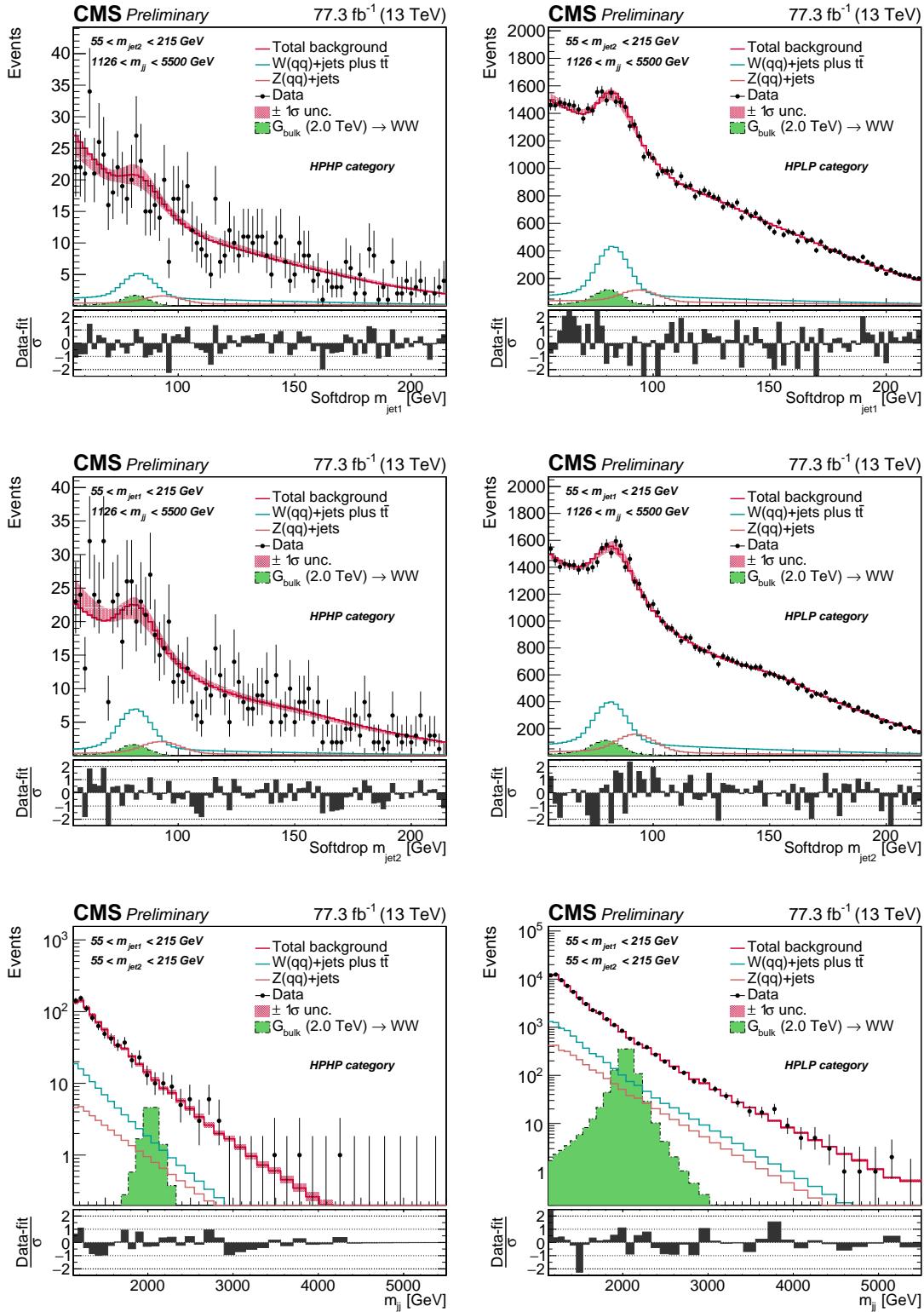


Figure 8.38: Postfit distributions after a combined fit to a toy datasets generated under the QCD PYTHIA8 template. The projections of M_{jet1} (top), M_{jet2} (middle) and M_{VV} (bottom) are shown for the high- (left) and low-purity (right) category.

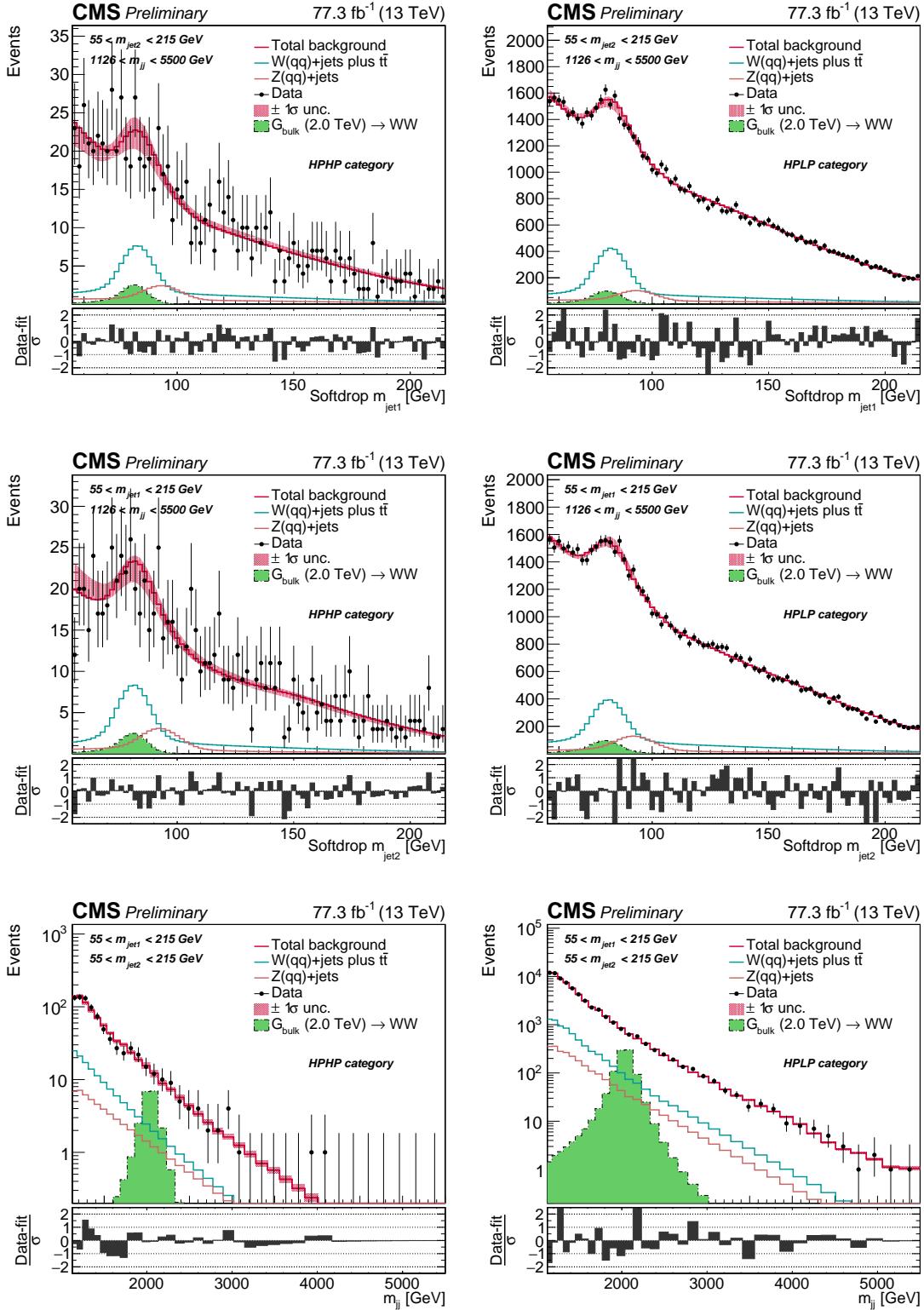


Figure 8.39: Postfit distributions after a combined fit to a toy dataset generated under the QCD HERWIG++ template. The projections of $M_{\text{jet}1}$ (top), $M_{\text{jet}2}$ (middle) and M_{ll} (bottom) are shown for the high- (left) and low-purity (right) category.

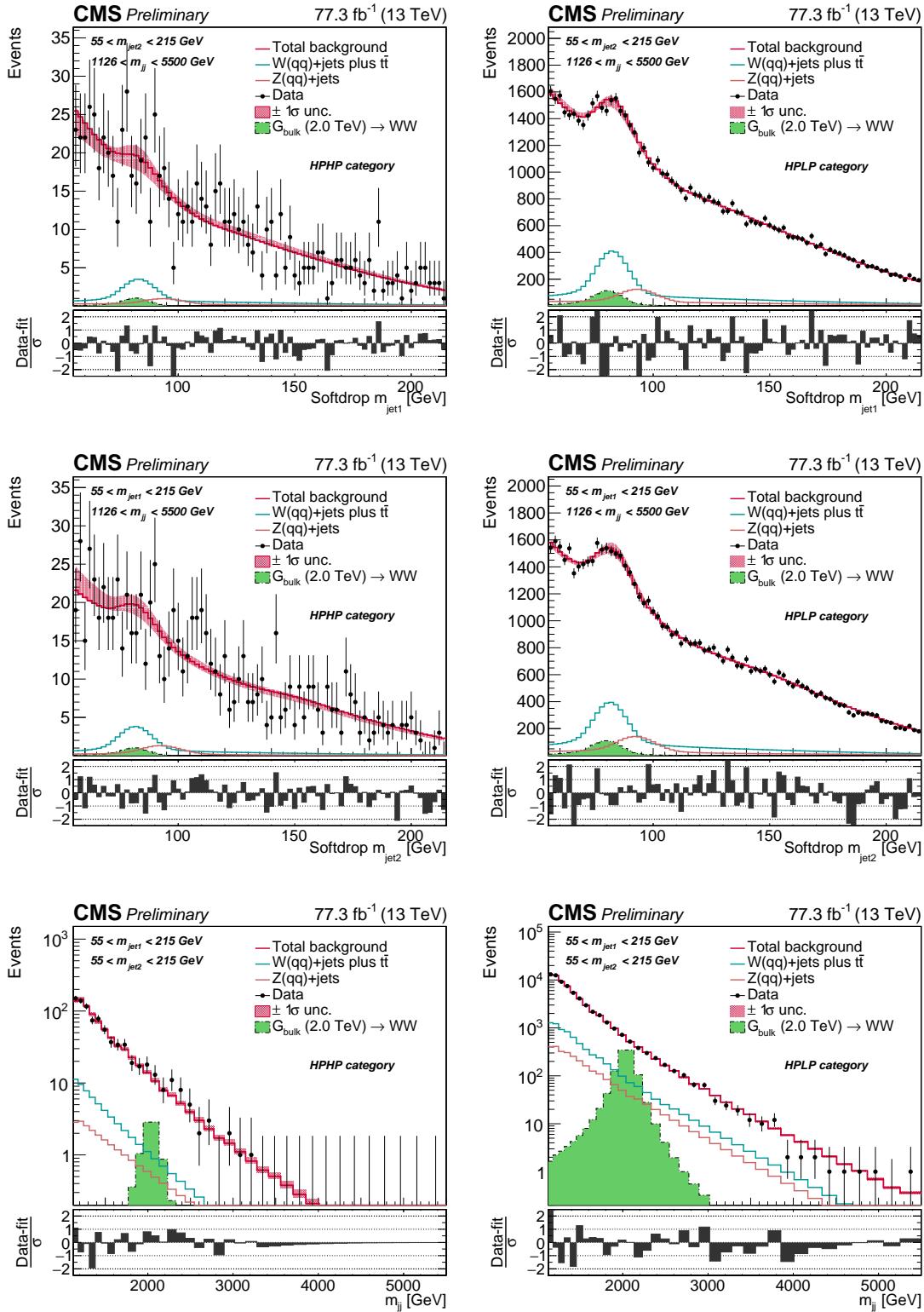


Figure 8.40: Postfit distributions after a combined fit to a toy datasets generated under the QCD QCD MADGRAPH+PYTHIA8 template. The projections of M_{jet1} (top), M_{jet2} (middle) and M_{ll} (bottom) are shown for the high- (left) and low-purity (right) category.

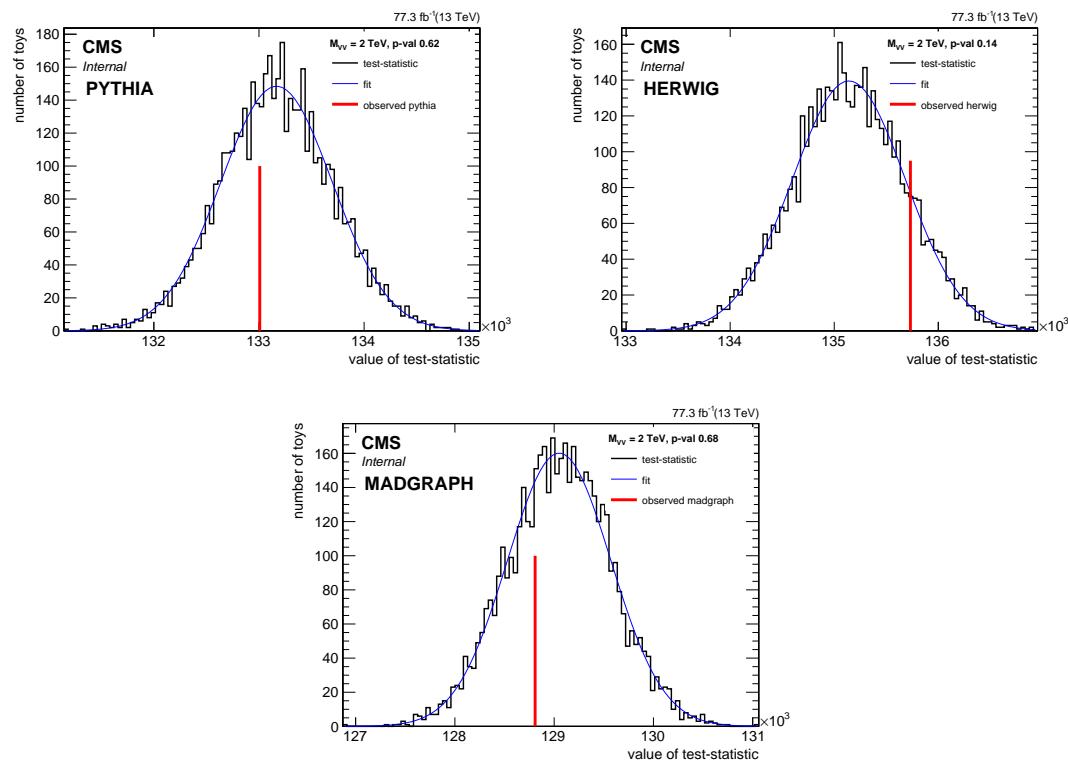


Figure 8.41: The likelihood for toys generated around the background-only hypothesis compared to the likelihood value of a toy dataset generated under the PYTHIA (top left), HERWIG++ (top right) and MADGRAPH+PYTHIA8 template (bottom).

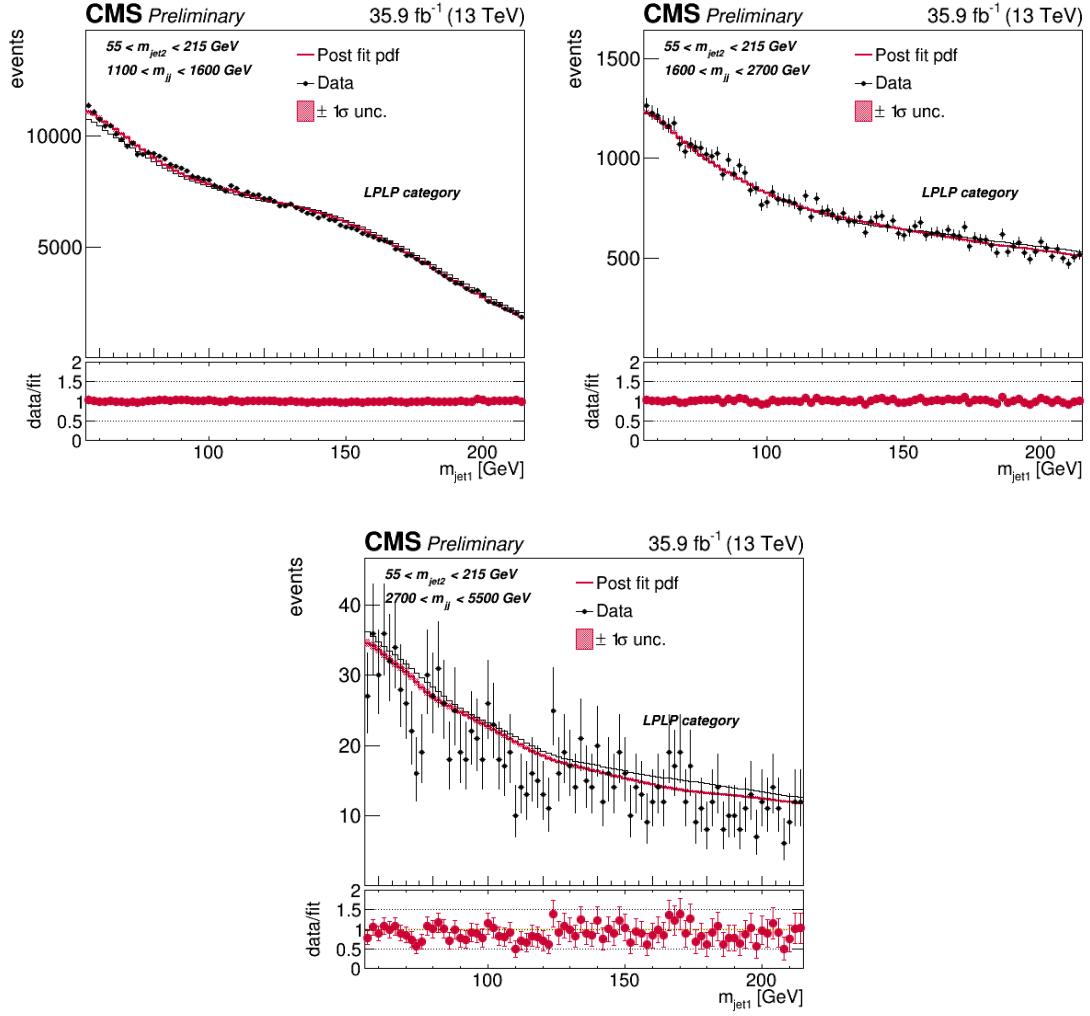


Figure 8.42: Distributions obtained from the fit to 2016 LPLP data. Here the projections of $M_{\text{jet}1}$ are shown for several ranges of M_{VV} , as labelled in the top left corner of each plot, and for the full $M_{\text{jet}2}$ range.

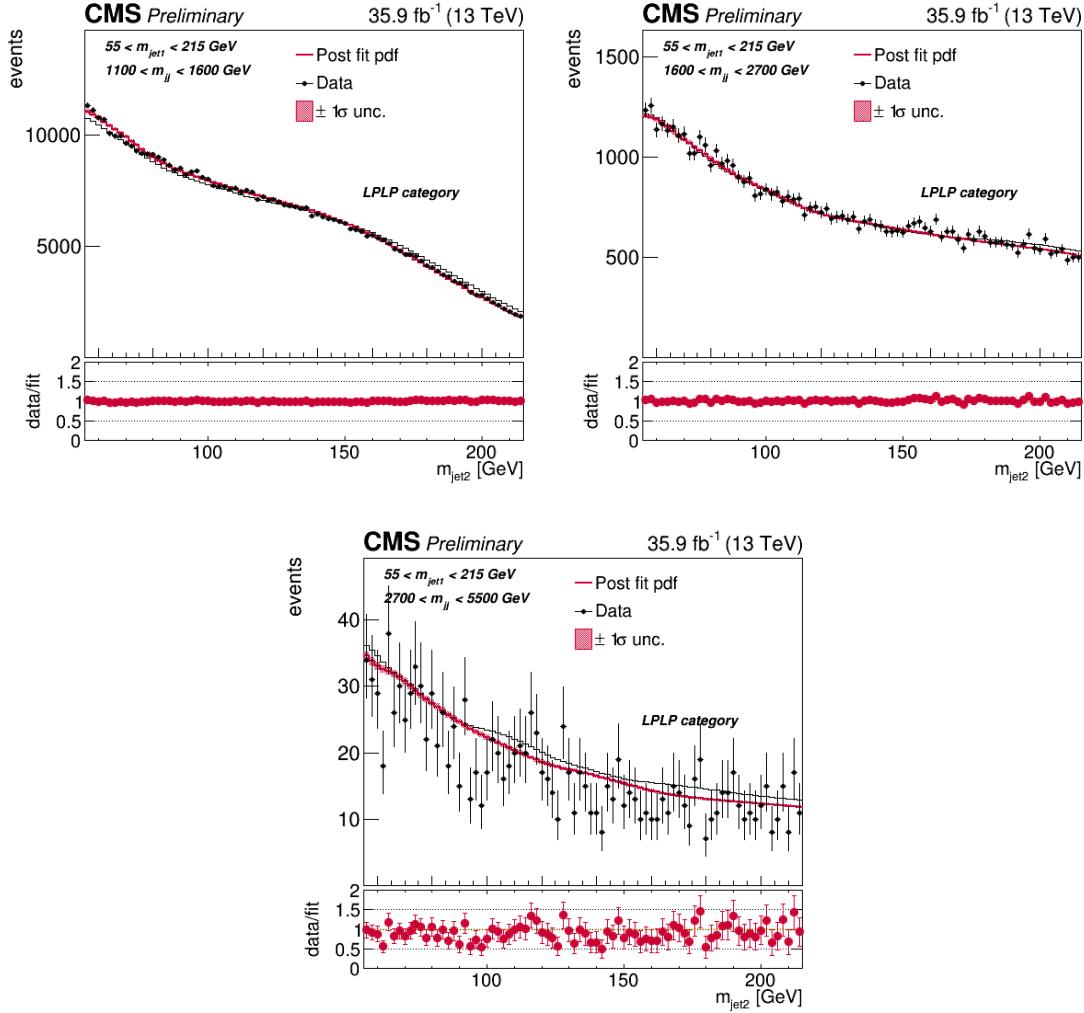


Figure 8.43: Distributions obtained from the fit to 2016 LPLP data. Here the projections of $M_{\text{jet}2}$ are shown for several ranges of M_{VV} , as labelled in the top left corner of each plot, and for the full $M_{\text{jet}1}$ range.

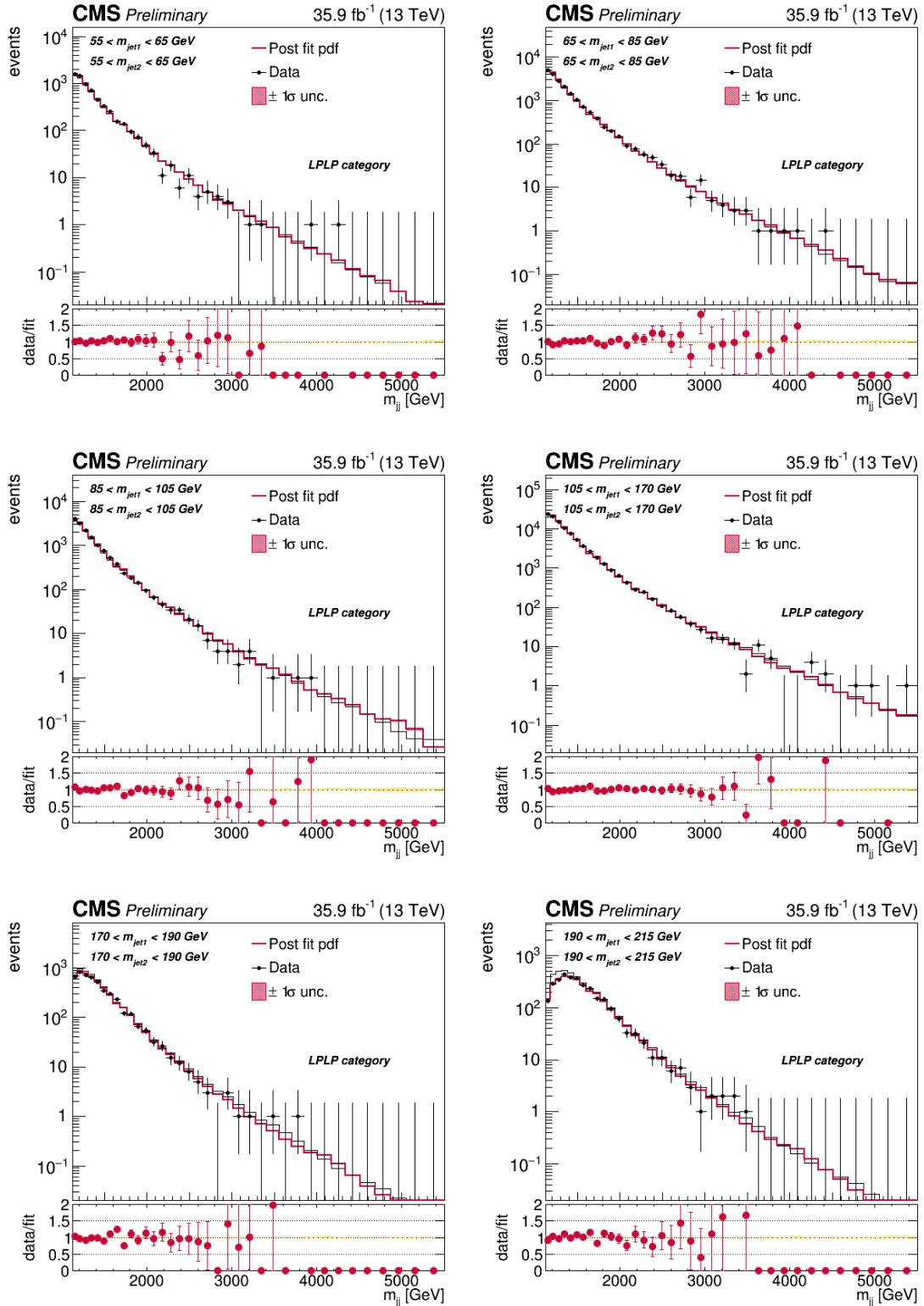


Figure 8.44: Distributions obtained from the fit to 2016 LPLP data. Here the projections of M_{VV} are shown for several equal ranges of M_{jet1} and M_{jet2} , as labelled in the top left corner of each plot.

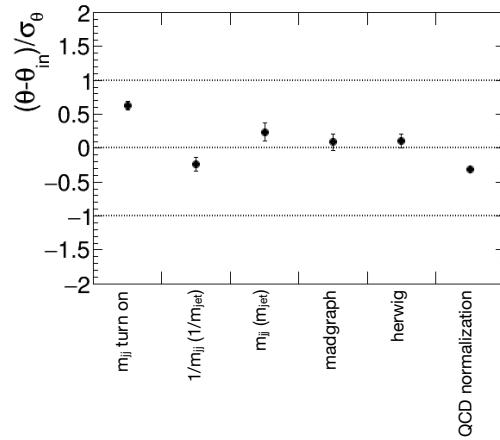


Figure 8.45: Pulls of the nuisance parameters for a background-only fit to LPLP data.

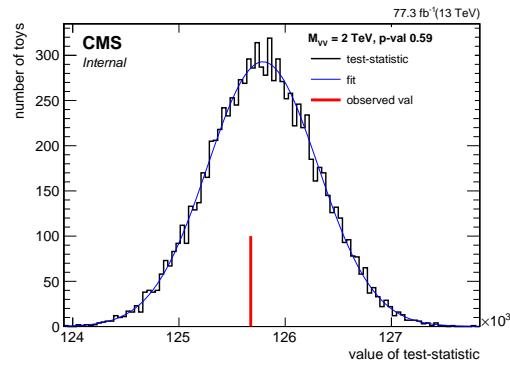


Figure 8.46: Distribution of the likelihood for toys generated around the background-only hypothesis compared to the likelihood on data for the full combination of 2016 and 2017 data.

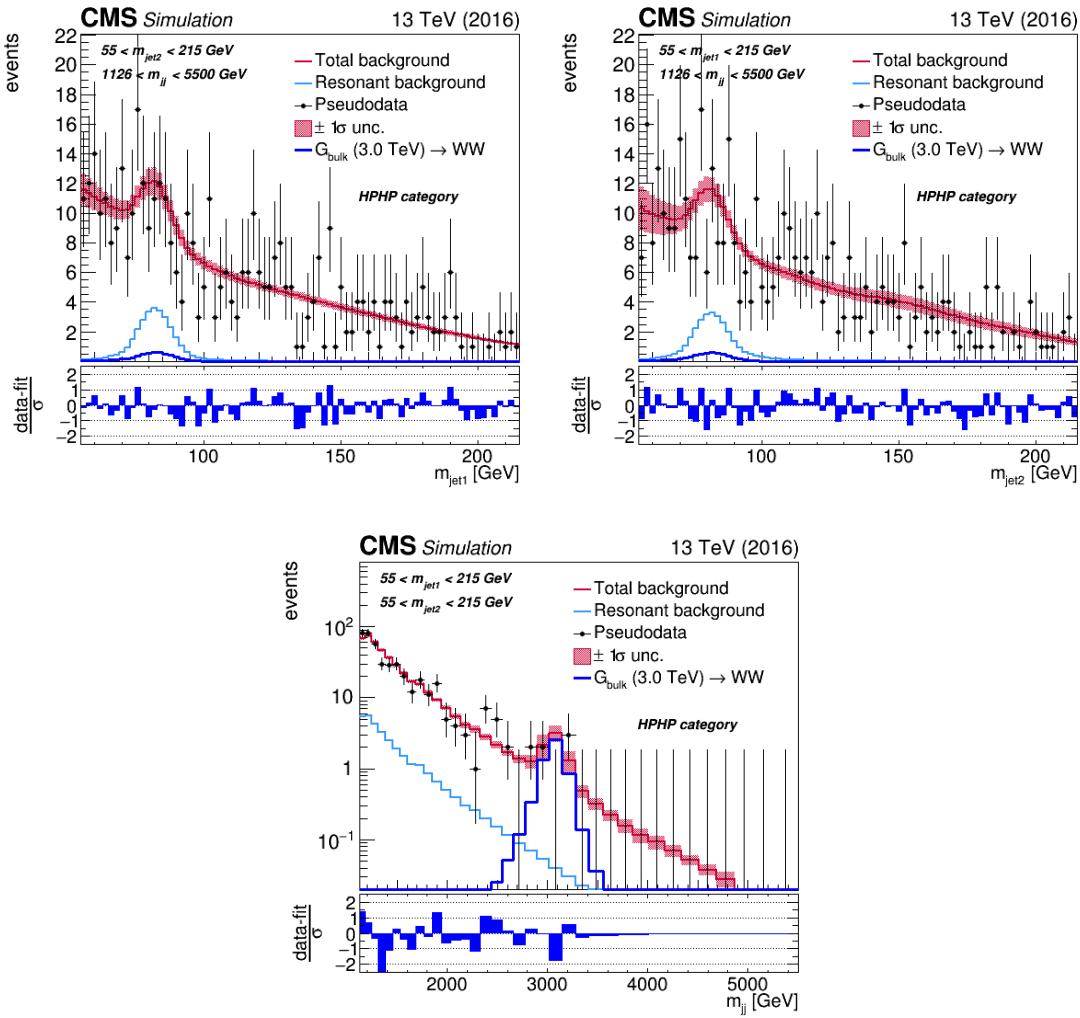


Figure 8.47: The post-fit distribution after injecting a signal with a mass of 3 TeV on top of the background in the HP category.

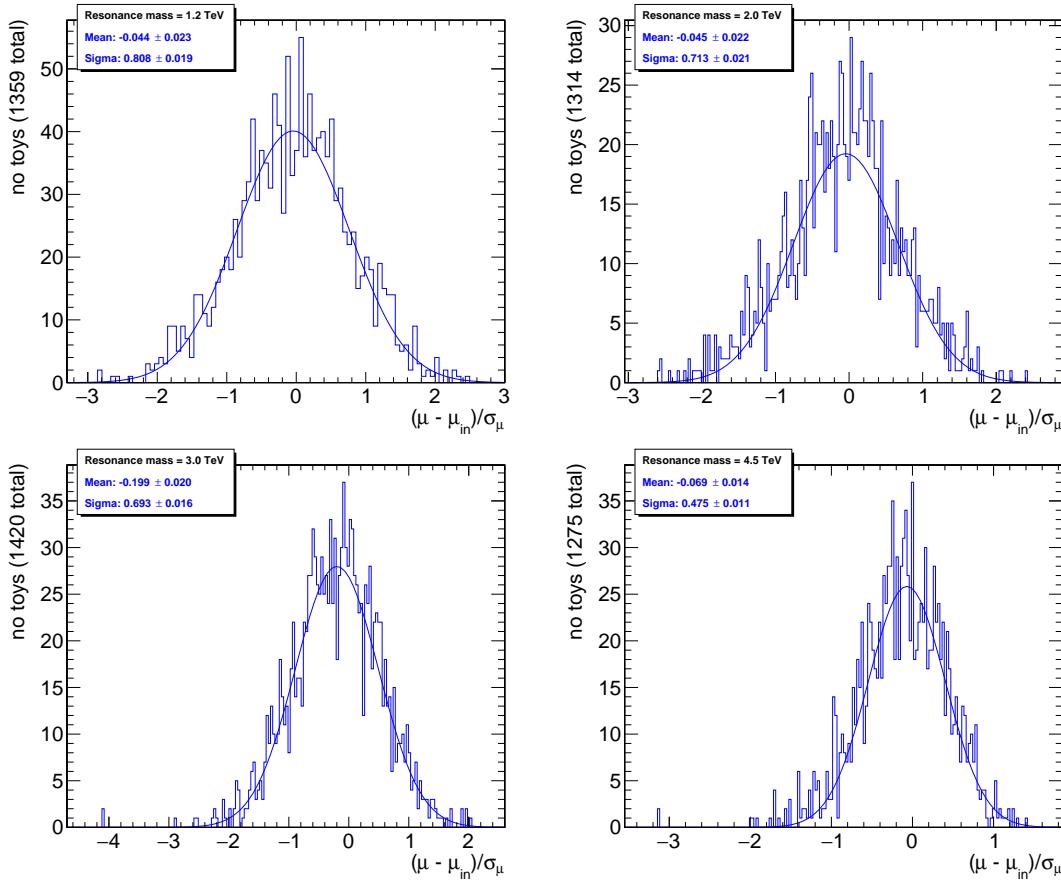


Figure 8.48: Cumulative distributions of the pulls for the signal strength for 4 different signal mass-points.

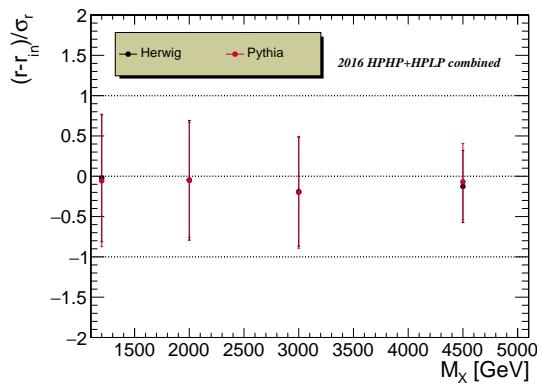


Figure 8.49: Estimated bias as a function of the resonance mass.

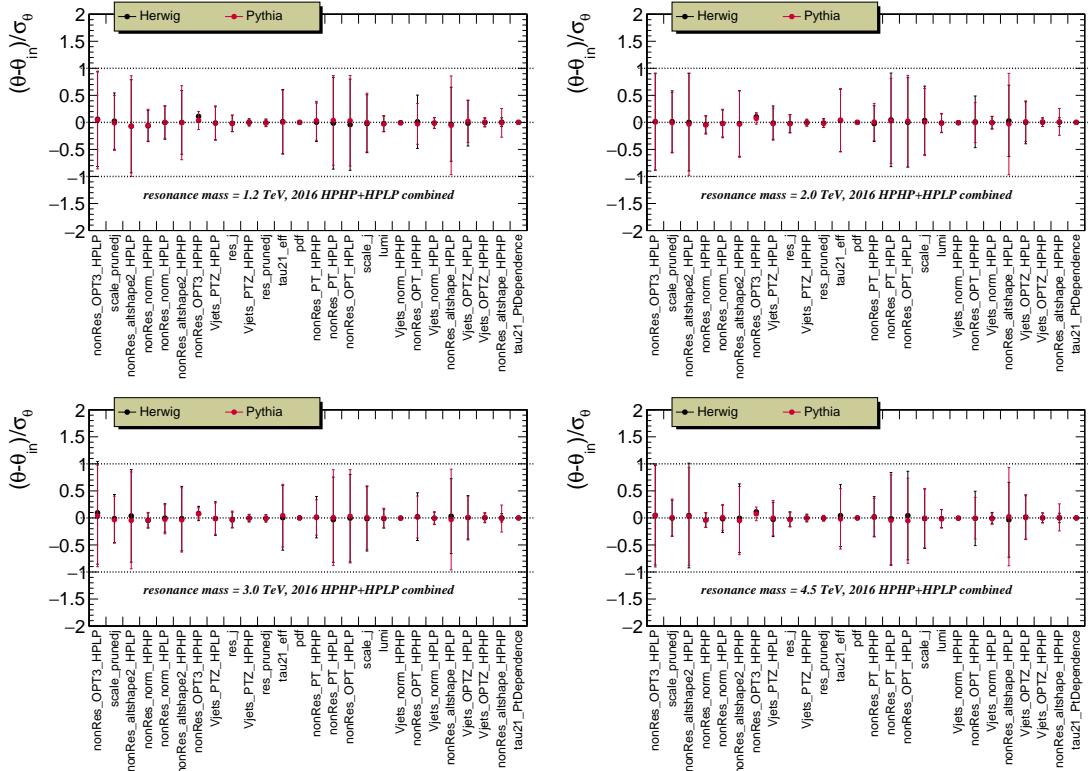


Figure 8.50: Pulls of the fit nuisance parameters after injecting signals of different mass and signal strength on top of a toy datasets generated from the nominal PYTHIA8 template (red markers) and for a toy generated under the HERWIG++ template (black markers).

8.10 Results

The distributions obtained from a combined fit to the observed data in 2016 and 2017 are shown in Figure 8.51 and 8.52, with the corresponding predicted and observed number of background events in the signal region summarized in Table 8.7. We observe a beautiful double peak from the $W(q\bar{q})$ and $Z(q\bar{q})+jets$ background, especially visible in the low-purity category. This allows us to, for the very first time, constrain the softdrop jet mass scale and resolution simultaneously from a $W(q\bar{q})+jets$ and $Z(q\bar{q})+jets$ SM background, which we will discuss in Section 8.10.2 (a similar measurement for $Z(bb)$ was done in Ref. [86]). No excess is observed and we proceed by setting upper limits on the signal cross section times the branching ratio in the same way as in Section 6.9.

	HPHP	HPLP
$W+jets$	113.3 ± 18.1	4257.4 ± 257.0
	100.4 (exp.)	4318.0 (exp.)
$Z+jets$	46.5 ± 8.3	1747.5 ± 163.7
	50.2 (exp.)	2159.0 (exp.)
QCD	651.6 ± 4.0	51190.5 ± 313.1
	684.4 (exp.)	53767.5 (exp.)
Observed yield	778 ± 28	57227 ± 239
Post-fit total background	811.4 ± 20.3	57195.5 ± 436.8

Table 8.7: Expected and observed yields and their total uncertainty (stat.+sys.) in the two purity categories.

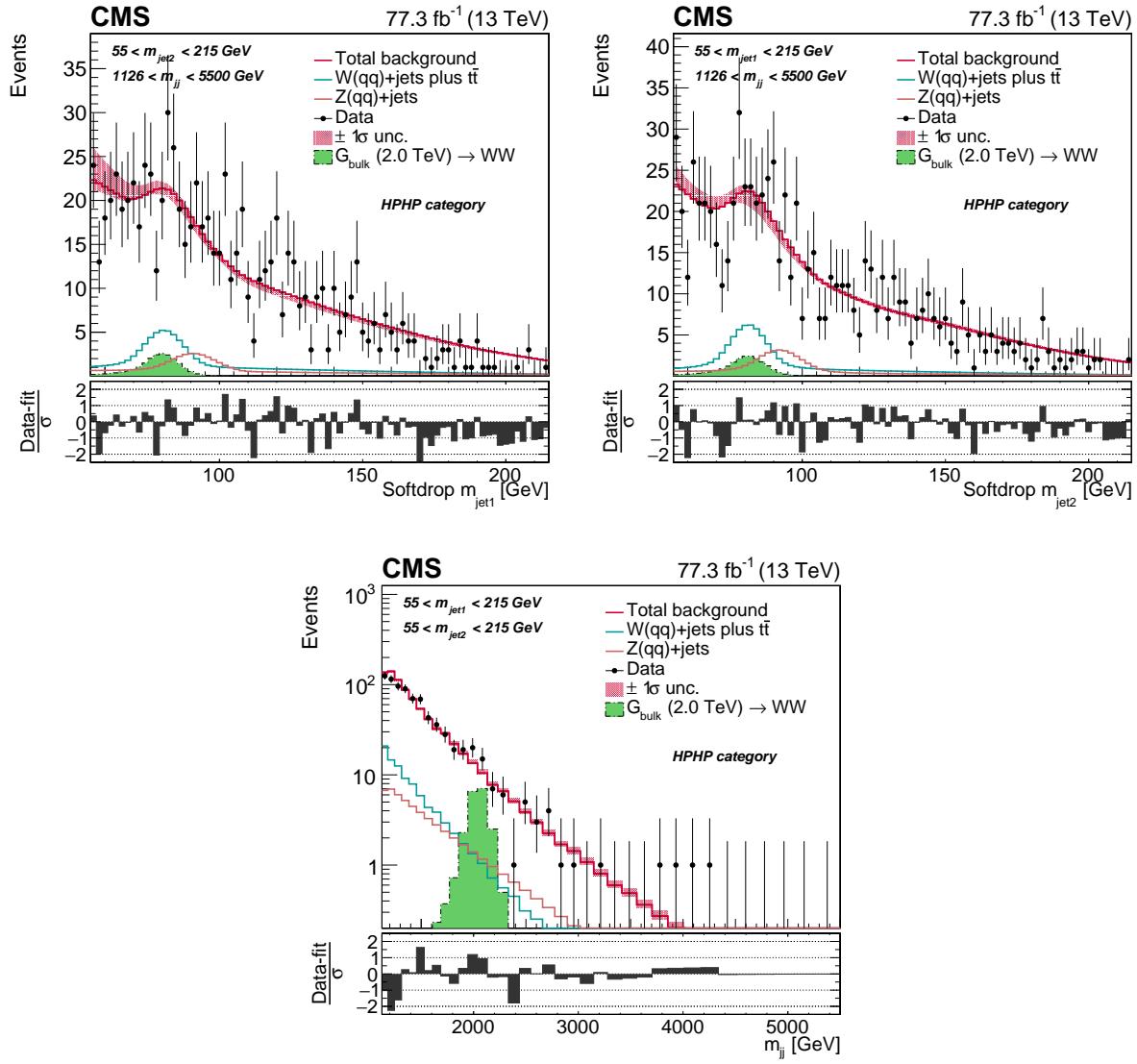


Figure 8.51: Postfit distribution after a fit to 2016 and 2017 data projected onto the $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle), and M_{VV} (right) axis for the high-purity category. The background shape uncertainty is shown as a red shaded band, and the statistical uncertainties of the data are shown as vertical bars. The overlaid signal distribution is arbitrarily normalized.

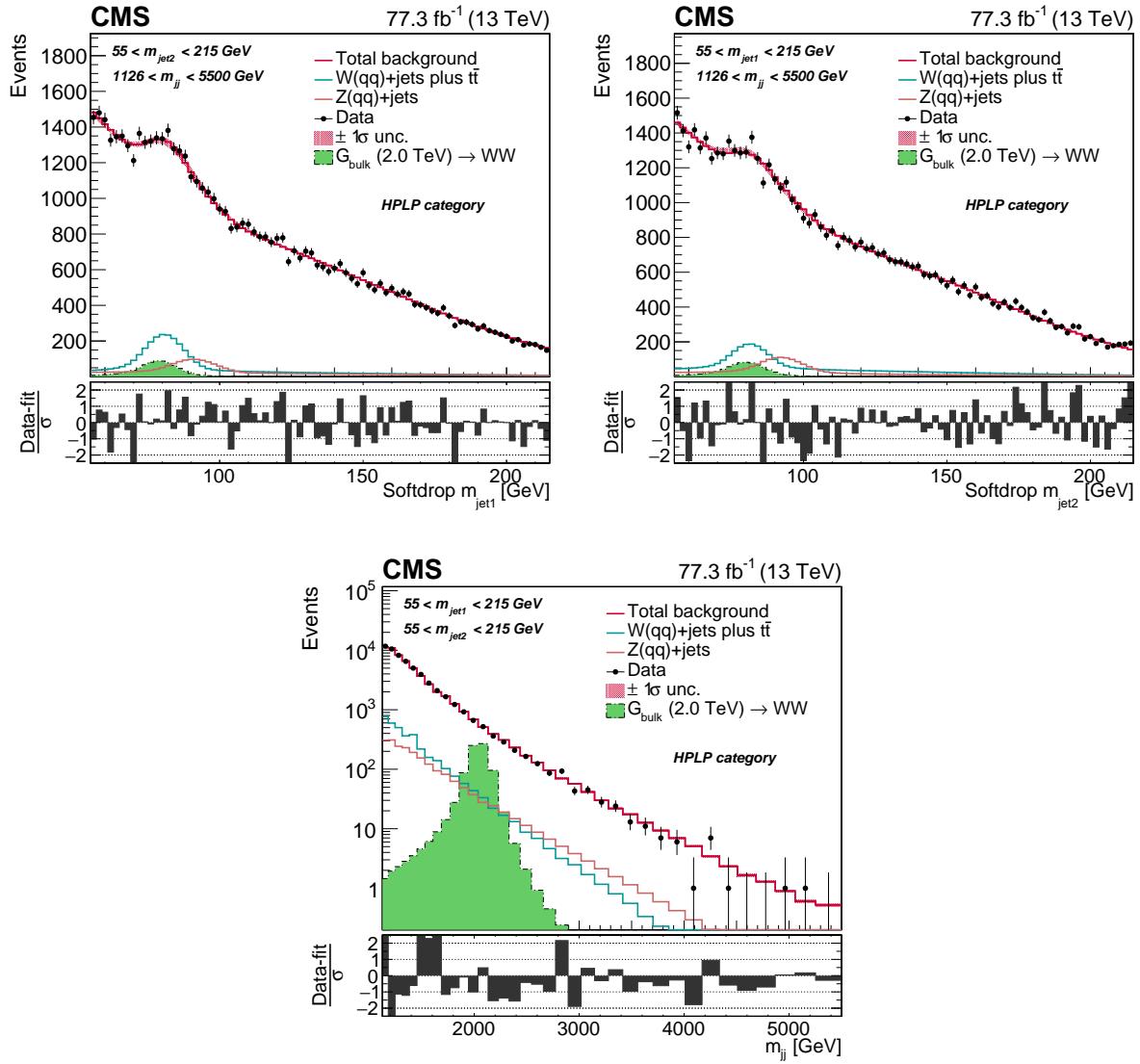


Figure 8.52: Postfit distribution after a fit to 2016 and 2017 data projected onto the M_{jet1} (left), M_{jet2} (middle), and M_{VV} (right) axis for the low-purity category. The background shape uncertainty is shown as a red shaded band, and the statistical uncertainties of the data are shown as vertical bars. The overlaid signal distribution is arbitrarily normalized.

8.10.1 Limits

As for Search I and Search II, exclusion limits on the cross section of the process $X \rightarrow VV$ are set in the context of the Bulk Graviton model and the HVT model B scenario (again obtained using the asymptotic CLs method). Figure 8.53 show the resulting expected and observed exclusion limits at 95 % confidence level on the signal cross section times branching fraction as a function of the resonance mass for a $G_{\text{bulk}} \rightarrow WW$ (top left), $G_{\text{bulk}} \rightarrow ZZ$ (top right), $W' \rightarrow WZ$ (bottom left) and $Z' \rightarrow WW$ (bottom right) signal. The obtained limits are compared with the resonance's production cross section times the branching fraction to WW, ZZ and WZ. For the heavy vector triplet model B, we exclude at 95% confidence level W' and Z' spin-1 resonances with masses below 3.8 and 3.5 TeV, respectively. In the narrow-width bulk graviton model, lower limits on the production cross sections are set in the range from 20 fb for a resonance mass of 1.2 TeV, down to 0.2 fb for high resonance masses above 5.2 TeV for $G_{\text{bulk}} \rightarrow WW$. In the case of $G_{\text{bulk}} \rightarrow ZZ$, lower cross section limits are 27 fb and 0.2 fb for bulk graviton masses of 1.2 and 5.2 TeV, respectively.

Comparison with 1D fit

The analysis of the 2016 dataset and corresponding limits obtained using the 3D fit method, can be compared to the limits using the 1D fit in Search II in order to estimate whether there is a sensitivity gain. One caveat of this comparison is that the analysis performed using the 1D method does not apply jet energy resolution smearing to the jets, and hence uses an overly optimistic assumption on the dijet invariant mass resolution. The difference in resonance mass resolution with and without jet energy resolution smearing is shown in Figure 8.54 for the 2016 (left) and 2017 (right) signal samples. With JER smearing applied, the resonance mass resolution is worse and the peak broadens by 5%. Figure 8.55 shows the expected limits based on analyses of the 2016 dataset, either using the new 3D fit method, where the jets have been smeared (red line) or not (yellow line), or using the 1D dijet method. We obtain a $\sim 30\%$ improvement with the 3D method. The JER smearing has no effect on the final limits.

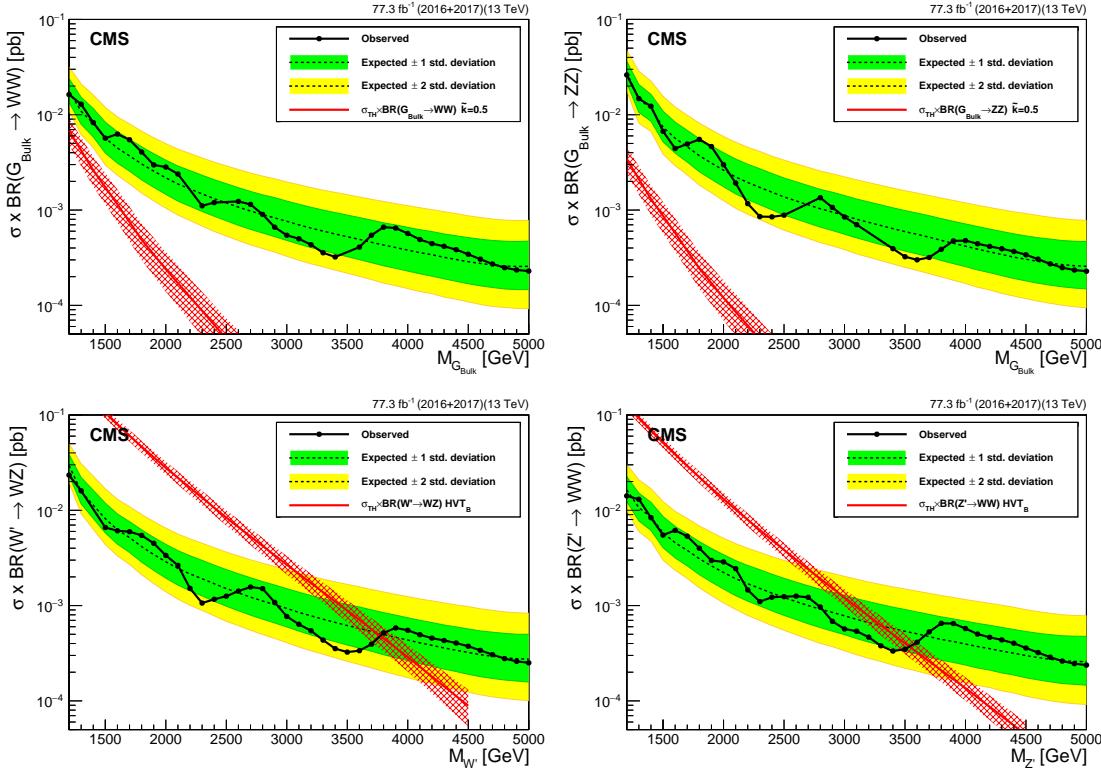


Figure 8.53: Expected limits obtained combining 35.9 fb^{-1} and 41.4 fb^{-1} of data of data after combining all purity categories for Bulk $G \rightarrow WW$ (top left), Bulk $G \rightarrow ZZ$ (top right), $W' \rightarrow WZ$ (bottom left) and $Z' \rightarrow WW$ (bottom right) signals.

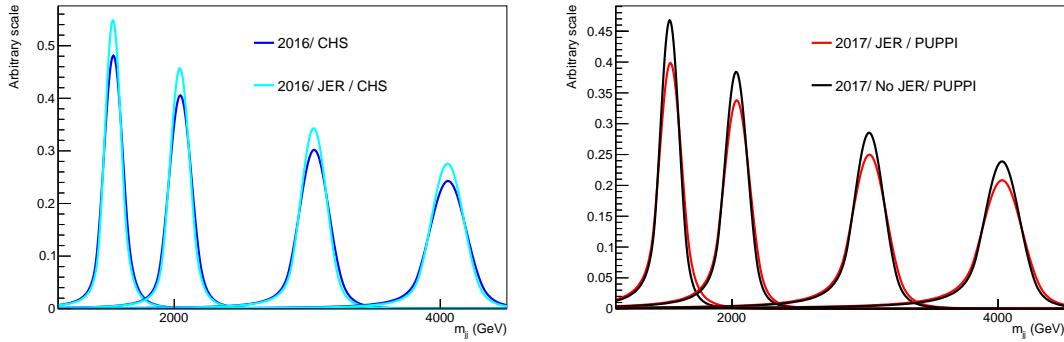


Figure 8.54: Right: A comparison of the resonance mass resolution with (turquoise) and without (blue) jet energy resolution (JER) smearing using 2016 signal samples. Left: A comparison of the resonance mass resolution with (red) and without (black) jet energy resolution (JER) smearing using 2017 signal samples. In both cases, the mass resolution is worse after smearing, and has a roughly 8-10% larger width.

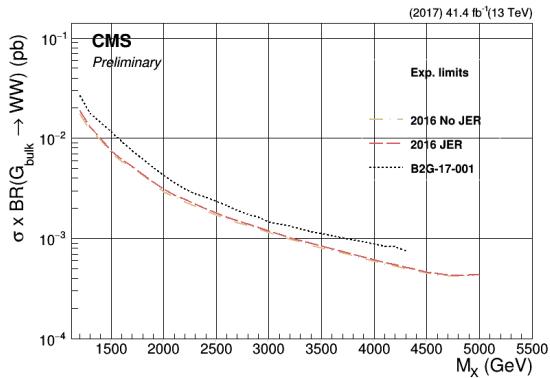


Figure 8.55: Expected limits using 35.9 fb^{-1} of data for a Bulk $G \rightarrow WW$ signal using the 3D fit method, where the jets have had JER smearing applied (red line) or not (yellow line), and using the 1D fit method (dotted black line).

Comparison between datasets

We additionally compare the sensitivity between the 2016 and 2017 datasets. Also here there is one important caveat: for the 2016 analysis, pileup was removed from the jets using the charged hadron subtraction algorithm, whereas in 2017 the PUPPI algorithm was used. In both cases, the softdrop jet mass was obtained from a PUPPI cleaned soft-dropped jet (in the 2016 analysis, the CHS jet was matched to a PUPPI soft-dropped jet in order to obtain the mass. The jet four-vector still relies on CHS). The reason for this switch is that, while for a pileup between 20-30 the CHS and PUPPI algorithms have a similar performance, once the average number of interactions per event is above 30 the PUPPI algorithm has been shown to achieve a better resolution for large-radius jet observables. In 2016 the average PU was 27 whereas in 2017 it was 38, a 40 % increase. However, despite being better in term of groomed jet-mass resolution, PUPPI pileup subtraction has a worse dijet invariant mass resolution than CHS pileup subtracted jets. This is most likely due to PUPPI removing far more (PU) tracks than CHS and, if these PU tracks happen to be associated to a large calorimeter energy deposit, some of the energy might be lost. This creates a tail at lower dijet invariant masses, as shown in Figure 8.56. Figure 8.57 shows

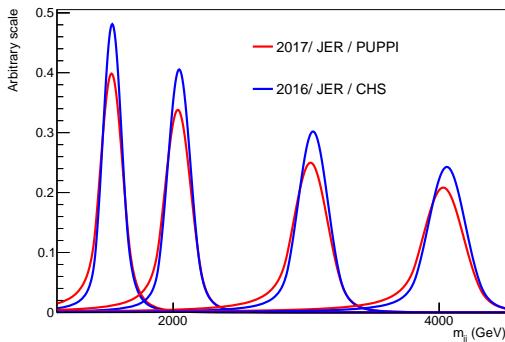


Figure 8.56: A comparison of the resonance mass resolution obtained from 2017 signal samples based on PUPPI jets (red) and 2016 signal samples based on CHS jets (blue).

the expected limits obtained with the new 3D fit method using 35.9 fb^{-1} of 2016 data (beige dashed line), 41.4 fb^{-1} of 2017 data (pink dashed line) and when combining the two yielding a total of 77.3 fb^{-1} of data (dotted black line). These are compared to the results obtained using the 1D fit in Search II (solid purple line). A 35—40% improvement is obtained when combining the two datasets with respect to the individual results, and we observe that the 2017 limits are slightly worse than the 2016 ones due to the above mentioned switch to PUPPI. For the analysis of the full Run 2 data, we will perform a thorough study comparing CHS to PUPPI.

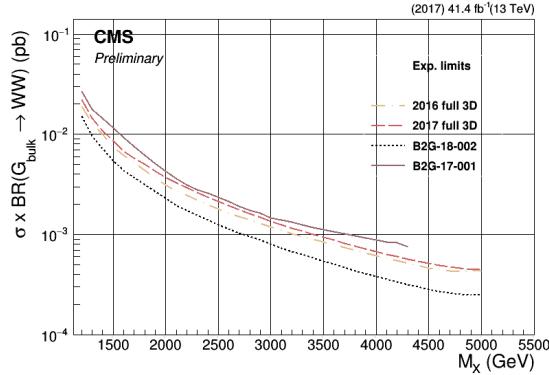


Figure 8.57: Expected limits for a Bulk $G \rightarrow WW$ signal using 35.9 (beige dashed line), 41.4 (pink dashed line), and 77.3 fb^{-1} of data using the 3D fit method (dotted black line), and using 35.9 fb^{-1} of 2016 data and the 1D fit method (solid purple line).

Comparison with ATLAS

Figure 8.58 shows the comparison with the expected limits obtained by the ATLAS collaboration in a similar search for a VV resonance decaying to hadrons [87], for W' and Z' signal hypotheses. For both signal hypotheses we obtain similar, or slightly better results by up to 35 %.

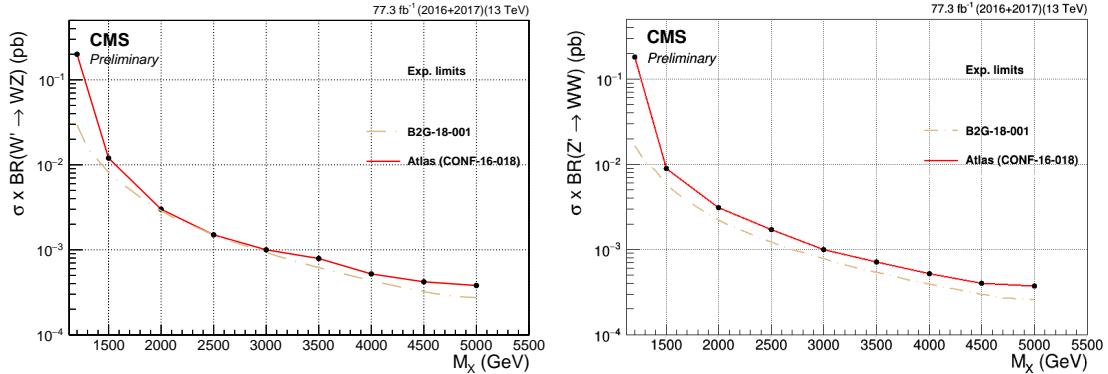


Figure 8.58: Comparison with the expected limits obtained by the ATLAS collaboration in a similar search for a VV resonance decaying to hadrons [87], for W' (left) and Z' (right) signal hypotheses.

Limits per category

In Fig. 8.59 we also show the expected limits for a Bulk $G \rightarrow WW$ signal separately for 2016 and 2017, and for the two HPHP and HPLP categories. As a result of the jet substructure selections

optimization, the HPHP category dominates at low masses while the HPLP adds sensitivity at higher values of the signal mass.

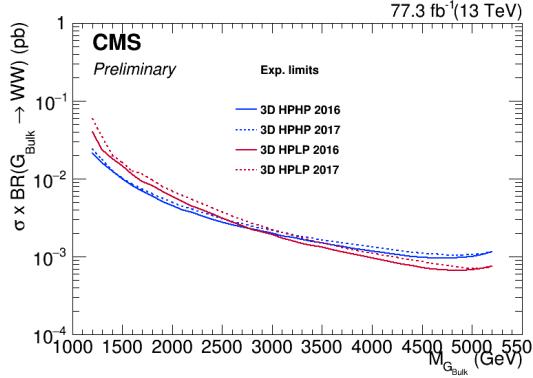


Figure 8.59: Comparison of the sensitivity of the two HPHP and HPLP categories for a Bulk $G \rightarrow WW$ signal hypothesis.

8.10.2 Pulls of nuisance parameters

As summarized in Section 8.8, we add a list of systematic uncertainties to the fit as nuisance parameters. To quantify the difference between the expected values of the systematic uncertainties and the observed values, we compute the pull of the nuisance parameters,

$$p_\theta = (\theta - \theta_{in})/\sigma_\theta, \quad (8.8)$$

where θ_{in} is the pre-fit value of the nuisance parameter under consideration, θ the corresponding parameter post-fit, and σ_θ the post-fit error. The error on p_θ is calculated as the ratio of post- and pre-fit uncertainties. Figure 8.60 shows the pulls after a signal and background fit to the combined 2016+2017 dataset (left) and when fitting the two separately (right), here using a signal hypothesis corresponding to a 2 TeV G_{bulk} . We observe that the softdrop jet mass scale (“CMS_scale_prunedj”) and the resolution (“CMS_res_prunedj”) gets pulled and constrained by the W+jets and Z+jets mass peaks. In addition, the QCD shape parameters (“CMS_VV_JJ_nonRes_”*) are significantly pulled and constrained by data because of their large pre-fit uncertainty and unknown a-priori pre-fit value (again, we do not know if Nature is PYTHIA8, HERWIG++ or MADGRAPH+PYTHIA8. Though from this measurement, HERWIG++ seems to be favored by data (“altshape2”)).

We also evaluate the impacts of each nuisance parameter θ on the signal strength r . This is defined as the shift, Δr , that is induced when θ is fixed and brought to its $+1\sigma$ or -1σ post-fit

values, with all other parameters profiled as normal. This is a measure of the correlation between θ and r , and is useful for determining which nuisance parameters have the largest effect on the signal strength uncertainty. The impacts are shown in Fig. 8.61 evaluated for the final combined fit and a G_{bulk} signal with a mass of 2000 GeV. The most limiting uncertainties are represented by the uncertainty on the W-tagging scale factor together with its pT-dependent extrapolation, and by the uncertainties on the scale and resolution of the reconstructed VV resonance mass.

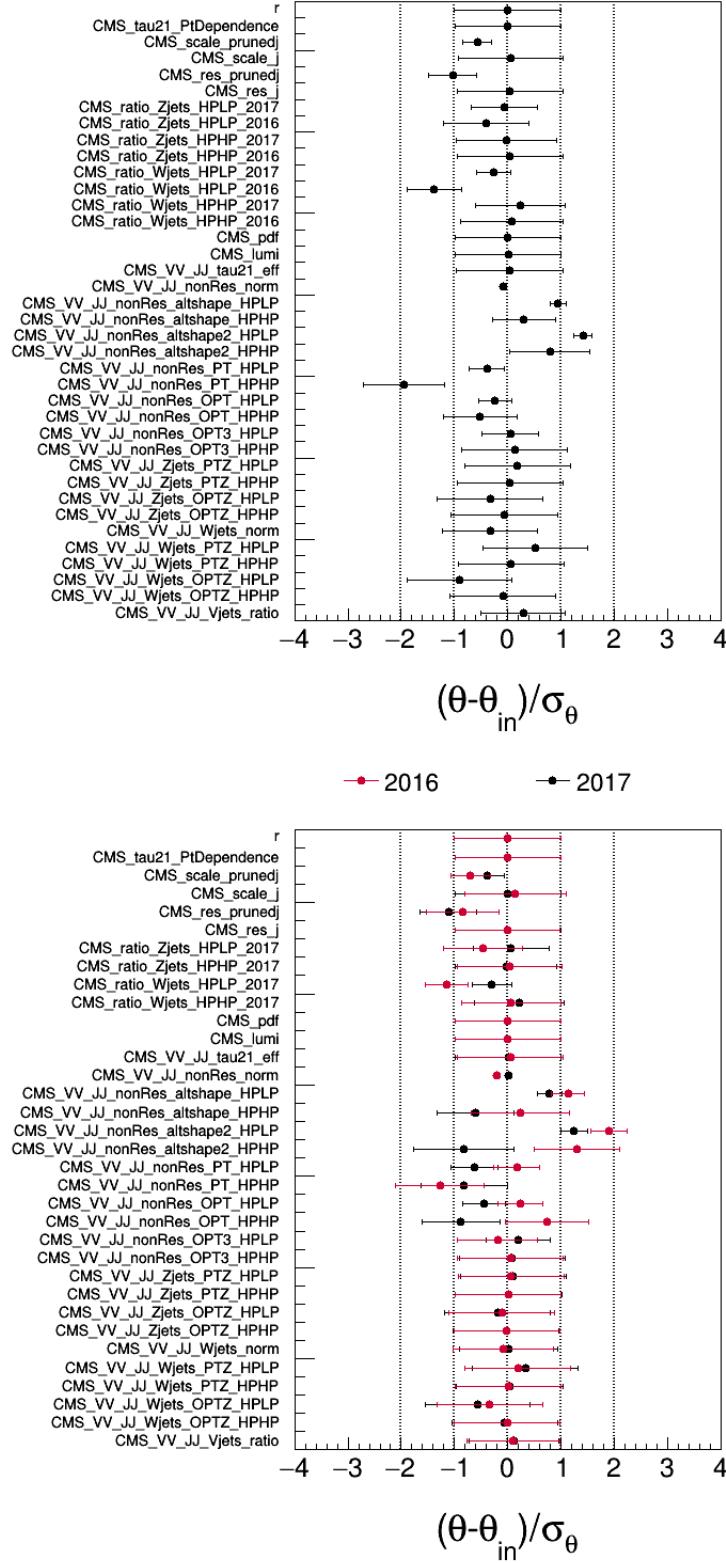


Figure 8.60: Pulls of each nuisance parameter for a combined signal+background fit to the combined 2016+2017 dataset (top) and when fitting the two separately (bottom).

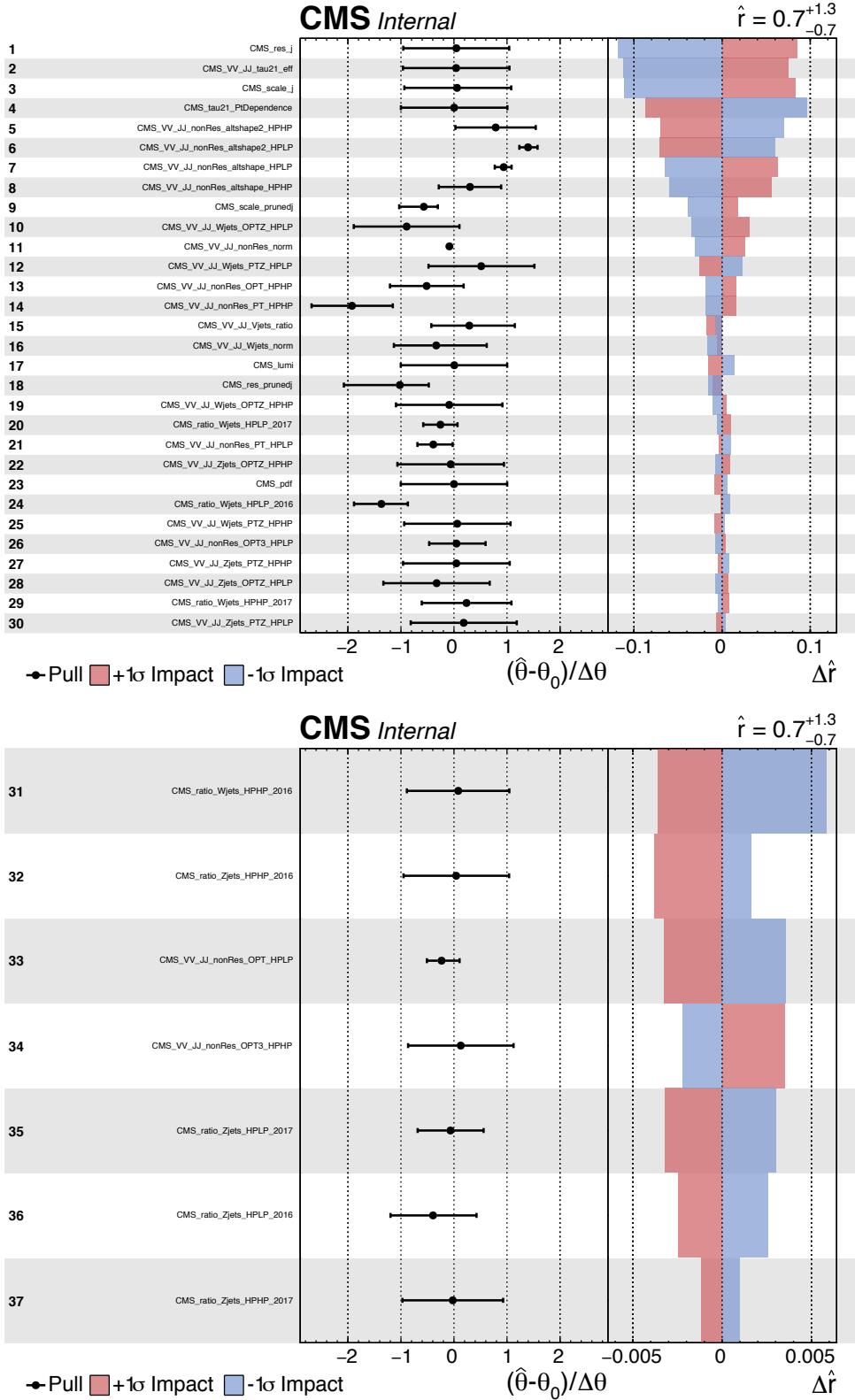


Figure 8.61: Impacts of each nuisance parameter θ on the signal strength r for a G_B signal with a mass of 2000 GeV.

CHAPTER 9

Summary and outlook

In chapters 6, 7, and 8, we have followed the search for VV resonances in the all-hadronic final state through three stages: from being one of the first-ever analyses in the “boosted” final state with a center-of-mass energy of 13 TeV, and the very first CMS analysis to take advantage of substructure at the trigger level, through leading the development of a new W-tagging algorithm and mass corrections now default in CMS, and finally ending with the development of a multi-dimensional fit for generic searches in the groomed jet mass and dijet invariant mass.

Each analysis has built on findings and improvements that came from the analysis before it: the substructure triggers and mass-corrected softdrop jet mass are both used for the 3D fit; the early discovery of the softdrop signal efficiency dependence on p_T led us to derive corrections for it; the observed lack of a constraint on the fit in the dijet invariant mass tail when statistics were low led to us exploring alternate methods that utilize simulation rather than data to create a more robust background prediction in the 3D fit; and the observation of large differences between predictions from different QCD MC generators led us to derive a clever way in which we can incorporate all of these predictions into one single fit. Now the question which remains is: *What comes next?*.

A few ideas were already mentioned in the introduction to Search III, Section 8. The natural next step for this search is an incorporation of the VH(bb) and H(bb)H(bb) searches into the three-dimensional fit. Orthogonality between the three is guaranteed through b-tagging categories, as illustrated in Figure 9.1. This process is already underway, and we are aiming for a publication based on the full Run 2 dataset (data collected in 2016, 2017 and 2018) in one common framework.

Secondly, there will be an increase in center-of-mass energy to 14 TeV in Run 3 (2020-2022). After this, there is not planned an increase in the collision energy. That means that better sensitivity will be achieved only through increases in integrated luminosity or better analysis technique, such as a better procedure for background estimation, a better fit procedure, or a

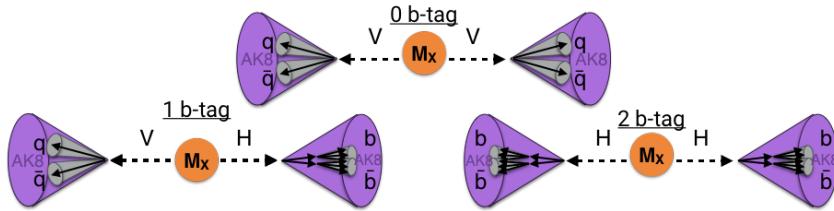


Figure 9.1: The VV , $VH(bb)$, and $H(bb)H(bb)$ analyses can all be incorporated into the multidimensional framework. Orthogonality between the analyses is ensured through b-tagging categories.

better V tagger.

Beyond that, and perhaps more interestingly, is the search for generic resonances peaking anywhere in the jet mass and dijet invariant mass spectrum, where the jets themselves could have other compositions other than two subjets (for instance, a scalar decaying to two vector bosons, whose decay products are merged into one jet, resulting in a 4-prong object). One caveat of the current analysis setup is that it only works for two-pronged signals due to its n-subjettiness cut. In order for the multi-dimensional framework to be truly generic, the τ_{21} tagger needs to be replaced by a generic anti-QCD tagger. Such a tagger works as an anomaly detector by encoding the probability density function for quark or gluon jets as a function of certain variables, variables for which signal jets are assumed to have a different probability density function. In our case, good variables would for instance be groomed jet mass or substructure, as any generic signal is assumed to be peaking in jet mass and have some (unknown) substructure. Such a tagger would not require information about what a potential signal would look like; it would return only the probability of any particular jet being a QCD jet. Taggers like these are usually based on deep neural networks (DNN), where the quark and gluon jet PDFs are obtained through training of the network, as demonstrated in Refs. [88, 89].

In order for such an encoding to work, the DNN needs access to the features distinguishing quark and gluon jets from signal jets without these features being biased towards any signal in particular. The network must learn how to encode “non-substructure”.

As a side project in parallel to working on the multi-dimensional fit, I spent the last half year of my PhD working on a deep neural network capable of discriminating quark and gluon jets from W jets in order to improve the W-tagging performance and improve the search sensitivity for VV analyses to come. Based solely on the four vectors of jet constituents, the idea is to let the neural network itself compute variables based on grooming and substructure, without feeding it any high-level features (like softdrop mass and τ_{21}). This type of architecture is, in addition to improving W-tagging performance, ideal for the purpose described above: encoding QCD in

terms of substructure-like features. The final chapter of this thesis is therefore dedicated to the two last points: how to improve W-tagging in CMS for future analyses and how to design a neural network capable of learning jet substructure in an unbiased way.

Part IV

Encoding jet substructure in a deep neural network

Infusing deep neural networks with physics

The previous chapter ended by mentioning two ingredients that will become important for future searches with the multi-dimensional fit: a better vector-boson tagger, and a generic anti-QCD tagger for signal-independent searches. As a side project during my final PhD semester, I worked on a solution for the first, which has the added benefit of being a stepping stone towards the latter. This is what I will cover in the final chapter of this thesis.

When applying machine learning to particle physics problems, the input has historically consisted of pre-computed high-level features (quantities based on lower-level variables and certain theoretical assumptions). With the rise of deep learning however, computational graphs have achieved an increased capability to find even the smallest correlations in datasets, allowing them to construct complex features on their own. The deep neural network (DNN) I will present in the following is based on the assumption that, given sufficient instructions about the laws of Nature, a neural network should be capable of reconstructing its own high-level features based on lower-level quantities only. In addition, if smartly designed, the network should be capable of finding novel correlations and physical features, *a-priori* unknown, by allocating a physical meaning to the training weights deep within the network. The deep neural network I will present here is trained to discriminate quark and gluon jets from W -jets. However, as I will discuss in the final section of this chapter, it is also the perfect starting point for developing a generic anti-QCD tagger.

The work presented in the following has not been published and still qualifies as work in progress. However, I believe developing taggers such as these is of great importance for future versions of the searches presented here, and it is something I hope to continue working on in the future.

COLA:

$$\begin{bmatrix} E_1 & E_2 \\ P_1^2 & P_2^2 \\ P_1^z & P_2^z \\ P_1^y & P_2^y \end{bmatrix}_{(4,2)}$$

$$\tilde{k}_{M,i} = k_{M,i}, \quad \tilde{k}_{M,j} = \begin{bmatrix} E_1 & E_2 \\ P_1^2 & P_2^2 \\ P_1^z & P_2^z \\ P_1^y & P_2^y \end{bmatrix}_{(4,2)}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}_{(4,4)}$$

$$\begin{bmatrix} E_1 & E_2 \\ P_1^2 & P_2^2 \\ P_1^z & P_2^z \\ P_1^y & P_2^y \end{bmatrix}_{(4,4)}$$

$$\sum E \quad \sum P_1^2 \quad \sum P_1^z \quad \sum P_1^y$$

$$\sum P_2^2 \quad \sum P_2^z \quad \sum P_2^y$$

linear combinations of momenta

“What can we teach the machine?” → “What can we learn from the machine?”.
Work in progress.

CHAPTER 10

A Lorentz invariance based Deep Neural Network for W-tagging

10.1 LoLa

LoLa is a deep neural network architecture which was first introduced for top tagging [90]. It is based on the idea that, given enough information about the laws of Nature, a neural network should be capable of calculating jet substructure observables on its own given only low-level information. The network is designed to discriminate between AK R=0.8 jets originating from W bosons from those originating from quarks or gluons, solely based on the jet constituent four-vectors (variables with little discriminating power on their own) as illustrated in Figure 10.1. Rather than being

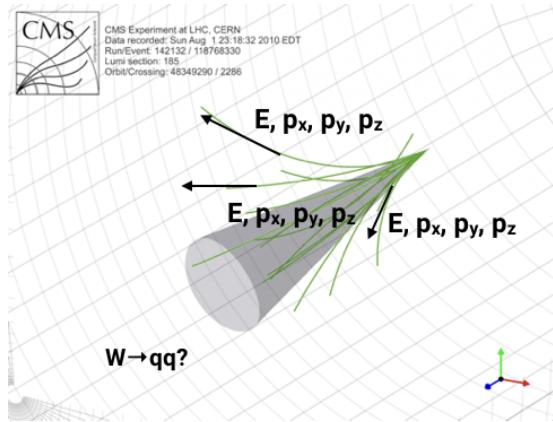


Figure 10.1: LoLa uses only jet constituent four-vectors as input to discriminate W jets from quark and gluon jets.

input high-level features, the neural network is given tools to perform calculations on Lorentz

vectors using the Minkowski metric. Through two novel layers, linear combinations similar to jet clustering and jet substructure algorithms are performed, allowing the algorithm to create its own substructure variables. Additionally, training weights deep within the network correspond to physical quantities reconstructed by the algorithm; distance between particles, masses and energies, linear combinations of particle four-vectors, etc. Besides the end goal of discriminating W bosons from quarks and gluons, one could therefore hope to learn of new correlations separating QCD from vector-boson jets.

10.1.1 Architecture

The LoLa architecture is designed as a four-layer deep, feed-forward sequential network doing supervised learning on fixed-size input vectors. Two novel layers are introduced, the Combination Layer (CoLa) and the Lorentz Layer (LoLa), which perform basic jet clustering and substructure calculations as well as implement the Minkowski metric. These two layers are followed by two fully connected layers, consisting of 100 and 50 nodes respectively, before the final output is computed yielding two output probabilities between 0 and 1 (using a Softmax activation function [91]). The loss function to be minimized is “categorical crossentropy” (or log loss) with two categories corresponding to the probability of a given jet to be W versus non-W like. Only the W-jet probability is stored. The optimizer used in the training is the ADAM optimizer [92], which adapts the learning rate of the model parameters during training. The code itself is written using the Keras [93] interface with a TensorFlow [94] backend. The full architecture with input and output dimension per layer is shown in Figure 10.2. The three first boxes, representing the input, CoLa and LoLa layers, correspond to tensors of rank 2, while the final four boxes correspond to tensors of rank 1. In the following, each layer will be explained in detail.

10.1.2 Input

This algorithm is trained to discriminate between fully merged hadronic W-jets coming from the process $G_{\text{bulk}} \rightarrow WW \rightarrow q\bar{q}q\bar{q}$ (where $M_{G_{\text{bulk}}} = 0.6 - 4.5 \text{ TeV}$), and quark or gluon jets from a QCD sample generated with PYTHIA8. All jets are clustered with the anti- k_T algorithm with a distance parameter of $R=0.8$, with the PUPPI pileup removal algorithm applied. In addition, they are required to have $p_T > 200 \text{ GeV}$ and $|\eta| < 2.5$. Jets are defined as W jets if they are matched to a generator-level, hadronically decaying W boson, with the following matching criteria. The generated vector boson needs to be within $\Delta R < 0.6$ of the jet axis, and the quark decay products need to be within $\Delta R < 0.8$ of the jet axis. The p_T and η distribution of signal and background jets, is shown in Figure 10.3. From these signal and background jets, only the four-vectors of the

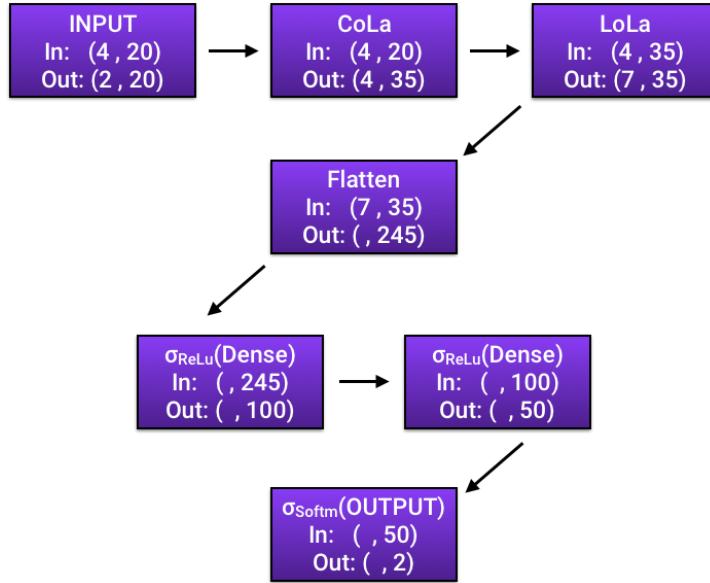


Figure 10.2: The full LoLa architecture. “In” denotes the dimension of the input tensor to the given layer, “Out” is the dimension of the output tensor.

$N = 20$ highest- p_T jet constituents are used as input to the deep neural network: E , p_x , p_y and p_z . I use 20 constituents since any larger number has a negligible effect on the performance, while performance tends to drop once going below 15. The input is therefore a $4 \times (N = 20)$ matrix for each signal and background jet, with one four-vector for each of the 20 jet constituents:

$$x_{\mu,i} = \begin{pmatrix} E^1 & E^2 & \dots & E^N \\ p_x^1 & p_x^2 & \dots & p_x^N \\ p_y^1 & p_y^2 & \dots & p_y^N \\ p_z^1 & p_z^2 & \dots & p_z^N \end{pmatrix}. \quad (10.1)$$

The total number of jet constituents are shown in Figure 10.4, and the input variables (here for all constituents) are shown in Figure 10.5. It is clear that the input variables provide little discriminating power on their own. Therefore, the network must learn how to derive other physical quantities where the signal and background PDFs differ to a larger extent. This is achieved through the two custom layers described in the following.

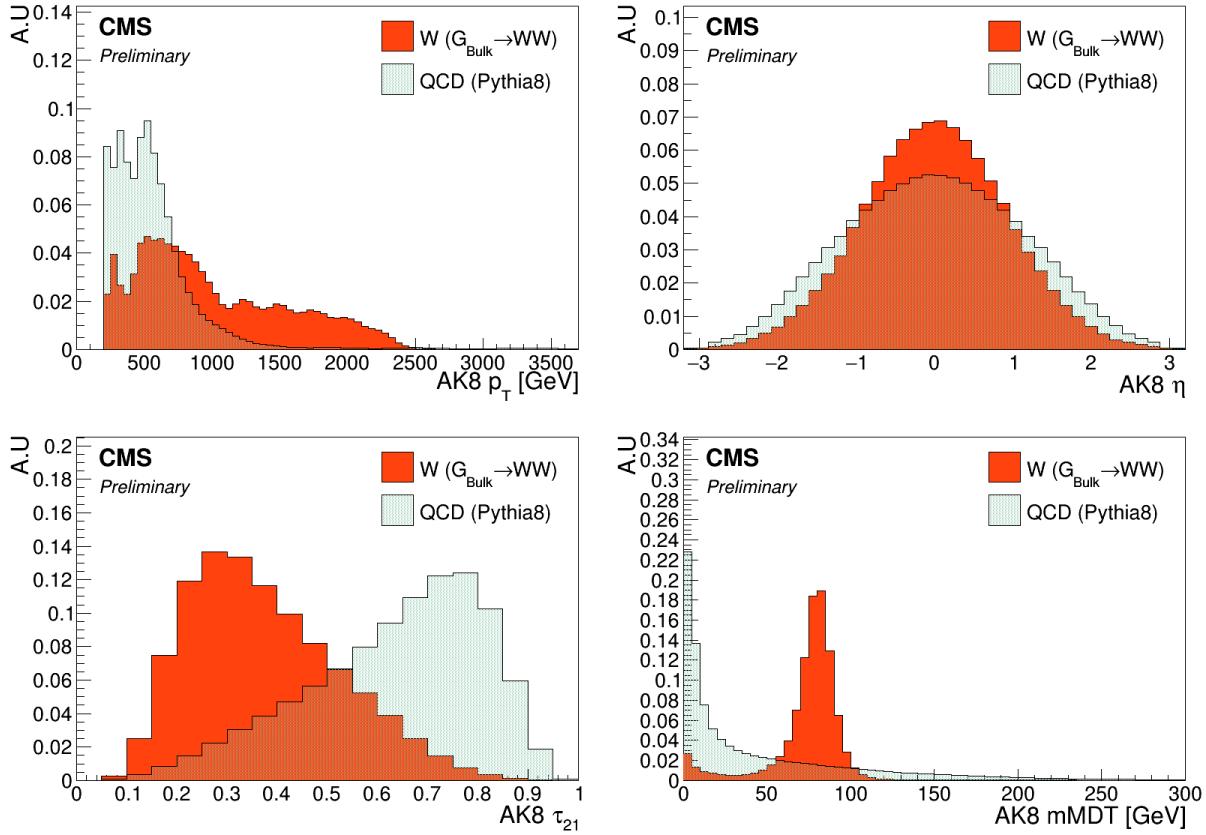


Figure 10.3: Jet p_T (top left), η (top right), τ_{21} (bottom left) and softdrop jet mass (bottom right) for signal and background jets.

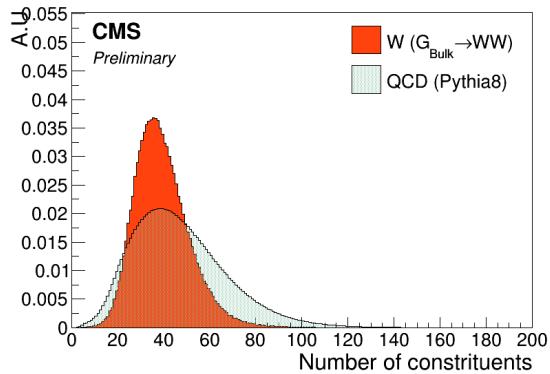


Figure 10.4: The number of jet constituents for signal (red) and background (blue). Only the 20 highest- p_T constituents are used during training.

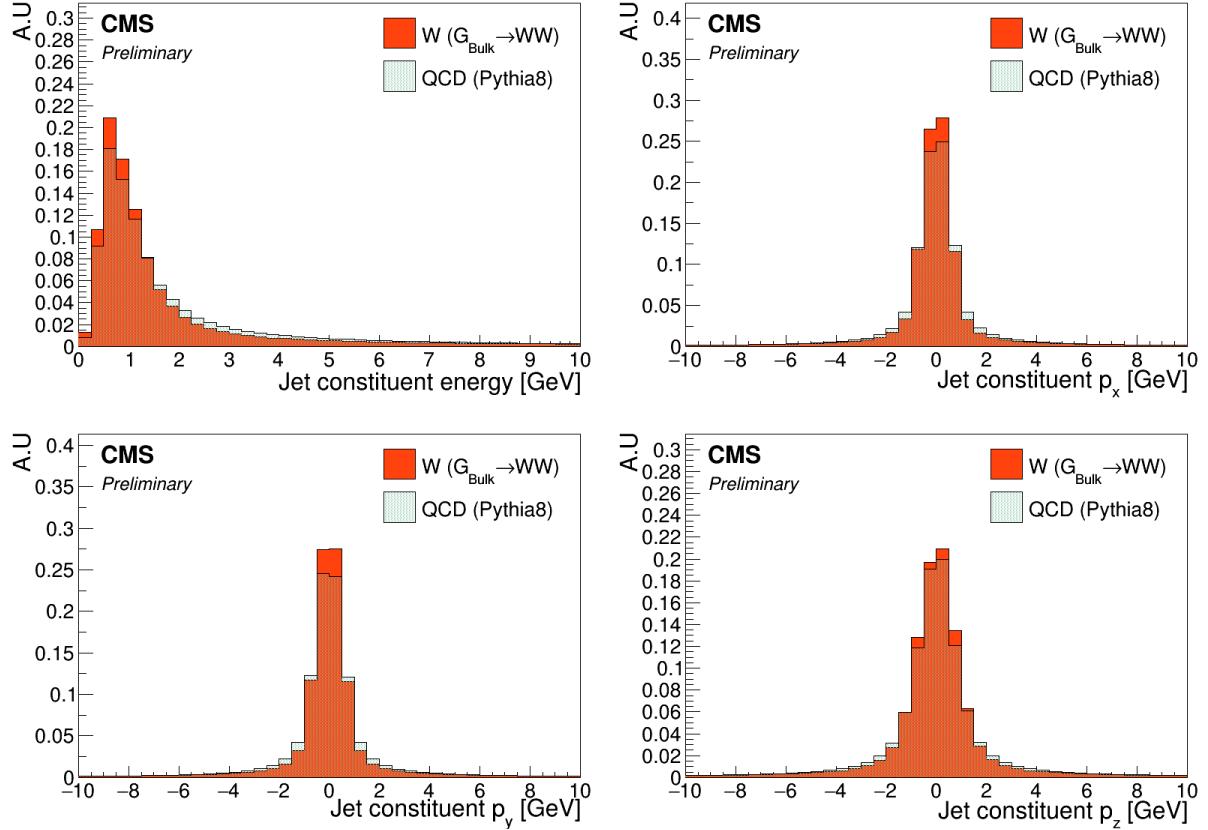


Figure 10.5: Energy (top left), p_x (top right), p_y (bottom left), and p_z (bottom right) for all jet constituents. These values are used as input to the neural network training.

10.1.3 The Combination Layer

The Combination Layer (CoLa) consists of a matrix which, when taking the scalar product with the input matrix, compute linear combinations of the jet constituents, similar to what is done in recombination jet algorithms. The main goal is to create additional four-vectors as input for the next layer. The CoLa matrix is a concatenation of the following: a vector of 1's of length N , the $N \times N$ identity matrix, and a matrix of $N \times M$ trainable weights:

$$C_{i,j} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & w_{1,N+2} & w_{1,N+3} & \dots & w_{1,(N+2)+M} \\ 1 & 0 & 1 & \dots & 0 & w_{2,N+2} & w_{2,N+3} & \dots & w_{2,(N+2)+M} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & 0 & 1 & w_{N,N+2} & w_{N,N+3} & \dots & w_{N,(N+2)+M} \end{pmatrix}. \quad (10.2)$$

When performing the following multiplication

$$x_{\mu,i}^C = x_{\mu,i} C_{i,j}, \quad (10.3)$$

the resulting output matrix will have dimensions $4 \times (1 + N + M)$ and consists of the following: a first column containing the sum of all constituent momenta, the four-momenta of each individual constituent, and $M=14$ different linear combinations of particles with trainable weights. The first corresponds to the neural network computing the four-vector of the “full” jet, at least the full jet in terms of its 20 highest- p_T constituents. The second simply passes each original constituent four-momentum to the next layer. The final, and most interesting part, lets the network construct alternative subjet four-vectors by letting it weigh constituents up and down as it sees fit, in order to reach optimal discrimination power. As an example, lets look at the effect of CoLa in the simple case of only two input jet constituents and two trainable linear combinations:

$$\begin{pmatrix} E^1 & E^2 \\ p_x^1 & p_x^2 \\ p_y^1 & p_y^1 \\ p_z^1 & p_z^2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & w_{1,4} & w_{1,5} \\ 1 & 0 & 1 & w_{2,4} & w_{2,5} \end{pmatrix} = \begin{pmatrix} E^1 + E^2 & E^1 & E^2 & w_{1,4}E^1 + w_{2,4}E^2 & w_{1,5}E^1 + w_{2,5}E^2 \\ p_x^1 + p_x^2 & p_x^1 & p_x^2 & w_{1,4}p_x^1 + w_{2,4}p_x^2 & w_{1,5}p_x^1 + w_{2,5}p_x^2 \\ p_y^1 + p_y^1 & p_y^1 & p_y^1 & w_{1,4}p_y^1 + w_{2,4}p_y^1 & w_{1,5}p_y^1 + w_{2,5}p_y^1 \\ p_z^1 + p_z^2 & p_z^1 & p_z^2 & w_{1,4}p_z^1 + w_{2,4}p_z^2 & w_{1,5}p_z^1 + w_{2,5}p_z^2 \end{pmatrix}.$$

In the two last columns, the neural network makes two “subjet” four-vectors by weighting the relative contribution of each particle as it sees fit. This is similar to jet grooming (Section 5.5.1) or PUPPI pileup subtraction (Section 5.3.2), and should allow the network to learn which constituents are part of the hard scatter and which are not. The $x_{\mu,i}^C$ matrix is finally passed on to the next layer, the Lorentz Layer.

10.1.4 The Lorentz Layer

The Lorentz Layer (LoLa) is responsible for encoding how particles move in space-time through a simple set of rules. Each column (four-vector) of $x_{\mu,i}^C$, is used to compute, and afterwards is replaced by, the following $k = 7$ features:

$$x_{k,i}^L = \begin{pmatrix} m^2(x_{\mu,i}^C) \\ p_T(x_{\mu,i}^C) \\ w_{ij}^E E(x_{\mu,j}^C) \\ w_{ij}^{s1} \sum d^2(x_{\mu,i}^C, x_{\mu,j}^C) \\ w_{ij}^{s2} \sum d^2(x_{\mu,i}^C, x_{\mu,j}^C) \\ w_{ij}^{m1} \min d^2(x_{\mu,i}^C, x_{\mu,j}^C) \\ w_{ij}^{m2} \min d^2(x_{\mu,i}^C, x_{\mu,j}^C) \end{pmatrix} \quad (10.4)$$

Going through from top to bottom, these are:

- the invariant mass and p_T of each four-vector,
- a linear combination of all four-vector energies where each is scaled by a trainable weight,
- two times the sum of distances between the four-vector under consideration and every other column reweighted with a trainable weight, and
- two times the minimum distance between the four-vector under consideration and every other column where each distance again is reweighted by a trainable weight.

The Minkowski metric enters explicitly in the first and in the last four calculations, where the neural network is told to abide by the rules

$$m^2(x_{\mu,i}^C) = g^{\mu\nu} x_{\mu,i}^C x_{\nu,i}^C \quad (10.5)$$

and

$$d^2(x_{\mu,i}^C, x_{\mu,j}^C) = (x_{\mu,i}^C - x_{\mu,j}^C)_\mu g^{\mu\nu} (x_{\mu,i}^C - x_{\mu,j}^C)_\nu \quad (10.6)$$

with $g^{\mu,\nu} = [-1, 1, 1, 1]$, when calculating the invariant mass and distance between all four-vectors produced by CoLa. This tells the neural network to use a space-time geometry in all its calculations and to respect Lorentz invariance. The four final rows of LoLa are the most interesting: Here the network computes quantities similar to n-subjettiness by summing up the distances between all constituents, the jet axis and the subjets produced by CoLa. If, for instance, the network

was able to reconstruct the two hard subjets in the final columns of CoLa, which create linear combinations of particles, it can create its own “ τ_2 ” variable by taking the distance between those subjets and all the jet constituents, and weighting down the column corresponding to the full jet four-vector in column one. Then it can do the same by calculating the distance between the full jet four-vector and all constituents, now weighing down the linear combinations made by CoLa, and compute “ τ_1 ”. The two custom layers, CoLa and LoLa, therefore come together in order to encode jet clustering and substructure in a novel way. They provide the network with the necessary tools in order to create its own physical quantities, through linear combinations with trainable weights, which then again are used to produce other physical quantities with new trainable weights. This allows the network full freedom to explore all interesting particle correlations, where the resulting output features have a physical meaning that can be probed. LoLa turns the question “What can we teach the machine?” around to “What can we learn from the machine?”.

10.2 Performance

The deep neural network is trained on 320,000 signal and background jets for up to 100 epochs, but allow for an early stopping after ten epochs if the loss is stable. The test sample consists of 60,000 W jets and 60,000 quark and gluon jets. To quantify the performance, we look at the signal efficiency versus mistagging rate comparing the performance of LoLa to that of the taggers used previously in this thesis: PUPPI softdrop with τ_{21} and PUPPI softdrop with τ_{21}^{DDT} . The performance of these three different taggers, is shown in Figure 10.6. The point where the blue curves end, represent the signal efficiency for a mass cut of $65 < \text{Softdrop jet mass} < 105$ GeV, here roughly 70%. We clearly see that LoLa performs significantly better than the current

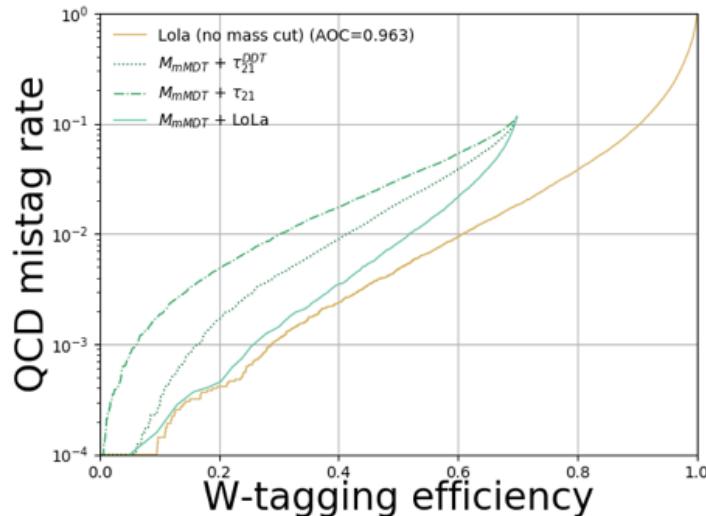


Figure 10.6: Performance of LoLa compared to other W-tagging discriminants in the background-signal efficiency plane: PUPPI softdrop + τ_{21} (dashed blue), PUPPI softdrop + τ_{21}^{DDT} (dotted blue), LoLa with a softdrop mass window applied (solid blue) and the nominal LoLa tagger with no mass cut applied.

baseline W-taggers based on τ_{21} and τ_{21}^{DDT} , with a roughly 20% higher signal efficiency at a given mistagging rate. LoLa also has a higher signal acceptance, as it can be used without a mass window applied. If LoLa were to replace the tagger used in Search II (a better comparison than Search III as the latter uses a rather unconventional mass window), which has a signal efficiency of $\sim 42\%$ at a 2% mistagging rate for a single jet, the signal efficiency for the same mistagging rate would be 65%, a 55% increase. For an analysis requiring two tagged jets, that would imply going from an 18 to a 43% total signal efficiency, a significant gain.

10.3 Dependence on jet mass and p_T

Despite being a key feature, absolute performance is not all that quantifies how good one tagger is compared to another. To use a tagger in a physics analysis, there are three key questions one needs to consider:

- Is the absolute performance better (compared to common methods)?
- Is the tagger p_T -dependent?
- Does the tagger sculpt the mass spectrum?

These three measures are important in quantifying performance and, in the following, I will attempt to explain why this is the case and which approaches are used here in order to tackle them.

Any deep neural network trained to distinguish W jets from quark and gluon jets will naturally learn that p_T and mass are discriminating features unless it is penalized for it. Figure 10.7 shows the LoLa discriminant as a function of jet p_T and softdrop jet mass. A strong correlation between

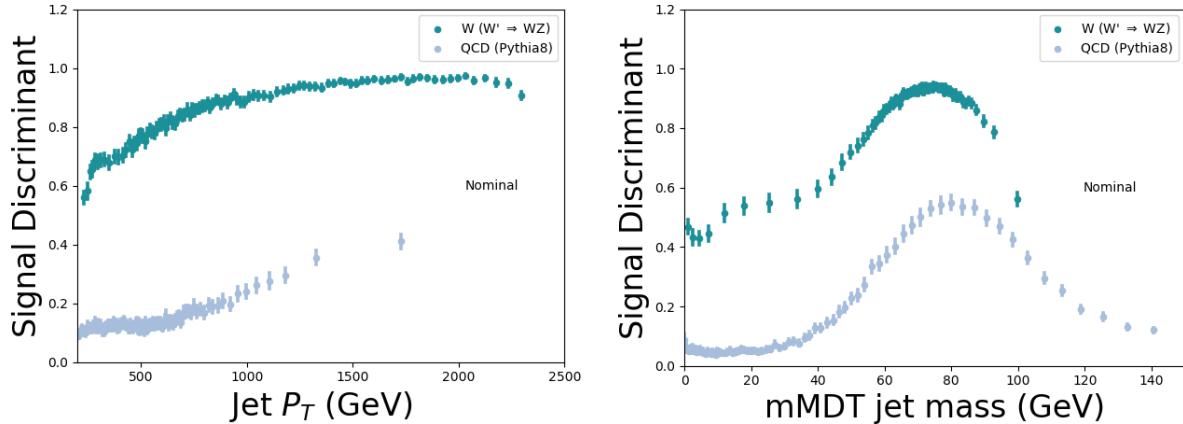


Figure 10.7: The LoLa discriminant as a function of jet p_T (left) and softdrop jet mass (right). A strong correlations with both variables is observed.

the LoLa discriminant and jet mass and p_T , is observed both for signal and for background jets (closer to 1 means more signal like), with a rising slope as a function of p_T (meaning the network interprets a higher jet p_T as being more signal-like) and a bump around the W mass for both signal and for background.

10.3.1 p_T -decorrelation

A tagger which is p_T -dependent is problematic due to the following reasons: firstly, the signal efficiency is variable, which requires a working point that scales with p_T . That in itself is not problematic and can easily be computed. However, it implies that when computing efficiency scale factors from data, a range of different scale factors for different working points is required. In addition, the scale factors are measured at low p_T , a region where the tagging efficiency can be substantially different from the analysis signal region due to the strong p_T -correlation present. Finally, the dijet invariant mass is intrinsically linked to the p_T -spectrum, meaning that any p_T -dependence can introduce sculpting of the dijet invariant mass spectrum. From the top left plot in Figure 10.3, one clearly sees that the jet p_T distribution is very different for signal and for background. In order to avoid that the network learns jet p_T to be a discriminating feature, I therefore compute a jet-by-jet weight intended to flatten the jet p_T spectrum. This weight is passed as a sample weight to the training set, reweighting the contribution from each jet to the total loss (making high-mass QCD jets and low-mass signal jets count more). Figure 10.8 shows the jet p_T -distribution without any p_T -reweighting applied (solid lines) and after applying a p_T -weight (dashed lines). The training is then repeated, this time passing a sample weight with each jet,

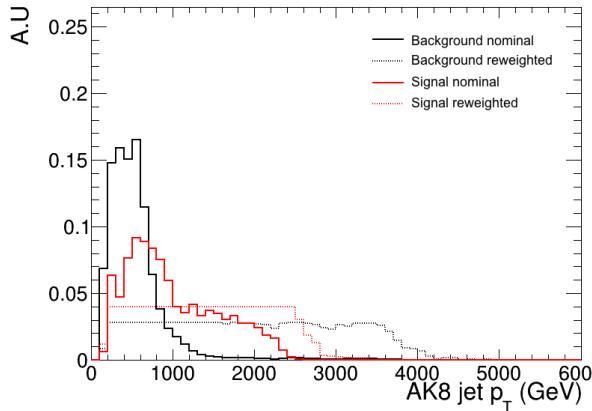


Figure 10.8: Jet p_T distribution before (solid lines) and after (dashed line) applying a weight intended to flatten the jet p_T spectrum.

and the final discriminant compared to the nominal training. Figure 10.9 shows the performance of the same taggers as above but with one additional line, which is LoLa p_T -reweighted. A clear drop in performance is observed, as expected when removing information from the training. However, when we again look at the discriminant output as a function of jet p_T in Figure 10.10, the correlation we observed before has vanished and we are left with a tagger not depending

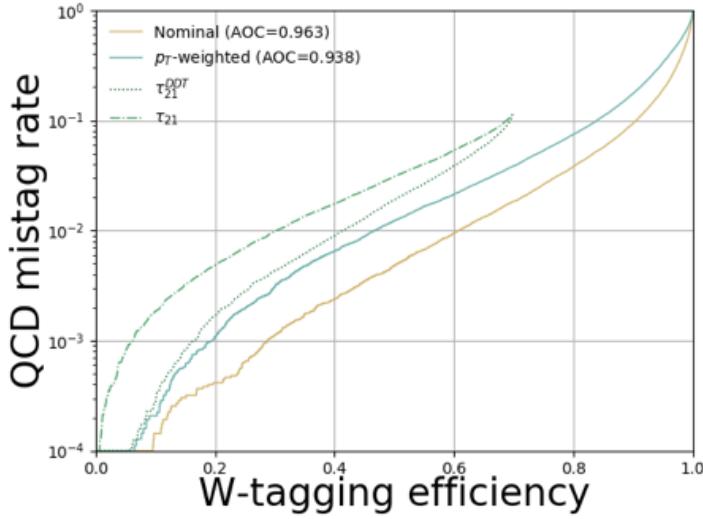


Figure 10.9: Performance of the p_T -reweighted LoLa tagger (solid blue) and the nominal LoLa tagger (solid yellow).

on the jet p_T . For completeness, Figure 10.11 shows the τ_{21} and τ_{21}^{DDT} discriminants versus jet

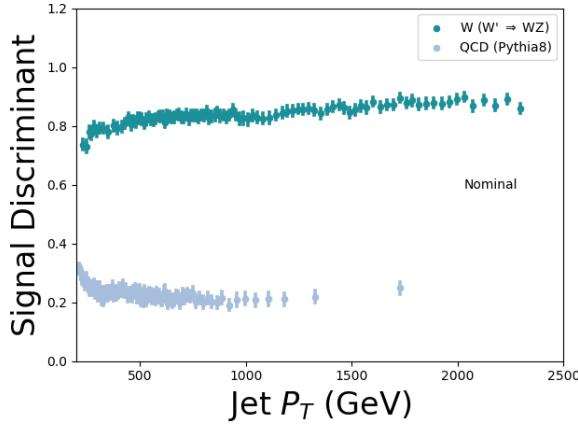


Figure 10.10: The LoLa discriminant as a function of jet p_T after training with a weight intended to flatten the sample p_T spectrum.

p_T . Whereas the nominal LoLa discriminant had a much larger correlation with jet p_T than the τ_{21} -based taggers, the p_T -reweighted version is as decorrelated from p_T as the τ_{21} variables while still exhibiting a better absolute performance than the baseline taggers. In summary, reweighting strategies such as the one described above yield a loss in overall performance, as expected when removing information from the training. However, the p_T dependence of the

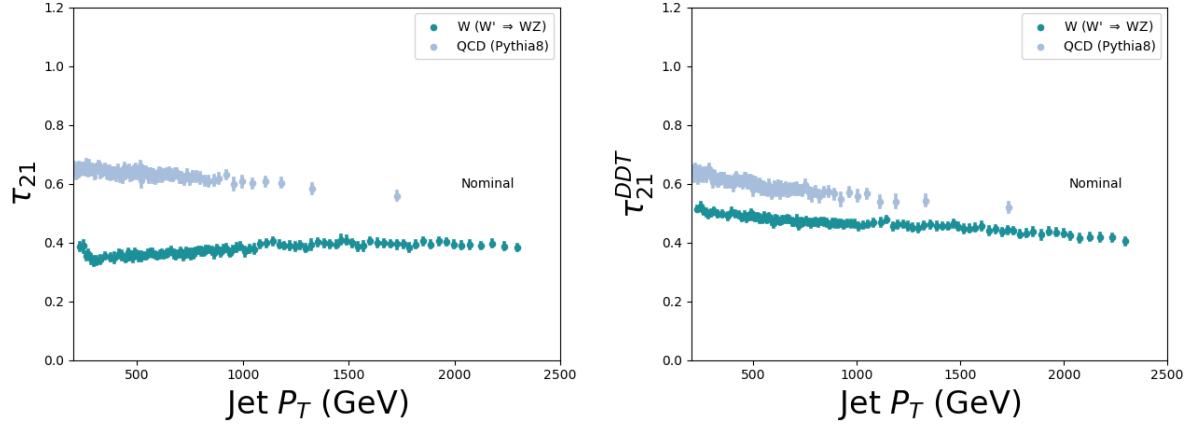


Figure 10.11: The τ_{21} (left) and τ_{21}^{DDT} (right) discriminant as a function of jet p_T .

tagger is strongly reduced, meaning that it might perform better overall in physics analysis when systematic uncertainties are taken into account. There is therefore no clear answer as to which method is better before running a full analysis including systematics for p_T -dependent tagging.

10.3.2 Mass sculpting

Any smart deep neural network intended to separate W bosons from quarks and gluons will inevitably learn the W boson mass as it clearly is very different from the quark or gluon mass. Unfortunately, as these taggers are meant to be used in physics analyses where we often estimate the background in mass sidebands, this has some undesired side effects. If a deep neural network has learned the mass then, after applying a cut on the discriminant, the background jet mass distribution becomes severely sculpted and difficult to constrain.

After applying a cut on the LoLa discriminant corresponding to a 1% mistagging rate, we see in the left plot of Figure 10.12 that the W-jet signal shape is nicely retained. In addition, there are no QCD jets left at low mass so no jet mass window is needed when using this tagger, leading to a significantly higher signal acceptance. However, when looking more closely at the QCD background on the right plot of Figure 10.12, where all histograms are normalized to unit area, we see that the bulk of the remaining 1% QCD jets is right below the W-boson mass peak and has been sculpted to look exactly like the signal. This mass sculpting is in and on its own not a problem, the tagger still manages to get rid of most of the background. However, in many physics analyses, in order to evaluate the background rate in the data signal region, mass sidebands are used. If the background distribution is peaky rather than smoothly falling, the shape and consequently the expected yield is very difficult to constrain. That leads to large uncertainties

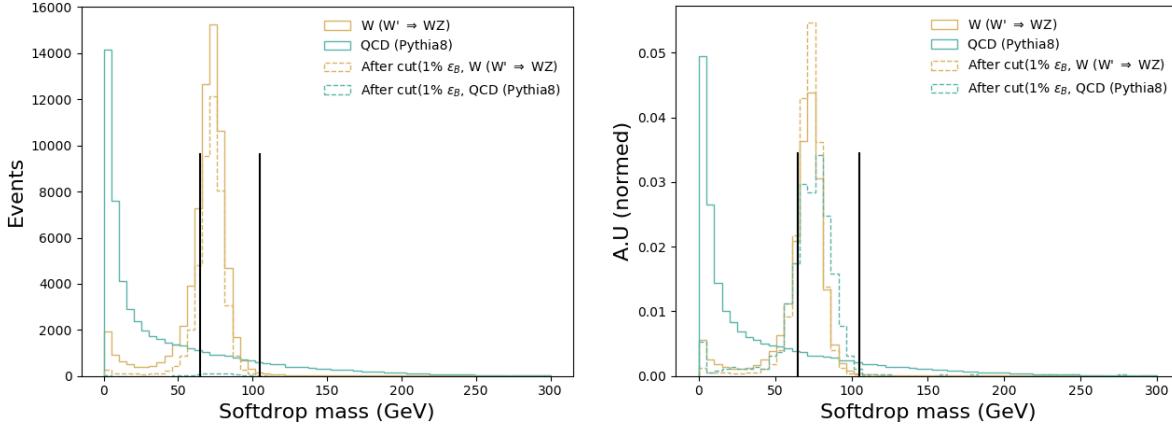


Figure 10.12: The softdrop jet mass distribution before (solid lines) and after (dotted lines) a cut on the LoLa discriminant corresponding to a 1% mistagging rate has been applied. The left plot shows the real number of events left after the cut, the right is normalized to area.

on the background rate and might eventually make an analysis less sensitive than when using a tagger with a worse absolute performance, but reduced mass correlation. In addition, if one were to search for peaks in the softdrop jet mass, as is the case for the multidimensional fit, this becomes increasingly difficult when attempting to fit a potential signal peak on top of a peaking background.

It should again be mentioned, that also for the baseline taggers based on τ_{21} , mass sculpting is a known feature. Figure 10.13 shown the same softdrop jet mass spectrum before and after a cut corresponding to a 1% mistagging rate on τ_{21} (left) and τ_{21}^{DDT} (right). Here τ_{21} clearly exhibits mass sculpting, but is not as peaky as was the case for LoLa. The τ_{21}^{DDT} exhibits the least amount of sculpting, but is also the tagger with the worst absolute performance.

I have not yet had the chance to implement a mass decorrelation strategy for LoLa, but I see two ways going forward: the first is, following the example of what was done to decorrelate LoLa from jet p_T , to pass a mass-dependent sample weight to the training. LoLa would then be trained with a weight derived to flatten the two dimensional jet mass - jet p_T plane. Another option would be to train LoLa together with an adversarial, a dedicated deep neural network running in parallel to LoLa and attempting to learn the jet mass from the LoLa output. The total loss function would then be a sum of the two, where the better the adversarial is in learning the mass, the worse the total loss function gets. Both of these options are something I would like to explore in the future.

In summary, mass- and p_T -dependence are in their own right not a problem for a tagger. The problem occurs when using these taggers in actual physics analyses where background rate

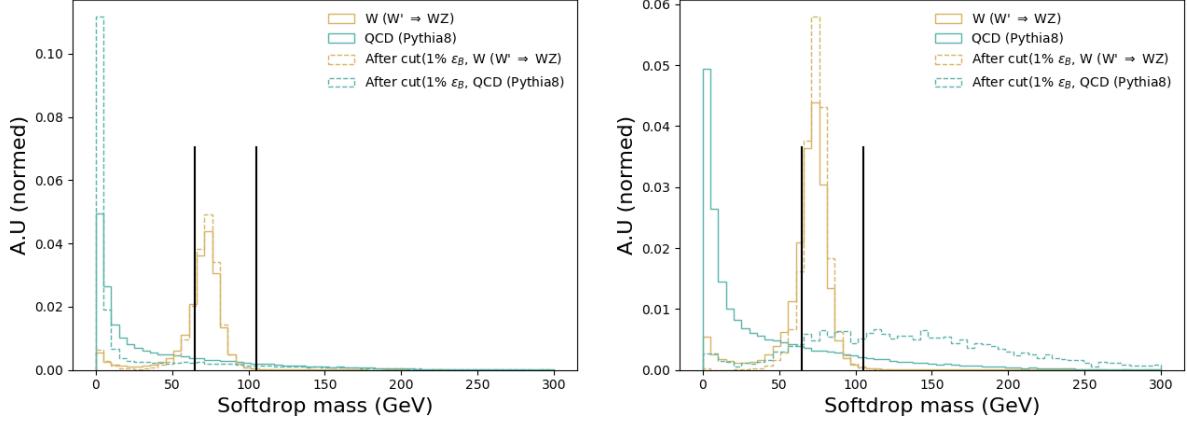


Figure 10.13: The softdrop jet mass distribution before (solid lines) and after (dotted lines) a cut on τ_{21} (left) and τ_{21}^{DDT} (right). All spectra are normalized to unit area.

uncertainties and tagging p_T -dependence uncertainties have a large impact on the final sensitivity. There is a trade-off between signal efficiency and (analysis-dependent) systematics. For LoLa, rather than choosing, I would like to provide two different taggers: a nominal tagger, where no mass/ p_T -decorrelation is attempted, and a decorrelated version. That would allow for both to be tested in a full analysis chain before deciding on which tagger to use when looking at data.

10.4 Validation

LoLa is additionally validated on independent samples as an unbiased measure of performance, allowing the comparison of different CMS W-tagging algorithms. The sample is based on a $Z' \rightarrow WW$ sample with $M_{Z'} = 3\text{ TeV}$ produced with MadGraph and a QCD PYTHIA 8 sample. Here, only jets with $1000\text{ GeV} < p_T < 1400\text{ GeV}$ and $|\eta| < 1.5$ are used. The signal efficiency versus mistagging rate for LoLa compared to the baseline PUPPI softdrop + τ_{21} tagger is shown in Figure 10.14. As was pointed out in Section 10.2, a mass selection is not necessary when using LoLa, but has been added to this plot for completeness. A significant improvement in tagging efficiency is observed for LoLa compared to the default tagger, also when being validated on a sample completely independent from the training sample. A selection corresponding to a 30% signal efficiency is used as a reference working point when studying the tagging performance as a function of jet p_T and pileup in the following, and is marked by triangles in Figure 10.14. The signal efficiency and mistagging rate as a function of jet p_T is shown in Figure 10.15. Again we observe the strong correlation between LoLa tagging efficiency and jet transverse momenta.

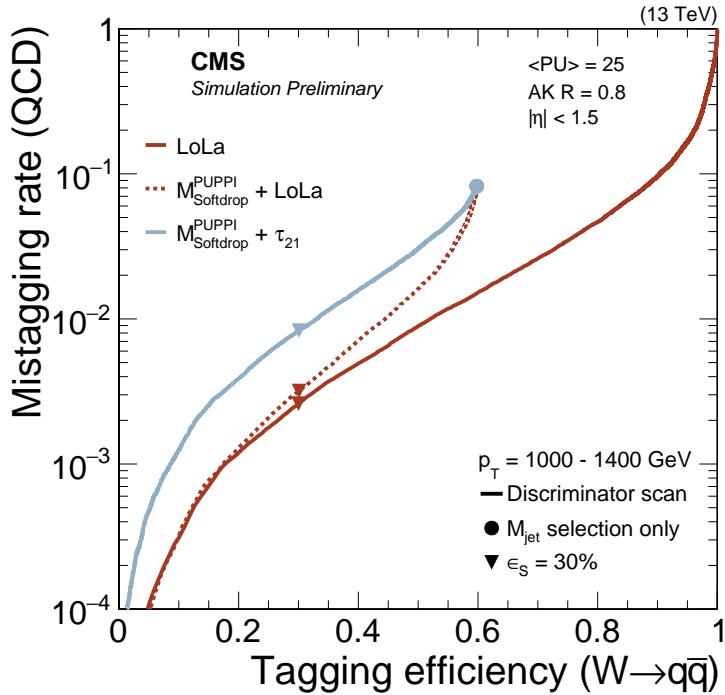


Figure 10.14: Performance of LoLa and PUPPI softdrop + τ_{21} in the background-signal efficiency plane. The PUPPI softdrop jet mass selection of $65 < M_{SD} < 105 \text{ GeV}$, and the 30 percent efficiency points are indicated with symbols.

There is, however, no value for which the τ_{21} tagger has a higher signal over background ratio than LoLa. LoLa performs its worst at very high jet p_T , but in this region the background is very small (dijet invariant masses around 2.5-3 TeV), so the absolute performance here matters less than at lower p_T . Figure 10.16 shows the tagging efficiency and mistagging rate as a function of pileup. Both taggers under study are more or less decorrelated from pileup, with a flat efficiency for up to 50 reconstructed primary vertices. In Run 3, this number is of course expected to be significantly higher and the study should be redone up to a higher number of reconstructed primary vertices.

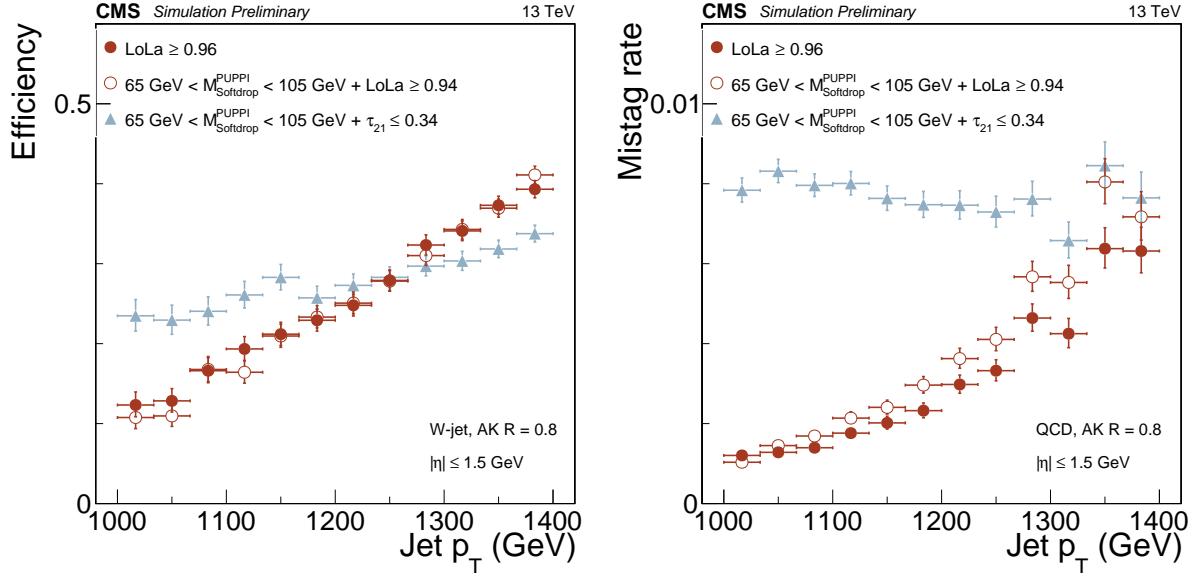


Figure 10.15: Efficiency (left) and mistagging rate (right) of the LoLa selection corresponding to a 30 percent signal efficiency as a function of jet p_T .

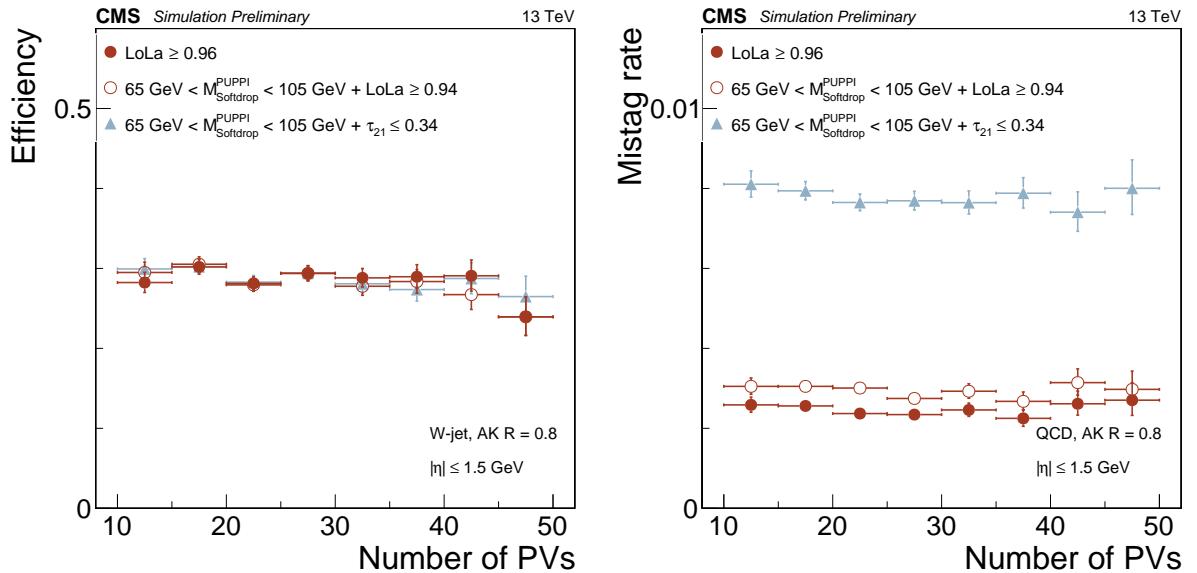


Figure 10.16: Efficiency (left) and mistagging rate (right) of the LoLa selection corresponding to a 30 percent signal efficiency as a function of the number of reconstructed vertices.

10.5 What does LoLa learn?

As a first study of what features LoLa learns, we remove features from the LoLa and CoLa layers. The first thing to investigate is what happens after removing variables sensitive to the jet mass and p_T . This is shown in Figure 10.17. The light blue line, no $\sum p_T$, corresponds to removing the first column of CoLa which passes the sum of all particle four-momenta to LoLa. This has little impact on the total performance, meaning the algorithm does not rely on the jet axis being defined. Removing the Lola variables that correspond to the invariant mass or p_T of the particles/pseudo-jets (green line) reduce the performance significantly, which explains the large drop in tagging efficiency that was observed in Section 10.3.1 when decorrelating the tagger from jet p_T . The worst performance is obtained when removing both the first CoLa column, which roughly corresponds to the jet axis, as well as the invariant mass and p_T LoLa variables (pink line). This implies that a significant part of what LoLa learns is, as suspected, related to the jet mass and p_T . We can also go through the CoLa and LoLa variables and systematically remove

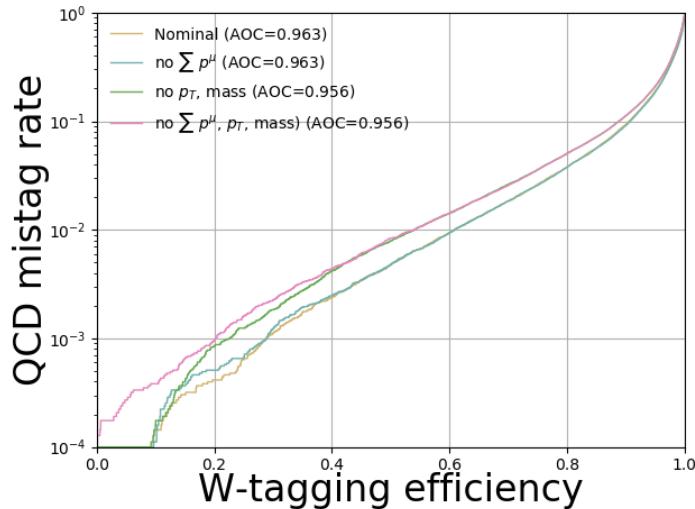


Figure 10.17: Performance after removing jet mass momentum and mass.

one after the other. This is shown in Figure 10.18. The first entry in the legend correspond to the full nominal training (not p_T -weighted) using all features of CoLa and LoLa (blue). The following four entries correspond to removing columns of the CoLa matrix: the first column corresponding to the full jet four-vector (orange), removing the identity matrix which passes through each and every jet constituent (green), allowing the construction of five rather than 15 pseudo-jets (pink) and a combination of all of the above (brown). Removing any of these has very little impact

on the absolute performance as quantified through the area under the curve (AUC) given in parenthesis. This indicates that the CoLa layer has little impact on the absolute performance in terms of tagging efficiency, though it might become important when attempting to train a signal-independent tagger. The next five entries correspond to removing quantities from the LoLa layer: the invariant mass (purple), transverse momentum (gray), energy (red), minimum of all distances (yellow), and the sum of all distances (turquoise). The largest drop in performance is observed when removing the transverse momenta from the training, as was also observed above, whereas removing any other quantity has a minor effect as long as the others are kept. The final legend entry, corresponding to the light green curve, is the performance after removing everything but the sum and minimum of distances between four-vectors. This leads to a large drop in efficiency. Except for the jet p_T , no single variable or CoLa column seems to contain a

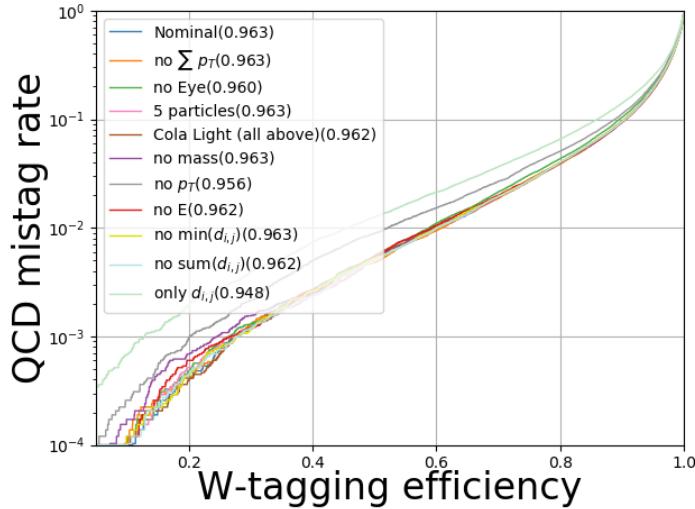


Figure 10.18: Performance after removing components of CoLa or LoLa.

significant amount of information that cannot be replaced by any of the other variables. However, when removing all information except for the distances between four-vectors, a significantly larger performance drop is observed than when removing p_T only. This implies that there is information in LoLa besides the jet p_T that improves the performance, but that this information can be replaced by one of the other variables. Different combinations of LoLa variables therefore need to be studied in the search for a clear performance drop or gain. A complimentary strategy is to look at the LoLa and CoLa weights post-training to check which ones have the largest magnitude, which would in a more direct way quantify the important features of the network. In order to really understand what LoLa learns, further studies are therefore needed and are, as of this writing, currently ongoing.

CHAPTER 11

Summary

In this part of the thesis, we have seen a promising new W-tagging algorithm for future VV searches. Its absolute performance is better than that of the baseline PUPPI softdrop + τ_{21} tagger up to a jet p_T of at least 1400 GeV, roughly corresponding to a dijet invariant mass of 2.5-3 TeV, and could lead to an increase in total signal efficiency from 18 to 43 % for the searches presented here. With a p_T decorrelation method already in place, it could already now be used for the one dimensional VV search presented in Search I and Search II. However, if to be used in the multidimensional search framework, a mass decorrelation method needs to be established. I have already outlined two possibilities of how to achieve this in Section 10.3.2, where one of these has already been shown to work in the context of p_T -decorrelation. This is, as of this writing, left to future studies.

When discussing the future of the multidimensional search, I mentioned how a deep neural network such as the one presented here could be used to encode jet substructure in a way useful in order to develop a generic anti-QCD tagger. This has already been achieved by a parallel analysis team through the use of auto-encoders, published ten days before this writing and documented in [88], and has shown very promising results. However, this strategy is to my knowledge, after discussing with the authors, no longer pursued after observing that the auto-encoder version of LoLa was very difficult to decorrelate from the jet mass. It is my belief that this can be overcome by changing some of the features calculated in the Lorentz Layer (in [88], only the invariant mass is calculated and the other features listed in Equation 10.4 are removed) and this is an idea I would like to pursue in future studies in order to achieve a truly generic search for boosted dijet resonances in the $M_{\text{jet}1}$ - $M_{\text{jet}2}$ - M_{VV} plane.

Part V

Final conclusions and outlook

In this doctoral thesis I have presented three searches for new, heavy resonances decaying to two vector bosons in the all-hadronic final state, using datasets collected by the CMS experiment corresponding to an integrated luminosity of 2.7 (2015), 35.9 (2016) and 77.3 (2016+2017) fb^{-1} . Due to the high energy (“boost”) of the vector bosons, their decay products are so collimated that they get merged into one single jet, leading to a dijet final state topology. Dedicated jet grooming and jet substructure techniques are therefore explored in order to discriminate vector bosons from the overwhelming QCD multijet background.

Each of the analyses presented has provided original contributions to the field: the first search was the first of its kind to ever be performed at $\sqrt{s} = 13 \text{ TeV}$, following an observed excess of $3.4 (1.3) \sigma$ by ATLAS (CMS) in the 8 TeV dataset, and the first time CMS demonstrated the efficiency of using jet-grooming techniques at trigger level. It, at the time, set the most stringent limits to date for the signal scenarios under consideration. The second search introduces a novel pileup-resistant and perturbative-safe vector boson tagging algorithm based on using PUPPI for pileup subtraction, and softdrop for jet grooming, ensuring a high and stable signal efficiency up to a pileup of at least 50 interactions per event. The optimization, validation, and full commissioning of the tagger was performed in the context of this search. Dedicated jet mass corrections, in order to account for an observed p_T and η dependence in PUPPI softdrop jet mass, due to the nature of the softdrop algorithm, were also developed. The tagger based on PUPPI with softdrop, together with the jet mass corrections, became, and still is, the recommended algorithm for W-tagging in CMS.

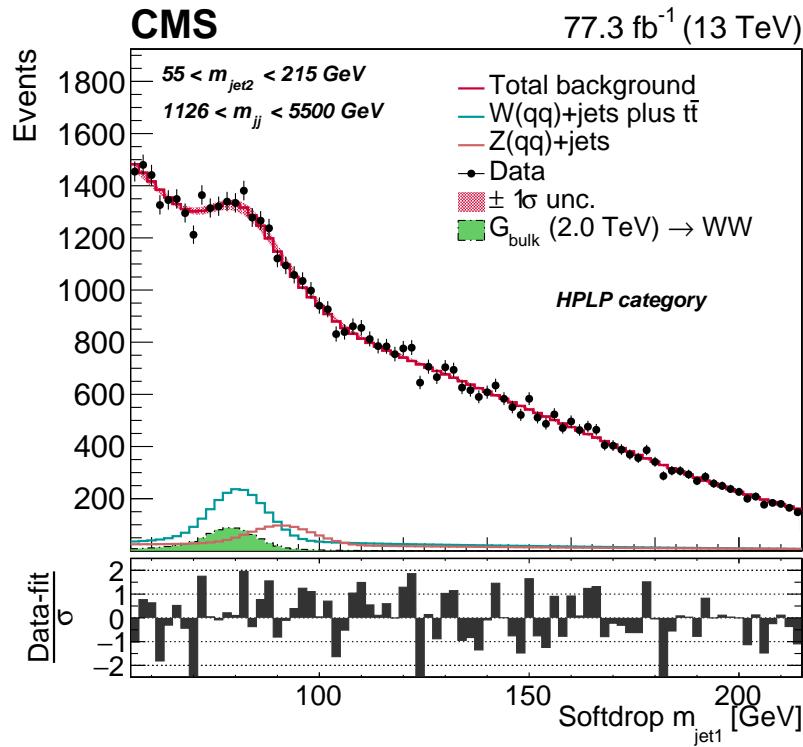
The final analysis introduces a brand new way of doing diboson resonance searches through a three-dimensional fit of the dijet invariant mass and the masses of the two jets. By optimizing the W-tagging algorithm used in this search and due to the nature of the search method, we have, for the first time, been able to constrain the jet mass scale and resolution from a boosted $W(\bar{q}q)$ and $Z(\bar{q}q)$ peak produced by the V+jets standard model process. The method itself leads to a 20-30 % higher sensitivity than the default search method.

The benefit of using a three-dimensional fit based on dijet invariant mass and the masses of the two jets, is that one can look for resonances peaking anywhere in the jet mass and dijet invariant mass spectrum. The natural next step for this search is therefore the incorporation of $VH(bb)$ and $H(bb)H(bb)$ final states into the three-dimensional fit framework, where orthogonality is guaranteed through b-tagging categories. This process is ongoing and planned for the full Run 2 dataset (including the data collected in 2018).

Going further, one can incorporate searches for generic resonances peaking anywhere in the softdrop jet mass and dijet invariant mass spectrum in the multi-dimensional fit, where the jets themselves can have substructure compositions other than two subjets. This type of model-

independent search requires a generic anti-QCD tagger in order to be truly model independent.

In the final chapter of this thesis I presented ongoing work on a W-tagging algorithm using a deep neural network for future searches, capable of more than doubling the analysis signal efficiency by incorporating jet substructure algorithms within the deep layers. As there will be no center-of-mass energy increase after the LHC reaches 14 TeV, achieving the best possible analysis sensitivity for the dataset to come will be of key importance. I also showed how such a deep neural network model is the ideal starting point for building a signal-independent anti-QCD tagger. With $\sim 80 \text{ fb}^{-1}$ of 13 TeV data analyzed and no excess observed, the future of this search therefore lies in increasing the analysis sensitivity through novel taggers, as the one I have presented here, and, making the search strategy as generic as possible through multi-dimensional scans and generic anti-QCD taggers, both of which have a foundation now built within this doctoral work.



APPENDIX A

V -tagging

A.1 mMDT and underlying event

Signal jets groomed with mMDT using a y_{cut} definition, $\min(p_{T,s_1}^2, p_{T,s_2}^2) \Delta R(s_1, s_2) / m_j^2 > y_{cut}$, rather than a z_{cut} definition (as is default in CMS), $\frac{\min(p_{T,i}, p_{T,j})}{p_{T,P}} > z_{cut}$, were found to be far more sensitive to the underlying event than pruned jets [74]. Figure A.1 shows the signal efficiency for pruning (left) and softdrop (right) as a function of jet transverse momenta when including final state radiation (FSR) only, final and initial state radiation (ISR), hadronization, and hadronization with an underlying event component added. At parton level, as well as after hadronization, the

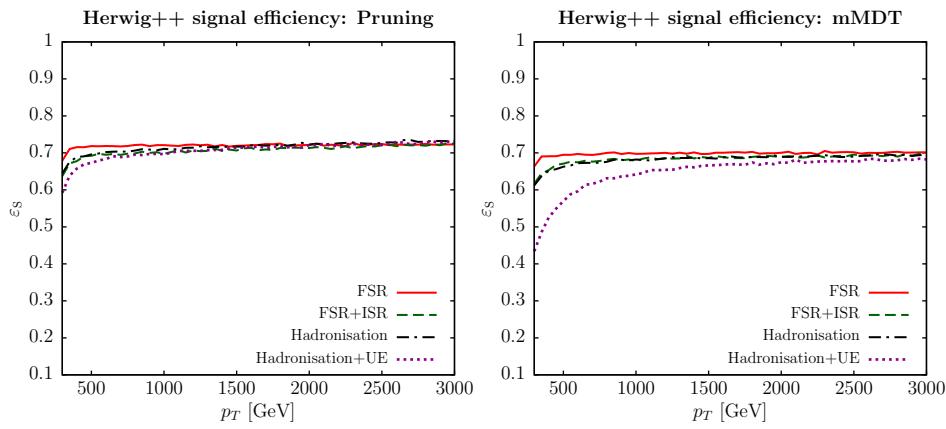


Figure A.1: The signal efficiency for pruning (left) and softdrop (mMDT) (right) as a function of jet p_T when adding FSR, ISR, hadronization and UE. The UE has a severe impact on the softdrop efficiency for signal jets [74].

two algorithms perform very similar as a function of p_T . However, once UE contamination is added, the softdrop tagging efficiency is severely affected. This can be explained by the larger effective radius considered by the mMDT algorithm ($\propto m_V/p_T \sqrt{z_{cut}(1-z_{cut})}$) in comparison to pruning ($\propto m_V/p_T$), as can be seen from Equations 5.9 and 5.8. This observation corresponds very well with the shift in jet mass we observed for softdrop as a function of p_T in Section 6.4.3: as the jet p_T decreases, the mMDT effective radius increases and the jet mass mean shifts to higher values, due to absorbing more background radiation.

A.2 N-subjettiness on groomed jets

In CMS, n-subjettiness is by default calculated on un-groomed jets. A proposal was made in Ref. [95] to, when calculating n-subjettiness, calculate τ_2 on softdrop $\beta = 2$ jets (less aggressive groomer) τ_1 on softdrop $\beta = 0$ jets (more aggressive groomer). This would emphasize the jet hard substructure when calculating 1-subjettiness and the full color radiation pattern when calculating 2-subjettiness. In the above mentioned reference, this was shown to increase the signal significance by 25% (at 2 TeV) for a given mistagging rate, as shown in Figure A.5 and also reduce non-perturbative effects by a factor of 2 to 3. In this section, we study the performance

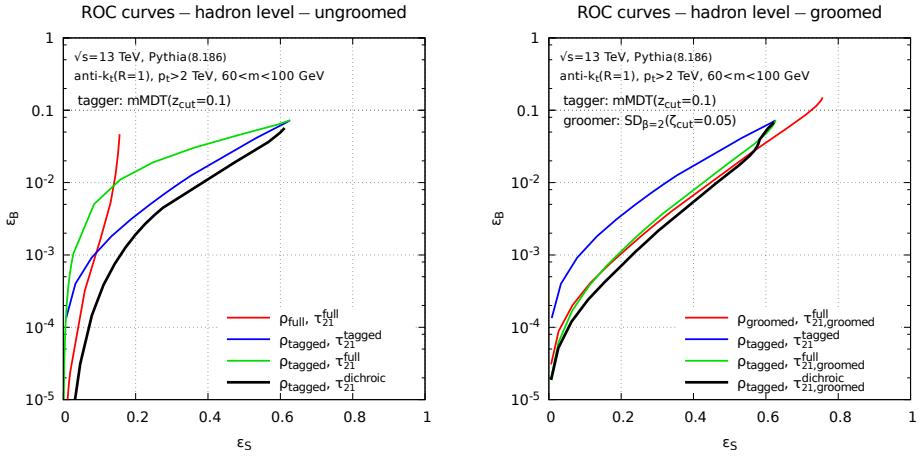


Figure A.2: Using a dichroic tagger has been to increase the signal significance by 25% at a jet p_T of 2 TeV and reduce non-perturbative effects by a factor of 2 to 3 [95].

of n-subjettiness when used in combination with different versions of softdrop, varying the β -parameter. Different values of β results in the algorithm removing more or less radiation. In this section we study three groomer definitions: mMDT (softdrop with $\beta = 0$), softdrop with $\beta = 2$ and the un-groomed (plain) mass. The mass distribution after applying these algorithms is shown in Figure A.3. The current CMS default for V-tagging at the time of this study was softdrop $\beta = 0$ and τ_{21} , where τ_{21} was calculated on the un-groomed jet. N-subjettiness can be calculated starting from un-groomed jets or jets groomed with different algorithms. Going from the most to the least aggressive, the ones studies here are τ_{21} on softdrop $\beta = 0$ jets, τ_{21} on softdrop $\beta = 2$ jets, dichroic τ_{21} (τ_2 from softdrop $\beta = 2$ and τ_1 from softdrop $\beta = 0$ jets) or τ_{21} calculated on un-groomed jets. The distributions of these are shown in Figure A.4. The performance in the signal efficiency versus mistagging rate plane for different versions of τ_{21} , is shown in Figure A.5. A cut on the softdrop $\beta = 0$ jet-mass of 65-105 GeV is applied by default

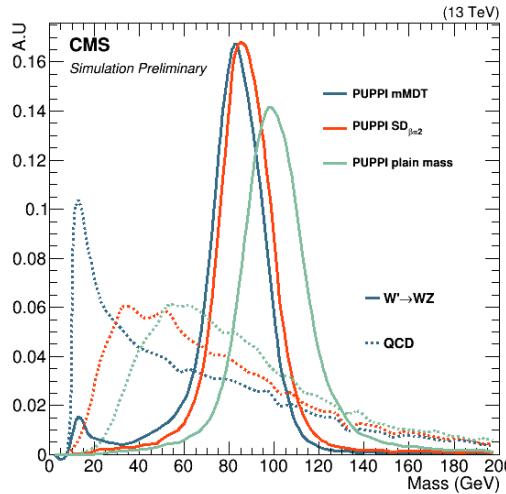


Figure A.3: The bare PUPPI (turquoise), softdrop $\beta = 0$ (blue), and softdrop $\beta = 2$ (red) jet-mass.

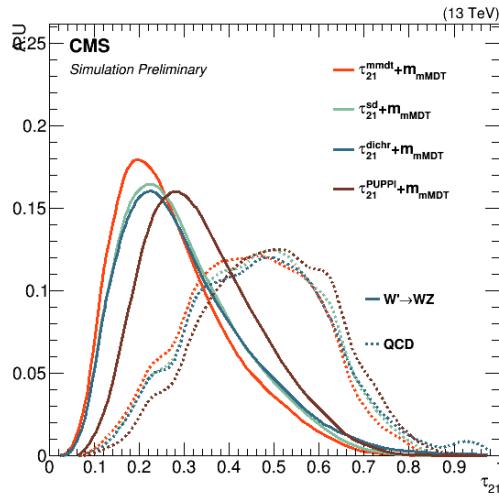


Figure A.4: N-subjettiness distributions after the algorithm has been applied to un-groomed jets (brown, CMS default), or on jets groomed with softdrop $\beta = 0$ (red), softdrop $\beta = 2$ (turquoise) or a mix of both, so-called dichroic (blue), where softdrop $\beta = 0$ is used for τ_1 and softdrop $\beta = 2$ is used for τ_2 .

for all taggers. The CMS default tagger, τ_{21} in combination with softdrop $\beta = 0$ has the lowest signal efficiency for a given mistagging rate. The best performance is achieved when calculating n-subjettiness on softdrop $\beta = 0$ groomed jets. No huge gain in performance is observed when using a dichroic tagger.

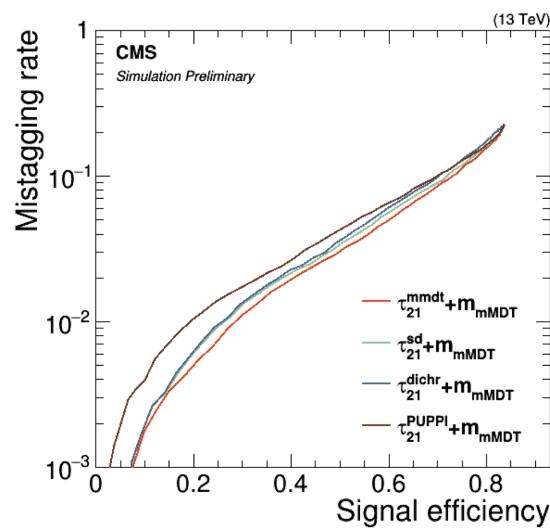


Figure A.5: Performance of a tagger consisting of a selection on the softdrop $\beta = 0$ jet mass of 65-105 GeV and a cut on n-subjettiness computed on groomed or un-groomed jets

APPENDIX B

Search I: Limits per mass category

B.1 Limits per mass category

The asymptotic limits obtained with 2.6 fb^{-1} of 13 TeV CMS data per mass and purity category.

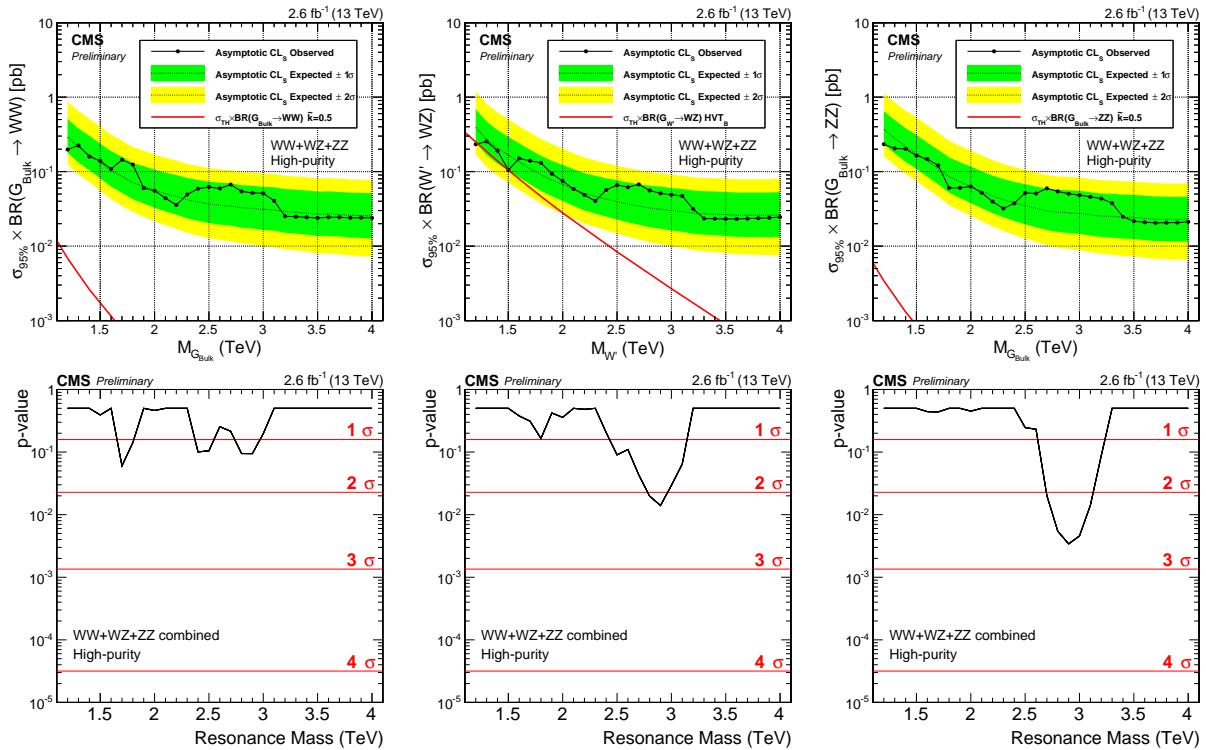


Figure B.1: Expected and observed limits at 95% CL and corresponding p-values obtained in the high purity category using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ (left), $W' \rightarrow WZ$ (middle) and $G \rightarrow ZZ$ (right) signal.

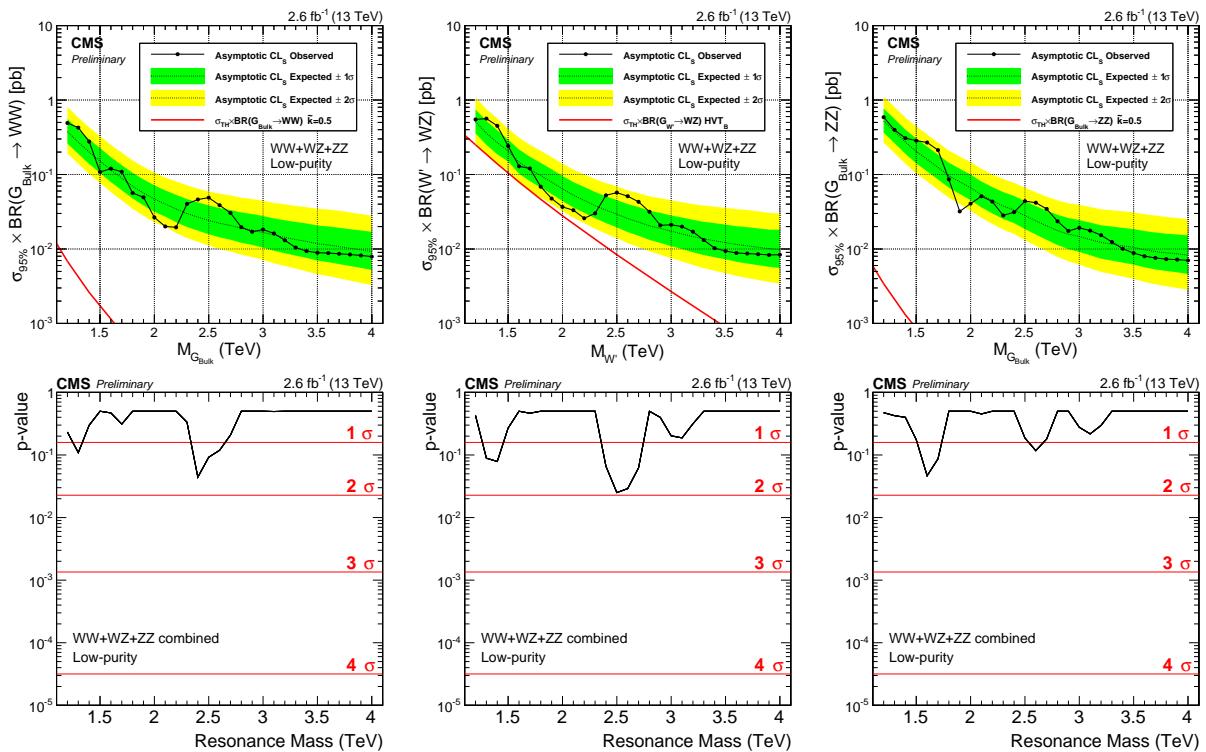


Figure B.2: Expected and observed limits at 95% CL and corresponding p-values obtained in the low purity category using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ (left), $W' \rightarrow WZ$ (middle) and $G \rightarrow ZZ$ (right) signal.

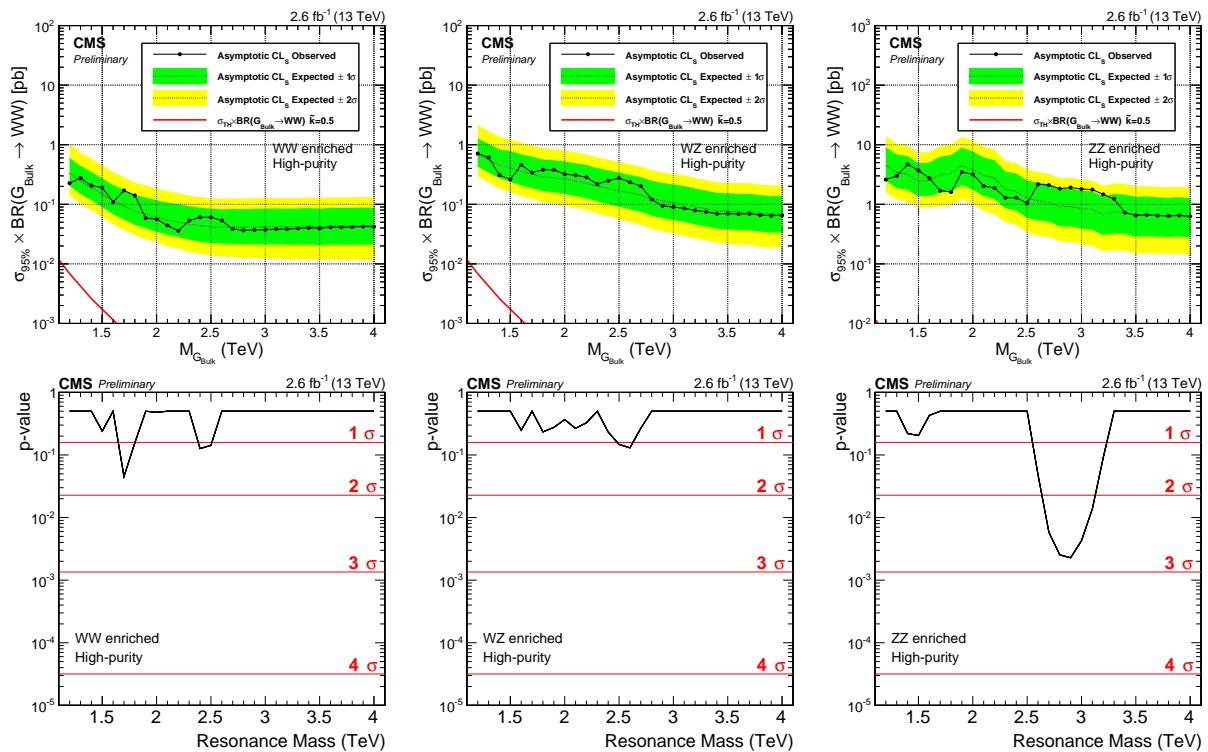


Figure B.3: Expected and observed limits at 95% CL and corresponding p-values obtained for the different mass categories using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ signal in the HP category

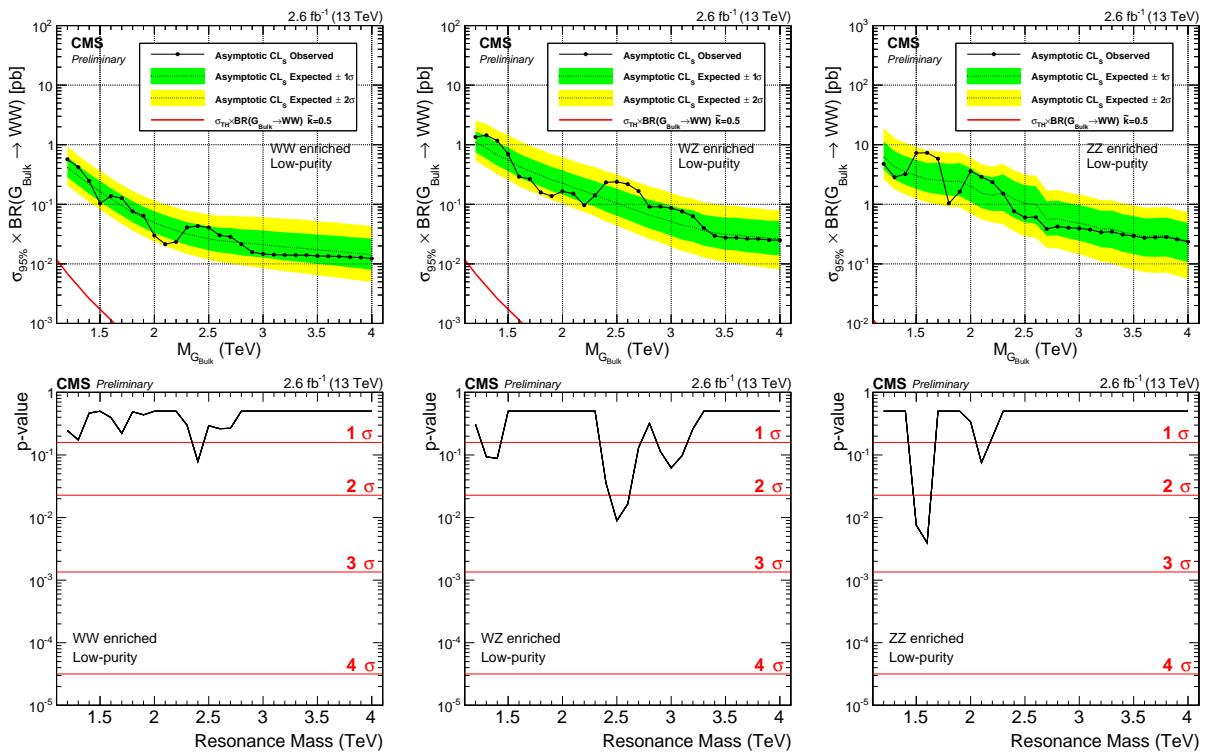


Figure B.4: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ signal in the LP category.

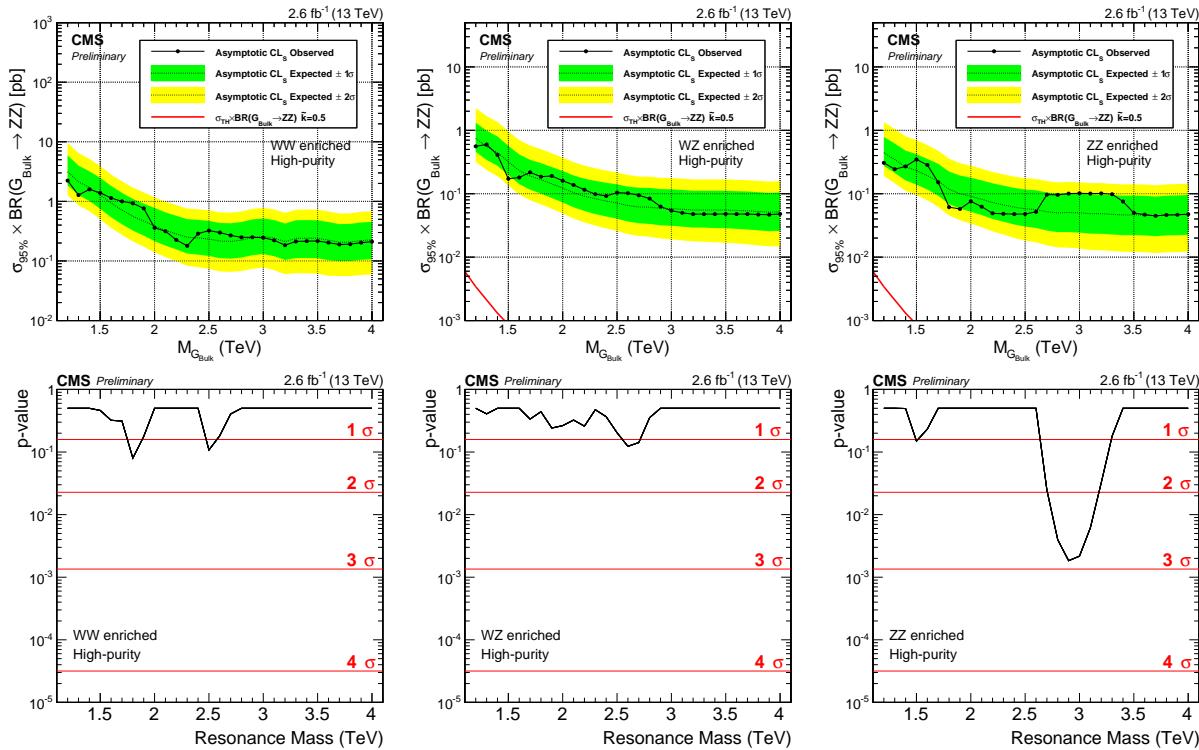


Figure B.5: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories. Here for a $G \rightarrow ZZ$ signal in the HP category.

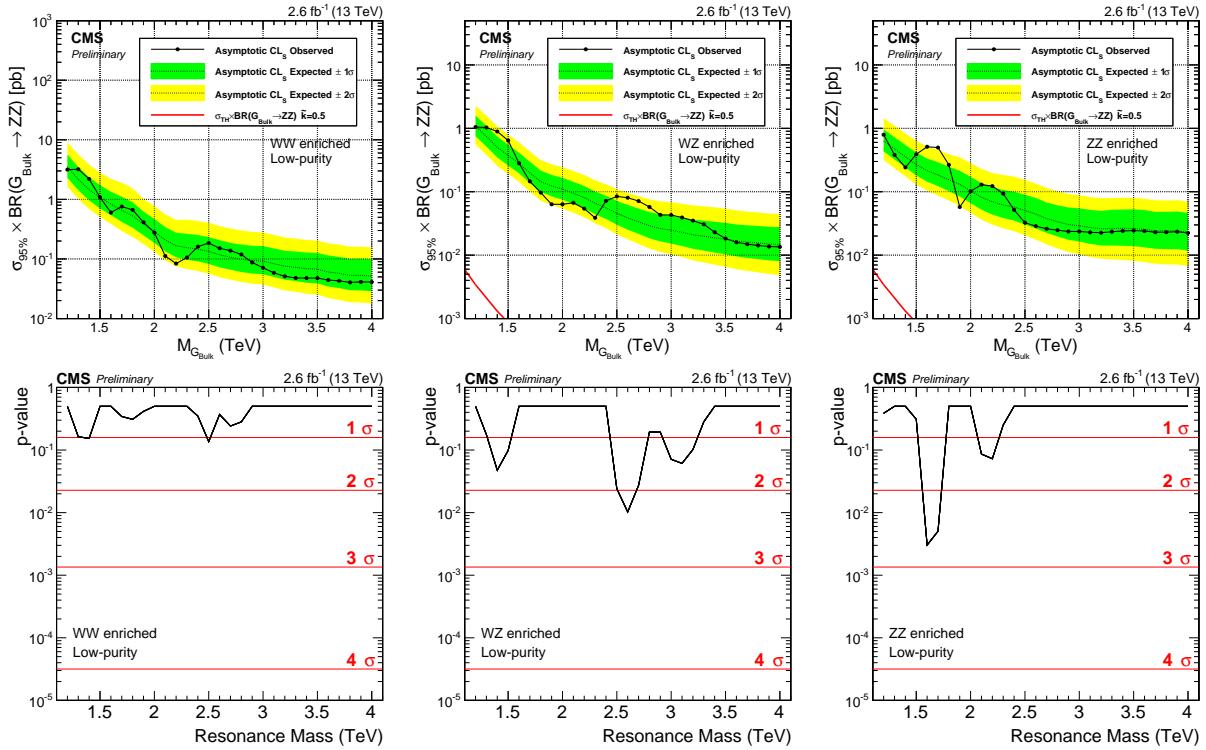


Figure B.6: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories. Here for a $G \rightarrow ZZ$ signal in the LP category.

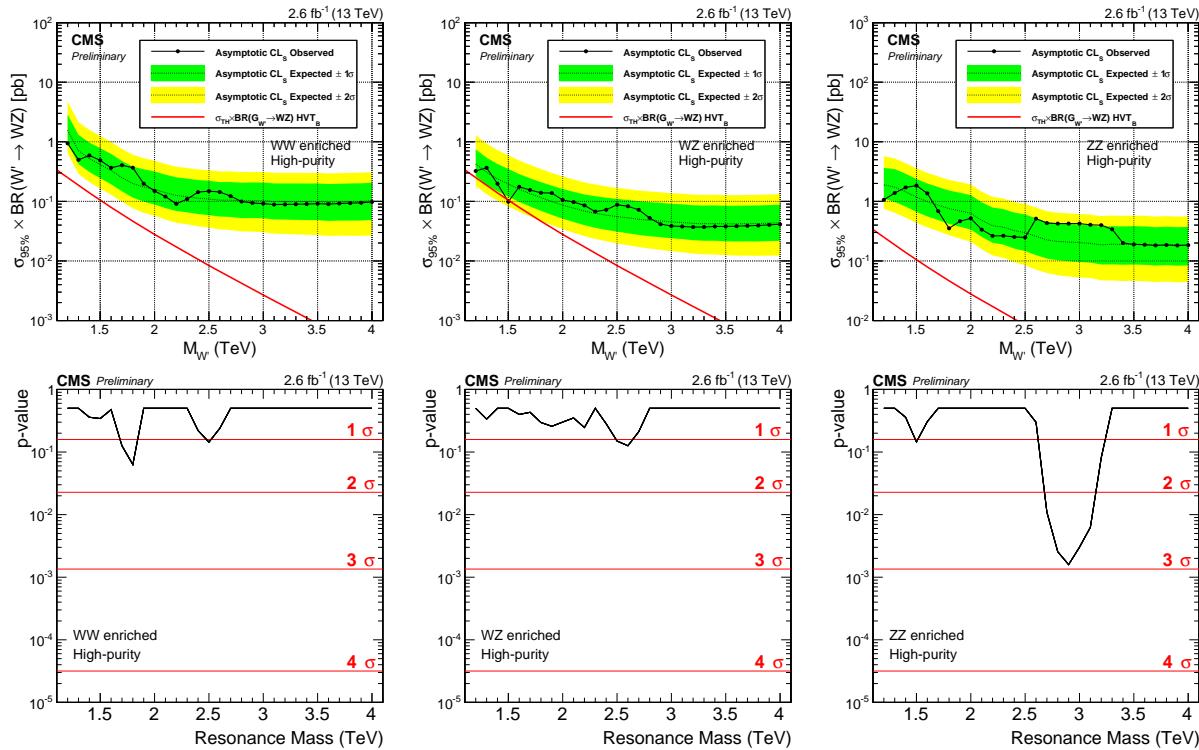


Figure B.7: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories. Here for a $W' \rightarrow WZ$ signal in the high-purity category.

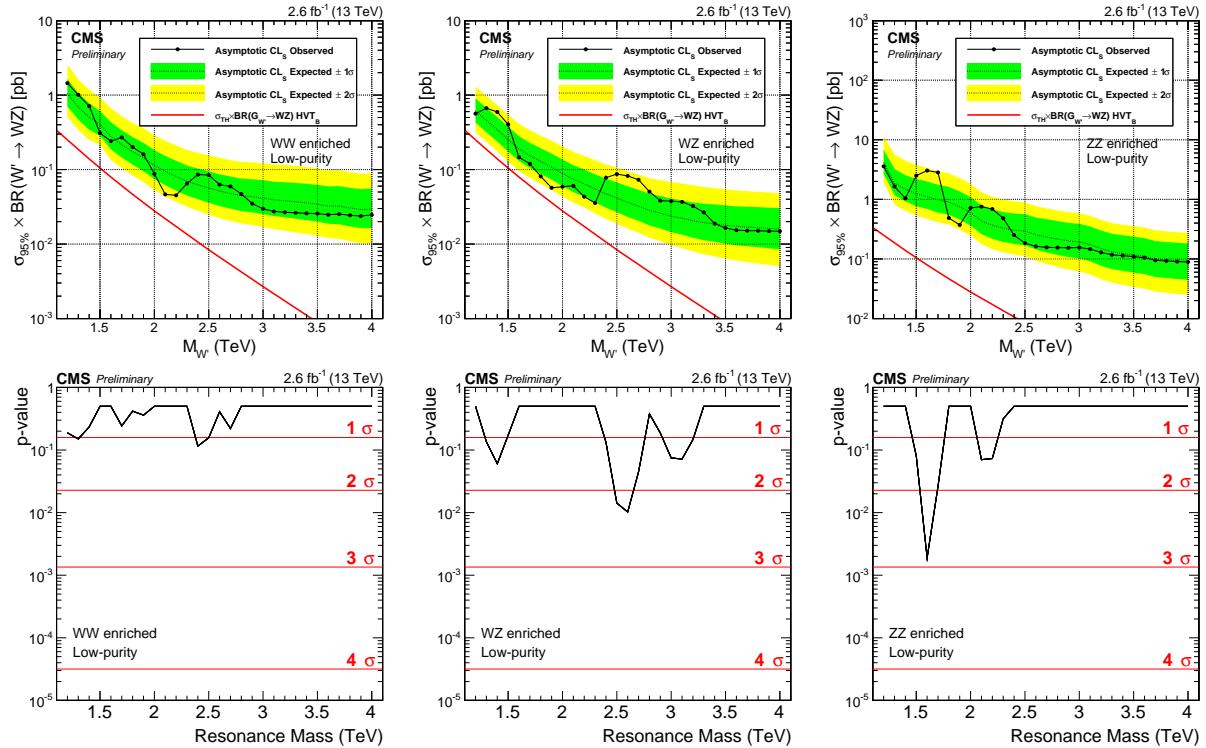


Figure B.8: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories. Here for a $W' \rightarrow WZ$ signal in the low purity category.

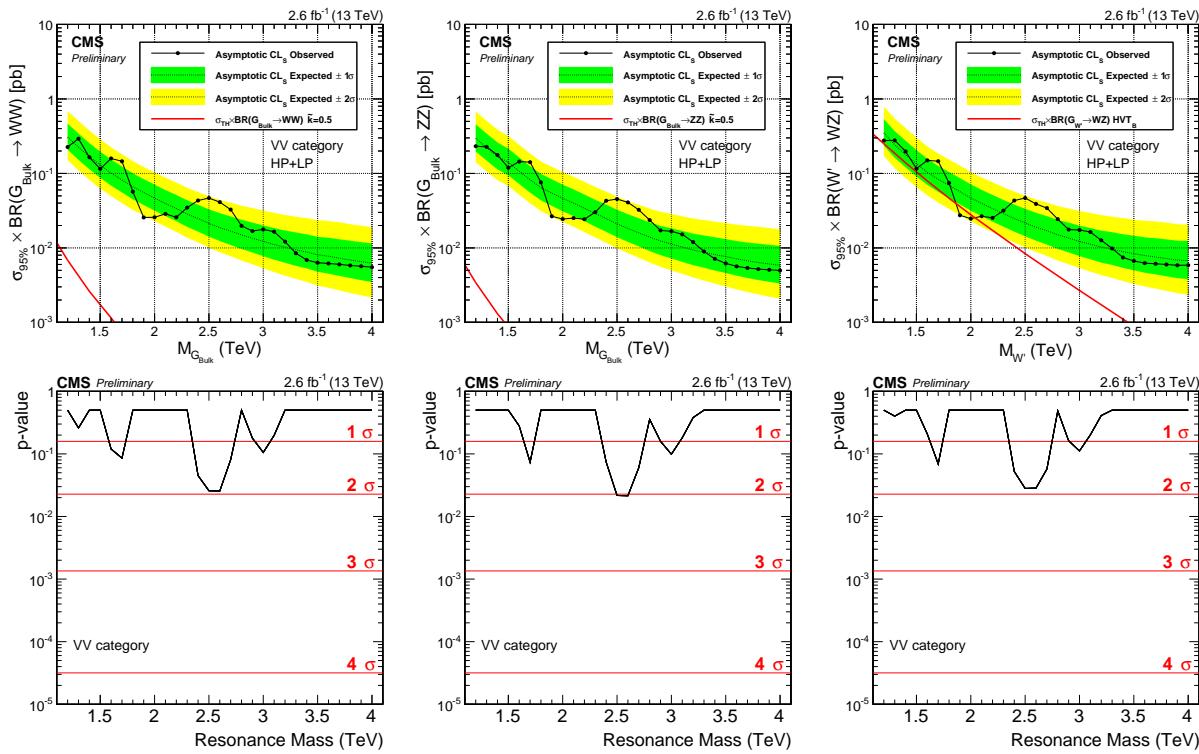


Figure B.9: Expected and observed limits at 95% CL and corresponding obtained without splitting into mass categories. This analysis is performed as a cross check analysis and directly compares with the method used in the corresponding Run 1 analysis. Here for a Bulk $G \rightarrow WW$ (left), $G \rightarrow ZZ$ (middle) and $W' \rightarrow WZ$ signal (right).

APPENDIX C

Search II

C.1 W-tagging scale factor: Additional plots

Figures C.1 shows the $t\bar{t}$ real W (top) and non-W (bottom) PUPPI softdrop jet mass distributions for jets that passed (left) and failed (right column) the N-subjettiness selections $\text{PUPPI } \tau_{21} < 0.40$. Figures C.2 shows the fitted PUPPI softdrop jet mass distributions for the non-dominant

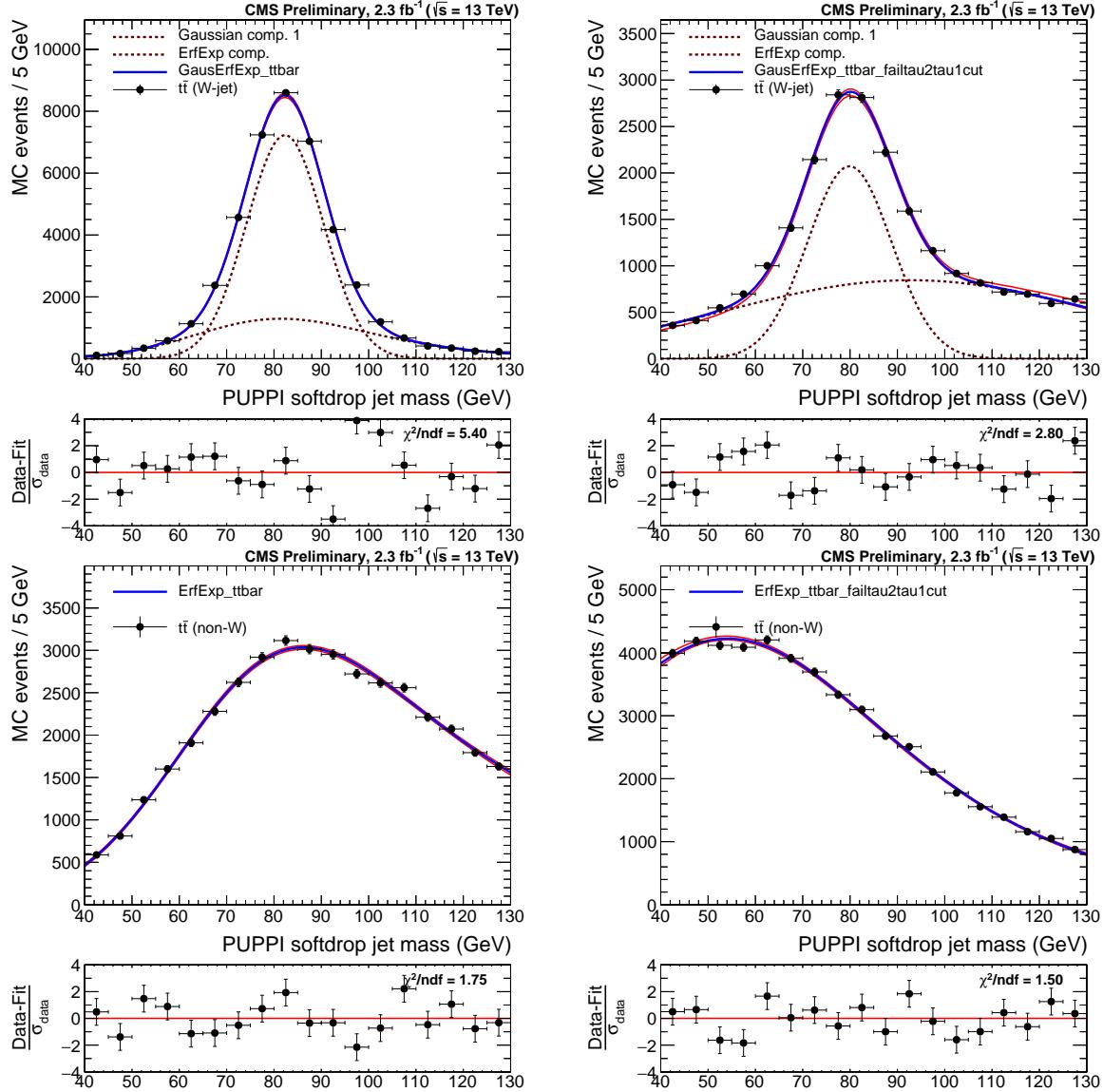


Figure C.1: Fit to the real W (top) and non-W (bottom) softdrop jet mass distribution for jets that pass (left) and fail (right) the cut on $\text{PUPPI } \tau_{21} < 0.4$.

backgrounds in the evaluation of the W-tagging scale factors. Here it is shown for jets that pass

(top) and failed (bottom) the N-subjettiness selections $\text{PUPPI } \tau_{21} < 0.40$.

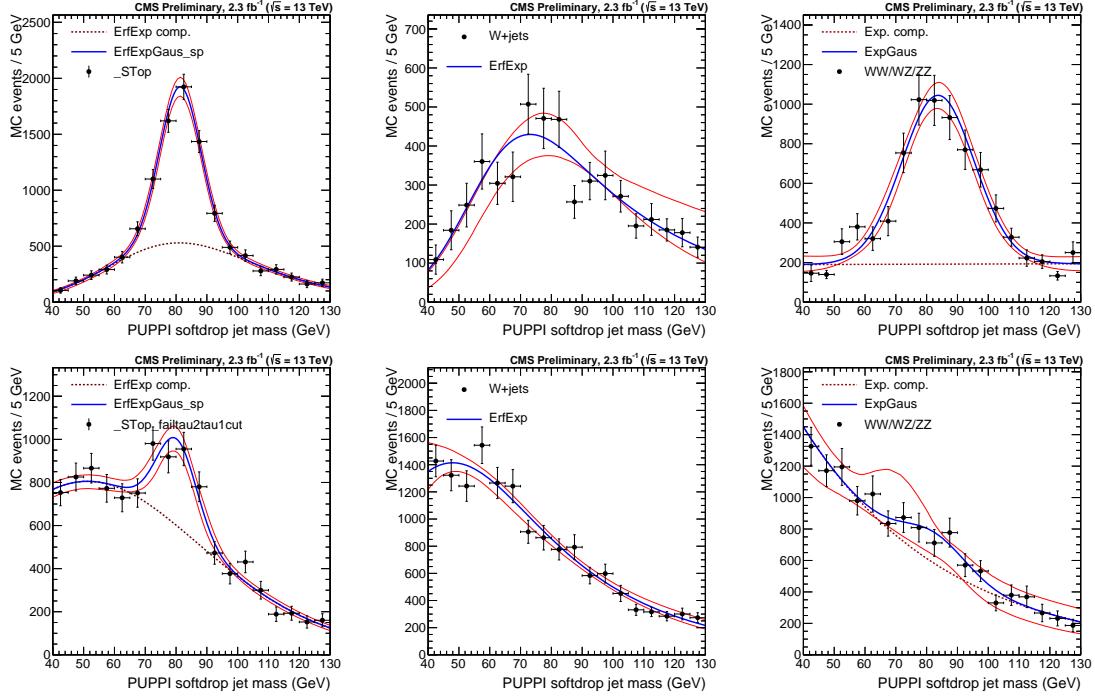


Figure C.2: Fits to the PUPPI softdrop jet mass spectrum for the non-dominant backgrounds (Single top, W+jets and VV respectively) in the pass (top) and fail (bottom) regions.

C.2 Efficiency scale factors for 2.5 fb^{-1}

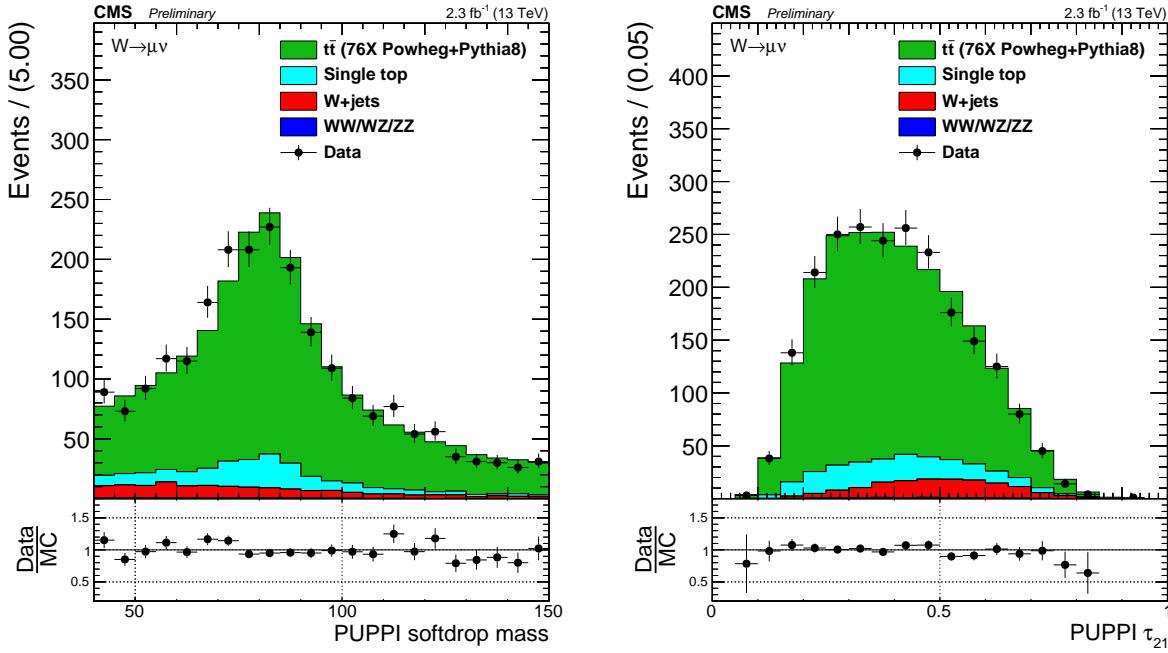


Figure C.3: Distribution of the PUPPI softdrop mass (left) and PUPPI n-subjettiness (right) distribution in the $t\bar{t}$ control sample.

Category	Working point	Eff. data	Eff. simulation	Scale factor
HP	$\tau_2/\tau_1 < 0.4$	0.785 ± 0.045	0.81 ± 0.01	$0.97 \pm 0.06 \text{ (stat)} \pm 0.04 \text{ (sys)} \pm 0.06 \text{ (sys)}$
LP	$0.4 < \tau_2/\tau_1 < 0.75$	0.215 ± 0.057	0.204 ± 0.041	$1.13 \pm 0.24 \text{ (stat)} \pm 0.17 \text{ (sys)} \pm 0.12 \text{ (sys)}$

Table C.1: W-tagging scale factors for high purity and low purity categories for a tagger based on PUPPI softdrop jet mass and τ_{21} .

Parameter	Data	Simulation	Data/Simulation
PUPPI softdrop $\langle m \rangle$	$80.3 \pm 0.8 \text{ GeV}$	$81.9 \pm 0.01 \text{ GeV}$	0.98 ± 0.01
PUPPI softdrop σ	$9.0 \pm 0.9 \text{ GeV}$	$8.5 \pm 0.4 \text{ GeV}$	1.07 ± 0.12

Table C.2: Summary of the fitted W-mass peak fit parameters.

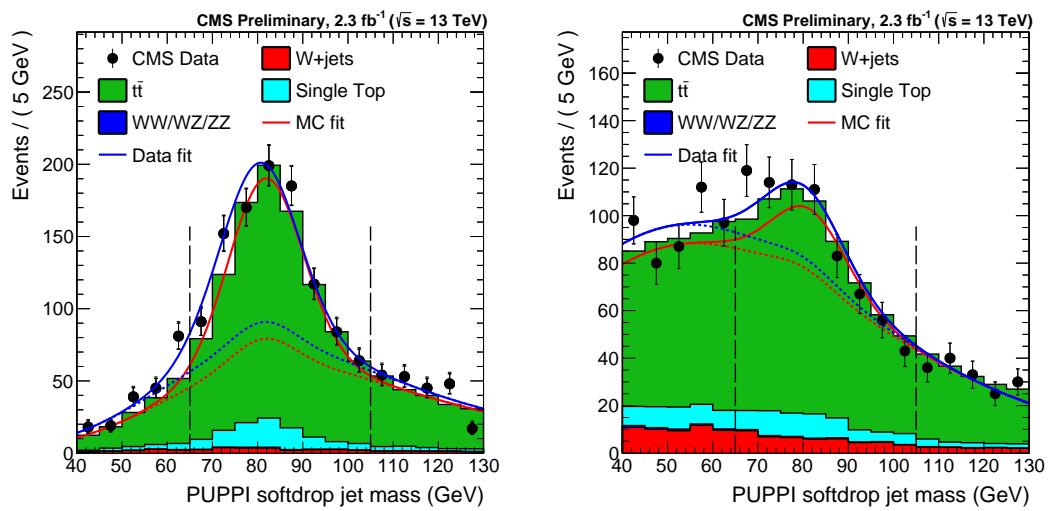


Figure C.4: PUPPI softdrop jet mass distribution for events that pass (left) and fail (right) the PUPPI $\tau_2/\tau_1 < 0.40$ selection. Results of both the fit to data (blue) and simulation (red) are shown and the background components of the fit are shown as short-dashed lines.

C.3 Background fit checks

The background from QCD multijet events is modeled by a smoothly falling distribution in each analysis category. The method consists of a smoothness test of the observed data where the background is assumed to be described by the following empirical probability density function:

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2}}, \quad (\text{C.1})$$

where m is the dijet invariant mass, \sqrt{s} the centre-of-mass energy, P_0 is a normalization parameter for the probability density function, and P_1 and P_2 describe the shape. To ensure that this function is sufficient to describe the data in all the different analysis categories, we first perform a test to check that no additional parameters are needed and to check the systematic uncertainty due to choice of fit function. For these studies we use a data sideband, where one of the two jets is required to have a mass between $20 < M_{\text{Softdrop}} < 65$ GeV. In order to quantify how many parameters are necessary, a Fishers F-test is performed for the fits to data in the data sideband. The critical value that the test statistic must exceed is chosen to be $\alpha > 10\%$. If the returned Confidence Level is larger than α , the simpler fit is preferred. The three parameter fit is compared with the following 2, 4 and 5 parameter functions:

$$\frac{dN}{dm} = \frac{P_0}{(m/\sqrt{s})^{P_2}}, \quad (\text{C.2})$$

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2 + P_3 \times \log(m/\sqrt{s})}}, \text{ and} \quad (\text{C.3})$$

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2 + P_3 \times \log(m/\sqrt{s}) + P_4 \times \log(m/\sqrt{s})^2}}. \quad (\text{C.4})$$

Additionally, fits with an alternative fit function have also been performed (for the single-tag categories we try both 4 and 5 parameter versions):

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s} + P_3(m/\sqrt{s})^2)^{P_1}}{(m/\sqrt{s})^{P_2}}, \text{ and} \quad (\text{C.5})$$

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s} + P_3(m/\sqrt{s})^2)^{P_1}}{(m/\sqrt{s})^{P_2 + P_4 \times \log(m/\sqrt{s})}} \quad (\text{C.6})$$

C.3.1 Background fit checks in data sideband

We perform a test in a data sideband to make sure the fit functions work on real data and to exercise the estimation of number of necessary fit parameters via an F-test. The sideband is constructed by requiring one of the two jets two have a mass in the low softdrop jet mass sideband, between $20 \text{ GeV} < M_{\text{Softdrop}} < 65 \text{ GeV}$, while the full W/Z-tag selections are applied to the other jet. The low-mass jet is also required to pass the τ_{21} selection corresponding to the given category. For the single-tag category, the sideband is constructed by requiring one of the two jets to have a mass in the low softdrop jet mass sideband, between $20 \text{ GeV} < M_{\text{Softdrop}} < 65 \text{ GeV}$ and the other in a high-mass sideband, between $105 < M_{\text{Softdrop}} < 200 \text{ GeV}$. One of the jets is also required to pass the τ_{21} selection corresponding to the given category. We first check whether the sideband can be used to exercise the F-test by checking whether or not there are features introduced in the dijet mass spectrum that may be hard to cover with the fit using QCD MC. To do so we look at the dijet invariant mass spectrum in the signal region divided by the distribution in the sideband. The obtained distributions are shown in Figure C.5 for three different generators, where the pure Pythia8 QCD samples (top left) have the highest number of events. The distributions are mostly smooth, but we do see features introduced in the “WW HP” and “ZZ HP” categories which might prove difficult to fit, as well as in the tail of the single-tag categories. These kinks shift around depending on what MC generator is used and do not seem to be a systematic feature caused by the selection requirements, but rather due to limited statistics. We proceed with exercising the F-test in a data sideband, with the caveat that there might be features introduced in the spectrum where statistics are low.

Figure C.6 shows the fit to data in the data sideband for the WW and ZZ mass categories, both in the HP and in the LP n-subjettiness categories. The corresponding residuals, χ^2 and F-test results are shown in Table C.7 through C.10. For the double-tag categories, a two or three parameter function is sufficient to describe the data and we conclude that the function as defined in C.1 is sufficient for all mass categories. For the single-tag category a five parameter fit seems to be required in order to describe the data and the fit quality is not optimal. To ensure that the fit functions with sufficient number of parameters are able to describe the shape in the single-tag categories, we have additionally looked at the fit quality in QCD MC (see below). Here, the default dijet function seems sufficient to describe the distributions. The sideband in the single-tag categories in QCD MC do not show the same features as the data sideband as shown in Figure C.7.

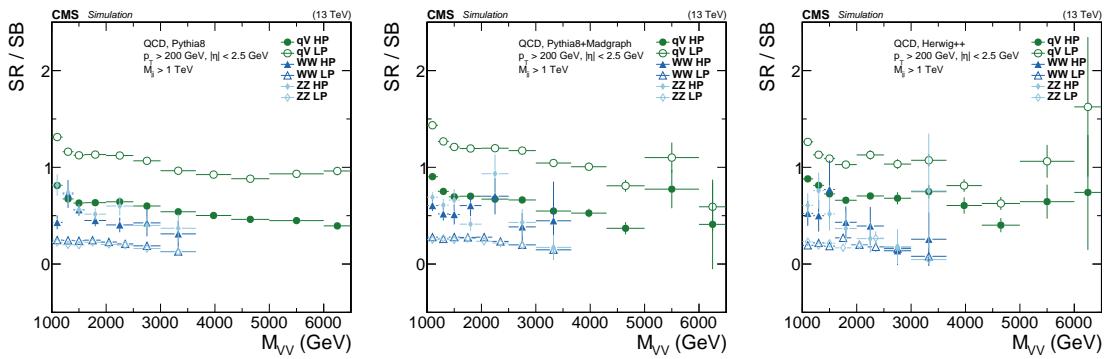


Figure C.5: Dijet mass spectrum in the signal region divided by the dijet mass spectrum in the sidebands using QCD Pythia8 (left), QCD Pythia8+Madgraph (middle) and QCD Herwig++ (right) simulated samples. Here for the double W/Z-tag and the single V-tagged HP and LP categories. Some jumps are observed in the high-mass tail of the dijet invariant mass distribution in the high-purity WW/ZZ categories, but otherwise no strange features seem to be induced by using the sideband.

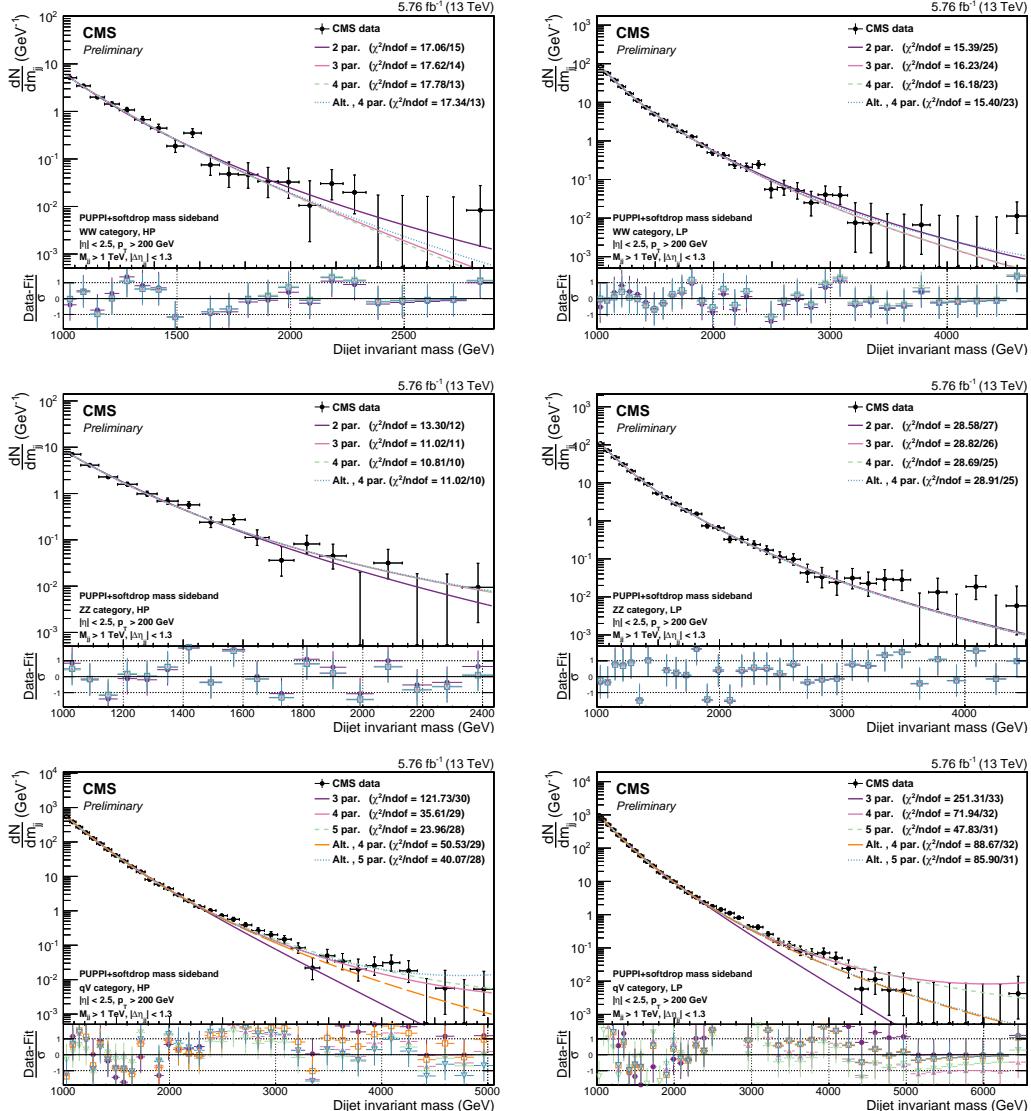
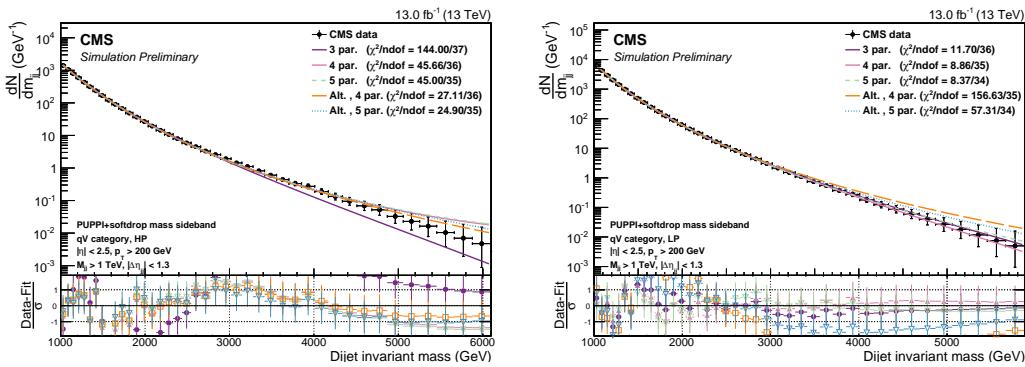


Figure C.6: Fitted dijet mass spectrum in the different mass and purity categories in a data sideband: WW high-purity (top left) and low-purity (top right), ZZ high-purity (middle left) and low-purity (middle right), qV high-purity (bottom left) and low-purity (bottom right).



WW category, HP			
Function	Residuals	χ^2	ndof
2 par	0.129	17.060	15
3 par	0.111	17.623	14
4 par	0.111	17.783	13
Fishers23	2.430	CL	0.140
Fishers34	0.012	CL	0.914

Table C.3: Residuals, χ^2 , and degrees of freedom for the WW category, HP category. A 2-parameter fit is needed to describe these data.

WW category, LP			
Function	Residuals	χ^2	ndof
2 par	0.908	15.388	25
3 par	0.279	16.225	24
4 par	0.263	16.178	23
Fishers23	56.395	CL	0.000
Fishers34	1.406	CL	0.247

Table C.4: Residuals, χ^2 , and degrees of freedom for the WW category, LP category. A 3-parameter fit is needed to describe these data.

ZZ category, HP			
Function	Residuals	χ^2	ndof
2 par	0.215	13.296	12
3 par	0.133	11.022	11
4 par	0.119	10.810	10
Fishers23	7.465	CL	0.018
Fishers34	1.304	CL	0.278

Table C.5: Residuals, χ^2 , and degrees of freedom for the ZZ category, HP category. A 3-parameter fit is needed to describe these data.

ZZ category, LP			
Function	Residuals	χ^2	ndof
2 par	2.459	28.583	27
3 par	2.363	28.817	26
4 par	2.175	28.694	25
Fishers23	1.107	CL	0.302
Fishers34	2.244	CL	0.146

Table C.6: Residuals, χ^2 , and degrees of freedom for the ZZ category, LP category. A 2-parameter fit is needed to describe these data.

qV category, HP			
Function	Residuals	χ^2	ndof
3 par	128.276	121.731	30
4 par	29.113	35.606	29
5 par	7.036	23.962	28
Alt. 4 par	37.232	50.528	29
Alt. 5 par	30.948	40.068	28
Fishers34	102.185	CL	0.000
Fishers45	90.988	CL	0.000
FishersAlt4Alt5	5.888	CL	0.022

Table C.7: Residuals, χ^2 , and degrees of freedom for the qV category, HP category. A 5-parameter fit is needed to describe these data.

qV category, LP			
Function	Residuals	χ^2	ndof
3 par	671.341	251.311	33
4 par	171.593	71.942	32
5 par	80.801	47.830	31
Alt. 4 par	215.431	88.666	32
Alt. 5 par	214.766	85.896	31
Fishers34	96.109	CL	0.000
Fishers45	35.957	CL	0.000
FishersAlt4Alt5	0.099	CL	0.755

Table C.8: Residuals, χ^2 , and degrees of freedom for the qV category, LP category. A 5-parameter fit is needed to describe these data.

C.3.2 Background fit checks in QCD MC

As an additional check, we look at the fit functions in the different signal categories using QCD MC. This is shown in Figure C.8 for the double and Figure C.9 for the single tag categories. Here all fit functions and their pull distributions are plotted. The fits are performed in a mass range corresponding to the expected distribution in the different categories for 13 fb^{-1} of data. We have adapted the error bars to correspond to the maximum of the expected Poisson error for 13 fb^{-1} of data and the pure simulation error (accounting for the different weights assigned to the p_T -binned QCD MC sample). The reason for this choice is to get an estimate of whether the set of fit functions we plan to use to fit the background distribution in data, and plan to use in order to understand the systematic uncertainty due to our choice of fit function, are appropriate and do not produce fake bumps/kinks. As this distribution is the pure MC simulation curve, whose variation at high masses is much smaller than the expected poisson error for 13 fb^{-1} of data, we expect the χ^2/ndof to be lower than one. In order to protect against the fact that the MC simulation at lower dijet masses does not have more statistics than the expected data for 13 fb^{-1} of data, we use the largest of Poisson and the MC errors. The resulting errors are therefore a mixture of Poisson and the MC errors and the χ^2/ndof for the QCD MC fits should not be considered. Fit quality in the form of χ^2/ndof should only be estimated from the fits to data sideband where pure Poisson errors are used. Overall the fits to QCD MC in the different categories describe the data well, with a two or three parameter function sufficient to describe the distributions. However, due to an under fluctuation of the first bin in the “ZZHP” category, the higher parameter fits are steered by the first bin leading to discrepancies in the tail. This is the lowest statistics category and the danger for under-fluctuations does exist. We have investigated the dijet invariant mass distribution down to a dijet invariant mass threshold of 800 GeV to make sure we are not seeing a turn-on effect. This is shown in Figure C.10. The fits in the data sideband for the same category (Figure C.6, middle left) do not show the same trend.

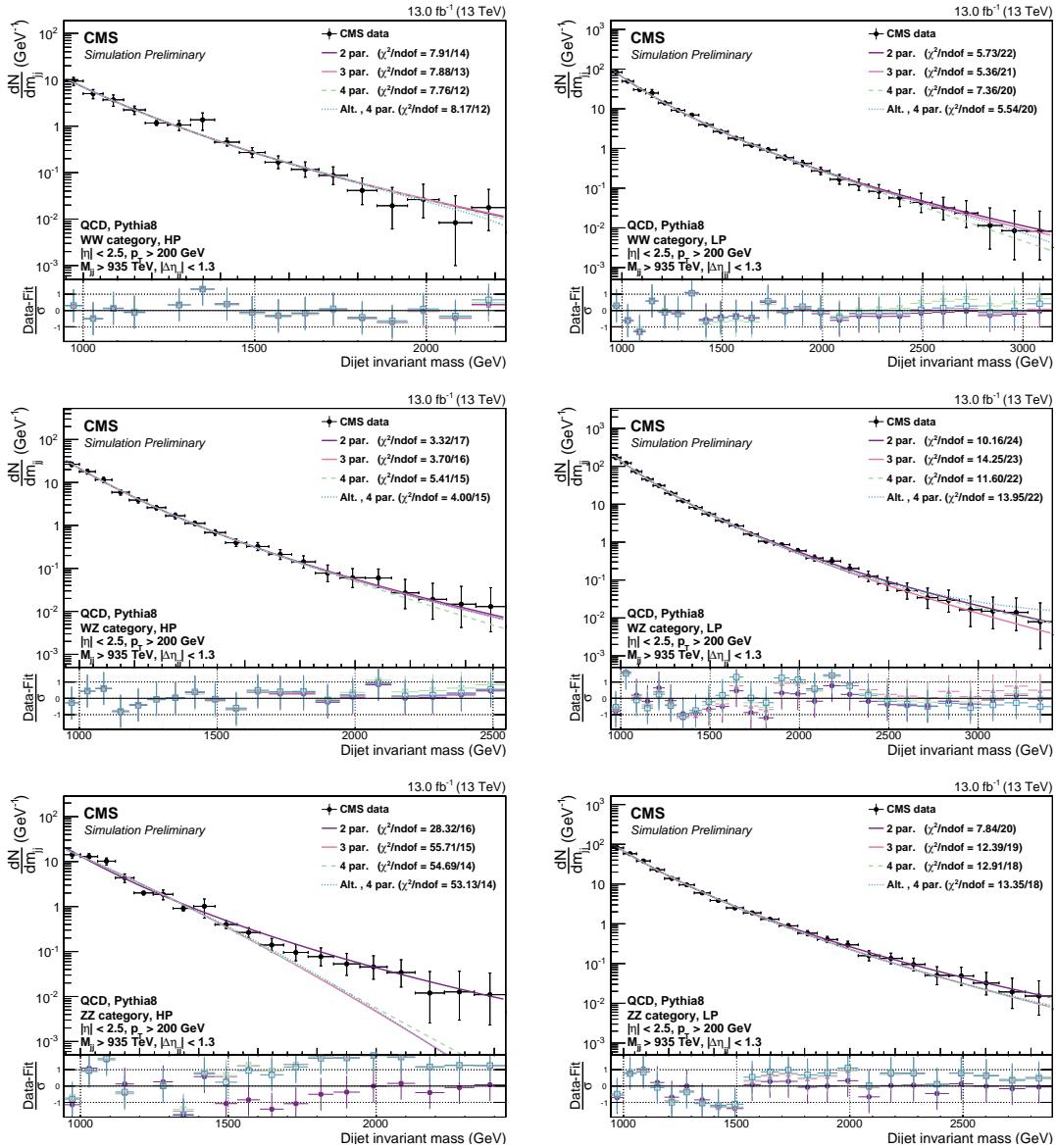


Figure C.8: Background fit for the M_{jj} distribution in QCD MC corresponding to 13 fb^{-1} . Shown here for the high- and low-purity double W/Z-tag category for the three different mass categories: WW category (top), WZ category (middle) and ZZ category (bottom).

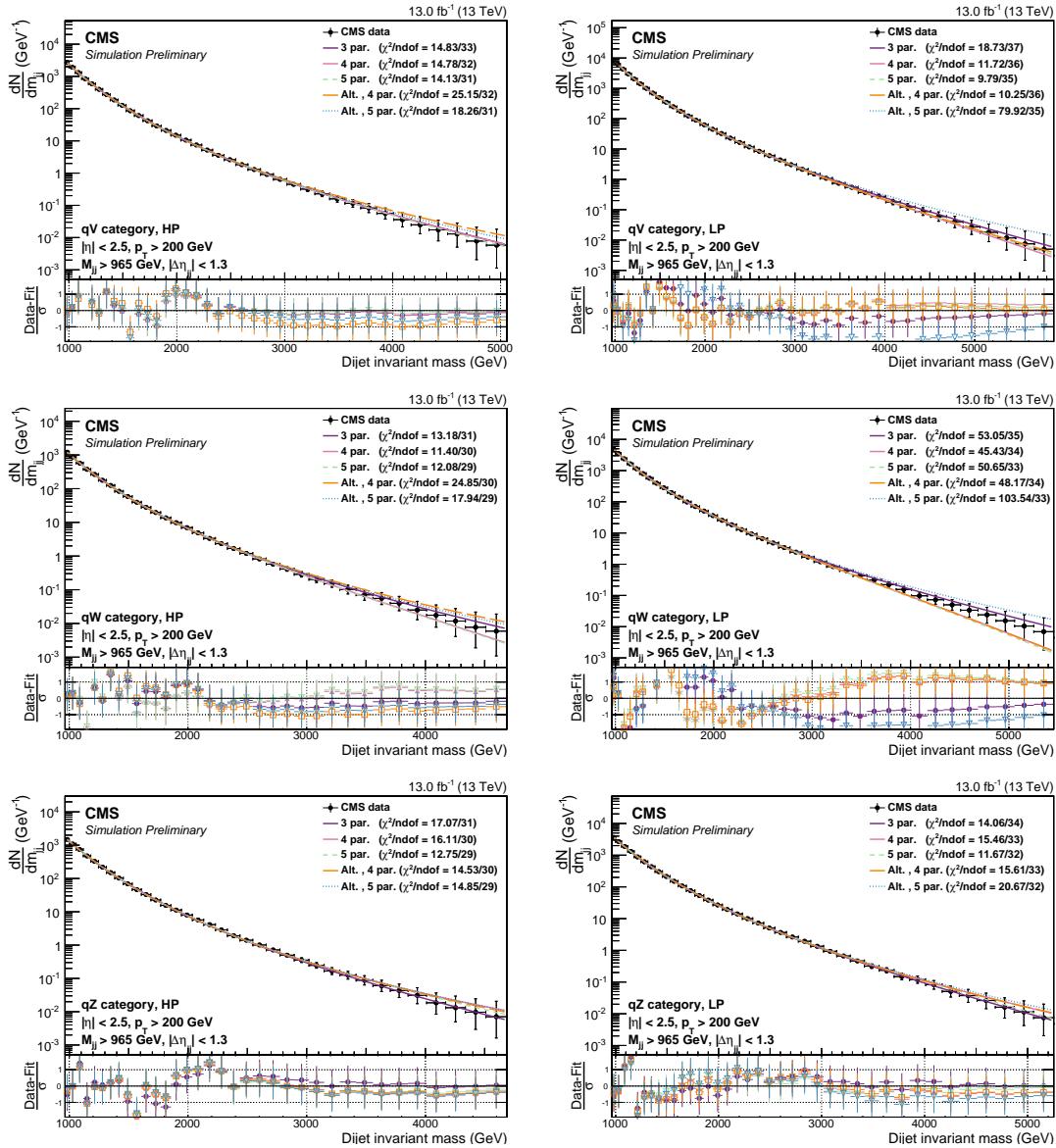


Figure C.9: Background fit for the M_{jj} distribution in QCD MC corresponding to 13 fb^{-1} . Shown here for the high- and low-purity single W/Z-tag category for the two different mass categories: qW category (top) and qZ category (bottom).

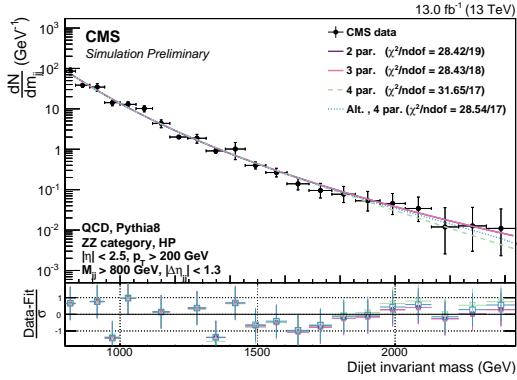


Figure C.10: Background fit for the M_{jj} distribution in QCD MC corresponding to 13 fb^{-1} . Shown here for the high-purity double Z-tag category using a dijet invariant mass threshold of 800 GeV. No turn-on effect at low invariant masses is observed.

C.4 F-test in signal region

The final F-test performed in order to define the number of fit parameters to be used to fit the background in each analysis category is performed in the signal region. The resulting fits and F-test values for the double tag categories are shown in Figure C.11 and Tables C.9 to C.14. The F-test results for the single-tag category are listed in Tables C.15 to C.18. Here, a three-parameter fit is sufficient for all categories except the “high-purity” qW category where a 5-parameter fit is preferred.

WW category, HP			
Function	Residuals	χ^2	ndof
2 par	0.251	17.673	16
3 par	0.187	14.863	15
4 par	0.183	14.618	14
Fishers23	5.454	CL	0.033
Fishers34	0.391	CL	0.541

Table C.9: Residuals, χ^2 , and degrees of freedom for the WW category, HP category. A 3-parameter fit is needed to describe these data.

WW category, LP			
Function	Residuals	χ^2	ndof
2 par	2.974	13.997	23
3 par	3.082	14.775	22
4 par	3.080	14.768	21
Fishers23	-0.805	CL	1.000
Fishers34	0.015	CL	0.905

Table C.10: Residuals, χ^2 , and degrees of freedom for the WW category, LP category. A 2-parameter fit is needed to describe these data.

WZ category, HP			
Function	Residuals	χ^2	ndof
2 par	2.333	17.562	17
3 par	2.158	16.952	16
4 par	2.114	16.842	15
Fishers23	1.372	CL	0.258
Fishers34	0.338	CL	0.569

Table C.11: Residuals, χ^2 , and degrees of freedom for the WZ category, HP category. A 2-parameter fit is needed to describe these data.

WZ category, LP			
Function	Residuals	χ^2	ndof
2 par	12.301	21.368	25
3 par	6.827	20.715	24
4 par	6.521	20.419	23
Fishers23	20.046	CL	0.000
Fishers34	1.126	CL	0.299

Table C.12: Residuals, χ^2 , and degrees of freedom for the WZ category, LP category. A 3-parameter fit is needed to describe these data.

ZZ category, HP			
Function	Residuals	χ^2	ndof
2 par	0.634	17.919	17
3 par	0.662	17.400	16
4 par	0.716	17.096	15
Fishers23	-0.720	CL	1.000
Fishers34	-1.197	CL	1.000

Table C.13: Residuals, χ^2 , and degrees of freedom for the ZZ category, HP category. A 2-parameter fit is needed to describe these data.

ZZ category, LP			
Function	Residuals	χ^2	ndof
2 par	9.293	19.452	22
3 par	6.884	20.118	21
4 par	6.598	20.076	20
Fishers23	7.701	CL	0.011
Fishers34	0.909	CL	0.351

Table C.14: Residuals, χ^2 , and degrees of freedom for the ZZ category, LP category. A 3-parameter fit is needed to describe these data.

qW category, HP			
Function	Residuals	χ^2	ndof
3 par	69.757	30.375	30
4 par	59.677	28.318	29
5 par	25.298	21.815	28
Alt. 4 par	35.610	22.810	29
Alt. 5 par	25.634	22.687	28
Fishers34	5.067	CL	0.032
Fishers45	39.409	CL	0.000
FishersAlt4Alt5	11.285	CL	0.002

Table C.15: Residuals, χ^2 , and degrees of freedom for the qW category, HP category. A 5-parameter fit is needed to describe these data.

qW category, LP			
Function	Residuals	χ^2	ndof
3 par	153.869	38.713	35
4 par	156.715	38.586	34
5 par	201.767	38.167	33
Alt. 4 par	189.434	39.327	34
Alt. 5 par	192.782	39.170	33
Fishers34	-0.636	CL	1.000
Fishers45	-7.592	CL	1.000
FishersAlt4Alt5	-0.590	CL	1.000

Table C.16: Residuals, χ^2 , and degrees of freedom for the qW category, LP category. A 3-parameter fit is needed to describe these data.

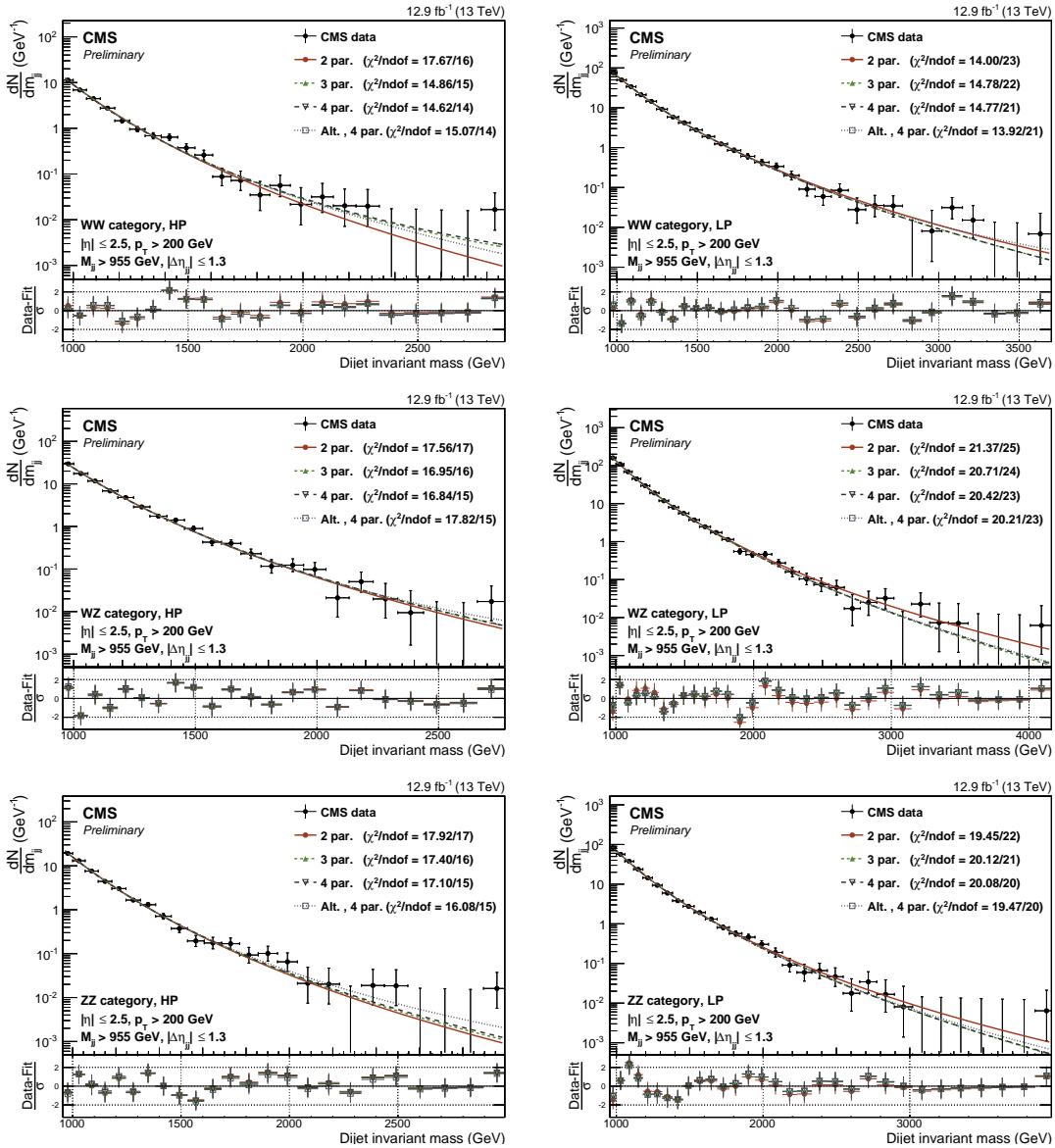


Figure C.11: Background fit for the M_{jj} distribution in the data signal region. Shown here are the high- (left) and low-purity (right) double W/Z-tag categories WW (top), WZ (middle) and ZZ (bottom).

qZ category, HP			
Function	Residuals	χ^2	ndof
3 par	12.963	21.252	30
4 par	12.961	21.252	29
5 par	9.256	19.644	28
Alt. 4 par	13.931	20.977	29
Alt. 5 par	9.739	20.344	28
Fishers34	0.004	CL	0.948
Fishers45	11.609	CL	0.002
FishersAlt4Alt5	12.484	CL	0.001

Table C.17: Residuals, χ^2 , and degrees of freedom for the qZ category, HP category. A 3-parameter fit is needed to describe these data.

qZ category, LP			
Function	Residuals	χ^2	ndof
3 par	369.554	47.426	36
4 par	369.554	47.426	35
5 par	298.358	46.525	34
Alt. 4 par	379.111	47.531	35
Alt. 5 par	379.120	47.531	34
Fishers34	0.000	CL	0.994
Fishers45	8.352	CL	0.007
FishersAlt4Alt5	-0.001	CL	0.000

Table C.18: Residuals, χ^2 , and degrees of freedom for the qZ category, LP category. A 3-parameter fit is needed to describe these data.

APPENDIX D

Search III

D.1 Signal fits

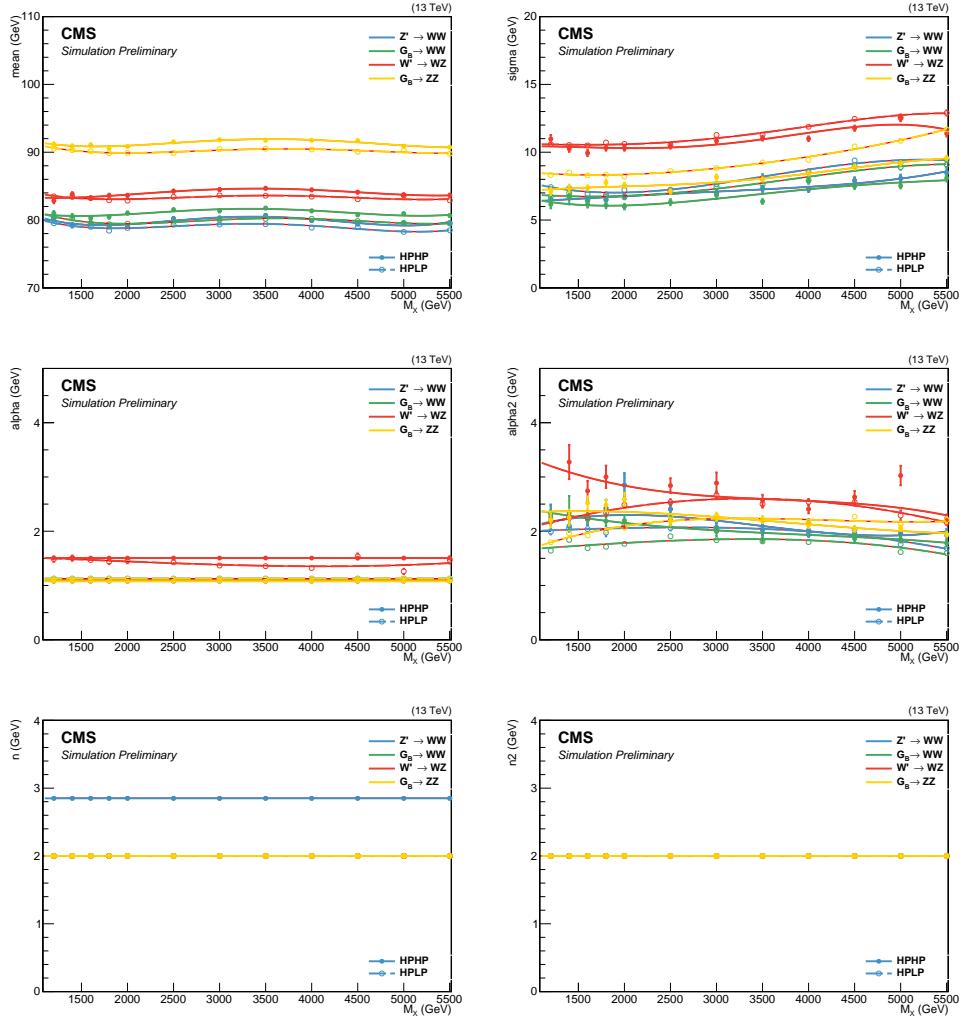


Figure D.1: The interpolated double Crystal-ball parameters for the softdrop jet mass of $M_{\text{jet}2}$ as a function of M_X . To improve the stability of the fit some parameters are set constant.

D.2 2016 kernels

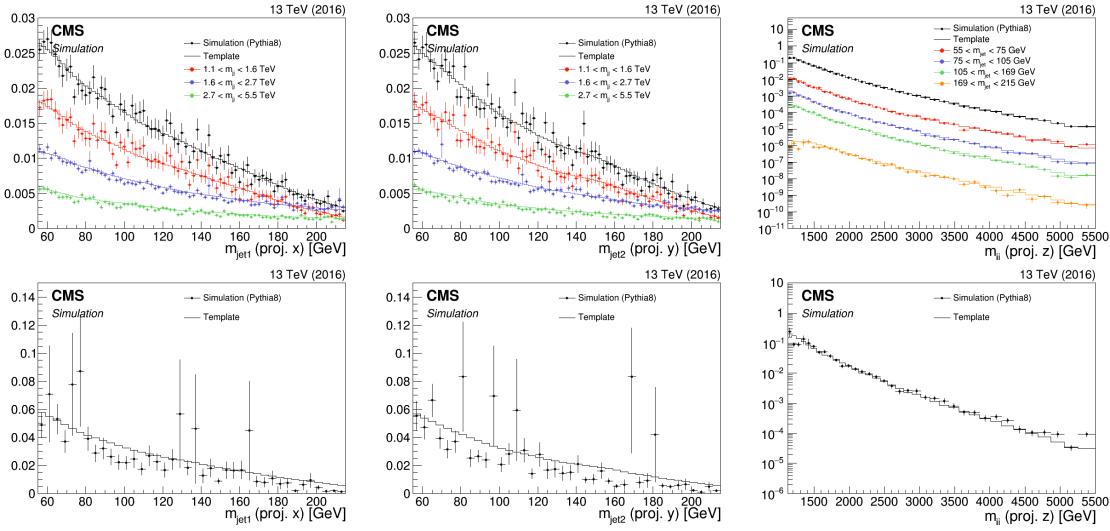


Figure D.2: Comparison between QCD MC simulation (markers) and kernels derived from generator-level quantities (lines) for the HP category (top) and the LP category (bottom) for 2016 MC. The kernels are shown for $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{VV} (right).

D.3 Resonant background

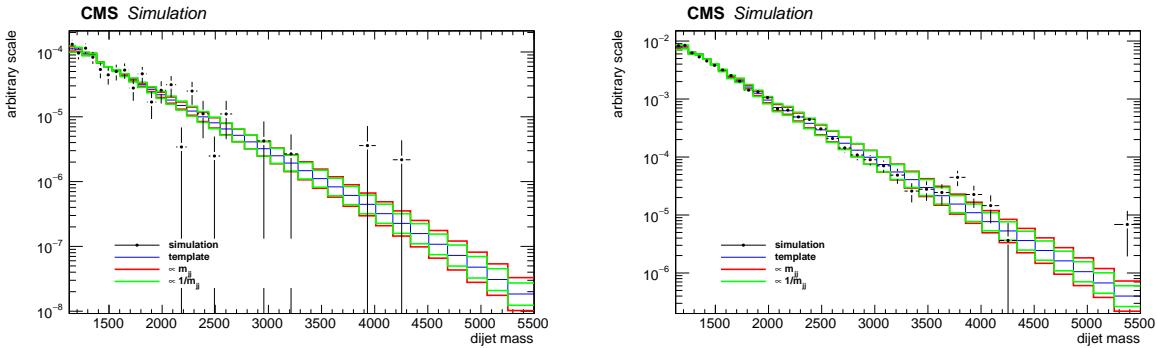


Figure D.3: One-dimensional M_{VV} kernels (solid line) compared to MC (markers) for the Z+jets background in the HP (left) and LP (right) categories. The nominal shape derived from the smoothing procedure can be seen as a blue line, alternative shapes derived from varying the slope of the M_{VV} spectrum are shown in green and red.

Bibliography

- [1] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys.* **C40** (2016), no. 10, 100001, doi:10.1088/1674-1137/40/10/100001.
- [2] P. Collaboration, “Planck intermediate results - XLVI. Reduction of large-scale systematic effects in HFI polarization maps and estimation of the reionization optical depth”, *A&A* **596** (2016) A107, doi:10.1051/0004-6361/201628890.
- [3] M. Gell-Mann, “Symmetries of Baryons and Mesons”, *Phys. Rev.* **125** (Feb, 1962) 1067–1084, doi:10.1103/PhysRev.125.1067.
- [4] W. Pauli, “Mathematical contributions to the theory of Dirac’s matrices”, *Ann. Inst. H. Poincaré Phys. Theor.* **6** (1936) 109–136.
- [5] R. Tumulka, “Pauli Spin Matrices”, pp. 470–472. 07, 2009.
doi:10.1007/978-3-540-70626-7_142.
- [6] S. Duplij and F. Klinkhamer, “Levi-Civita Symbol”, pp. 227–227. Springer Netherlands, Dordrecht, 2004. doi:10.1007/1-4020-4522-0_298.
- [7] S. Weinberg, “Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions”, *Rev. Mod. Phys.* **52** (1980) 515–523, doi:10.1103/RevModPhys.52.515. [543(1979)].
- [8] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- [9] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- [10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, doi:10.1103/PhysRevLett.13.585.

- [11] P. W. Anderson, “Coherent Excited States in the Theory of Superconductivity: Gauge Invariance and the Meissner Effect”, *Phys. Rev.* **110** (May, 1958) 827–835, doi:10.1103/PhysRev.110.827.
- [12] J. Goldstone, “Field Theories with Superconductor Solutions”, *Nuovo Cim.* **19** (1961) 154–164, doi:10.1007/BF02812722.
- [13] J. Ellis, “Higgs Physics”, arXiv:1312.5672.
- [14] B. Bellazzini, C. Csáki, and J. Serra, “Composite Higgses”, *Eur. Phys. J.* **C74** (2014), no. 5, 2766, doi:10.1140/epjc/s10052-014-2766-x, arXiv:1401.2457.
- [15] R. Contino, D. Marzocca, D. Pappadopulo, and R. Rattazzi, “On the effect of resonances in composite Higgs phenomenology”, *JHEP* **10** (2011) 081, doi:10.1007/JHEP10(2011)081, arXiv:1109.1570.
- [16] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, “Heavy Vector Triplets: Bridging Theory and Data”, *JHEP* **09** (2014) 060, doi:10.1007/JHEP09(2014)060, arXiv:1402.4431.
- [17] L. Randall and R. Sundrum, “Large Mass Hierarchy from a Small Extra Dimension”, *Phys. Rev. Lett.* **83** (Oct, 1999) 3370–3373, doi:10.1103/PhysRevLett.83.3370.
- [18] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, “Warped gravitons at the CERN LHC and beyond”, *Phys. Rev. D* **76** (Aug, 2007) 036006, doi:10.1103/PhysRevD.76.036006.
- [19] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, “Searching for the Kaluza-Klein Graviton in Bulk RS Models”, *JHEP* **09** (2007) 013, doi:10.1088/1126-6708/2007/09/013, arXiv:hep-ph/0701150.
- [20] CERN, “Accelerators and Schedules”, 2018 (accessed November 12, 2018), <https://beams.web.cern.ch/content/accelerators-schedules>.
- [21] CMS, “CMS Detector Design”, 2018 (accessed November 21, 2018), <https://cms.cern/news/cms-detector-design>.
- [22] T. Lenzi, “Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors”, Master’s thesis, U. Brussels (main), 2013.
- [23] A. Dominguez et al., “CMS Technical Design Report for the Pixel Detector Upgrade”, Technical Report CERN-LHCC-2012-016. CMS-TDR-11, CERN, Sep, 2012.

- [24] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [25] P. Adzic et al., “Energy resolution of the barrel of the CMS electromagnetic calorimeter”, *JINST* **2** (2007) P04004, doi:10.1088/1748-0221/2/04/P04004.
- [26] C. Biino, “The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections”, *Journal of Physics: Conference Series* **587** (2015), no. 1, 012001.
- [27] L. Veillet, “Assembly of the 7th wedge”, (Aug, 2000). CMS Collection.
- [28] S. Sharma, “Understanding the performance of CMS calorimeter”, *Pramana* **69** (Dec, 2007) 1069–1074, doi:10.1007/s12043-007-0229-8.
- [29] CMS Collaboration, “CMS reconstruction improvement for the muon tracking by the RPC chambers”, *PoS RPC2012* (Sep, 2012) 045. 9 p. Presented by Minsuk Kim at the XI workshop on Resistive Plate Chambers and Related Detectors - RPC2012, INFN Laboratori Nazionali di Frascati Italy, February 5-10, 2012.
- [30] CMS, “CMS Luminosity - public results”, 2018 (accessed December 11, 2018), <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [31] P. Billoir, “Progressive track recognition with a Kalman-like fitting procedure”, *Computer Physics Communications* **57** (1989), no. 1, 390 – 394, doi:[https://doi.org/10.1016/0010-4655\(89\)90249-X](https://doi.org/10.1016/0010-4655(89)90249-X).
- [32] A. S. et. Al., “Particle-flow reconstruction and global event description with the CMS detector”, *Journal of Instrumentation* **12** (2017), no. 10, P10003.
- [33] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear and Particle Physics* **31** (2005), no. 9, N9.
- [34] T. C. collaboration, “Performance of CMS muon reconstruction in pp collision events at s = 7 TeV”, *Journal of Instrumentation* **7** (2012), no. 10, P10002.
- [35] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup per particle identification”, *Journal of High Energy Physics* **2014** (2014), no. 10, doi:10.1007/JHEP10(2014)059.
- [36] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better jet clustering algorithms”, *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.

- [37] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev.* **D48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.
- [38] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [39] J. POG, “Jet Identification: Recommendations for 13TeV data analysis”, 2018 (accessed December 18, 2018), <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID>.
- [40] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.
- [41] D. Krohn, J. Thaler, and L.-T. Wang, “Jet Trimming”, *JHEP* **02** (2010) 084, doi:10.1007/JHEP02(2010)084, arXiv:0912.1342.
- [42] CMS Collaboration Collaboration, “Pileup Removal Algorithms”, Technical Report CMS-PAS-JME-14-001, CERN, Geneva, 2014.
- [43] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, “Jet substructure as a new Higgs search channel at the LHC”, *Phys. Rev. Lett.* **100** (2008) 242001, doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.
- [44] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet substructure”, *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.
- [45] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.
- [46] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP* **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.
- [47] J. Thaler and K. Van Tilburg, “Maximizing boosted top identification by minimizing N-subjettiness”, *Journal of High Energy Physics* **2012** (Feb, 2012) 93, doi:10.1007/JHEP02(2012)093.
- [48] J. Dolen et al., “Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure”, arXiv:1603.00027.
- [49] S. Bolognesi et al., “Spin and parity of a single-produced resonance at the LHC”, *Phys. Rev. D* **86** (Nov, 2012) 095031, doi:10.1103/PhysRevD.86.095031.

- [50] CMS Collaboration, “Identification techniques for highly boosted W bosons that decay into hadrons”, *JHEP* **12** (2014) 017, doi:10.1007/JHEP12(2014)017, arXiv:1410.4227.
- [51] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur. Phys. J.* **C58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.
- [52] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.
- [53] The ATLAS collaboration, “Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $s=8$ TeV with the ATLAS detector”, *Journal of High Energy Physics* **2015** (Dec, 2015) 1–39, doi:10.1007/JHEP12(2015)055.
- [54] S. Frixione and B. R. Webber, “Matching NLO QCD computations and parton shower simulations”, *JHEP* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.
- [55] S. Frixione, P. Nason, and B. R. Webber, “Matching NLO QCD and parton showers in heavy flavor production”, *JHEP* **08** (2003) 007, doi:10.1088/1126-6708/2003/08/007, arXiv:hep-ph/0305252.
- [56] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.
- [57] S. A. et. al., “Geant4—a simulation toolkit”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250 – 303, doi:[https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [58] CMS Collaboration, “Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s}=8$ TeV”, *JHEP* **08** (May, 2014) 173. 37 p.
- [59] F. Dias et al., “Combination of Run-1 Exotic Searches in Diboson Final States at the LHC”, *JHEP* **04** (2016) 155, doi:10.1007/JHEP04(2016)155, arXiv:1512.03371.
- [60] CMS Collaboration, “Search for massive WH resonances decaying into the $\ell\nu b\bar{b}$ final state at $\sqrt{s}=8$ TeV”, *Eur. Phys. J.* **C76** (2016), no. 5, 237, doi:10.1140/epjc/s10052-016-4067-z, arXiv:1601.06431.
- [61] CERN, “Collider Reach”, 2019 (accessed January 12, 2019), <http://collider-reach.web.cern.ch/>.

- [62] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [63] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “A brief introduction to PYTHIA 8.1”, *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.
- [64] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [65] S. C. et al., “Search for resonances in the dijet mass spectrum from 7 TeV pp collisions at CMS”, *Physics Letters B* **704** (2011), no. 3, 123 – 142, doi:<https://doi.org/10.1016/j.physletb.2011.09.015>.
- [66] G. Punzi, “Sensitivity of searches for new signals and its optimization”, *eConf* **C030908** (2003) MODT002, arXiv:physics/0308063. [,79(2003)].
- [67] CMS Collaboration, “Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at 7 TeV”, *Phys. Lett.* **B723** (2013) 280–301, doi:10.1016/j.physletb.2013.05.040, arXiv:1212.1910.
- [68] CMS Collaboration Collaboration, “Search for Narrow Resonances using the Dijet Mass Spectrum with 19.6fb-1 of pp Collisions at sqrt(s)=8 TeV”, Technical Report CMS-PAS-EXO-12-059, CERN, Geneva, 2013.
- [69] D. J. Hand, “Statistical Concepts: A Second Course, Fourth Edition by Richard G. Lomax, Debbie L. Hahs-Vaughn”, *International Statistical Review* **80** (2012), no. 3, 491–491.
- [70] T. C. collaboration, “Identification of b-quark jets with the CMS experiment”, *Journal of Instrumentation* **8** (2013), no. 04, P04013.
- [71] T. C. collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *Journal of Instrumentation* **13** (2018), no. 05, P05011.
- [72] CMS Collaboration Collaboration, “Search for massive resonances decaying into WW, WZ, ZZ, qW and qZ in the dijet final state at $\sqrt{s} = 13$ TeV using 2016 data”, Technical Report CMS-PAS-B2G-16-021, CERN, Geneva, 2016.

- [73] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *The European Physical Journal C* **71** (Feb, 2011) doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727.
- [74] M. Dasgupta, A. Powling, and A. Siodmok, “On jet substructure methods for signal jets”, *JHEP* **08** (2015) 079, doi:10.1007/JHEP08(2015)079, arXiv:1503.01088.
- [75] U. Baur, I. Hinchliffe, and D. Zeppenfeld, “Excited quark production at hadron colliders”, *Int. J. Mod. Phys. A* **2** (1987) 1285, doi:10.1142/S0217751X87000661.
- [76] U. Baur, M. Spira, and P. M. Zerwas, “Excited-quark and -lepton production at hadron colliders”, *Phys. Rev. D* **42** (1990) 815, doi:10.1103/PhysRevD.42.815.
- [77] CMS Collaboration Collaboration, “Search for massive resonances decaying into WW , WZ , ZZ , qW , and qZ with dijet final states at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **97** (Apr, 2018) 072006, doi:10.1103/PhysRevD.97.072006.
- [78] CMS Collaboration Collaboration, “Jet algorithms performance in 13 TeV data”, Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.
- [79] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J. C* **53** (2008) 473–500, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.
- [80] J. Gallicchio and M. D. Schwartz, “Quark and Gluon Tagging at the LHC”, *Phys. Rev. Lett.* **107** (2011) 172001, doi:10.1103/PhysRevLett.107.172001, arXiv:1106.3076.
- [81] J. A. Aguilar-Saavedra, “Stealth bosons and where to find them”, 2018, <https://indico.cern.ch/event/649482/contributions/2993323/attachments/1688557/2716149/SBAWTFD.pdf>.
- [82] J. A. Aguilar-Saavedra, “Running bumps from stealth bosons”, *Eur. Phys. J. C* **78** (2018), no. 3, 206, doi:10.1140/epjc/s10052-018-5717-0, arXiv:1801.08129.
- [83] S. Kallweit et al., “NLO electroweak automation and precise predictions for W +multijet production at the LHC”, *JHEP* **04** (2015) 012, doi:10.1007/JHEP04(2015)012, arXiv:1412.5157.
- [84] S. Kallweit et al., “NLO QCD+EW predictions for V +jets including off-shell vector-boson decays and multijet merging”, *JHEP* **04** (2016) 021, doi:10.1007/JHEP04(2016)021, arXiv:1511.08692.

- [85] S. Kallweit et al., “NLO QCD+EW automation and precise predictions for V+multijet production”, in *Proceedings, 50th Rencontres de Moriond, QCD and high energy interactions*, p. 121. 2015. [arXiv:1505.05704](https://arxiv.org/abs/1505.05704).
- [86] CMS Collaboration Collaboration, “Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair”, *Phys. Rev. Lett.* **120** (Sep, 2017) 071802. 18 p. Replaced with the published version. Added the journal reference and DOI. All the figures and tables can be found at
<http://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-17-010> (CMS Public Pages).
- [87] ATLAS Collaboration Collaboration, “Search for diboson resonances in hadronic final states in 79.8 fb^{-1} of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector”, Technical Report ATLAS-CONF-2018-016, CERN, Geneva, Jun, 2018.
- [88] T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson, “QCD or What?”, [arXiv:1808.08979](https://arxiv.org/abs/1808.08979).
- [89] J. A. Aguilar-Saavedra, J. H. Collins, and R. K. Mishra, “A generic anti-QCD jet tagger”, *JHEP* **11** (2017) 163, doi:10.1007/JHEP11(2017)163, [arXiv:1709.01087](https://arxiv.org/abs/1709.01087).
- [90] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, “Deep-learned Top Tagging with a Lorentz Layer”, *SciPost Phys.* **5** (2018), no. 3, 028, doi:10.21468/SciPostPhys.5.3.028, [arXiv:1707.08966](https://arxiv.org/abs/1707.08966).
- [91] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning”. The MIT Press, 2016.
- [92] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, *CoRR* **abs/1412.6980** (2014) [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [93] F. Chollet et al., “Keras”. <https://keras.io>, 2015.
- [94] M. A. et al, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems”, 2015. Software available from tensorflow.org.
- [95] G. P. Salam, L. Schunk, and G. Soyez, “Dichroic subjettness ratios to distinguish colour flows in boosted boson tagging”, *JHEP* **03** (2017) 022, doi:10.1007/JHEP03(2017)022, [arXiv:1612.03917](https://arxiv.org/abs/1612.03917).